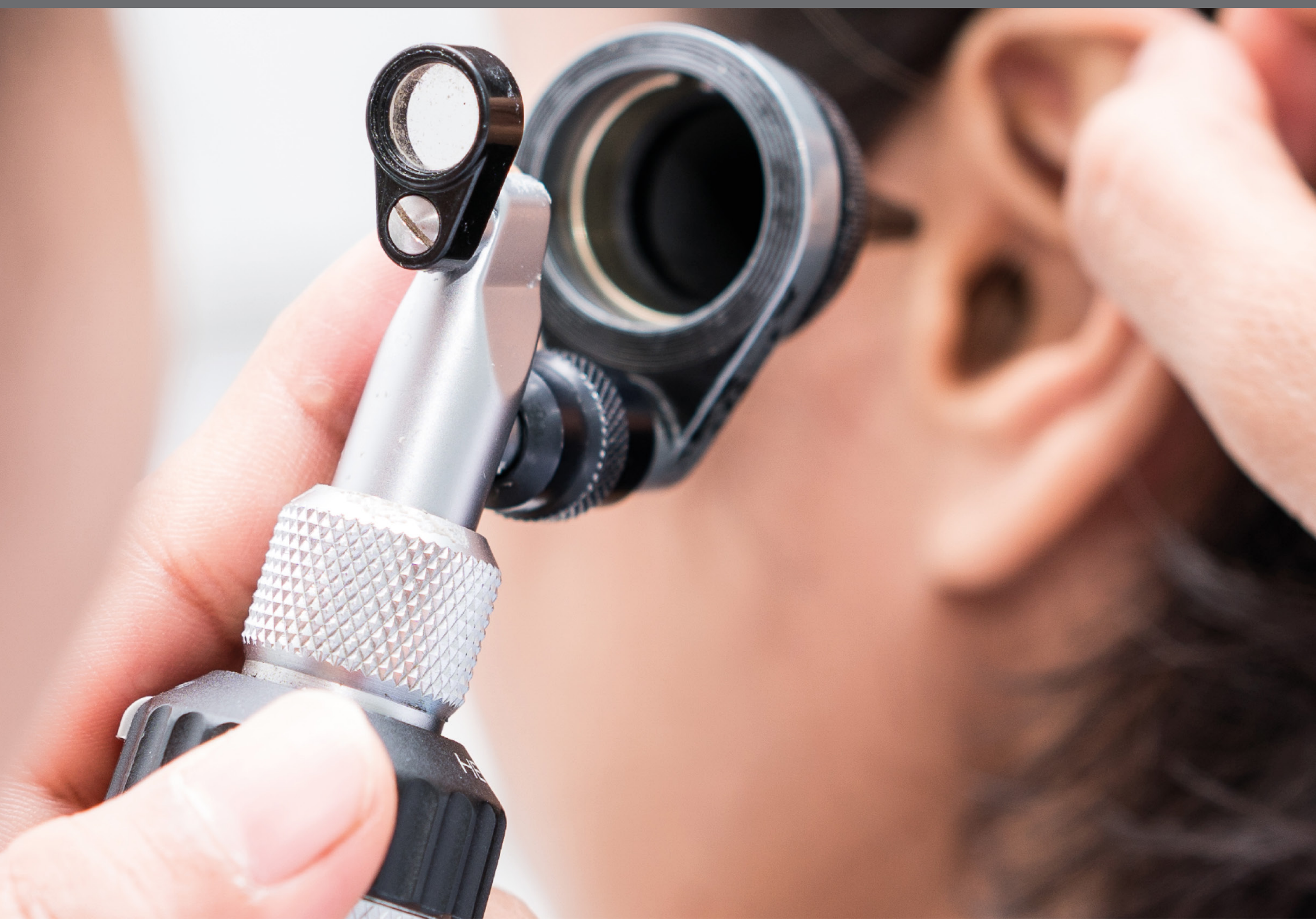
DIGITAL HEARING HEALTHCARE

EDITED BY: Qinglin Meng, Jing Chen, Changxin Zhang, Dennis L. Barbour and Fan-Gang Zeng

PUBLISHED IN: Frontiers in Digital Health, Frontiers in Psychology, Frontiers in Neuroscience, Frontiers in Neurology, Frontiers in Medicine and Frontiers in Public Health



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88976-814-1
DOI 10.3389/978-2-88976-814-1

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

DIGITAL HEARING HEALTHCARE

Topic Editors:
Qinglin Meng, South China University of Technology, China
Jing Chen, Peking University, China
Changxin Zhang, East China Normal University, China
Dennis L. Barbour, Washington University in St. Louis, United States
Fan-Gang Zeng, University of California, Irvine, United States

We would like to acknowledge VCCA2020-Organizer Jan-Willem Wasmann, who has acted as coordinator and has contributed to the preparation of the proposal for this Research Topic.

Dr. Qinglin Meng is working on an audio project for Huawei Technologies Co., Ltd. Dr. Jing Chen is working on research projects with Sonova AG.
Dr. Fan-Gang Zeng owns stock in Axonics, Nurotron, Syntiant, Velox and Xsense.
Dr. Dennis Barbour founded and owns equity in Bonauria. All other Topic Editor declare no conflicts of interest.

Citation: Meng, Q., Chen, J., Zhang, C., Barbour, D. L., Zeng, F.-G., eds. (2022). Digital Hearing Healthcare. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88976-814-1

Table of Contents

TYPE Editorial
PUBLISHED 13 July 2022
DOI 10.3389/fdgth.2022.959761

OPEN ACCESS

EDITED BY
Constantinos S. Pattichis,
University of Cyprus, Cyprus

REVIEWED BY
Alexander James Casson,
The University of Manchester,
United Kingdom
Siti Anom Ahmad,
Putra Malaysia University, Malaysia

*CORRESPONDENCE
Qinglin Meng
mengqinglin@scut.edu.cn

SPECIALTY SECTION
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

RECEIVED 02 June 2022
ACCEPTED 27 June 2022
PUBLISHED 13 July 2022

CITATION
Meng Q, Chen J, Zhang C, Wasmann J-WA, Barbour DL and Zeng F (2022) Editorial: Digital hearing healthcare. *Front. Digit. Health* 4:959761. doi: 10.3389/fdgth.2022.959761

Editorial: Digital hearing healthcare

Qinglin Meng[1,2*], Jing Chen[3,4], Changxin Zhang[5], Jan-Willem A. Wasmann[6], Dennis L. Barbour[7] and Fan-Gang Zeng[8]

[1]Acoustics Laboratory, School of Physics and Optoelectronics, South China University of Technology, Guangzhou, China, [2]n[3] Hearing Laboratory, Guangzhou, China, [3]Key Laboratory of Machine Perception (Ministry of Education), School of Artificial Intelligence, Speech and Hearing Research Center, Peking University, Beijing, China, [4]National Biomedical Imaging Center, College of Future Technology, Peking University, Beijing, China, [5]Faculty of Education, East China Normal University, Shanghai, China, [6]Department of Otorhinolaryngology, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center Nijmegen, Nijmegen, Netherlands, [7]Laboratory of Sensory Neuroscience and Neuroengineering, Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO, United States, [8]Department of Otolaryngology - Head and Neck Surgery, Center for Hearing Research, University of California, Irvine, Irvine, CA, United States

KEYWORDS

digital hearing healthcare, computational audiology, tele-audiology, audiometry, hearing aid, hearing loss, cochlear implant, tinnitus

Editorial on the Research Topic
Digital Hearing Healthcare

The Digital Hearing Healthcare or DHH Research Topic consists of 30 articles using modern digital methods to address a wide range of interesting and important hearing healthcare related issues. The interdisciplinary nature of the DHH Research Topic spans five Frontiers journals and six Frontiers sections. As the host journal, Frontiers in Digital Health published 16 articles, while Frontiers in Neuroscience, Medicine, Psychology, Public Health, and Neurology published 7, 3, 2, 1, and 1 articles, respectively. The Research Topic was initiated in March 2020, opened for submission from August 2020 to Octobor, 2021, with a total of 38 submissions being received.

The World Health Organization (WHO) estimates that "more than 1.5 billion people experience some decline in their hearing capacity during their life course, of whom at least 430 million will require care," but only a fraction of them are receiving the care, with unaddressed hearing loss resulting in an annual global cost of US$980 billion (1). We believe that digital health methods will play a significant role not only in the prevention, diagnosis, treatment and rehabilitation of hearing loss, but also in increasing the universal coverage of and decreasing the cost and burden of hearing healthcare. Here the term "Digital" is used to signify a much broader context than the traditional "digital signal processing" concept. Digital Hearing Healthcare uses a wide range of digital technologies to address hearing healthcare problems in ways complementary to, or different from, the conventional clinical processes in hospitals or clinics. This Editorial summarizes the key findings and contributions of the DHH Research Topic.

To show the full scope of the Research Topic, Figure 1 displays the 30 articles (middle column) that are conceptually categorized according to the target hearing healthcare issues (left column) and digital techniques (right column). The hearing health issues include basic audiometry tests, hearing devices, advanced hearing ability, and tinnitus therapeutics. The digital techniques include machine learning and big data, smartphone and wearables, tele-audiology, automatic speech recognition, and signal processing.

Basic audiometry testing

Audiometry consists of methods used in a standard audiology center to identify and quantitatively measure the degree and pathogenesis of hearing loss. Pure-tone audiometry (PTA) is the most widely used method, resulting in a quantitative estimate of the audiogram. The audiogrm reflects the hearing threshold as a function of frequency (2). Wang et al. studied phenotypes of noise-induced hearing loss by clustering analysis on audiograms of more than 10 thousand shipyard employees. Cox and de Vries proposed an improved probabilistic model-based PTA procedure that combines more prior information about the patients. This line of research on machine learning-based methods is expected to facilitate automatic PTA by shortening the testing time while maintaining accuracy (3, 4). Ellis and Souza examined the performance of a previously patented method for measuring an audiogram without precise sound level calibration in remote testing. That method combined the audiometric slope between pure-tone thresholds estimated at 2 and 4 kHz and questionnaire information. The latter two methods were examined in simulation experiments based on large databases. Frank et al. validated an iPad-based PTA app in 25 individuals with mild cognitive impairment and mild dementia. Beyond PTA, speech audiometry is critical for assessing the ability to hear and comprehend speech. Ratnanather et al. developed an automated program to calculate the phoneme confusion pattern based on the records from word or sentence-level tests. Ceccato et al. developed a French version of the antiphasic digits-in-noise (DIN) test for smartphone hearing screening. DIN, which does not require precise sound level calibration and tests digit-triplet reception thresholds in noise, has been developed in many languages. The antiphasic DIN may be more sensitive to unilateral hearing loss and conductive hearing loss than traditional diotic DIN (5). Other than the behavioral tests, novel technologies for otoacoustic emission (OAE) and auditory brainstem response

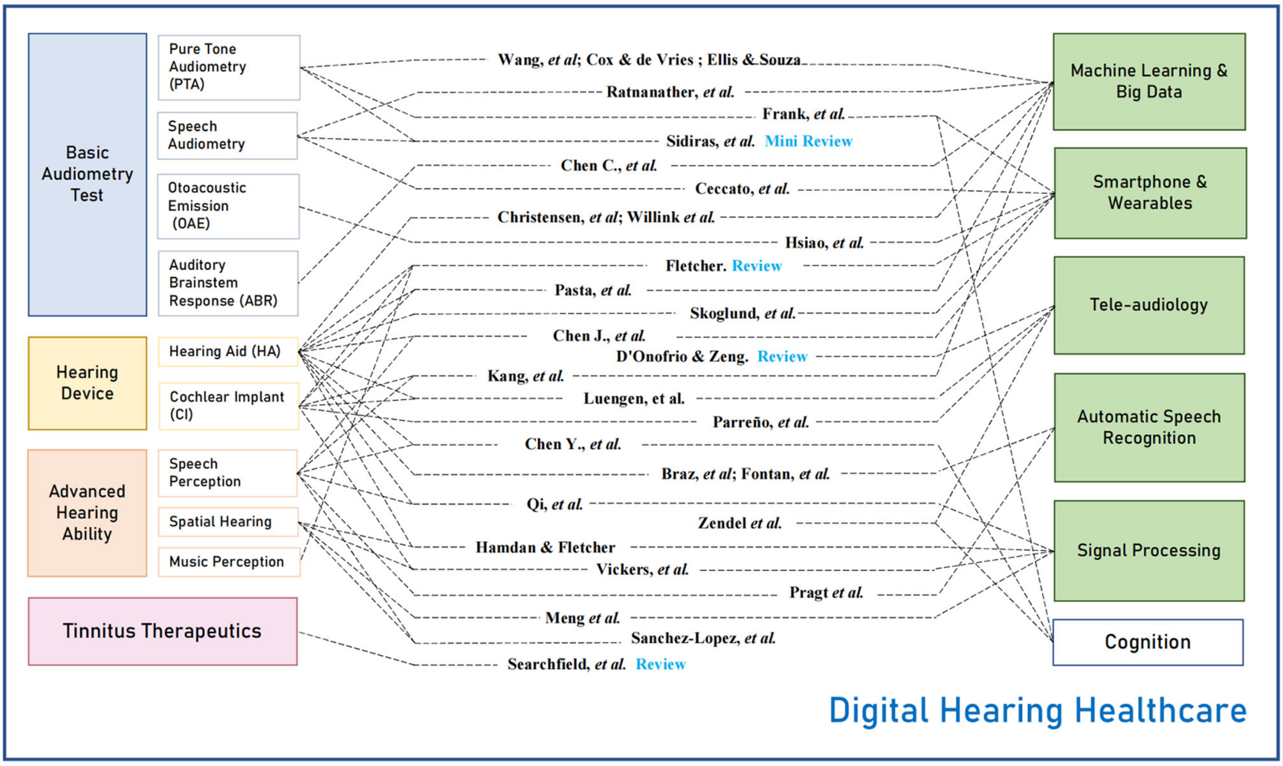


FIGURE 1
Distribution of the 30 articles in this Digital Hearing Healthcare (DHH) Research Topic.

(ABR) detection were also studied. Hsiao et al. introduced the stimulus design and signal processing to measure distortion-product OAE with a single loudspeaker in the ear. The research targets intelligent consumer earphones with integrated hearing health monitoring functions. Chen C. et al. proposed a machine learning method to recognize ABR waveforms automatically, and the results showed its feasibility in saving time and helping make diagnoses.

Hearing aid

As one of the most widely-used treatments for hearing impaired (HI) listeners, hearing aids (HAs) were explored and discussed adequately in this specific issue, including more than ten papers in which HA-related research was conducted. Willink et al. focused on the alternative pathway for hearing care by examining the HI population who do not use HAs via describing their characteristics and health care utilization patterns with the sample size of 7,361 Medicare beneficiaries. Two papers (Pasta et al.; Christensen et al.) tried to provide a deeper insight into the adoption of hearing care treatments and individual HA usage patterns by analyzing objective HA use data logged from real-world users. Braz et al. and Fontan et al. developed new fitting methods combining automatic speech recognition (ASR) to optimize compression parameters for HA. One of the most important objectives of using HAs is to improve the speech perception of HA customers, and the perception of the speech signals processed by HA technology is probably related to listeners' language experience (6). Three papers investigated the speech perception of HA users speaking Mandarin Chinese. Qi et al. aimed to examine the effects of an adaptive non-linear frequency compression algorithm on speech perception and sound quality in native Mandarin-speaking adult listeners with sensorineural hearing loss. They found significant perceptual benefit from the adaptive non-linear frequency compression algorithm in detecting high frequency sounds at 8 kHz, in consonant recognition, and in an overall sound quality rating. Chen J. et al. aimed to evaluate the improvement of speech recognition in noise measured with signal-to-noise ratio by using wireless remote microphone technology for HI listeners speaking Mandarin Chinese. The experiment results confirmed the significant improvement. Chen Y. et al. investigated the relationships between cognitive and hearing functions in older Chinese adults with HAs and untreated hearing loss. They concluded that speech perception in noise is significantly associated with different cognitive functions. In addition to speech perception, researchers also tried to extend HA functions to make HI listeners' hearing experience close to normal hearing, such as spatial hearing and music perception. Hamdan and Fletcher proposed a compact two-loudspeaker virtual sound reproduction system for clinical testing of spatial hearing with hearing-assistive devices. They found such a system can give broad access to advanced and ecologically valid audiological testing, and it could substantially improve the clinical assessment of hearing-assistive device users. Meng et al. investigated a novel minimum audible angle (MAA) test using virtual sound source synthesis and found it to be a suitable alternative to more complicated and expensive setups. Sanchez-Lopez et al. developed the Better hEAring Rehabilitation (BEAR) project to measure various hearing abilities, especially supra-threshold hearing deficits.

Cochlear implant

For most patients with severe-to-profound hearing impairment, HAs cannot provide enough benefits for their speech communication. The cochlear implant (CI) is a good option for them to partially (re)gain speech perception abilities, at least in a quiet environment. However, CI users still face significant challenges in advanced hearing functions, e.g., speech-in-noise perception, spatial hearing, and music perception (7). Many efforts have been made to improve the CI performance on these issues gradually. Three papers in this Research Topic researched CI improvements. Kang et al. used recent progress to improve noise suppression or speech enhancement of CIs and found that the intelligibility of the denoised speech can be significantly enhanced when a neural network is trained with a loss function bias toward more noise suppression than that with equal attention on noise residue and speech distortion. Vickers et al. developed a package of virtual-reality games to train spatial hearing in children and teenagers with bilateral CIs. The virtual-reality implementation demonstrated to be more engaging than traditional rehabilitation methods. As introduced in the HA section, Fletcher reviewed the effects of haptic stimulation methods on enhancing music perception in HA and CI listeners. New technologies such as advanced algorithms, virtual reality, and multimodal stimulation should be up-and-coming to positively integrate future CI devices with CI users' lives.

Tinnitus

Tinnitus, or ringing in the ears, affects 15% of the general population. At present, tinnitus mechanisms remain unclear (8). Although no cure is available for tinnitus, safe and effective therapy and counseling have been developed to help alleviate its symptoms. Searchfield et al. reviewed the first three generations of digital technologies for tinnitus management, ranging from digital hearing aids and apps to stand-alone, customized digital devices. Most interestingly, they forecast the fourth-generation digital technology that incorporates physiological sensors, multi-modal transducers, and AI for personalized tinnitus therapy.

Tele-audiology

Accelerated by COVID-19, tele-audiology has become a necessary means of delivering hearing care service, especially for the most-vulnerable elderly population. D'Onofrio and Zeng examined technological and regulatory barriers for a wide range of audiological services from audiometry to hearing device fitting and rehabilitation. Most of these barriers can be overcome not only to provide reliable and effective tele-audiological service, but more importantly, to improve access to care, increase follow-up rates, and reduce travel time and costs. Luengen et al. envisioned an innovative tele-audiological model that consists of (1) one-to-one remote interaction between a patient and an audiologist, (2) a one-to-many model that relies on automated service provided by AI, and (3) a several-to-many application that can fit hearing devices not only based on the patient's audiological profile but also their listening environments. One critical piece of information in tele-audiology is reliable monitoring of hearing status and device functionality, especially for a complicated device such as a CI. Parreño et al. developed a self-monitoring method of measuring CI electrode impedance. Making use of a computer, the device interface provided by the manufacturer, and secure internet connectivity, they were able to record, transfer and monitor impedance at home without any adverse events. Another critical piece of information is accurate and reliable communication between the patient and providers in tele-audiology. Zendel et al. found that current teleconferencing is less reliable than in-person instruction in terms of patients' recollection of the healthcare messages delivered. Speech and video quality, as well as communication methods, need to be improved in order to reduce the memory deficit associated with current telehealth technologies.

Machine learning and big data

Machine Learning (ML) represents a set of tools that reveal complex trends in data that would be difficult or impossible to discern otherwise. Often these tools are powered by large amounts of data ("big data"), which provide more opportunities to observe interesting trends. Ellis and Souza took this approach to train an ML classifier of audiograms using almost 10,000 individuals from a large national auditory and demographic information database. Wang et al. followed a similar approach, analyzing over 10,000 audiograms for notch appearance, identifying three noise-induced hearing loss subtypes in the process. Using a different type of data, Chen C. et al. employed a recurrent neural network and signal processing to recognize potential waveforms in Auditory Brainstem Response (ABR) signals. Big Data can be leveraged in different ways, as well. For example, Cox and de Vries learned from existing databases how to shorten the test time of future audiograms computed in real time using an ML algorithm.

Ongoing data collection enabled by always-connected devices offers the opportunity to trade up for more informative algorithms as more data become available. Christensen et al. clustered the usage activity of 64 hearing aid users over several days with a straightforward K-means algorithm. Pasta et al. approached a similar problem with nearly 16,000 users and several month's worth of data. They used a more sophisticated multilayer neural network to reveal finer trends.

ML methods have great potential for therapeutics as well as diagnostics. Kang et al. describe a deep learning (i.e., many-layered neural network) approach to improve speech enhancement in cochlear implant (CI) encoding algorithms. Braz et al. used knowledge of audiograms and a genetic algorithm to search the many configurations of hearing aid program settings to optimize the device for a particular patient.

Smartphone and wearables

Miniaturization has led to smaller sensors and stimulators to create haptic devices worn as gloves or bracelets. Fletcher contributed a review paper to discuss "Can haptic stimulation enhance music perception in HI listeners." It has been reported that "electro-haptic stimulation" improves melody recognition and pitch discrimination, as well as speech-in-noise performance and sound localization. This review paper focused on the use of haptic stimulation to enhance music perception in HI listeners. One of his conclusions is that haptic devices, in concert with other modalities, can enhance the music experience in hearing impaired listeners. Using several sensor technologies, modern HAs strive to become better, more personalized, and self-adaptive devices that can handle environmental changes and cope with the day-to-day fitness of the users. Skoglund et al. measured the accuracy of activity tracking, e.g., step detection, through small accelerometers embedded in hearing aids. They showed classification of activities was similar to conventional activity tracking techniques, which is encouraging for applications in hearing health care. In noisy conditions, small microphones can be employed to improve the signal-to-noise ratio. Chen J. et al. confirmed that wireless microphones improve speech recognition in Chinese hearing impaired listeners when the target speaker is at a larger distance.

Automatic speech recognition

Within digital hearing health care, ASR has various applications, including the use of ASR as a tool to understand conversations. Pragt et al. examined the speech recognition performance of four ASR apps on smartphone using conventional Dutch audiological speech tests. They compared human speech recognition performance and performance by ASR apps. They found that the performance of the apps was limited on audiological tests that provide little linguistic context or use low signal-to-noise levels. They concluded that conventional performance metrics and conventional hearing tests are insufficient to assess the benefits of ASR apps for the deaf and proposed that adding new performance metrics including the semantic difference as a function of SNR and reverberation time could help monitor ASR performance. Another strategy uses ASR, hearing loss models, and hearing aid signal processing simulations to mimic impaired hearing listeners. As mentioned in the Hearing Aid section, Braz et al. and Fontan et al. demonstrated that ASR, in combination with random search algorithms, can be used to find optimal subsets of parameter settings for hearing aids. These optimizations should subsequently be validated in actual hearing aid users, especially in case of severe hearing loss. Recently, ASR has also been used in predicting CI performance (9) and, together with the above HA studies, demonstrates good potential in modeling hearing device performance.

Discussion and future outlook

It is worth noting that several articles present the progress of their projects driven by long-term interdisciplinary collaborations. They are SHOEBOX Audiometry for hearing screening using sound-attenuating headphones (Frank et al.), Better hEAring Rehabilitation (BEAR) to provide a test battery for hearing loss characterization (Sanchez-Lopez et al.), User-Operated Audiometry (UAud) to introduce an automated system for user-operated audiometric testing into everyday clinical practice (Sidiras et al.), BEARS (Both EARS) to develop virtual reality training suite for improving spatial hearing for 8–16 year-olds with bilateral CIs (Vickers et al.), OPRA for developing Objective Prescription Rule based on ASR (Braz et al.; Fontan et al.). These long-term, collaborative projects may result in improved efficiency while lowering the cost for hearing healthcare.

The interdisciplinary nature of the DHH field also provides great challenges and opportunities for both academic research and industrial development. For example, the Interdisciplinary Technologies for Auditory Prostheses (iTAP) conference series (http://www.itap.org.cn/) was founded in 2017 in China and has a vast interest overlap with this DHH Research Topic. Computational audiology, the application of complex models to clinical care (10), is an important part of digital hearing health care. Many contributors to this Research Topic first discussed their projects at the Virtual Conferences on Computational Audiology (VCCA) (https://computationalaudiology.com/), which have provided a platform to share progress in e-research, machine learning, big data, models, virtual reality, and other developments. The Research Topic, iTAP, and VCCA all share the same goal of providing a platform for facilitating communication between experts from different fields and accelerating research and development.

In conclusion, the contributions in this Research Topic have demonstrated that novel digital technologies in machine learning, big data, signal processing, telehealth, and mobile health are being actively applied toward hearing healthcare applications. Improving the accessibility and performance of audiometry, hearing-assist devices and tinnitus therapeutics stands out as successful application of these technologies. This Research Topic provides general readers a glimpse of the emerging Digital Hearing Healthcare field and hopefully will inspire more people, companies, and organizations to develop and deploy digital health techniques for better hearing, and as a result, a better world.

Author contributions

All authors except J-WW were guest associate editors of the Research Topic. J-WW has made significant contributions to the call-for-paper through the VCCA network. All authors wrote the paper text and approved the submitted version.

Funding

QM was supported by Guangdong Basic and Applied Basic Research Foundation Grant (2020A1515010386), Science and Technology Program of Guangzhou (202102020944). JC was supported by the National Key Research and Development Program of China (2021ZD0201503) and National Natural Science Foundation of China (12074012). CZ was supported by National Natural Science Foundation of China (31900801). DB was supported by the Advanced Education Research & Development Fund and the National Institues of Health (R21EY033553). F-GZ was supported by the National Institues of Health (5R01DC015587 and 1R01AG067073).

Acknowledgments

We thank the authors of the manuscripts submitted to this research topic for their valuable contributions and the referees for their rigorous review. We thank Wang Yamin for designing Figure 1.

Conflict of interest

DB reports equity ownership in Bonauria. F-GZ reports equity ownership in Axonics, DiaNavi, Neocortix, Nurotron, Syntiant, Velox and Xense.

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

1. WHO. *WHO, World report on hearing.* (2021).

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

3. Song XD, Wallace BM, Gardner JR, Ledbetter NM, Weinberger KQ, Barbour DL. Fast, continuous audiogram estimation using machine learning. *Ear Hearing.* (2015) 36:e326. doi: 10.1097/AUD.0000000000000186

4. Barbour DL, Howard RT, Song XD, Metzger N, Sukesan KA, DiLorenzo JC, et al. Online machine learning audiometry. *Ear Hearing.* (2019) 40:918. doi: 10.1097/AUD.0000000000000669

5. Sousa KCDe, Smits C, Moore DR, Myburgh HC, Swanepoel DW. Diotic and antiphasic digits-in-noise testing as a hearing screening and triage tool to classify type of hearing loss. *Ear Hearing.* (2022) 43:1037–48. doi: 10.1097/AUD.0000000000001160

6. Chen J, Huang Q, Wu X. Frequency importance function of the speech intelligibility index for Mandarin Chinese. *Speech Commun.* (2016) 83:94–103. doi: 10.1016/j.specom.2016.07.009

7. Zeng F-G. Challenges in improving cochlear implant performance and accessibility. *IEEE T Bio-Med Eng.* (2017) 64:1662–4. doi: 10.1109/TBME.2017.2718939

8. Zeng F-G. Tinnitus and hyperacusis: central noise, gain and variance. *Curr Opin Physiol.* (2020) 18:123–9. doi: 10.1016/j.cophys.2020.10.009

9. Bance M, Brochier T, Vickers D, Goehring T. From microphone to phoneme: an end-to-end computational neural model for predicting speech perception with cochlear implants. *IEEE Tran Biomed Eng.* (2022). doi: 10.1109/TBME.2022.3167113

10. Wasmann J-WA, Lanting CP, Huinck WJ, Mylanus EA, van der Laak JW, Govaerts PJ, et al. Computational audiology: new approaches to advance hearing health care in the digital age. *Ear Hearing.* (2021) 42:1499. doi: 10.1097/AUD.0000000000001041

frontiers in Digital Health

ORIGINAL RESEARCH
published: 09 November 2020
doi: 10.3389/fdgth.2020.582562

Toward Self-Measures in Cochlear Implants: Daily and "Homemade" Impedance Assessment

Matias Parreño [1], Federico A. Di Lella [1], Florencia Fernandez [1], Carlos M. Boccio [1] and Sebastian A. Ausili [2]**

[1] Department of Otolaryngology, Hospital Italiano, Buenos Aires, Argentina, [2] Department of Otolaryngology, University of Miami, Coral Gables, FL, United States

Introduction: Cochlear implant (CI) impedance reflects the status of the electro neural interface, potentially acting as a biomarker for inner ear injury. Most impedance shifts are diagnosed retrospectively because they are only measured in clinical appointments, with unknown behavior between visits. Here we study the application and discuss the benefits of daily and remote impedance measures with software specifically designed for this purpose.

Methods: We designed software to perform CI impedance measurements without the intervention of health personnel. Ten patients were recruited to self-measure impedance for 30 days at home, between CI surgery and activation. Data were transferred to a secured online server allowing remote monitoring.

Results: Most subjects successfully performed measurements at home without supervision. Only a subset of measurements was missed due to lack of patient engagement. Data were successfully and securely transferred to the online server. No adverse events, pain, or discomfort was reported by participants.

Discussion: This work overviews a flexible and highly configurable platform for self-measurement CI impedance. This novel approach simplifies the CI standard of care by reducing the number of clinical visits and by proving useful and constant information to CI clinicians.

Keywords: cochlear implants, telehealth, objective measurements, electrical impedance, self-measures

OPEN ACCESS

Edited by:
Fan-Gang Zeng, University of California, Irvine, United States

Reviewed by:
Jace Wolfe, Hearts for Hearing, United States
Ming Huang, Nara Institute of Science and Technology (NAIST), Japan

***Correspondence:**
Federico A. Di Lella
federico.dilella@hospitalitaliano.org.ar
Sebastian A. Ausili
s.ausili@miami.edu

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: *29 July 2020*
Accepted: *15 October 2020*
Published: *09 November 2020*

Citation:
Parreño M, Di Lella FA, Fernandez F, Boccio CM and Ausili SA (2020) Toward Self-Measures in Cochlear Implants: Daily and "Homemade" Impedance Assessment. Front. Digit. Health 2:582562. doi: 10.3389/fdgth.2020.582562

INTRODUCTION

Cochlear implants are the most successful sensory prosthetic device in medicine. Research has demonstrated that CIs typically provide significant improvement in speech recognition for persons with severe to profound hearing loss (1, 2). Like all neural prostheses, the interface between electrodes and neural tissue is a critical aspect for adequate functioning (3). The measurement of intracochlear electrode impedance provides an indication of the status of the electrode–tissue interface, which may give important information for the clinician providing CI management (4). Normally, this is a quick (i.e., 1–2 min) and safe procedure, because it involves the use of subthreshold- or near-threshold level stimulation (5, 6). During the postoperative period, impedance measurement is routinely performed for CI-programming guidance, detecting device failures, and extrusion of electrode contacts. Moreover, this value is a biomarker for inner ear injury (i.e., fibrosis and ossification) that may help predict residual hearing loss or vertigo events (7, 8).

While electrode impedance measurement provides essential information regarding the CI and cochlear status, it is only performed in the patient's clinical appointments, and not much is known on the behavior of this parameter between visits (7–9). Thus, most impedance variations are diagnosed retrospectively when little can be done to correlate them with clinical presentation or to start pharmacological treatment. More frequent monitoring of CI impedance with available methods is not feasible, but currently the use of telemedicine can be used to improve clinical practice by performing constant monitoring of electrode impedance values.

Telemedicine made its way into the cochlear implant clinic in the last 10 years with advances in connectivity. This progress was supported with the development of remote-access applications and new telecommunication systems (10, 11). Audiologists started performing remote fitting and monitoring of implanted patients, allowing medical care while keeping patients at their homes with no significant differences regarding standard programming sessions (10–18).

Here, we study the application and discuss the benefits of assessing daily and remote impedance measures with a software specifically designed for CIs. Patients measured themselves at home for 30 days, and data were automatically uploaded to an encrypted cloud database. The procedure did not require supervision of any clinicians and was easily performed in all patients. During the 30 days over which a recipient's electrode impedance values were measured in this study, the researchers could retrieve this data at any time, which enables true remote monitoring of CI status and cochlear health.

METHODS

Hardware

The measurement setup was designed so that patients only had to connect the audio-processor coil to their implant. It included a Freedom® speech processor with research firmware (ver. 0102E00F02), a clinical programming interface (Pod, Cochlear Ltd.), and the patient's personal computer (**Figure 1**). Note that it has the same number of elements as for a normal clinical fitting appointment.

Software

We designed a software that performs CI impedance measurements at home without the intervention of health personnel (**Figure 2A**). We used the Delphi platform and a dynamic-link library (DLL) provided by Cochlear Ltd. to develop a patient-oriented software that runs under a Windows® operating system and in personal computers. The application was distributed as a single-file installer. Upon installation, it runs in the background automatically detecting the programming Pod connection. The app launches manually or automatically when the Pod is detected. Once running, the main window automatically pops up showing an intuitive and simple graphical user interface (GUI) front end (**Figure 2B**). The GUI provides the instructions to start the impedance measurements and offers help if it detects incorrect connection to CI (**Figures 2D,E**). Once an adequate connection is obtained and after user confirmations (**Figures 2C,F**), it performs impedance measurements (**Figure 2G**). Each electrode impedance (Z_e) measurement is assessed by streaming a constant current (I) pulse of 74.21 μA with a phase width of 25 μs. Using the active intracochlear electrode and the extracochlear reference electrode [MP1 coupling mode (8)], voltage (V) at the trailing edge of the pulse is recorded. Finally, the impedance value is calculated through Ohm's law as follows:

$$Z_{e_n}[\Omega] = \frac{V}{I} = \frac{measured\ voltage\ [volt]}{74.21 \cdot 10^{-6}[A]} \quad (1)$$

with n being the intracochlear electrode number.

Values are temporarily stored into a secured local database (**Figure 2I**) by means of the industry-standard SQLite and automatically exported to a web-based secure server (**Figure 2J**). This process is automated and requires no intervention of the patient. Last, the health professionals with granted access to the cloud database can analyze the CI user data from the hospital.

Patient Safety

CI impedance measurement is a safe procedure performed as routine in the clinic. It involves stimulating with a low current stimulus which is inaudible for most patients and causes no discomfort (5). Unexpected problems such as connection failure, computer, or software lagging are unable to produce a current level that exceeds defined parameters. This is mainly due to the transmission commands: data packages or tokens carry the information required to perform a required task. In the event of loss of a package, the processor automatically stops all tasks. However, for extra caution or in case of discomfort, the patient was instructed by the investigator and software on how to terminate the procedure immediately by removing the coil (see GUI in **Figure 2G**).

Data Security

Both software and connections were developed to ensure maximum protection of the patient's data and anonymity. Information required by the software to register on first use does not include any of the Health Insurance Portability And Accountability Act (HIPAA) identifiers (19). Data transfer to the cloud database only includes deidentified measurements and registration number, which are not associated with the patient's ID. Moreover, the exchanged traffic between the web server and both patient and investigator uses Hypertext Transfer Protocol Secure over transport layer security (HTTPS over TLS). These protocols provide encryption, data integrity, and authentication; thus, reasonable protection is ensured (20).

Patient measurements are temporarily stored in a local encrypted SQLite database (**Figure 2I**) until data is transmitted to the web server (**Figure 2J**). Upon transfer to the server, local information is deleted to mitigate risk of local breach. When an internet connection is available, transfer is immediate; otherwise, periodic attempts every 3 min are performed. Impedance acquisition parameters, such as current level and pulse width, are embedded in the software, making its alteration highly improbable.

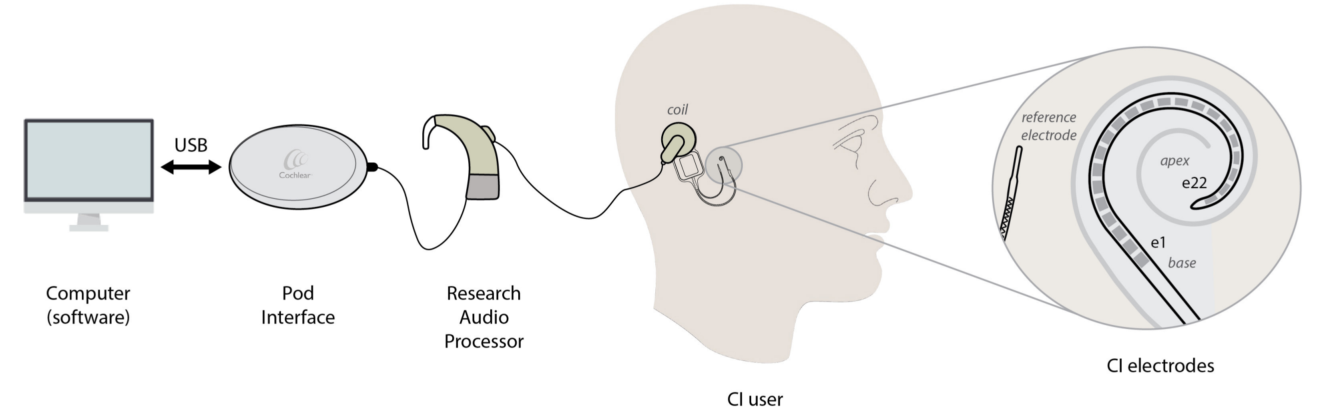


FIGURE 1 | Hardware setup for CI impedance measurement and electrode illustration.

CI Subjects

Impedance measurements were conducted after approval of the local Ethics Committee in concordance with international standards for human research. Written informed consent was provided to all participants. A total of 10 CI users were recruited for this study. All patients were implanted at the Hospital Italiano in Buenos Aires, Argentina. A Cochlear Nucleus CI24RE Contour Advance™ electrode array with Freedom or Profile platform (Cochlear Ltd., Australia) was used for all subjects. This CI consists on an array of 22 active electrodes tonotopically arranged inside the cochlea and uses an extracochlear (reference) electrode for MP1 coupling mode (see **Figure 1**).

The average patient age was 34 (range 1–67). **Table 1** shows patient description, including age at implantation, supervisor's age in case of a minor CI user, gender, CI side, and etiology.

Impedance Measurement

To assess CI impedances, the pulse characteristics (i.e., amplitude, phase duration, and interphase gap) were configured according to values used in Custom Sound Suite Software (Cochlear Ltd.) (21). The impedance coupling mode was limited only to Monopolar 1 (MP1), where the circuit is closed using an intra-cochlear and extra-cochlear electrode, where the last operates as the reference for all measures. This external electrode is normally positioned between the skull and the temporal muscle, also referred to as "ball" electrode in Cochlear Ltd. devices. Every time the patient runs the measurement session, a stream of 22 pulses is sent (one for each electrode), and each corresponding voltage telemetry measurement is retrieved. This procedure is performed in accordance with the predefined parameters embedded in the software code. The stream of pulses and recording of each electrode voltage is completed in ∼10 s. Note that the entire procedure also includes the connection of the POD and change of audio processor (see **Figure 1**), which extends the overall time to ∼1–2 min.

All subjects were provided with the previously described custom software and measurement hardware. The research team instructed subjects (and/or supervisors) on how to self-perform the measurements at home twice a day—with ∼12 h difference—for 30 days before CI activation. A printed brochure on how to connect and measure was also provided as support. A training measurement session was performed under supervision before the patient went home with the equipment. The first appointment (day 0) was measured postoperatively with help and supervision at the hospital, until the activation day (day 1–day 29) subjects measured themselves at home and the last appointment (day 30) was performed at the hospital again.

RESULTS

A total of 450 measurement sessions were performed, accounting for 75% of the total expected measurements (2 times × 30 days × 10 subjects = 600 measures). From the non-performed measurements (150 measures), only 1 was due to software or hardware issues and the rest correspond to skipped measurement sessions. Subject 7 did not measure for 24 consecutive days, accounting for 48 lost sessions. Although not all subjects were measured twice a day as required, at least one session per day was performed. No adverse events, pain, or discomfort was reported by participants.

Figure 3 shows two example subjects over time for all electrodes and its overall mean. The average patter of the group is represented by S8 (**Figure 3A**). All electrodes showed an initial decrease (days 0–3), a continuous growth (4–17 days), and a final stabilization period (days>18). Interestingly, an atypical variation between days 5 and 10 was captured for S1 (**Figure 3B**), where higher values were observed for the basal electrodes (e1, e2, and e3). Despite the differences across electrodes, they all converge to a more stable value from day 18 approximately to the activation day.

To illustrate the overall impedance behavior over time, we computed the average electrode impedance values across subjects over each electrode contact. As shown in **Figure 4**, average electrode impedance values increased until reaching a plateau at approximately day 15. However, small variations can be observed between electrodes and daily shifts are present. Overall, the

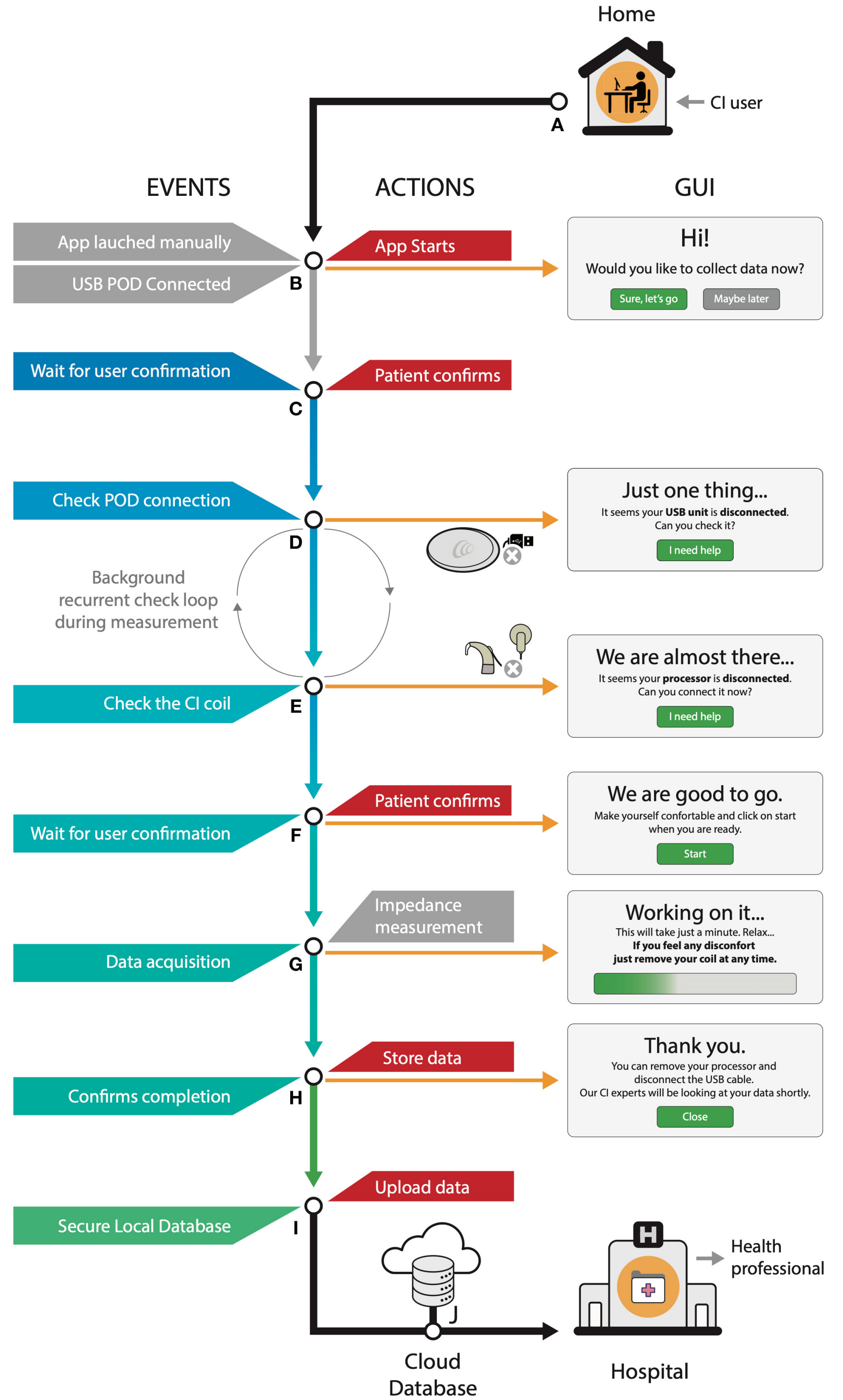


FIGURE 2 | Software workflow **(A–J)** and graphical user interface.

TABLE 1 | Demographic and general information about participants.

Subject	Age at implantation	Supervisor age	Gender	Implanted ear	Etiology
S1	1	38	F	Right	Preterm—ototoxicity
S2	13	40	M	Right	Ototoxicity
S3	34	–	M	Right	Viral parotitis
S4	16	43	M	Left	Unknown
S5	49	–	F	Left	Otosclerosis
S6	63	–	M	Left	Unknown
S7	41	–	F	Left	Genetic
S8	59	–	F	Left	Unknown
S9	6	30	F	Right	Ototoxicity
S10	67	–	M	Right	Unknown

group showed impedances of 6.1 kΩ on the surgical day along electrodes and 13.7 kΩ on the activation day.

DISCUSSION

A Novel Method

To the best of our knowledge, this is the first report of daily patient remote self-objective measurement in cochlear implants. All CI users (and supervisor) were able to self-perform measurements effectively and in little time. Adherence to measurement was high, allowing precise tracking of clinical impedance evolution on a daily basis.

Given the increasing number of implanted patients and the geographical spread all around the world, the possibility of acquiring remote measurements saves travel costs, time, and physical requirements in clinical care centers. Furthermore, this approach can generate extensive data collection helping to understand overall trends, hidden patterns, unknown correlations, etc.

The presented platform is highly versatile, enabling the integration of other measurements. For example, a more complex measurement of impedance includes polarization impedance and access resistance, which helps to reveal the underlying cochlear pathophysiology mechanism of these changes (7, 22, 23).

In our study, CI users performed their measurement with their personal computer, using specialized research hardware (POD and research CI processor, see methods). However, actual and future connectivity of personal mobile devices (i.e., mobile phones or tablets) allows for streaming of telemetry data, enabling impedance measurements protocols as well as other rehabilitation practices (e.g., audiometry test, speech in noise evaluation, questionnaires). These devices connect wirelessly to the patient audio processors which also can simplify measurements, especially in the pediatric population. More "homemade" measurements in the CI population will substantially improve the CI standard of care, simplifying actual unnecessary procedures and benefiting both the CI user and clinical care centers.

It is important to highlight that an important limitation of this procedure is the requirement of patient collaboration. Most of the lost measurements in our study were due to lack of subject cooperation. Considering that this investigation was carried out during the first month after implantation, where patient expectation on the CI is high, it is likely that this cooperation is further diminished with time. Although we did not assess user's feedback or satisfaction (e.g., via surveys), overall subjects positively agreed with the benefits of "homemade" measures. However, a systematic assessment of user's experience would certainly gain knowledge toward an optimized patient-oriented design. Recently Cochlear Ltd. released a smartphone app to perform remote impedance measurements and other tests in CI users (Cochlear's Remote Check). The benefit of this tool is

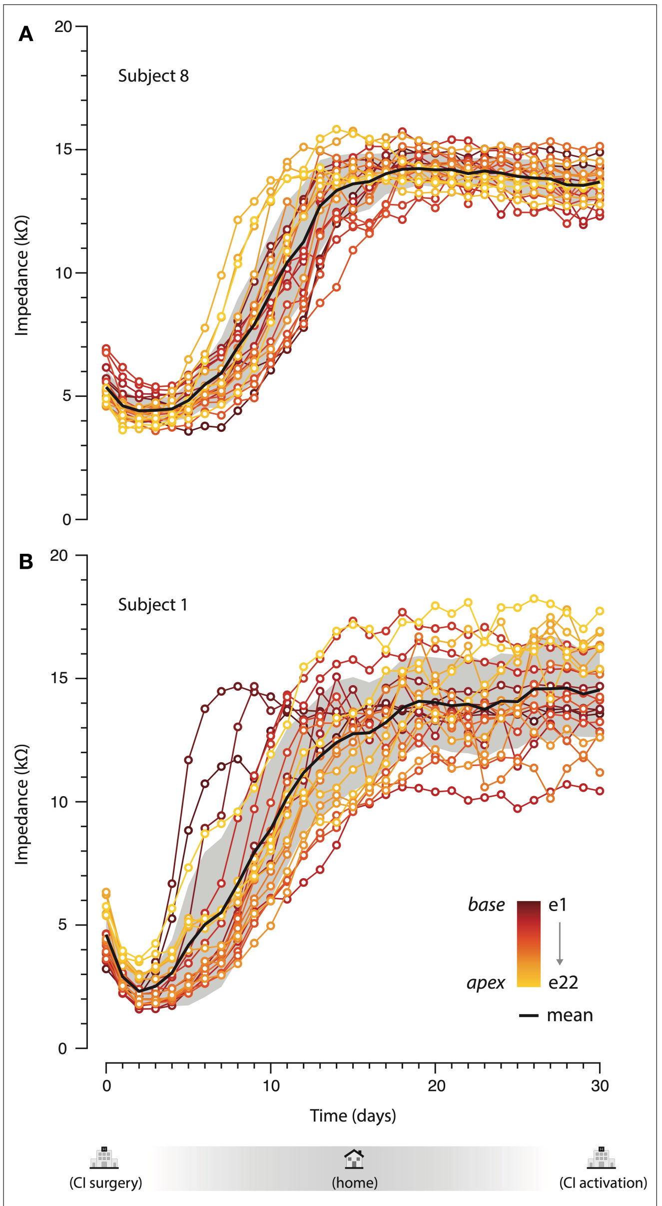


FIGURE 3 | Impedance measurements over time of Subject 8 **(A)** and Subject 1 **(B)**. Colors indicate the electrode number. Black line shows the overall mean over time and the gray patch its standard deviation.

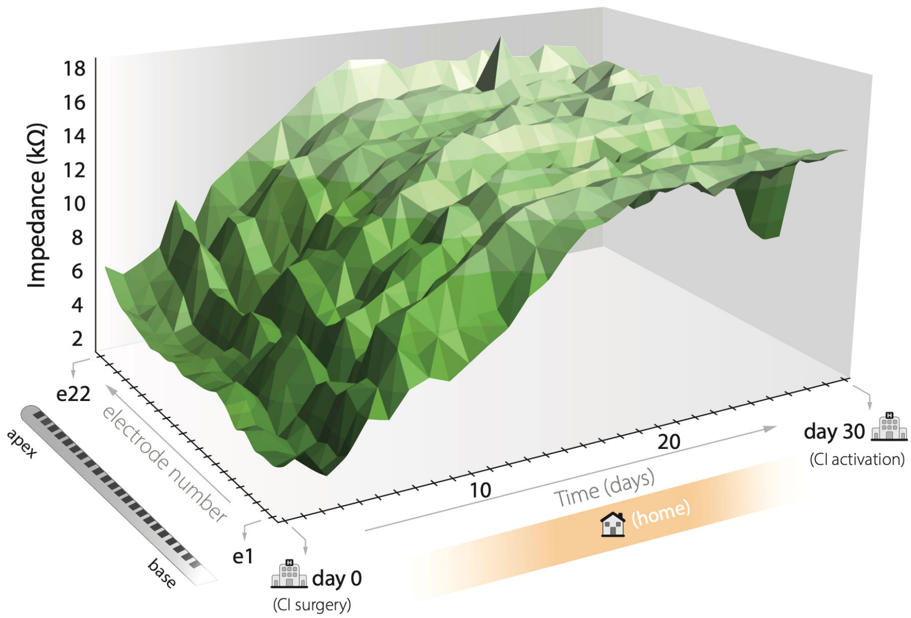


FIGURE 4 | Average impedance value progression pooled for all subjects. First, electrodes are located at the base of the cochlea with higher electrode numbers in the apex. Day 0 was measured postoperatively at the hospital, days 1–29 at home by the patient, and day 30, CI activation, at the hospital again.

the portability and wireless connectivity to the CI, potentially increasing the user's engagement. However, one could imagine that future applications with constant background impedance monitoring will rule out any cooperation-related issue and substantially increase the data availability.

Interestingly, the actual epidemiological context due to the COVID-19 pandemic imposed on us the challenge of considering new clinical approaches while practicing social distancing. As we continue to navigate the coronavirus pandemic and its economic consequences, telemedicine approaches like the one presented in this study not only promote the needed social distancing but also help to build the future of the CI standard of care.

About CI Impedance Daily Monitoring

To the moment, impedances in cochlear implants are a series of isolated values in time measured by audiologists during the fitting process. Daily home monitoring brings a whole new field of opportunities for audiologists, surgeons, and researchers. Impedance shifts may relate to clinical manifestations such as vertigo, Meniere-like symptoms, tinnitus, and loss of residual hearing. Unfortunately, the majority of studies are retrospective; thereby, it is difficult to establish a correlation between the symptoms and impedance variations (7–9). More sophisticated methods, such as the one presented in this paper, may allow rapid diagnosis of the impedance variations and a better correlation with the clinical manifestations. When detecting unusual impedance variations (like the one observed on S1; **Figure 3B**), automatic alerts could be directed to the CI center for further clinical decision and follow-up. These impedance shifts may be responsive to steroids; thus, detecting them on an early basis may allow prompt treatment and outcome improvement (7, 9, 24). Furthermore, the surgical approach adopted by the surgeon and the electrode insertion itself can cause trauma at the basal turn of the cochlea, which might elicit higher impedances due to its inflammatory process (25, 26).

It is noteworthy that even after impedance stabilization values continue to vary (see **Figure 3**), which could affect hearing perception even over the course of the same day. Continuous real-time measurement may also improve our results by the development of future auto fitting algorithms and automatic medical referral when values exceed defined parameters.

Impedance Dynamics Over Time

During the following 2–3 weeks from the surgery, the body's immune response is evidenced by a fibrous tissue encapsulation of the electrode array, which is reflected in a systematic overall increase on the impedance (3, 4, 8, 27–30). Once the CI is activated, the provided electrical current has major implications on the electrode–electrolyte interface (28). Typically, the impedance decreased and then stabilized within the first few months of device use (8, 28, 29, 31–36).

Hu et al. (37) showed the impedance dynamics when activating the CI 1 day after surgery and measuring intraoperatively and postoperatively. This study was performed with the same CI device and shared the first period of measurements as the presented in this paper. Overall, measurements started with a mean 7.9 kΩ intraoperatively and showed an average decrease of 1.9 kΩ at the activation day and a subsequent rise reaching 8.9 kΩ after 8 weeks. Interestingly, the initial impedance drop at the activation day was substantially higher than that observed in our data (mean of 200 Ω; **Figure 4**). This could be associated with the difference of electrical current provided between studies, since we delivered sub-threshold stimulation which potentially reduced the polarization effect on the inner ear medium. Moreover, Hu et al. reported that 28 days postoperatively the group showed an average of 8.7 kΩ while in our case values reached a mean of 13.6 kΩ. We also argue that this difference could also be due to the interaction of the natural inflammatory process (observed in this study) with the increasing electrical stimulation provided after CI activation.

In conclusion, the method in this paper could be of potential use to better understand the different factors that can play a role on the impedance dynamics over time by offering two main advantages: increased amount of data and measurement simplicity for the CI users and centers.

CONCLUSION

This work overviews a flexible and configurable software platform for CI users, which allows self-measures of CI impedance. The outcome enables a remote check of CI status, substantially reducing patients' clinical appointments. All patients performed the measurements in a very short time and without complications. This novel approach can be used to quickly relate a change in the objective measures with a clinical manifestation. Further advances in the method to fully automate measurements are required.

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/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Hospital Italiano, Buenos Aires, Argentina. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

FD, MP, and FF designed the methodological approach and collected the data. SA performed the data analysis. FD, MP, FF, and SA wrote the articule. CB supervised the findings and revised final manuscript. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We thank Diego De Marco for his assistance in software development. We also thank Tecnosalud, the Cochlear representative in Argentina, for the hardware support.

REFERENCES

1. Wilson BS, Dorman MF. Cochlear implants: a remarkable past and a brilliant future. *Hear Res.* (2008) 242:3–21. doi: 10.1016/j.heares.2008.06.005
2. Naples JG, Ruckenstein MJ. Cochlear implant. *Otolaryngol Clin North Am.* (2020) 53:87–102. doi: 10.1016/j.otc.2019.09.004
3. Paasche G, Bockel F, Tasche T, Lesinski-Schiedat A, Lenarz T. Changes of postoperative impedances in cochlear implant patients: the short-term effects of modified electrode surfaces and intracochlear corticosteroids. *Otol Neurotol.* (2006) 27:639–47. doi: 10.1097/01.mao.0000227662.88840.61
4. Xu J, Shepherd RK, Millard RE, Clark GM. Chronic electrical stimulation of the auditory nerve at high stimulus rates: a physiological and histopathological study. *Hear Res.* (1997) 105:1–29. doi: 10.1016/S0378-5955(96)00193-1
5. Wolfe J, Schafer EC. Programming cochlear devices. In: *Programming Cochlear Implants.* San Diego, CA: Plural Publishing (2015). p. 167–83.
6. Hughes ML. *Objective Measures in Cochlear Implants*. San Diego, CA: Plural Pub (2013).
7. Choi J, Payne MR, Campbell LJ, Bester CW, Newbold C, Eastwood H, et al. Electrode impedance fluctuations as a biomarker for inner ear pathology after cochlear implantation. *Otol Neurotol.* (2017) 38:1433–9. doi: 10.1097/MAO.0000000000001589
8. Hughes ML, Vander Werff KR, Brown CJ, Abbas PJ, Kelsay DM, Teagle HF, et al. A longitudinal study of electrode impedance, the electrically evoked compound action potential, and behavioral measures in nucleus 24 cochlear implant users. *Ear Hear.* (2001) 22:471–86. doi: 10.1097/00003446-200112000-00004
9. Shaul C, Bester CW, Weder S, Choi J, Eastwood H, Padmavathi KV, et al. electrical impedance as a biomarker for inner ear pathology following lateral wall and peri-modiolar cochlear implantation. *Otol Neurotol.* (2019) 40:e518–26. doi: 10.1097/MAO.0000000000002227
10. McElveen JT Jr, Blackburn EL, Green JD Jr, McLear PW, Thimsen DJ, Wilson BS. Remote programming of cochlear implants: a telecommunications model. *Otol Neurotol.* (2010) 31:1035–40. doi: 10.1097/MAO.0b013e3181d35d87
11. Rodríguez C, Ramos A, Falcon JC, Martínez-Beneyto P, Gault A, Boyle P. Use of telemedicine in the remote programming of cochlear implants. *Cochlear Implants Int.* (2010) 11(Suppl. 1):461–4. doi: 10.1179/146701010X12671177204624
12. Wesarg T, Wasowski A, Skarzynski H, Ramos A, Falcon Gonzalez JC, Kyriafinis G, et al. Remote fitting in nucleus cochlear implant recipients. *Acta Otolaryngol.* (2010) 130:1379–88. doi: 10.3109/00016489.2010.492480
13. Goehring JL, Hughes ML, Baudhuin JL, Valente DL, McCreery RW, Diaz GR, et al. The effect of technology and testing environment on speech perception using telehealth with cochlear implant recipients. *J Speech Lang Hear Res.* (2012) 55:1373–86. doi: 10.1044/1092-4388(2012/11-0358)
14. Hughes ML, Sevier JD, Choi S. Techniques for remotely programming children with cochlear implants using pediatric audiological methods via telepractice. *Am J Audiol.* (2018) 27:385–90. doi: 10.1044/2018_AJA-IMIA3-18-0002
15. Eikelboom RH, Jayakody DM, Swanepoel DW, Chang S, Atlas MD. Validation of remote mapping of cochlear implants. *J Telemed Telecare.* (2014) 20:171–7. doi: 10.1177/1357633X14529234
16. Goehring JL, Hughes ML. Measuring sound-processor threshold levels for pediatric cochlear implant recipients using conditioned play audiometry via telepractice. *J Speech Lang Hear Res.* (2017) 60:732–40. doi: 10.1044/2016_JSLHR-H-16-0184
17. Kuzovkov V, Yanov Y, Levin S, Bovo R, Rosignoli M, Eskilsson G, et al. Remote programming of MED-EL cochlear implants: users' and professionals' evaluation of the remote programming experience. *Acta Otolaryngol.* (2014) 134:709–16. doi: 10.3109/00016489.2014.892212
18. Schepers K, Steinhoff H-J, Ebenhoch H, Böck K, Bauer K, Rupprecht L, et al. Remote programming of cochlear implants in users of all ages. *Acta Otolaryngol.* (2019) 139:251–7. doi: 10.1080/00016489.2018.1554264
19. Health Information Privacy. *HHS.gov.* (2020) Available online at: https://www.hhs.gov/hipaa/index.html (accessed October 1, 2020).
20. Rescorla E. *HTTP over TLS.* (2000). Available online at: https://tools.ietf.org/html/rfc2818#page-6 (accessed October 1, 2020).

21. Wolfe J. *Cochlear Implants: Audiologic Management and Considerations for Implantable Hearing Devices*. San Diego, CA: Plural Publishing, Incorporated (2017).
22. Di Lella FA, De Marco D, Fernández F, Parreño M, Boccio CM. *In vivo* real-time remote cochlear implant capacitive impedance measurements: a glimpse into the implanted inner ear. *Otol Neurotol.* (2019) 40:S18–22. doi: 10.1097/MAO.0000000000002214
23. Newbold C, Mergen S, Richardson R, Seligman P, Millard R, Cowan R, et al. Impedance changes in chronically implanted and stimulated cochlear implant electrodes. *Cochlear Implants Int.* (2014) 15:191–9. doi: 10.1179/1754762813Y.0000000050
24. Wolfe J, Baker RS, Wood M. Clinical case study review: steroid-responsive change in electrode impedance. *Otol Neurotol.* (2013) 34:227–32. doi: 10.1097/MAO.0b013e31827b4bba
25. Bas E, Bohorquez J, Goncalves S, Perez E, Dinh CT, Garnham C, et al. Electrode array-eluted dexamethasone protects against electrode insertion trauma induced hearing and hair cell losses, damage to neural elements, increases in impedance and fibrosis: a dose response study. *Hear Res.* (2016) 337:12–24. doi: 10.1016/j.heares.2016.02.003
26. Wilk M, Hessler R, Mugridge K, Jolly C, Fehr M, Lenarz T, et al. Impedance changes and fibrous tissue growth after cochlear implantation are correlated and can be reduced using a dexamethasone eluting electrode. *PLoS ONE.* (2016) 11:e0147552. doi: 10.1371/journal.pone.0147552
27. Tykocinski M, Duan Y, Tabor B, Cowan RS. Chronic electrical stimulation of the auditory nerve using high surface area (HiQ) platinum electrodes. *Hear Res.* (2001) 159:53–68. doi: 10.1016/S0378-5955(01)00320-3
28. Newbold C, Richardson R, Huang CQ, Milojevic D, Cowan R, Shepherd R. An *in vitro* model for investigating impedance changes with cell growth and electrical stimulation: implications for cochlear implants. *J Neural Eng.* (2004) 1:218–27. doi: 10.1088/1741-2560/1/4/005
29. Busby PA, Plant KL, Whitford LA. Electrode impedance in adults and children using the nucleus 24 cochlear implant system. *Cochlear Implants Int.* (2002) 3:87–103. doi: 10.1179/cim.2002.3.2.87
30. Clark GM, Shute SA, Shepherd RK, Carter TD. Cochlear implantation: osteoneogenesis, electrode-tissue impedance, and residual hearing. *Ann Otol Rhinol Laryngol Suppl.* (1995) 166:40–2.
31. Henkin Y, Kaplan-Neeman R, Muchnik C, Kronenberg J, Hildesheimer M. Changes over time in electrical stimulation levels and electrode impedance values in children using the nucleus 24M cochlear implant. *Int J Pediatr Otorhinolaryngol.* (2003) 67:873–80. doi: 10.1016/S0165-5876(03)00131-9
32. Henkin Y, Kaplan-Neeman R, Kronenberg J, Migirov L, Hildesheimer M, Muchnik C. A longitudinal study of electrical stimulation levels and electrode impedance in children using the Clarion cochlear implant. *Acta Otolaryngol.* (2006) 126:581–6. doi: 10.1080/00016480500443391
33. Zeng F-G, Rebscher SJ, Fu Q-J, Chen H, Sun X, Yin L, et al. Development and evaluation of the nurotron 26-electrode cochlear implant system. *Hear Res.* (2015) 322:188–99. doi: 10.1016/j.heares.2014.09.013
34. De Ceulaer G, Johnson S, Yperman M, Daemers K, Offeciers FE, O'Donoghue GM, et al. Long-term evaluation of the effect of intracochlear steroid deposition on electrode impedance in cochlear implant patients. *Otol Neurotol.* (2003) 24:769–74. doi: 10.1097/00129492-200309000-00014
35. Dorman MF, Smith LM, Dankowski K, McCandless G, Parkin JL. Long-term measures of electrode impedance and auditory thresholds for the Ineraid cochlear implant. *J Speech Hear Res.* (1992) 35:1126–30. doi: 10.1044/jshr.3505.1126
36. Saunders E, Cohen L, Aschendorff A, Shapiro W, Knight M, Stecker M, et al. Threshold, comfortable level and impedance changes as a function of electrode-modiolar distance. *Ear Hear.* (2002) 23:28S—40. doi: 10.1097/00003446-200202001-00004
37. Hu H-C, Chen JK-C, Tsai C-M, Chen H-Y, Tung T-H, Li LP-H. Evolution of impedance field telemetry after one day of activation in cochlear implant recipients. *PLoS ONE.* (2017) 12:e0173367. doi: 10.1371/journal.pone.0173367

Conflict of Interest: The authors declare that this study received equipment from Cochlear Ltd. They were not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

frontiers in Medicine

ORIGINAL RESEARCH
published: 11 January 2021
doi: 10.3389/fmed.2020.613708

Automatic Recognition of Auditory Brainstem Response Characteristic Waveform Based on Bidirectional Long Short-Term Memory

Cheng Chen[1†], *Li Zhan*[2†], *Xiaoxin Pan*[1], *Zhiliang Wang*[1], *Xiaoyu Guo*[1], *Handai Qin*[2], *Fen Xiong*[2], *Wei Shi*[2], *Min Shi*[2], *Fei Ji*[2], *Qiuju Wang*[2], *Ning Yu*[2*] and *Ruoxiu Xiao*[1,3*]

[1] School of Computer and Communication Engineering, University of Science & Technology Beijing, Beijing, China, [2] College of Otolaryngology Head and Neck Surgery, National Clinical Research Center for Otolaryngologic Diseases, Key Lab of Hearing Science, Ministry of Education, Beijing Key Lab of Hearing Impairment for Prevention and Treatment, Chinese PLA General Hospital, Beijing, China, [3] Institute of Artificial Intelligence, University of Science and Technology Beijing, Beijing, China

OPEN ACCESS

Edited by:
Changxin Zhang,
East China Normal University, China

Reviewed by:
Chun Liang,
Shenzhen Maternity and Child Healthcare Hospital, China
Kirupa Suthakar,
National Institute on Deafness and Other Communication Disorders (NIDCD), United States

***Correspondence:**
Ning Yu
yuning@301hospital.org
Ruoxiu Xiao
xiaoruoxiu@ustb.edu.cn

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Translational Medicine, a section of the journal Frontiers in Medicine

Received: 03 October 2020
Accepted: 03 December 2020
Published: 11 January 2021

Citation:
Chen C, Zhan L, Pan X, Wang Z, Guo X, Qin H, Xiong F, Shi W, Shi M, Ji F, Wang Q, Yu N and Xiao R (2021) Automatic Recognition of Auditory Brainstem Response Characteristic Waveform Based on Bidirectional Long Short-Term Memory. Front. Med. 7:613708. doi: 10.3389/fmed.2020.613708

Background: Auditory brainstem response (ABR) testing is an invasive electrophysiological auditory function test. Its waveforms and threshold can reflect auditory functional changes in the auditory centers in the brainstem and are widely used in the clinic to diagnose dysfunction in hearing. However, identifying its waveforms and threshold is mainly dependent on manual recognition by experimental persons, which could be primarily influenced by individual experiences. This is also a heavy job in clinical practice.

Methods: In this work, human ABR was recorded. First, binarization is created to mark 1,024 sampling points accordingly. The selected characteristic area of ABR data is 0–8 ms. The marking area is enlarged to expand feature information and reduce marking error. Second, a bidirectional long short-term memory (BiLSTM) network structure is established to improve relevance of sampling points, and an ABR sampling point classifier is obtained by training. Finally, mark points are obtained through thresholding.

Results: The specific structure, related parameters, recognition effect, and noise resistance of the network were explored in 614 sets of ABR clinical data. The results show that the average detection time for each data was 0.05 s, and recognition accuracy reached 92.91%.

Discussion: The study proposed an automatic recognition of ABR waveforms by using the BiLSTM-based machine learning technique. The results demonstrated that the proposed methods could reduce recording time and help doctors in making diagnosis, suggesting that the proposed method has the potential to be used in the clinic in the future.

Keywords: auditory brainstem response, characteristic waveform recognition, neural network model, bi-directional long short-term memory, wavelet transform

INTRODUCTION

Auditory brainstem response (ABR) is a global neural activity in the auditory brainstem centers evoked by acoustic stimulations. It can observe the functional status of the auditory nerve and lower auditory center and reflect the conduction ability of the brainstem auditory pathway (1, 2). Given that patient's hearing impairment can be diagnosed without his active cooperation, ABR has become one of the routine methods for adult hearing recording (3–5). The ABR waveform usually has a range of interwave latency, and its potential in microvolts is recorded. Normal ABR usually has five peaks visible, i.e., waves I, II, III, IV, and V. Wave V usually appears as the largest peak in the ABR. In clinical diagnosis, the minimum intensity of sound stimulation to be capable of evoking a recognized ABR is defined as ABR threshold, which is usually dependent on wave V or wave III (6, 7). **Figure 1** shows the annotated ABR waveforms, which are mainly identified as waves I, III, and V clinically. Other characteristic waves are usually not displayed clearly because of small amplitude, two-wave fusion, and noise interference. Thus, they are rarely used as a basis for diagnosis.

In clinical diagnosis, the minimum stimulation intensity of wave V is usually used as ABR threshold. Sometimes, when wave III is greater than wave V, the ABR threshold is judged by stimulation intensity of wave III (8). In determining lesions, the location can be judged according to the interwave latency of waves I, III, and V and the interwave latency between waves and binaural waves (9). Furthermore, the types of deafness of a patient can be judged by observing the change characteristics of ABR waveform latency and the special shape of the ABR waveform in the same patient under different stimulation levels. Thus, the ABR threshold and interwave latency of waves I, III, and V, which are of great significance in clinical applications, can be obtained by identifying the position of the characteristic wave of ABR. Usually, the potential obtained from each stimulation is weak. In a clinical testing, multiple stimulations must be performed to superimpose, average, and obtain relatively stable waveform results. This process is susceptible to interference by electrical noise arising from stray myogenic potentials or movement artifact. In addition, performing multiple tests on patients and comparing results to avoid unobvious peaks, overlapping peaks, and false peaks, which not only consume a lot of time but are also prone to subjective judgment errors, are usually necessary. Thus, identifying the waveform characteristics of ABR and avoiding interference caused by unclear differentiation, fuzzy characteristics, and abnormal waveforms are important issues that need to be solved urgently and correctly in clinical ABR recording.

The application of computer technology in assisting medical diagnosis can effectively reduce errors caused by repetitive work and complex waveform characteristics. This research direction has been important for ABR consultation for a long time (10). For example, Wilson (11) discussed the relationship between ABR and discrete wavelet transform reconstructed waveforms, indicating that the discrete wavelet transform waveform of ABR can be used as an effective time–frequency representation of normal ABR but with certain limitations. Especially in some

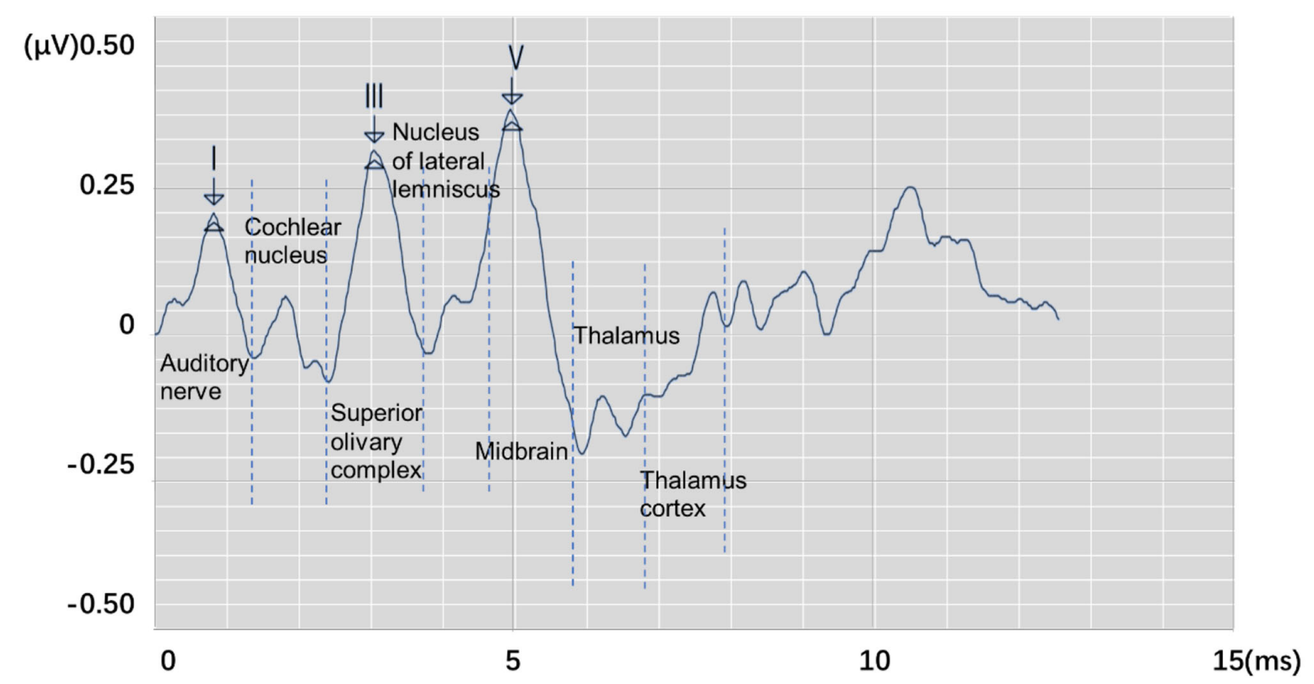


FIGURE 1 | The annotated ABR waveform (legend data is selected from the datasets applied in this work).

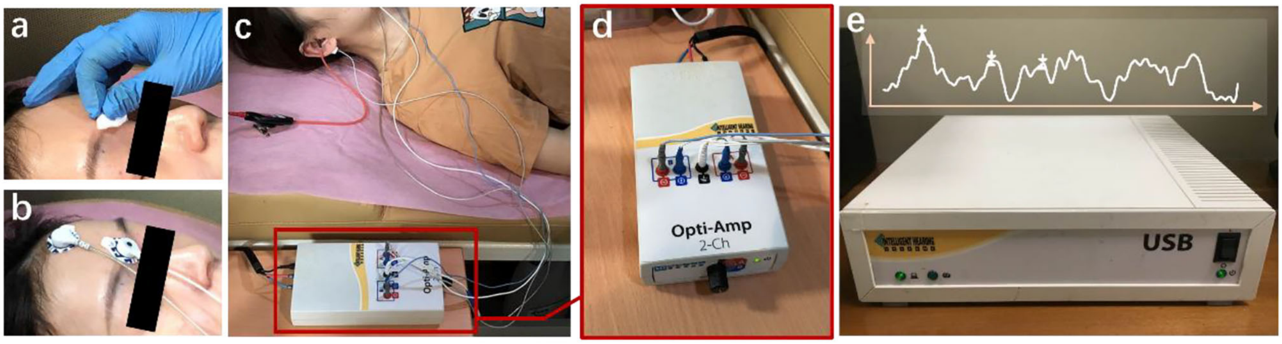


FIGURE 2 | The ABR hearing diagnosis clinical collection process. **(a)** Skin degreasing to enhance conductivity; **(b)** the position of the forehead and earlobe electrodes; **(c)** the positional relationship diagram of the preamplifier, electrodes, and plug-in earphones; and **(d)** the details of the preamplifier. The collected waveform is stored in a server **(e)** and can be observed with the monitor.

cases, the reconstructed ABR discrete wavelet transform wave is missing because of the invariance of discrete wavelet transform shift. Bradly and Wilson (12) further studied the method of using derivative wavelet estimation to automatically analyze ABR, which improved the accuracy of the main wave identification to a high level. However, they also mentioned the need for further research on the performance of waveform recognition of abnormal subjects, and manual judgment of abnormal waveforms is still required under clinical conditions. Zhang et al. (13) proposed an ABR classification method that combined wavelet transform and Bayesian network to reduce the number of stimulus repetitions and avoid nerve fatigue of the examinee. Important features are extracted through image thresholding and wavelet transform. Subsequently, features were applied as variables to classify using Bayesian networks. Experimental results show that the ABR data with only 128 repetitive stimulations can achieve an accuracy of 84.17%. Compared with the clinical test that usually requires 2,000 repetitions, the detection efficiency of ABR is improved greatly. However, wave I and wave V are always prolonged by about 0.1 ms and cause wave range changes. Therefore, III–V/I–III would be inaccurate as an indicator.

Thus, automatic recognition of ABR waveforms through computer-assisted methods can assist clinicians and audiologists in ABR interpretation effectively. It also reduces the errors caused by subjective factors, the interference of complex waveforms, and the burden of a large number of repetitive tasks for the medical staff. This study proposes a method of using the long short-term memory (LSTM) network to identify waves I, III, and V in the ABR waveform and proposes a new idea for the recognition of ABR characteristic waveforms by neural networks. The structure of the study is organized as follows: The experimental data and the detailed description of the proposed method are presented in the Materials and Methods section. The Results section presents the experimental design and the corresponding results. Finally, the Discussion section provides an elaboration of the findings of this work.

MATERIALS AND METHODS

Data Source

The data are provided by the Department of Otolaryngology Head and Neck Surgery, Chinese PLA General Hospital. The SmartEP evoked potential test system developed by the American Smart Listening Company is used for measurement and acquisition. **Figure 2** shows the clinical collection process, where **Figure 2a** represents skin degreasing to enhance conductivity; **Figure 2b** represents the position of the forehead and earlobe electrodes; **Figure 2c** represents the positional relationship diagram of the preamplifier, electrodes, and plug-in earphones; and **Figure 2d** shows the details of the preamplifier. The collected waveform is stored in a server **Figure 2e** and can be observed with the monitor. Six hundred and fourteen subjects' clinical click stimuli ABR data were collected at 96 dB nHL stimulation intensity after 1,024 repeated stimulations, which contain 181 normal and 433 abnormal hearing. The clinical dataset comprises 348 men and 266 women aged 18 to 90 years old. For data structure, the data contain 1,024 sampling points that range from −12.78 to 12.80 ms with an average interval of 0.025 ms between every two sampling points. All data were marked by three clinical audiologists with characteristic waves: wave I, wave III, and wave V, and cross-validated. Finally, the data were randomly divided into training and test sets. A total of 491 training sets were used to train the network model, and 123 test sets were used for the final recognition accuracy test.

Data Processing

In this work, a new data processing method is proposed. To quantify waveform and label points, two $1,024 \times 1$ matrices A and B were generated as the classification train and label, respectively. A represents the potential of the input ABR data. The position of the serial number corresponds to the position of the ABR data sampling point. B represents nonfeature (0) and feature points (1), respectively. Thus, according to the position of the label value of the label data, the data that corresponded to the position of the label matrix was changed to 1 to meet the binary

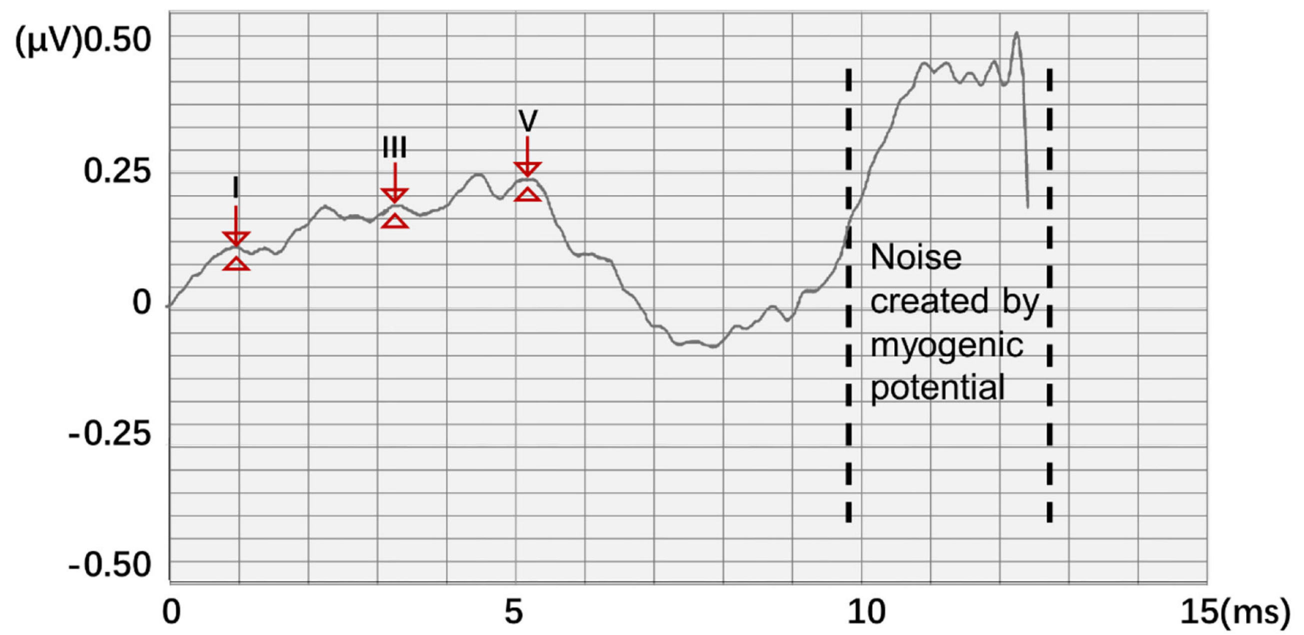


FIGURE 3 | Abnormal ABR waveform and data quantization method.

classification requirements of all sampling points. However, noise created by myogenic potential is observed in some experimental data (**Figure 3**). In this ABR clinical test data, the ABR waveform has an unusual increase in the sampling point at the end because of the fluctuation of characteristic waves VI and VII and the result of the external interference. To prevent the interference caused by abnormal data, the data up to 8 ms were selected uniformly to identify the characteristic waves.

On the other hand, the starting point of the actual stimulation is 0 ms. The final potential value input data and the corresponding training label both retained only 321 sampling points of 0–8 ms to avoid interference with neural network training and reduce the amount of calculation in the neural network training process. Thus, A and b_f are updated as follows:

$$\begin{cases} A\,(321) = \{y_1, y_2, ..., y_{321}\}^T \\ B\,(321) = \{t_1, t_2, ..., t_{321}\}^T \end{cases} \tag{1}$$

In actual processing, the loss function value can easily reach a low level, and sufficient information cannot be learned because the ratio of the labeled value to the unlabeled value in the 321 sample points is only 3:318. The manually labeled information may also bring certain errors. Thus, this study adopted the method of augmenting the position of the identification point in the training label. The four points (0.1 ms) before and after the original marking point were marked as the characteristic area, which expands the marking range of the characteristic waveform.

Network Structure

LSTM is a recurrent neural network and mainly improved on the basis of the time step unit by adding the output of memory cells to carry information that needs to be transmitted for a long time. Three gate structures are also added. These gate structures are used to select the retention of the memory cell C_{t-1} value passed from the previous time step, add new information into the memory cell V, and predict and output the information transmitted by the memory cell and continue to pass it to the next time step.

Figure 4 is a schematic diagram of the LSTM structure. First, to control the proportion of the input information retained by the memory cells at the previous time step, the output is calculated as follows:

$$f_t = \sigma\left(W_f h_{t-1} + U_f x_t + b_f\right) \tag{2}$$

h_{t-1} is the hidden state value passed at the previous time step; and W_f, ..., and b_f are the corresponding weights and biases. The activation function usually uses the sigmoid function to map the activation value between [0, 1]. To control the proportion of information updated into the memory cell, the sigmoid activation function was first applied to obtain the output i_i. Then, the tan$h$ activation function is applied to obtain, and the product of the two is used as the information to update the memory cell. i_t and a_t are calculated as follows:

$$i_t = \sigma\left(W_i h_{t-1} + U_i x_t + b_i\right) \tag{3}$$

$$a_t = \tanh\left(W_a h_{t-1} + U_a x_t + b_a\right) \tag{4}$$

where W_i, U_i, b_i, W_a, U_a, and b_a are the weights and biases. Finally, the memory cell C_t is calculated to the next time step by using Equation (5):

$$C_t = C_{t-1} \odot f_t + i_t \odot a_t \tag{5}$$

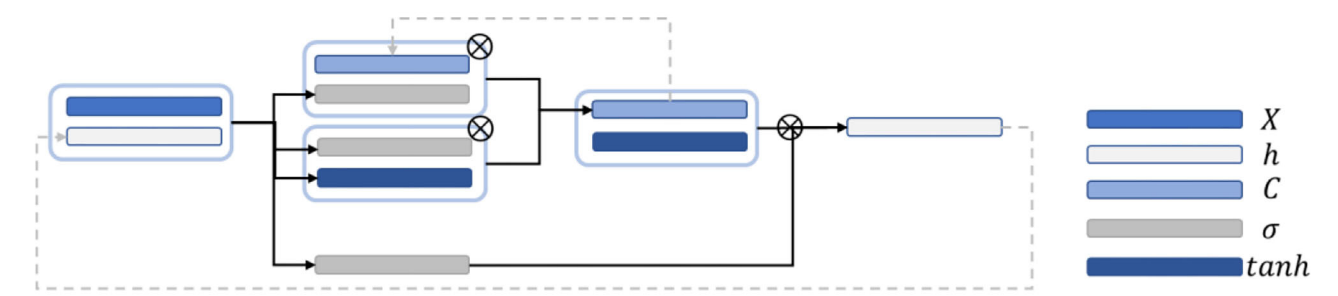


FIGURE 4 | Schematic diagram of the LSTM network structure.

where $\odot$ is the Hadamard product, which indicates that the corresponding positions of the matrix are multiplied. The right side refers to the output gate, and the output of the output gate is calculated by using Equation (6):

$$o_t = \sigma\left(W_o h_{t-1} + U_o x_t + b_o\right) \tag{6}$$

where W_o, U_o, and b_o are the weights and offsets. Finally, the output value h_t at the time step is obtained through using Equation (7):

$$h_t = o_t \odot \tanh(C_t) \tag{7}$$

The predicted output weight and bias are applied to activate the output value to obtain the predicted value, as shown in Equation (8):

$$\hat{y}_t = \sigma\left(V h_t + c\right) \tag{8}$$

Finally, the loss values δ_h^t and δ_C^t of the hidden state are calculated as follows:

$$\delta_h^t = V^T\left(\hat{y}_t - y_t\right) + \left(\frac{\partial h_{t+1}}{\partial h_t}\right)^T \delta_h^{t+1} \tag{9}$$

$$\delta_C^t = \delta_C^{t+1} \odot f_{t+1} + \delta_h^t \odot o_t \odot \left(1 - \tanh^2\left(C_t\right)\right) \tag{10}$$

In this work, BiLSTM is established as the network structure to enable the input sequence to have a bidirectional connection with one another (14). **Figure 5** shows that another LSTM layer that propagates backward in time is added on the basis of the unidirectional LSTM forward propagation in time sequence. The final output is determined by the output of the two LSTM layers: forward and backward. Compared with the one-way LSTM, the final output avoids the prediction at each time to only be affected by the input of the previous time. Moreover, it can reflect the information characteristics before and after each prediction point better, thereby making more accurate predictions.

Wavelet Transform

In the traditional mode, wavelet transform is a commonly used method in ABR extraction and recognition research (15). In ABR extraction, wavelet transform can achieve the effect of eliminating noise by selecting the detailed components of specific frequencies for reconstruction and to make the ABR waveform smoother. Obtaining relatively clear waveforms while reducing repetitive stimulation is also possible. Generally, continuous wavelet transform is defined as (16):

$$WT(a,\tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) * \psi\left(\frac{t-\tau}{a}\right) \mathrm{d}t \tag{11}$$

where $f(t)$ is the signal in the time domain, and the part of $\frac{1}{\sqrt{a}}\psi\left(\frac{t-\tau}{a}\right)$ is a wavelet function, which can also be denoted as $\psi_{a,\tau}(t)$. Two variables, namely, scale a and translation τ, are available. Scale a is applied to control the expansion and contraction of the wavelet function, and the translation amount τ controls the translation of the wavelet function. Scale a is inversely proportional to its equivalent frequency, which is defined as $\varphi(t)$. The complete wavelet expansion is as follows:

$$f(t) = \sum_{k=-\infty}^{\infty} c_k \varphi(t-k) + \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} d_{j,k} \psi\left(2^j t - k\right) \tag{12}$$

where c and d are the coefficients of the corresponding function, j is the frequency domain parameter that determines the frequency characteristics of the wavelet, and k is the time domain parameter that controls the position of the wavelet base in the time domain. Although the scale and wavelet functions are complex and have different characteristics, the process of wavelet decomposition can be regarded as using a low-pass filter and a high-pass filter to decompose the signal by frequency. The low-frequency components decomposed in each layer are called approximate components, and the high-frequency components are called detailed components. Thus, approximate components and detailed components were applied to the reconstructed waveform.

RESULTS

Experimental Procedure

In this study, three sets of experiments, namely, (1) comparison between various network structures, (2) comparison experiment of wavelet transform, and (3) comparison experiment of different hidden layer nodes, were designed. **Figure 6** shows the experimental flowchart. The sequence input layer was used

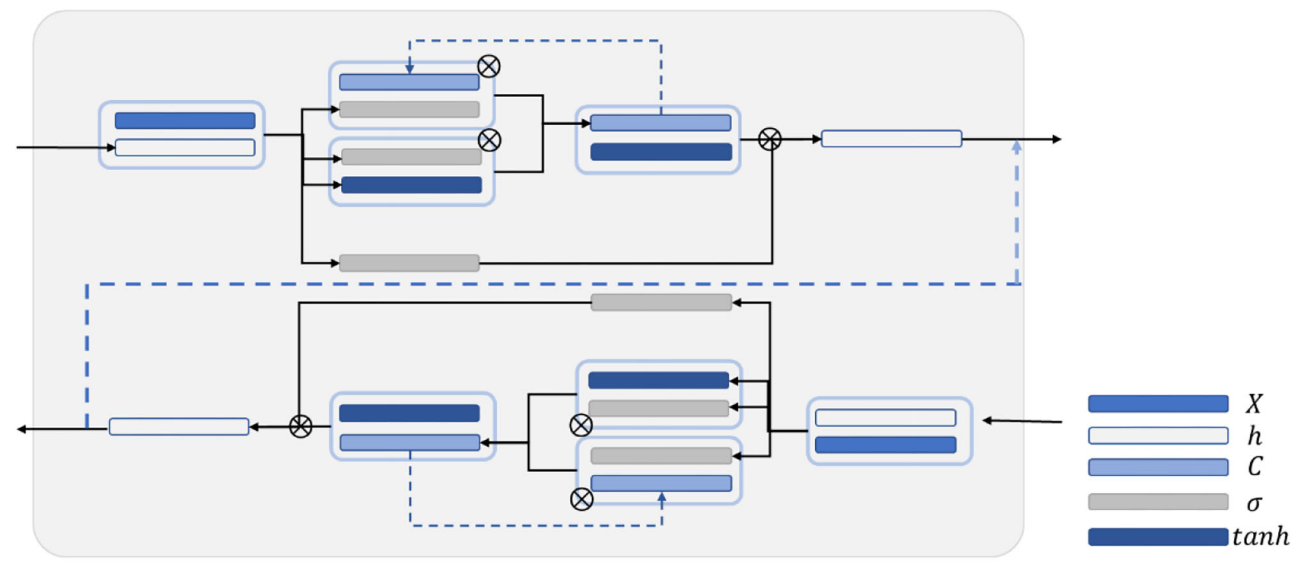


FIGURE 5 | Schematic diagram of the BiLSTM structure.

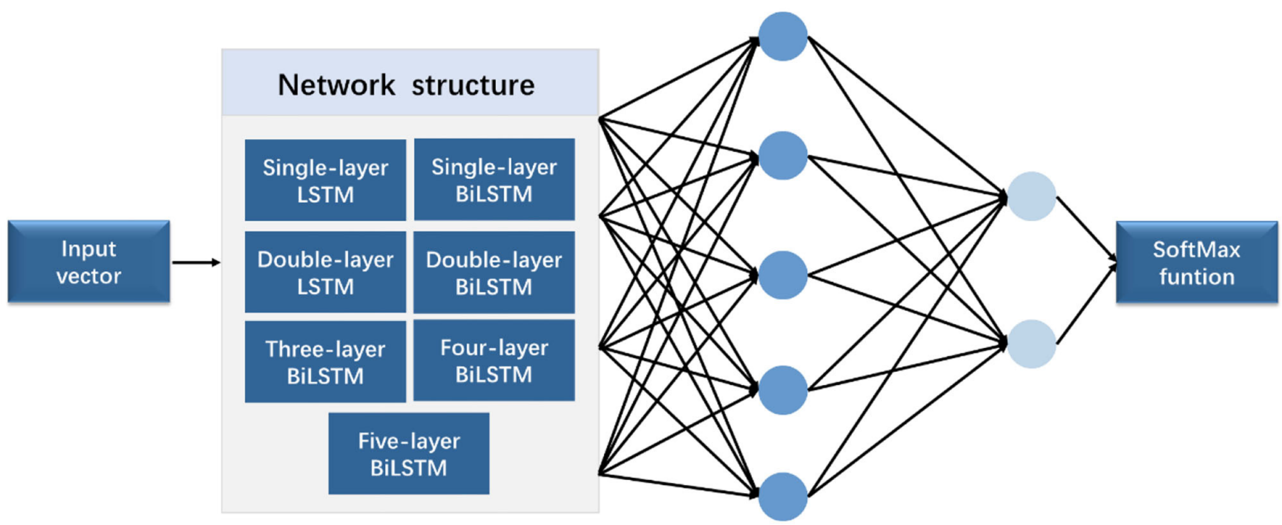


FIGURE 6 | Experimental flowchart.

as the input of the potential value of 321 sampling points, and the data were passed to several LSTM or BiLSTM layers. Subsequently, the fully connected layer was connected. The classification probability of each time point was calculated using the softmax function. Finally, the classification layer was connected. The cross-entropy function (17) was used to calculate the loss function of each time point and the overall loss function of the sequence. Then, the time sequence was classified.

In the comparison experiment of multiple network structures, seven network structures, namely, (1) single-layer LSTM, (2) double-layer LSTM, (3) single-layer BiLSTM, (4) double-layer BiLSTM, (5) three-layer BiLSTM, (6) four-layer BiLSTM, and (7) five-layer BiLSTM network layers, were selected. In the comparative experiment of different hidden layer nodes, a three-layer bidirectional LSTM network was used for training, and different numbers of hidden neurons were applied. The experiment applied four groups of different numbers of hidden neurons, namely, 64, 128, 256, and 512.

In the comparative experiment of the wavelet transform, all data added noise as interference. Seven different network structures were used for testing. For instance, the training data preprocessed by wavelet transform were used as the experimental group, and the training data trained using the original data were used as the control group. In this experiment, ABR data were decomposed in six layers, and the approximate and detailed components of the sixth layer and the fourth, fifth, and sixth layers were retained to reconstruct the waveform, respectively. The parameter configuration is consistent. The network was trained with five K-fold cross-validation ($K = 9$), and the test was performed to obtain the average value.

The output results are in the form of "region." **Figure 7** expresses the output visualization, where the curve is the original ABR used for identification, and the red labels are the network prediction classification results reduced by four times. The ABR of the first 8 ms is clearly divided into two different labels. The part with 1 is the identified peak, and the other part is

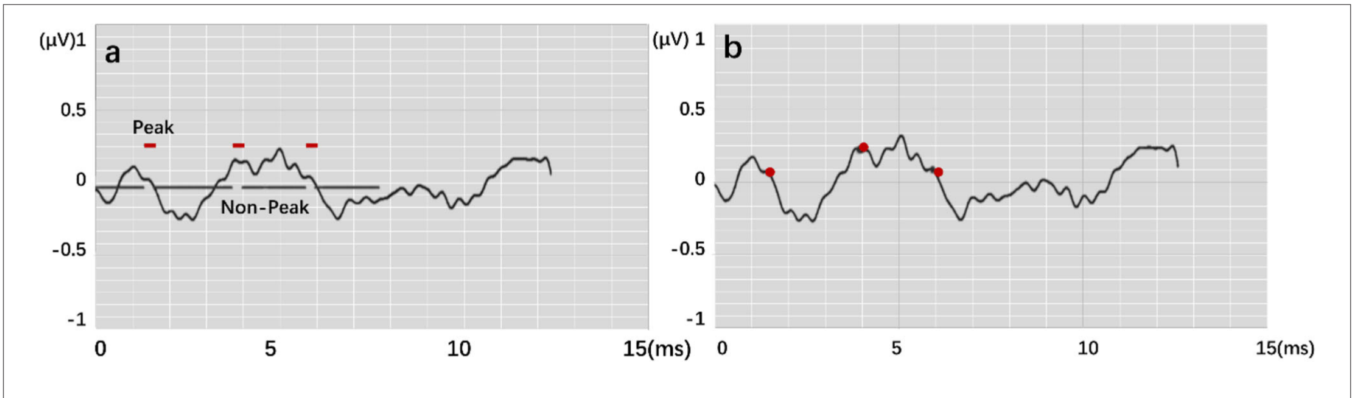


FIGURE 7 | Feature labeling on the ABR, where **(a)** shows output by modes. **(b)** is result by postprocessing.

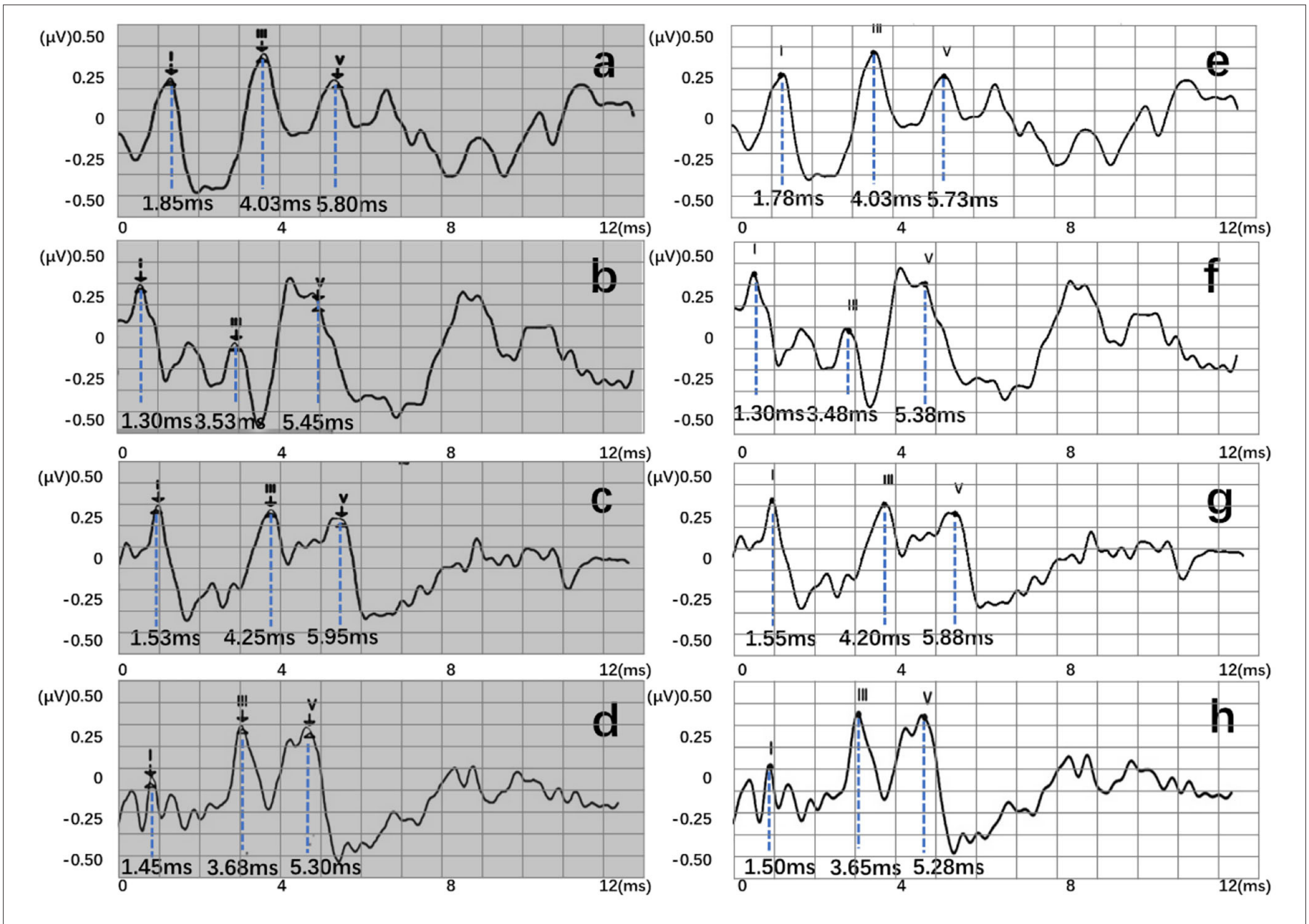


FIGURE 8 | Recognition results of four data, where **(a–d)** are manual labels. Also, **(e–h)** represent outputs of the proposed three-layer BiLSTM model.

the identified characteristic nonpeak. Postprocessing is defined as follows: A total of 20 sampling points (0.5 ms) are set as the threshold. The area within 20 sampling points between the beginning and the end is the same characteristic wave area. Finally, the time mean value of the first and last points is calculated as the time value of the recognized characteristic wave. The similar sampling points are calculated to obtain the unique characteristic wave value. Finally, the recognition accuracy rate is calculated according to the identified ABR feature wave position.

Four recognition results of ABR data were randomly selected and presented in **Figure 8**. After postprocessing, output vectors from models were converted to feature points. The identified feature points are almost identical to those selected using manual labeling techniques, illustrating the potential utility of this method in clinical settings. Even in some complex ABR data, manual annotation usually records multiple sets of data to determine the correct peak (**Figure 8d**). However, the model can directly and accurately identify the peak of the waveform from a single waveform (**Figure 8h**). Therefore, they also verify the possibility of the proposed method. To better verify the accuracy of recognition, this work has carried out a quantitative discussion from different network structures, wavelet transform processing, and number of hidden neurons. However, the model may also lead to some misjudgments. For example, **Figure 9a** shows an incorrect recognition result. Since wave I and wave III of the waveform are not obvious, enough continuous identification points cannot be obtained. Therefore, only relatively obvious wave V is obtained after postprocessing (**Figure 9c**). Also, **Figure 9b** presents another wrong result. In this case, the obtained error of wave I reached 0.67 ms. This is because the model has judged the wrong wave I (**Figure 9d**). Thus, in future work, improving the model's ability to analyze complex waveforms is still an important direction.

Comparison Between Multiple Network Structures

Generally, an error scale of 0.2 ms is applied as a scale range of clinically marked points. Three criterion values for the maximum allowable error value (ME) were tested: −0.1, 0.15, and 0.2 ms. The prediction result was deemed acceptable if the prediction point and the manually identified point were within the ME criteria range. According to the number of correct prediction points r_p and the total marked points p_n, the accuracy (ACC) rate is calculated using r_p/p_n, as shown in Equation (13):

$$ACC = r_p/p_n \tag{13}$$

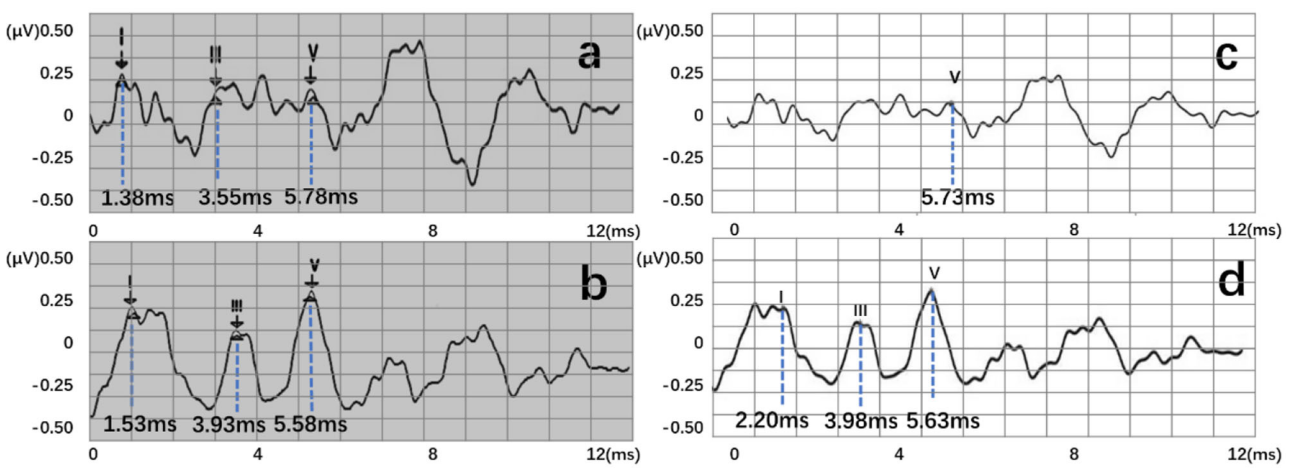


FIGURE 9 | Two error recognition results, where **(a,b)** are manual labels. Also, **(c,d)** represent outputs of the proposed three-layer BiLSTM model.

TABLE 1 | Loss value and ACC of each network structure.

Network structure	Training loss	Validation loss	Accuracy (0.1 ms) (%)	Accuracy (0.15 ms) (%)	Accuracy (0.2 ms) (%)
LSTM	0.1463	0.1635	37.08	44.92	50.37
LSTMx2	0.1123	0.1625	58.61	65.75	70.59
BiLSTM	0.1264	0.1562	61.96	72.03	77.60
BiLSTMx2	0.0849	0.1285	78.74	84.88	86.84
BiLSTMx3	0.0704	0.1275	85.46	91.06	92.91
BiLSTMx4	0.0651	0.1342	82.48	88.32	90.20
BiLSTMx5	0.0617	0.1467	83.31	88.80	90.90

In this study, three error scales (ME) of 0.1, 0.15, and 0.2 ms were calculated, respectively, to further explore the recognition accuracy and other related laws. Loss value of training results with different network structures and the ACC under different error scales are revealed in **Table 1**.

Figure 10A indicates data distribution to observe correlation with different network structures visually. Notably, the ACC of the BiLSTM network is higher than that of the LSTM network. In addition, the ACC of the single-layer BiLSTM network and the double-layer LSTM network is similar. The reason is due to the fact that the two-way LSTM network has a similar structure to the double-layer LSTM network. However, information in the BiLSTM network has the characteristics of propagating in forward and reverse directions, whereas the two-layer LSTM network only propagates in the forward sequence over time. This phenomenon leads to differences in the ACC between the two models. The LSTM and BiLSTM networks increase ACC with the number of superimposed layers. After the BiLSTM network reaches three layers, the ACC will no longer increase significantly. Network structure will gradually reach an over-fitting state and increase computational pressure because of excessive parameters. Thus, the three-layer BiLSTM network is a better choice.

Wavelet Transform Experiment

When testing the ACC of the wavelet transform, ABR data was decomposed in six layers. Also, approximate components of the sixth layer and detailed components of the fourth, fifth, and sixth layers were retained to reconstruct the waveform. **Figure 11** expresses an instance of filtered result by wavelet transform. The curve processed by wavelet transform becomes smoother. Then, unprocessed ABR data served as a control experiment. In this work, detection and comparison were carried out based on two error scales of 0.1 and 0.2 ms (**Table 2**). The results of recognition ACC are shown in **Figure 12**.

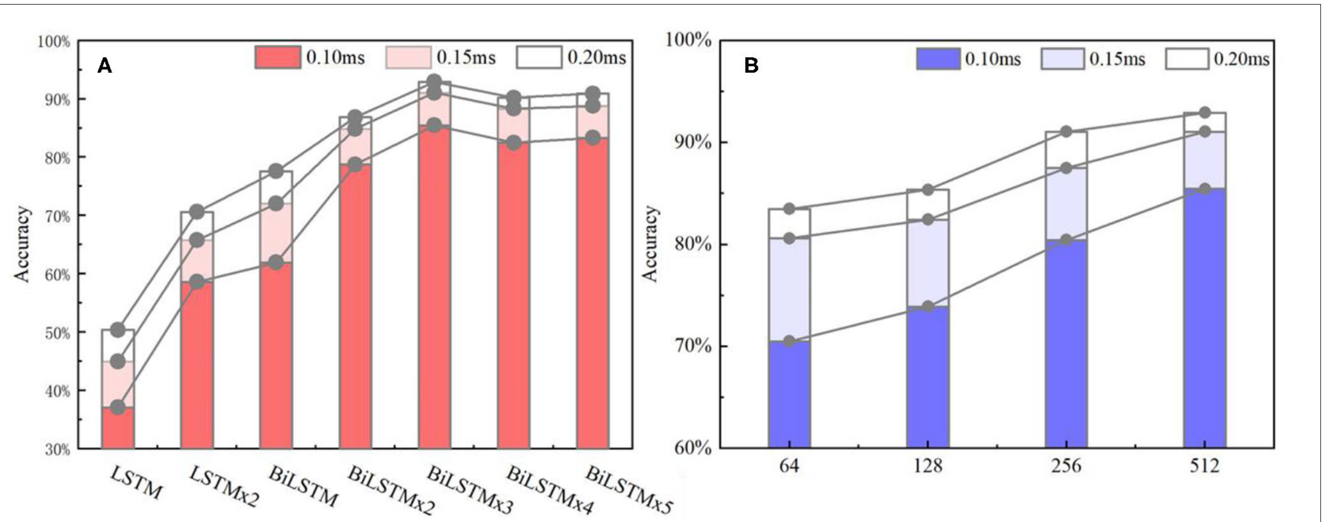


FIGURE 10 | (A) ACC metrics with different network structures. In the statistical results, the three-layer BiLSTM network reached 92.91% and is the highest index among all the networks. The single-layer LSTM, which has the lowest index, is about half of it. **(B)** ACC metrics with different hidden nodes, where the 512 nodes ranked first, and the 256 and 128 quantities stood at the second and third positions. Also, the 64 nodes ranked last.

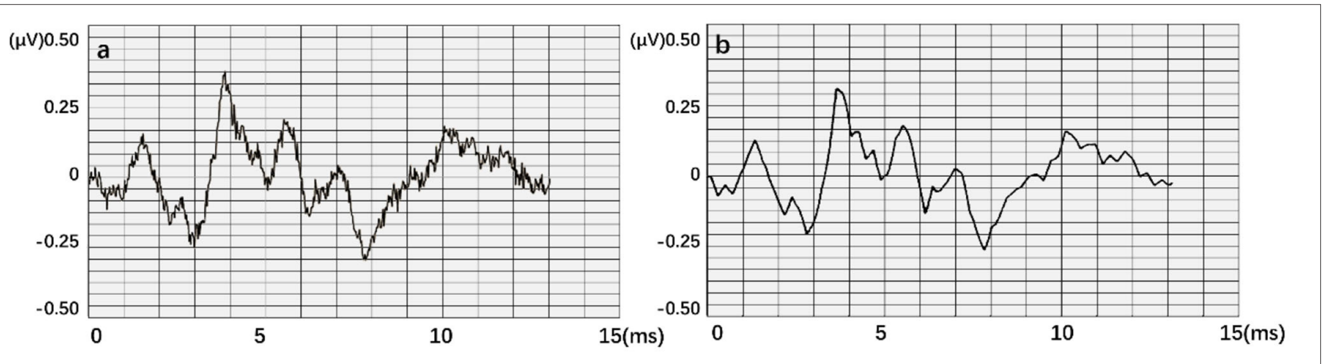


FIGURE 11 | An instance result from the wavelet transform, where **(a)** is the original data. An obvious interference occurred in this waveform. **(b)** is obtained after smoothing.

TABLE 2 | The ACC of each network structure with original data and wavelet transform data.

Network structure	Original data (0.1 ms) (%)	Wavelet transform data (0.1 ms) (%)	Original data (0.2 ms) (%)	Wavelet transform data (0.2 ms) (%)
LSTM	37.08	37.95	50.37	52.94
LSTMx2	58.61	55.47	70.59	72.46
BiLSTM	61.96	59.17	77.60	76.25
BiLSTMx2	78.74	73.03	86.84	84.71
BiLSTMx3	85.46	79.00	92.91	90.50
BiLSTMx4	82.48	77.73	90.20	89.67
BiLSTMx5	83.31	78.09	90.90	89.17

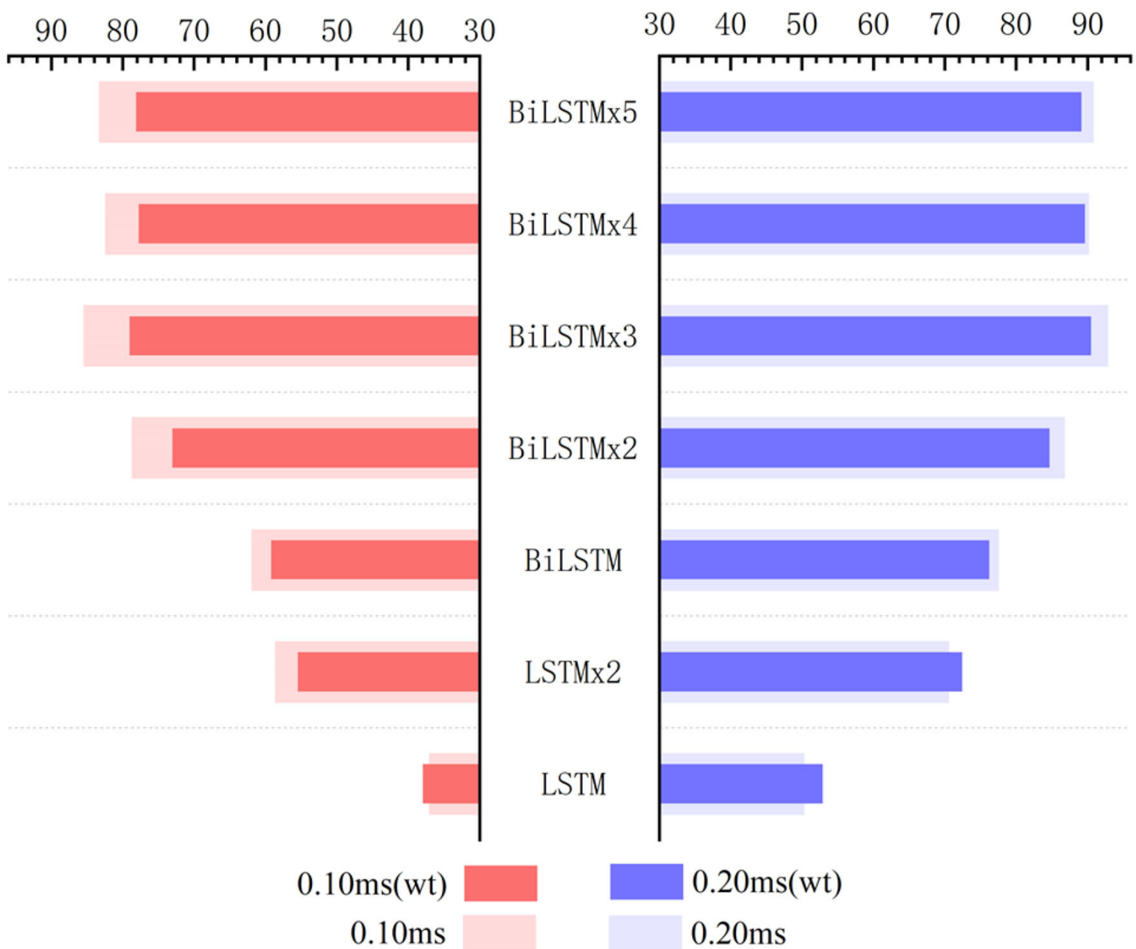


FIGURE 12 | Influence of wavelet transform preprocessing on accuracy. wt represents the results obtained by wavelet transform preprocessing.

Recognition ACC values of preprocessing in the LSTM network using wavelet transform are slightly higher than those of the control group. However, they are not as good as those in the control group in the BiLSTM network. Especially, the highest ACC difference reaches 6.46% when calculated with a 0.1-ms error scale. Also, the difference reduces to <3% when calculated with a 0.2-ms error scale. Results indicate that wavelet transform preprocessing does not obtain a higher ACC by smoothing curves. Due to wavelet decomposition and reconstruction, a slight deviation was created in the position of wave crest. Some information was destroyed in the ABR waveform; therefore, the results of training and recognition were affected. This means that the BiLSTM network has noise immunity and can handle low-quality ABR data.

TABLE 3 | The ACC with different hidden layer nodes.

Hidden layer nodes	Accuracy (0.1 ms) (%)	Accuracy (0.15 ms) (%)	Accuracy (0.2 ms) (%)
64	70.50	80.61	83.48
128	73.90	82.44	85.36
256	80.44	87.49	91.07
512	85.46	91.06	92.91

Comparative Experiments of Different Hidden Layer Nodes

Based on the above results, the three-layer BiLSTM network is a better choice. The ACC results with different hidden node numbers were discussed in this work (**Table 3**). **Figure 10B** expresses the ACC results with different hidden layer nodes of 64, 128, 256, and 512. Obviously, recognition ACC increases with the number of hidden nodes, because enough parameters make network fitting accurately. Also, the ACC of the 0.2-ms error scale increases slowly during the change process of 256–512 nodes and is basically saturated. Considering accuracy standard in practical applications and time cost of training that may be brought by the increasing number of hidden nodes, a network of 512 hidden nodes is a better choice.

Furthermore, this work mainly discusses the characteristic wave recognition process of a click ABR with a 96-dB nHL stimulus. Also, only parameters such as latency and wave interval can be obtained. In clinical applications, many indicators can still be used as a diagnostic basis, such as the relationship between potential values of different stimulus sizes, response and disappearance of wave V, and change of interwave latency of each characteristic wave. This also provides a new idea for the subsequent computer-assisted ABR diagnosis and treatment.

DISCUSSION

This work proposes an automatic recognition method for ABR characteristic waveforms using the BiLSTM network. The main purpose is to identify positions of characteristic waves I, III, and V, which assist the medical staff in obtaining relevant clinical test parameters, such as interwave latency and wave interval. A data quantification process is designed to analyze the characteristic waveform of ABR, including selection area of potential signal and expansion of label position. An optimal network model structure is obtained through multiple sets of comparative experiments. In 614 sets of clinically collected ABR waveform experiments, the network's overall recognition of characteristic waves showed an ACC of 92.91%.

Experimental results express that the method proposes a new idea for the identification of ABR characteristic waveforms, and helps professionals to obtain interwave latency parameters in ABR waveforms. Therefore, a computer automatic identification method can obtain deeper information, avoid subjective judgment error by the medical staff in the manual identification process effectively, reduce the number of repeated stimulations during a test, and also avoid vision fatigue of the tested person. Because of noise immunity of the proposed network model, it can effectively reduce repetitive detection of patients. In the process of large-scale identification, the average time of each data by using the method only takes approximately 0.05 s, which is much faster than the speed of manual identification. Thus, it has great advantages in repeatable work.

Some efforts have been proposed to analyze ABR waveforms using deep learning methods. For example, Fallata and Dajani (18) proposed a new detection method of ABR based on ANN to reduce detection time. Before ANN calculation, discrete wavelet transform was processed to extract features of ABR. The reduction in recording time was expected to promote the application of this measurement technique in clinical practice. McKearney and MacKinnon (19) divided ABR data into clear response, uncertain, or no response. In their work, they constructed a deep convolutional neural network and fine-tuned it to realize ABR classification. Results showed that the network may have clinical utility in assisting clinicians in waveform classification for the purpose of hearing threshold estimation. Different from the existing works, this research proposed a new data processing method and established an end-to-end deep learning model. The model can also be directly calculated without complicated mathematical transformations, so it provides a new idea for deep learning in signal processing.

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/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The ethic committee of the PLA General Hospital. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

CC and LZ: conceptualization and writing—original draft preparation. CC: methodology. XP: software and data curation. HQ, FX, and WS: validation. MS: formal analysis. FJ: investigation. QW: resources. RX and NY: writing—review and editing. LZ: visualization. NY: supervision. ZW and XG: project administration. RX: funding acquisition. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was funded by the National Key Research and Development Program (2017YFB1002804 and

2016YFC0901304), National Natural Science Foundation of China (61701022), PLA General Hospital (QNC19051), the Active Health Project of the Ministry of Science and Technology (2020YFC2004001), the Fundamental Research Funds for the Central Universities (FRF-BD-20-11A), and the Beijing Top Discipline for Artificial Intelligent Science and Engineering, University of Science and Technology Beijing.

ACKNOWLEDGMENTS

This manuscript has been released as a pre-print at bioRxiv (20).

REFERENCES

1. Marina SA, Antonio EF, Tobias R. Individual differences in the attentional modulation of the human auditory brainstem response to speech inform on speech-in-noise deficits. *Sci Rep.* (2019) 9:14131. doi: 10.1038/s41598-019-50773-1
2. Etard O, Kegler M, Braiman C, Frote AE, Reichenbach T. Decoding of selective attention to continuous speech from the human auditory brainstem response. *NeuroImage.* (2019) 200:1–11. doi: 10.1016/j.neuroimage.2019.06.029
3. Kurihara T, Igarashi Y, Kobai K, Mizobuchi T, Yokota H. Diagnosis and prediction of prognosis for Bickerstaff's brainstem encephalitis using auditory brainstem response: a case report. *Acute Med Surg.* (2020) 7:e517. doi: 10.1002/ams2.517
4. Feng S, Li X, Luo Y, Li W, Wang Z, Jiang X. Characteristics and clinical significance of auditory brainstem response in tinnitus with normal auditory thresholds. *Chin J Otol.* (2019) 17:209–13. doi: 10.3969/j.issn.1672-2922.2019.02.013
5. Jiang Y, Wang D, Liu, Z; Tan J, Li G. Comparing auditory brainstem responses evoked by click and sweep-tone in normal-hearing adults. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. Berlin: IEEE (2019). p. 5237–40.
6. Qu L, Tao, L, Zeng, M. Analysis of the characteristics of auditory brainstem response in high-risk neonates of different gestational age. *Mater Child Health Care China.* (2013) 28:4322–4. doi: 10.7620/zgfybj.j.issn.1001–4411.2013.28.22
7. Sara MK, MadsenJames M, HarteClaus ED. Accuracy of averaged auditory brainstem response amplitude and latency estimates. *Int J Audiol.* (2018) 57:345–53. doi: 10.1080/14992027.2017.1381770
8. Lewis JD, Kopun J, Neely ST, Schimid KK, Gorga MP. Tone-burst auditory brainstem response wave V latencies in normal-hearing and hearing-impaired ears. *J Acoust Soc Am.* (2015) 138:3210–9. doi: 10.1121/1.4935516
9. Wang Y, Lan Y, Wang L, Yu K, Zhao W, Wang T. ABR in early diagnosis and its prognosis of patients with dysaudia after craniocerebral trauma. *Prog Mod Biomed.* (2016) 27:5336–9.
10. Krumbholz K, Hardy AJ, Boer J. Automated extraction of auditory brainstem response latencies and amplitudes by means of non-linear curve registration. *Comput Methods Prog Biomed.* (2020) 196:105595. doi: 10.1016/j.cmpb.2020.105595
11. Wilson WJ. The relationship between the auditory brain-stem response and its reconstructed waveforms following discrete wavelet transformation. *Clin Neurophysiol.* (2004) 115:1129–39. doi: 10.1016/j.clinph.2003.11.019
12. Bradley AP, Wilson WJ. Automated analysis of the auditory brainstem response using derivative estimation wavelets. *Audiol Neurotol.* (2005) 10:6–21. doi: 10.1159/000081544
13. Zhang R, McAllister G, Scotney B, Mcclean S, Houston G. Combining wavelet analysis and Bayesian networks for the classification of auditory brainstem response. *IEEE Trans Inform Technol Biomed.* (2006) 10:458–67. doi: 10.1109/TITB.2005.863865
14. Cross J, Huang L. Incremental parsing with minimal features using bi-directional LSTM. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. Berlin (2016) doi: 10.18653/v1/P16-2006
15. Sun Y, Chen ZX. Fast extraction method of auditory brainstem response based on wavelet transformation. *Int Conf Wave Anal Patt Recogn.* (2007) 4:1862–4. doi: 10.1109/ICWAPR.2007.4421758
16. Rushaidin MM, Salleh SH, Hafizi O, Mahyar H, Ariff AK. Wave V detection using continuous wavelet transform of auditory brainstem response signal. *Prog Electromag Res Symp.* (2012) 2012:1889–93.
17. Nasr GE, Badr EA, Joun C. *Cross Entropy Error Function in Neural Networks: Forecasting Gasoline Demand*. Alberta, CA: The Florida ai Research Society. (2002).
18. Fallatah A, Dajani HR. Accurate detection of speech auditory brainstem responses using a spectral feature-based ANN method. *Biomed Sign Process Control.* (2018) 44:307–13. doi: 10.1016/j.bspc.2018.05.007
19. McKearney RM, MacKinnon RC. Objective auditory brainstem response classification using machine learning. *Int J Audiol.* (2019) 58:224–30. doi: 10.1080/14992027.2018.1551633
20. Chen C, Zhan L, Pan XX, Wang ZL, Guo XY, Qin HD, et al. Automatic recognition of auditory brainstem response characteristic waveform based on BiLSTM. *bioRxiv. [Preprint]*. (2020). doi: 10.1101/2020.10.03.324665

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.

BRIEF RESEARCH REPORT
published: 24 March 2021
doi: 10.3389/fpsyg.2021.604074

Memory Deficits for Health Information Provided Through a Telehealth Video Conferencing System

Benjamin Rich Zendel[1,2]*, *Bethany Victoria Power*[1], *Roberta Maria DiDonato*[1,2] and *Veronica Margaret Moore Hutchings*[1,2]

[1] Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL, Canada, [2] Aging Research Centre-Newfoundland and Labrador, Memorial University, Corner Brook, NL, Canada

It is critical to remember details about meetings with healthcare providers. Forgetting could result in inadequate knowledge about ones' health, non-adherence with treatments, and poorer health outcomes. Hearing the health care provider plays a crucial role in consolidating information for recall. The recent COVID-19 pandemic has meant a rapid transition to videoconference-based medicine, here described as telehealth. When using telehealth speech must be filtered and compressed, and research has shown that degraded speech is more challenging to remember. Here we present preliminary results from a study that compared memory for health information provided in-person to telehealth. The data collection for this study was stopped due to the pandemic, but the preliminary results are interesting because the pandemic forced a rapid transition to telehealth. To examine a potential memory deficit for health information provided through telehealth, we presented older and younger adults with instructions on how to use two medical devices. One set of instructions was presented in-person, and the other through telehealth. Participants were asked to recall the instructions immediately after the session, and again after a 1-week delay. Overall, the number of details recalled was significantly lower when instructions were provided by telehealth, both immediately after the session and after a 1-week delay. It is likely that a mix of technological and communication strategies by the healthcare provider could reduce this telehealth memory deficit. Given the rapid transition to telehealth due to COVID-19, highlighting this deficit and providing potential solutions are timely and of utmost importance.

Keywords: telehealth, memory, hearing, healthcare delivery, aging

OPEN ACCESS

Edited by:
Fan-Gang Zeng, University of California, Irvine, United States

Reviewed by:
Elaine Hoi Ning Ng, Oticon, Denmark
Soyun Kim, University of California, Irvine, United States

***Correspondence:**
Benjamin Rich Zendel
bzendel@mun.ca

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology

Received: 25 September 2020
Accepted: 26 February 2021
Published: 24 March 2021

Citation:
Zendel BR, Power BV, DiDonato RM and Hutchings VMM (2021) Memory Deficits for Health Information Provided Through a Telehealth Video Conferencing System. Front. Psychol. 12:604074. doi: 10.3389/fpsyg.2021.604074

INTRODUCTION

In early 2020 many countries introduced physical distancing protocols to limit the spread of COVID-19. Health authorities around the world encouraged health care providers to move to virtual care when possible. In many cases, this meant using telehealth, where a health care provider meets with a patient using voice alone (i.e., telephone) or voice and video (i.e., video conferencing). For the purposes of this paper, telehealth will refer exclusively to video conferencing between a patient and a health care provider. The rapid nature of this transition left many health care providers without training on how to deliver health care through a video conferencing system, and without supports to make this transition work. While this transition was occurring, our lab was conducting a study on how hearing status and age impact memory for health information provided through a telehealth system. Due to the COVID-19 pandemic, data collection for this study was stopped with fewer than half of the total number of participants we had planned to collect. A preliminary analysis of the data revealed a significant memory deficit for health information provided through a telehealth system compared to when the same information was provided in-person. Given the rapid transition to telehealth due to COVID-19, we felt that this preliminary data would be of interest. The next section will briefly summarize the background for the original study.

Difficulties with hearing are one of the most commonly reported health issues in older adults (Gates and Mills, 2005). Hearing difficulties make understanding speech more difficult, particularly when there is background noise, or when the speech is degraded. Interestingly, even when older adults fully comprehend mildly degraded speech, there are long-term memory deficits for the content of that speech (Pichora-Fuller et al., 1995). The proposed reason for this deficit is that a limited amount of cognitive resources that can be used at any given time, and when speech is degraded, even mildly, additional cognitive resources are needed to comprehend the speech (Pichora-Fuller et al., 1995; Schneider et al., 2010). This reduces the cognitive resources available to encode information into long-term memory (Pichora-Fuller et al., 1995; Schneider et al., 2010). Recently a series of studies demonstrated that the memory for health instructions was improved when the speech quality was enhanced, and memory for the same information was reduced when the speech was degraded (DiDonato, 2014; DiDonato and Surprenant, 2015). In the same series of studies, older adults with hearing loss benefited most when the speech was enhanced (DiDonato, 2014; DiDonato and Surprenant, 2015). This series of studies supports the idea that increased listening effort reduces memory, and demonstrates that this memory deficit occurs specifically for health information. Thus, even when older adults understand mildly degraded speech, they have more difficulty remembering what was said compared to younger adults.

One situation where this memory deficit could have major implications for older adults is for users of telehealth. Using this technology may have a significant impact on older adults, because video-conferencing systems rely on audio-compression algorithms that distort the audio signal so that information can be transmitted efficiently over the internet. The amount of distortion in the audio signal is usually dependant on the overall network bandwidth available, the quality of the microphone encoding the audio signal, and the quality of the speaker reproducing the audio signal. Proprietary digital compression algorithms can reduce the bandwidth needed to transmit the audio signal by down sampling the digital encoding and reducing the bit depth through amplitude compression. These proprietary compression algorithms, along with non-audiological grade microphones, speakers, or earphones, and non-ideal room acoustics degrade the speech. Degraded speech may not reduce the ability for an older patient to understand the healthcare provider during the session, because older adults will use additional cognitive resources in order to overcome hearing difficulties. The real problem may emerge hours or days later, when an older adult tries to remember what was said during the telehealth session. This memory deficit likely occurs because the use of additional cognitive resources to aid comprehension takes cognitive resources away from the memory system (Pichora-Fuller et al., 1995; Schneider et al., 2010; DiDonato, 2014). A recent study confirmed that older adult users of telehealth with hearing loss report more difficulty remembering what was said during a telehealth session compared to an in-person session (Willoughby and Zendel, 2017).

The goal of the original study was to experimentally test the hypothesis that older adults, particularly those with hearing loss, would have more difficulty remembering health information presented through telehealth compared to in-person after a 1-week delay. Before the COVID-19 pandemic, identifying potential memory deficits for health information presented through telehealth was critical for at least two reasons. First, telehealth use is increasing. In Canada there was a 54.6% increase in telehealth use between 2010 and 2012 (Canada's Health Informatics Association, 2013). Second, Canada's population is aging. Between 2006 and 2011, there were 1.1% more Canadians over the age of 65 (Statistics Canada, 2011). Most critical, the number of Canadians over 65, living in rural areas far from urban centres, increased 50% more than the national average between 2006 and 2011. Aging rural populations will put additional stress onto telehealth systems across Canada because older adults are more frequent users of healthcare. In 2011, older adults made up 14.8% of the Canadian population, but accounted for 45% of healthcare expenditures (Canadian Institute for Health Information, 2011). Moreover, health outcomes for older adults are more positive when they have an increased sense of control over their day-to-day lives, and are not forced to move to new communities or into long-term care facilities (Rodin, 1986). This supports the idea that aging at home benefits the health of older adults. For older adults whose home is in a remote community, this means increased reliance on telehealth. With physical distancing in place for COVID-19, identifying issues that could impact digital healthcare delivery is of utmost importance.

Unfortunately, this research was interrupted by the COVID-19 pandemic, and data collection is not complete. During the research stoppage, we explored the data and found a significant memory deficit for health information when the health information was presented through telehealth compared to in-person across all participants. Given that healthcare for many is likely to be delivered remotely for the near future, we thought these preliminary findings should be presented. Once data collection can continue, we plan to complete the study and publish the full results.

METHOD

Participants

To date, 27 participants have been recruited into the study. Ten of these participants were Younger [$M_{\text{age}} = 27.1$ ($SD = 5.9$), range 20–37; 6 female], and 17 of these participants were Older

[$M_{age} = 67.4$ ($SD = 8.8$) range 51–81; 10 female]. All participants were native English speakers, and self-reported good health. All participants completed the Montreal Cognitive Assessment (MOCA; Nasreddine et al., 2005) and scored 23 or above ($M = 28$, $SD = 2.17$), the revised cut-off for mild cognitive impairment (Carson et al., 2018). On the MOCA there are two subtests associated with verbal recall: *Verbal Fluency* and *Delayed Recall*. On the Verbal Fluency participants generated 19.2 ($SD = 5.5$) words that start with the letter F, and on the Delayed Recall, participants recalled 3.5 ($SD = 1.3$) out of five words. For the full study we plan to test 60 participants (20 younger, 40 older), based on a power analysis that assumed a small-medium effect size of 0.15, an alpha of 0.05 and beta of 0.8. We planned to test a larger sample of older adults so the group could be split based on their audiological thresholds.

Stimuli and Task

For the purpose of this brief report, only preliminary results from the experimental task will be reported. Participants also completed other audiometric and cognitive assessments, and a questionnaire about their hearing, memory, and education. These data will be reported in the final analysis. This study was approved by the Health Research Ethics Board (HREB) in Newfoundland and Labrador, and all participants provided written informed consent prior to participating. The study took part across two sessions that were 1 week apart from each other.

Session 1-Encoding

All testing took place in 2 adjacent teleconferencing rooms that each included a large table, and a Polycom teleconferencing system located in the Health Sciences Centre in St. John's, Newfoundland and Labrador. This type of teleconferencing system is commonly used by telehealth services in Newfoundland and Labrador. Each participant was presented with two vignettes about how to use two different medical devices (inhaler; medipatch) adapted from DiDonato and Surprenant (2015; see below for more information). One was presented in-person and the other via telehealth. In order to minimize potential practice effects, the order of presentation (telehealth or in-person, and inhaler or medipatch) was fully counter balanced between participants. Before the experimental tasks, participants completed a written informed consent, and a demographics questionnaire administered orally. Participants were instructed that they would be asked to recall the instructions of both vignettes immediately after they were presented, and again in 1 week, at the beginning of Session 2. Participants were further instructed that no information from the vignettes would be repeated, not to ask questions during the presentation of the vignette but to otherwise behave as they would during a visit with a health-care provider.

In-person

For the in-person condition, the participant and researcher were both seated facing each other across a conference table, about 1 meter apart. The participant could see the researchers face, arms, hands, and upper torso. The telehealth screen was off during this session to avoid distractions. The researcher read aloud the vignette about how to use one of the medical devices. The researcher spoke slowly and clearly, at a typical conversational sound level, ∼60–65 dB SPL for the listener. Before testing any participants, the researcher practiced speaking at this level, using a portable sound level meter placed where a participants' ears would be to ensure they could maintain a constant level. Immediately following the vignette, participants were asked to verbally recall as much detail as they could. During recall, participants were not assisted by the researcher, and were not provided with any feedback about their accuracy.

Telehealth

For a vignette conducted via telehealth, the researcher moved into the adjacent room, that was nearly identical to the room the participant was in. During this session, the teleconferencing system was turned on. The teleconferencing system was a Polycom HDX6000. This type of system is used throughout Newfoundland and Labrador for telehealth sessions. Video was presented at a resolution of 1080p at 30 frames per second, and audio was presented at a 22 kHz sampling rate. The participant was seated approximately 1 meter from the screen, and 0.5 meters from the free-field speaker. The participant was able to see the researchers face, arms, hands, and upper torso on the screen. The researcher was positioned so that their image was approximately "life size" on the screen. The volume on the speaker was adjusted so that the researcher's voice was ∼60–65 dB SPL where the participant was sitting. Other than being presented through the telehealth system, the task was identical to the in-person task.

Medical Device Vignettes

The medical device vignettes were adapted from DiDonato and Surprenant (2015). One vignette featured information on how to use a medipatch to deliver pain medication, and the other vignette featured information on how to use an asthma puffer. The vignettes were matched on many linguistic and non-linguistic aspects of speech to equate them as much as possible on the complexity of the stimuli, while at the same time maintaining their ecological validity (see DiDonato and Surprenant, 2015 for more details). Both vignettes were 10 sentences long, and included 37 details; the medipatch vignette was 154 words long, and the asthma puffer vignette was 151 words long. Reading the vignette took ∼70 s, for an average speaking rate of about 2.2 words per second, which was slower and easier to understand than conversational speech (Baker and Bradlow, 2009). No visual aids or demonstrations of how to use the medipatch or asthma inhaler were provided. The 37 details were content words within each phrase that carried the most critical meaning for the purpose of using these medical devices. Details may have been a single word, compound word, or multiple words (e.g., breathe out, out of reach, etc.). The distribution of the details throughout the vignette were arranged so that each third of the vignette had a similar number and distribution of details to recall.

Immediate Recall

After the presentation of the vignette, the participant was asked to recall the instructions as best they could. The researcher recorded the number of details the participant correctly recalled using a scoring sheet. There was no time restriction on how long a participant could take to recall the instructions. Details did not have to be remembered in the correct order, and each detail was worth 1 point. Participants were given a score out of 37 based on how many details they correctly recalled. This score was converted to a percentage and used as a measure of Immediate Recall.

Session 2

After a 7-day delay, participants returned for a second session. The focus of this session was to examine the delayed recall of the medical vignettes, and to collect audiometric data. This session took place in the Cognitive Aging and Auditory Neuroscience Laboratory, a large, quiet room that contains a sound-attenuating booth, and equipment for audiometric assessments. Participants were seated at a table in an office chair facing the researcher. The researcher sat across from the participant, ∼1 meter away. At this point the researcher reminded the participant that they were asked to remember the instructions given for both the medipatch and the asthma puffer. Participants were asked to describe these instructions in as much detail as they could remember. The order in which they were asked to recall the vignettes was the same as in Session 1 (i.e., if in session 1 the participant heard the medipatch vignette first and the asthma puffer second, then in session 2 the participant was asked to recall the instructions for the medipatch first, and the asthma puffer second). This order was maintained regardless of the order of presentation in session 1 was in-person or telehealth first. Scoring was identical to scoring for immediate recall, and like the immediate recall session, there was no restriction on how long a participant could take to recall the instructions. Upon completion of the delayed recall task for both vignettes, participants were given a 5-min break. Participants were then given a series of auditory and cognitive assessments. For the purpose of this brief report, this data is not presented.

RESULTS

Data was analyzed using a mixed-design ANOVA that included Mode of Delivery (In-person, Telehealth) and Memory (Immediate, Delayed) as within-subject variables, and Age Group as a between-subject variable. The Order of presentation (i.e., Telehealth or In-person first) was initially included as a factor, but its main effect ($p = 0.99$) and interactions were not significant ($p = 0.15$–0.93), so it was removed from the analysis. There were main effects of Memory and Mode of Delivery. Overall, more details were recalled immediately after the session compared to after a 1-week delay [$F_{(1, 23)} = 19.54$, $p < 0.001$, $\eta^2 = 0.08$] (**Figure 1A**). Additionally, more details were recalled when delivered In-person compared to via Telehealth [$F_{(1, 23)} = 5.70$, $p = 0.026$, $\eta^2 = 0.05$] (**Figure 1A**). There was no main effect of Age Group [$F_{(1, 23)} = 2.79$, $p = 0.11$, $\eta^2 = 0.06$], and there were no significant interactions between the three variables; however, the interaction between Memory and Age Group approached significance [$F_{(1, 23)} = 3.90$, $p = 0.060$, $\eta^2 = 0.02$] (**Figure 1B**). Follow-up tests, using pairwise comparisons revealed that for Younger Adults, compared to older adults, there were larger differences in the number of details recalled immediately compared to the number of details recalled after the 1-week delay ($p = 0.002$). For Older Adults, differences in the number of details recalled immediately compared to after a 1-week-delay were smaller and not significant ($p = 0.20$).

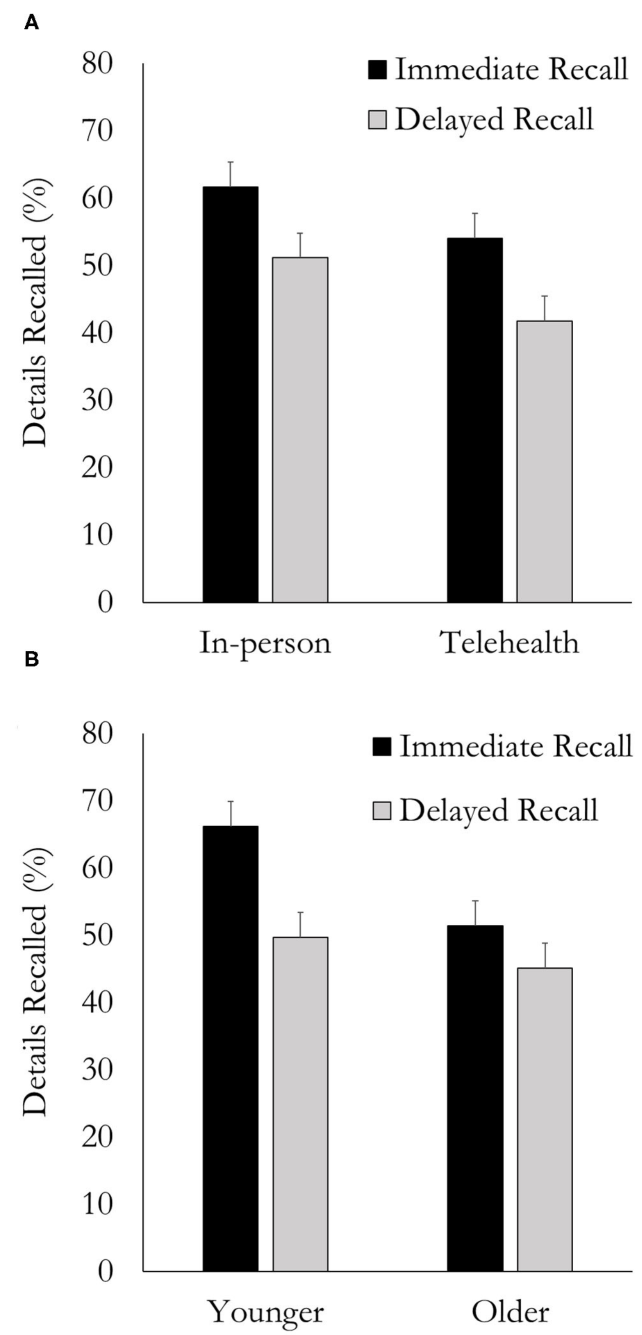


FIGURE 1 | Number of details recalled immediately after the session and after a 1-week delay as a function of **(A)** Mode-of-Delivery and **(B)** Age group. Error bars are one standard error of the mean.

DISCUSSION

Overall, both groups (younger and older) demonstrated a memory deficit when health information was presented via Telehealth compared to In-person. One interesting observation was that people remember about the same amount of information immediately after a Telehealth session (54.3%) compared to an In-Person session after a 1-week delay (51.6%), suggesting that the impact of getting health information through telehealth was similar to the impact of a 1-week delay in verbal memory recall (herein referred to as simply "recall"). Furthermore, the decrease in recall performance immediately after the session, and 1 week later was similar for both In-Person and Telehealth delivery (10.5 & 12.3%, respectively). This preliminary data suggests comprehension difficulties (i.e., impaired speech perception or cognitive-linguistic processes) due to the Telehealth mode of delivery led to a reduction in the number of details that were initially encoded. The telehealth memory deficit is unlikely to be related to differences in retention because the number of details "forgotten" during the 1-week delay was similar for both the In-person and Telehealth conditions. Once the full data set is collected a detailed exploration of the impact of aging and hearing status on immediate and delayed recall for health information provided either through telehealth or in-person will be possible.

Differences between Telehealth and In-person mode-of-delivery led to an immediate recall deficit in the Telehealth condition. The main goal of this study was to compare immediate- and delayed-recall for health information presented in two realistic situations. The main challenge was that by using a real telehealth system, we were unable to control or manipulate a number of factors that could reduce both audio and video fidelity in the Telehealth condition. The deficit we observed could therefore be related to reduced audio fidelity, reduced video fidelity or a combination of both. In a review, Mattys et al. (2012) describes this as a form of environmental/transmission degradation in communication. Mattys et al. (2012) highlights that this type of degradation could reduce speech recognition, reduce the ability to attend to the telehealth session, and reduce memory for information from the session.

There are a number of theories that can be used to interpret these results in terms of the reduced quality of the message, including The *Ease of Language Understanding* theory (Rönnberg et al., 2010), and the *Effortful Listening* theory, first introduced by Rabbitt (1968, 1991), and more recently developed into the *Framework for Understanding Effortful Listening (FUEL)* by Pichora-Fuller et al. (2016). These theories state that working memory load will increase when there is a difficulty matching incoming speech to ones' mental lexicon. This increased working memory load during listening inhibits the ability to encode details into long-term memory. Findings from the current study suggest that audio and video degradation in a commercial grade telehealth system increases working memory processing in order to accurately match incoming speech to the mental lexicon (i.e., comprehension of the message). In turn this reduced the number of details that were encoded into long-term memory, which led to a deficit in recalling health information immediately after the encoding session was complete.

Based on perceptual factors alone, one might not predict significant recall deficits due to the high quality of the telehealth system used in the study; however, a minimally degraded message can act in an insidious way. When communication is degraded minimally, an individual may not recognize the degradation, and may not engage compensation mechanisms to properly attend to, and remember what was said (Bäckman and Dixon, 1992). It is this lack of awareness that a communication event is sub-optimal that then interferes with the automatic or explicit use of those to-be-employed compensations for mitigating the effects of the degraded message (Bäckman and Dixon, 1992). In a recent review of best practices for tele-mental-health Hilty et al. (2019), specifically highlight the importance of minimizing distractions and optimizing speech clarity in order to improve the therapeutic relationship. Accordingly, in this situation, the slight imperfections in the communication associated with the Telehealth condition may have contributed to difficulty remembering what was said due to both the reduced quality of the message, and a lack of awareness of the reduced quality of the message. Previous work in visual perception has shown that target degradation increases distractor effects (Lavie and De Fockert, 2003). In the current study, this effect might be amplified because the telehealth system used was high quality, and thus the participant may not have engaged possible compensatory mechanisms (see: Bäckman and Dixon, 1992). The mildly degraded auditory and visual information in the Telehealth condition may have made minor, unavoidable distractions more salient for the individual compared to the In-Person condition. In-turn this would have reduced the number of details that were comprehended and then encoded into long-term memory.

Interestingly, this distractibility hypothesis and the Effortful Listening hypotheses make somewhat different predictions about comprehension. The Effortful Listening hypotheses predicts that immediate comprehension may indeed be similar for In-Person and Telehealth sessions with the impact of the degradation only demonstrated later as reduced recall of the message, while the distractibility hypothesis would predict that immediate comprehension may be reduced for the Telehealth session compared to the In-Person session. Unfortunately, due to the ecological nature of the study, comprehension of each statement was not measured; immediate recall was measured at the end of the vignette. Overall, it is likely that both increased distractibility and reduced fidelity of the signal play a role in reduced immediate recall. It is therefore likely that improving the fidelity of the telehealth signal, could improve memory due to reduced cognitive load, and a reduction of distractibility. Unfortunately, the current study was not designed to tease apart contributions of effortful listening and distractibility on immediate recall. This should be explored in future research.

Laboratory-based studies have shown that modifying the fidelity of speech impacts both comprehension and memory. Degraded speech has been shown to negatively impact recall compared to normal speech, including when speech is sped up (DiDonato, 2014), when it is presented with background noise (Pichora-Fuller et al., 1995) or when the speech is noise vocoded (Ward et al., 2016). Speech clarity can be improved by using inserted earphones instead of loudspeakers, and this difference has been shown to enhance recall (DiDonato and Surprenant, 2015). It is likely that providing the patient with high quality insert earphones could improve their memory from a telehealth session. The healthcare providers' speech patterns are also important for understanding. Speaking with normal prosody has been shown to improve recall compared to speaking with a flat prosody (Stine and Wingfield, 1987). One interesting finding was that older adults recalled more information when health information was presented slower, with shorter utterances, and more varied and higher pitched intonation (McGuire et al., 2000). While slower speaking has been shown to improve recall (Thompson, 1995), it has been shown that "self-paced" listening can further improve recall (Piquado et al., 2012). Self-paced listening is when the listener is allowed to pause the speech when they want. From a telehealth perspective this means that healthcare providers would be best to pause regularly. During these pauses recall could likely be further improved if patients are asked to repeat what the healthcare provider just said, as repetition is a well-known memory aid. Another important finding is that meaningful phrases are recalled better than random words (Stine and Wingfield, 1987; Thompson, 1995). This suggests that healthcare providers should take extra care to use colloquial speech, and to be thoughtful of their word choices so that each statement they make is understood and meaningful to the patient.

When speech is paired with a visual representation of the person speaking (i.e., video), comprehension improves, suggesting that videoconferencing is superior to telephone in situations where comprehension is critical (Grant et al., 1998; Grant and Seitz, 2000; Sommers et al., 2005). In these studies, the video was limited to the head and neck of the speaker, and the authors explain the audio-visual enhancement compared to auditory alone was due to an integration of audio and visual cues that facilitate phonetic and lexical decisions (Grant et al., 1998; Grant and Seitz, 2000; Sommers et al., 2005). In addition to the face, hand gestures have also been shown to improve speech understanding, although the benefit of seeing gestures seems to be reduced in older adults (Thompson, 1995; Hilty et al., 2019). In the current study, the researcher did not use their hands to demonstrate how to use the device, and based on previous work, this may have reduced the differences in recall between the in-person and telehealth conditions. Accordingly, it is important for users of telehealth to be able to see the face and hands of their healthcare provider clearly in order to see gesturing and other non-verbal forms of communication that occur through the hands.

Summary

The current study found that immediate and long-term recall of health information was lower when that information was presented through Telehealth, compared to In-Person. Importantly, this telehealth memory deficit was similar for both immediate and delayed recall, which suggests that the best way to improve memory for health information provided through telemedicine is to improve a patients' immediate recall of the session. This is likely to be effective because the decline in recall performance after a 1-week delay was about the same regardless of the mode-of-delivery. One way to accomplish this would be to facilitate understanding so that more efficient memory encoding occurs. It is therefore likely that optimizing comprehension in telehealth situations would mitigate the listening effort and enhance immediate recall of health information. Improved immediate recall should in turn improve long-term recall. From a practical perspective, health care providers should be mindful that their patients may not recall the information presented through telehealth as they would in person, unless the health-care provider takes steps to enhance the immediate recall of information they present. The limitations of the suggestions presented here to minimize the memory deficit are based on in-person healthcare studies, or laboratory-based speech perception studies, thus they may not be generalizable to telehealth. However, given the dearth of studies on memory for information presented through telehealth, the recent need to rapidly transition healthcare into an online format due to COVID-19, and the relatively easy and low risk techniques that could improve memory, we suggest that healthcare providers attempt to use as many of the techniques reported above to improve memory for health information when they are conducting telemedicine and to review information that was presented in previous sessions. These preliminary results could be a very positive finding for telehealth, as they suggest that supplementing telehealth sessions with communication supports including high-quality earphones, and communication guidance for healthcare providers could eliminate the recall difference between in-person and telehealth sessions.

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 Health Research Ethics Board, Newfoundland and Labrador. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

BZ: designed study, analyzed data, and wrote manuscript. BP: designed study, collected data, analyzed data, and wrote manuscript. RD: designed study and wrote manuscript. VH: analyzed data and wrote manuscript.

FUNDING

This research was supported by a Canada Research Chair held by BZ.

REFERENCES

Bäckman, L., and Dixon, R. A. (1992). Psychological compensation: a theoretical framework. *Psychol. Bull.* 112:259. doi: 10.1037/0033-2909.112.2.259

Baker, R. E., and Bradlow, A. R. (2009). Variability in word duration as a function of probability, speech style, and prosody. *Language Speech* 52, 391–413. doi: 10.1177/0023830909336575

Canada's Health Informatics Association (2013). *Canadian Telehealth Report*. Toronto, ON: National Office.

Canadian Institute for Health Information (2011). *Health Care in Canada, 2011: A Focus on Seniors and Aging*. Ottawa, ON: CIHI.

Carson, N., Leach, L., and Murphy, K. J. (2018). A re-examination of Montreal Cognitive Assessment (MoCA) cut off scores. *Int. J. Geriatr. Psychiatr.* 33, 379–388. doi: 10.1002/gps.4756

DiDonato, R. (2014). *Effortful and effortless listening: how age-related hearing loss and cognitive abilities interact and influence memory performance in older adults.* [Doctoral dissertation], Memorial University of Newfoundland. St. John's, NL, Canada.

DiDonato, R. M., and Surprenant, A. M. (2015). Relatively effortless listening promotes understanding and recall of medical instructions in older adults. *Front. Psychol.* 6:778. doi: 10.3389/fpsyg.2015.00778

Gates, G. A., and Mills, J. H. (2005). Presbycusis. *Lancet* 366, 1111–1120. doi: 10.1016/S0140-6736(05)67423-5

Grant, K. W., and Seitz, P. F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *J. Acoust. Soc. Am.* 108, 1197–1208. doi: 10.1121/1.1288668

Grant, K. W., Walden, B. E., and Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: consonant recognition, sentence recognition, and auditory-visual integration. *J. Acoustical Soc. Am.* 103, 2677–2690 doi: 10.1121/1.422788

Hilty, D. M., Randhawa, K., Maheu, M. M., McKean, A. J. S., and Pantera, R. (2019). Therapeutic relationship of telepsychiatry and telebehavioral health: Ideas from research on telepresence, virtual reality and augmented reality. *Psychol. Cognit. Sci. Open J.* 5, 14–29. doi: 10.17140/PCSOJ-5-145

Lavie, N., and De Fockert, J. W. (2003). Contrasting effects of sensory limits and capacity limits in visual selective attention. *Percept. Psychophys.* 65, 202–212. doi: 10.3758/BF03194795

Mattys, S. L., Davis, M. H., Bradlow, A. R., and Scott, S. K. (2012). Speech recognition in adverse conditions: a review. *Language Cognit. Proc.* 27, 953–978. doi: 10.1080/01690965.2012.705006

McGuire, L. C., Morian, A., Codding, R., and Smyer, M. A. (2000). Older adults' memory for medical information: influence of elderspeak and note taking. *Int. J. Rehab. Health* 5, 117–128. doi: 10.1023/A:1012906222395

Nasreddine, Z. S., Phillips, N. A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., and Chertkow, H. (2005). The montreal cognitive assessment, MoCA: a brief screening. *J. Am. Geriatr. Soc*. 53:695e699. doi: 10.1111/j.1532-5415.2005.53221.x

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

Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *J. Acoust. Soc. Am.* 97, 593–608. doi: 10.1121/1.412282

Piquado, T., Benichov, J. I., Brownell, H., and Wingfield, A. (2012). The hidden effect of hearing acuity on speech recall, and compensatory effects of self-paced listening. *Int. J. Audiol.* 51, 576–583. doi: 10.3109/14992027.2012.684403

Rabbitt, P. (1991). Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. *Acta Oto-Laryngol.* 111, 167–176. doi: 10.3109/00016489109127274

Rabbitt, P. M. (1968). Channel-capacity, intelligibility and immediate memory. *Quarterly J. Exp. Psychol.* 20, 241–248. doi: 10.1080/14640746808400158

Rodin, J. (1986). Aging and health: effects of the sense of control. *Science* 233, 1271–1276. doi: 10.1126/science.3749877

Rönnberg, J., Rudner, M., Lunner, T., and Zekveld, A. A. (2010). When cognition kicks in: working memory and speech understanding in noise. *Noise Health* 12:263. doi: 10.4103/1463-1741.70505

Schneider, B. A., Pichora-Fuller, K., and Daneman, M. (2010). "Effects of senescent changes in audition and cognition on spoken language comprehension," in *The Aging Auditory System* (New York, NY: Springer), 167–210.

Sommers, M. S., Tye-Murray, N., and Spehar, B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear Hear.* 26, 263–275. doi: 10.1097/00003446-200506000-00003

Statistics Canada (2011). *Age and Sex, 2011 Census*. Ottawa, On: Minister of Industry Telehealth (2017). Newfoundland and Labrador Centre for Health Information. Retrieved from: https://www.nlchi.nl.ca/index.php/ehealth-systems/telehealth

Stine, E. L., and Wingfield, A. (1987). Process and strategy in memory for speech among younger and older adults. *Psychol. Aging* 2:272.

Thompson, L. A. (1995). Encoding and memory for visible speech and gestures: a comparison between young and older adults. *Psychol. Aging* 10:215.

Ward, C. M., Rogers, C. S., Van Engen, K. J., and Peelle, J. E. (2016). Effects of age, acoustic challenge, and verbal working memory on recall of narrative speech. *Exp. Aging Res;.* 42, 97–111. doi: 10.1080/0361073X.2016.1108785

Willoughby, K., and Zendel, B. R. (2017). Hearing and memory deficits in older adults using telehealth. *J. Can. Acoust.* 45, 188–189. Available online at: https://jcaa.caa-aca.ca/index.php/jcaa

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.

ORIGINAL RESEARCH
published: 25 March 2021
doi: 10.3389/fmed.2021.662045

Audiometric Phenotypes of Noise-Induced Hearing Loss by Data-Driven Cluster Analysis and Their Relevant Characteristics

Qixuan Wang[1,2,3†], *Minfei Qian*[1,2,3†], *Lu Yang*[1,4], *Junbo Shi*[1,4], *Yingying Hong*[1,4], *Kun Han*[1,2,3], *Chen Li*[5], *James Lin*[5], *Zhiwu Huang*[1,2,3,4*] and *Hao Wu*[1,2,3*]

[1] Department of Otolaryngology-Head and Neck Surgery, Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [2] Ear Institute, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [3] Shanghai Key Laboratory of Translational Medicine on Ear and Nose Diseases, Shanghai, China, [4] Hearing and Speech Center, Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [5] Network and Information Center, Shanghai Jiao Tong University, Shanghai, China

OPEN ACCESS

Edited by:
Qinglin Meng,
South China University of Technology, China

Reviewed by:
Wei Qiu,
State University of New York, United States
Fei Ji,
PLA General Hospital, China

***Correspondence:**
Zhiwu Huang
huangzw86@126.com
Hao Wu
wuhao@shsmu.edu.cn

[†] These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Family Medicine and Primary Care, a section of the journal Frontiers in Medicine

Received: 31 January 2021
Accepted: 02 March 2021
Published: 25 March 2021

Citation:
Wang Q, Qian M, Yang L, Shi J, Hong Y, Han K, Li C, Lin J, Huang Z and Wu H (2021) Audiometric Phenotypes of Noise-Induced Hearing Loss by Data-Driven Cluster Analysis and Their Relevant Characteristics. Front. Med. 8:662045. doi: 10.3389/fmed.2021.662045

Background: The definition of notched audiogram for noise-induced hearing loss (NIHL) is presently based on clinical experience, but audiometric phenotypes of NIHL are highly heterogeneous. The data-driven clustering of subtypes could provide refined characteristics of NIHL, and help identify individuals with typical NIHL at diagnosis.

Methods: This cross-sectional study initially recruited 12,218 occupational noise-exposed employees aged 18–60 years from two factories of a shipyard in Eastern China. Of these, 10,307 subjects with no history of otological injurie or disease, family history of hearing loss, or history of ototoxic drug use were eventually enrolled. All these subjects completed health behavior questionnaires, cumulative noise exposure (CNE) measurement, and pure-tone audiometry. We did data-driven cluster analysis (k-means clustering) in subjects with hearing loss audiograms ($n = 6{,}599$) consist of two independent datasets ($n = 4{,}461$ and $n = 2{,}138$). Multinomial logistic regression was performed to analyze the relevant characteristics of subjects with different audiometric phenotypes compared to those subjects with normal hearing audiograms ($n = 3{,}708$).

Results: A total of 10,307 subjects (9,165 males [88.9%], mean age 34.5 [8.8] years, mean CNE 91.2 [22.7] dB[A]) were included, 3,708 (36.0%) of them had completely normal hearing, the other 6,599 (64.0%) with hearing loss audiograms were clustered into four audiometric phenotypes, which were replicable in two distinct datasets. We named the four clusters as the 4–6 kHz sharp-notched, 4–6 kHz flat-notched, 3–8 kHz notched, and 1–8 kHz notched audiogram. Among them, except for the 4–6 kHz flat-notched audiogram which was not significantly related to NIHL, the other three phenotypes with different relevant characteristics were strongly associated with noise exposure. In particular, the 4–6 kHz sharp-notched audiogram might be a typical subtype of NIHL.

Conclusions: By data-driven cluster analysis of the large-scale noise-exposed population, we identified three audiometric phenotypes associated with distinct NIHL subtypes. Data-driven sub-stratification of audiograms might eventually contribute to the precise diagnosis and treatment of NIHL.

Keywords: noise-induced hearing loss, audiometric phenotype, notched audiogram, unsupervised learning, data-driven cluster analysis, multivariate characteristics

INTRODUCTION

Noise-induced hearing loss (NIHL) is one of the most common hearing loss in adults (1), with increasing incidence in children and adolescents (2) due to widespread recreational and transport noise exposure (3, 4). The World Health Organization (WHO) estimates that 10% of the world population is exposed to sound levels that could potentially cause NIHL (5). To date, treatment options for NIHL are limited, while ∼50% of this burden could be prevented by early detection of NIHL, avoidance of noise exposure, prompt intervention, etc. (6).

It is widely accepted that the noise exposure usually causes high-frequency sensorineural hearing impairment (7, 8). Despite several previously concluded abstract phenotypes of NIHL including the high-frequency audiometric notch and the bulge downwards audiogram (9), there are still no clear audiometric criteria on stratifications of NIHL, which makes it difficult to specifically evaluate NIHL during clinical and primary health care (10, 11). One reason for this is the heterogeneous audiometric phenotypes of NIHL, involving complex confounding influencing factors. The majority of studies have adopted different definitions of high-frequency hearing loss (12, 13) and notched audiogram (14–16), which were chosen mainly by specialized intuition or clinical experience, rather than by data-driven analysis. These inconsistent assessment methods were manifested by various ranges of frequency and degrees of hearing loss, which may represent different subtypes of NIHL with inconsistent responses to intervention, and inevitably result in incomparable conclusions between studies.

Generally, descriptions of NIHL phenotypes are limited by subjectivity and poor data support. A data-driven classification that incorporates the multifrequency audiogram of NIHL is needed to identify subtypes with consistent patterns and characteristics. Cluster analysis is an unsupervised exploratory data mining technique able to group the most similar individuals with multiple specified variables in the same group called "cluster" without any previously defined hypothesis (17). Since audiogram stratification is based on the complex non-linear combination of thresholds at several frequencies, unbiased data-driven cluster analysis has recently been found to be a useful method for the identification of audiometric phenotypes (18, 19). We postulated that cluster analysis could be applied for classifying audiograms of NIHL.

In the current study, based on audiograms of 10,307 Chinese shipyard employees with various noise exposure levels, we used the k-means clustering algorithm to classify subtypes of NIHL in two distinct noise-exposed populations from different factories. The confounding influencing factors related to these subtypes were further analyzed to optimize the assessment for different subtypes of NIHL, which could provide a powerful tool to identify those individuals at great risk of NIHL and guide optimal prevention of noise exposure.

METHODS

Study Population

We conducted this hearing and health investigation in a shipyard in eastern China from August 1, 2017, to June 30, 2018. A total of 12,218 subjects aged 18–60 years were initially recruited, and 10,307 from two steel factories (6,631 from factory 1, and 3,676 from factory 2) were included in the analysis based on the following criteria: (1) completed questionnaire and audiometric data, (2) no history of otological injuries or diseases, (3) no family history of hearing loss, (4) no history of ototoxic drug use, (5) no profound hearing loss (average threshold at 0.5–2 kHz frequencies >70 dB HL in any ear), and (6) no perforation of tympanic membrane or abnormal tympanogram. Sex and race were self-reported. **Figure 1** shows the flowchart of this cross-sectional study, which was in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guidelines and approved by the ethics committee of the Ninth People's Hospital affiliated to Shanghai Jiao Tong University School of Medicine. All the participants signed written informed consent forms.

Audiometry

Pure-tone air-conduction audiometry at frequencies of 0.5, 1, 2, 3, 4, 6, and 8 kHz in both ears was performed by certified audiological technicians using an audiometer (Otometrics Madsen, Xeta, Denmark) with TDH-39P headsets in a soundproof booth in accordance with the regulations of ISO 8253-1: 2010. The subjects were not exposed to occupational noise or loud sounds within 16 h before being examined. The average threshold of the left and right ears at each frequency was calculated for subsequent analysis without the age-correction according to ISO 7029: 2017, in order to avoid the artificial modification on the subsequent cluster analysis. Normal hearing was defined as hearing threshold ≤25 dB HL over 0.5–8 kHz frequencies. Hearing loss was defined as hearing threshold >25 dB HL at any frequency.

Questionnaire

Demographic variables (sex, age, race, job type, working time-length) and behavioral characteristics, including hearing

FIGURE 1 | Flowchart of this cross-sectional study.

protection device (HPD) use (<4 h/work-day, ≥4 h/work-day), personal earphone use (<1 h/day, ≥1 h/day), tobacco (<10 cigarettes/day, ≥10 cigarettes/day) and alcohol (<50 g/day, ≥50 g/day) consumption, and auditory-related symptoms (hearing difficulty and tinnitus), were collected through a self-reported questionnaire. Body mass index (BMI) was measured and calculated by investigators, and then categorized into non-obese (<28 kg/m^2) and obese (≥28 kg/m^2) groups.

Noise Exposure Dose

A composite quantitative noise exposure index, the cumulative noise exposure (CNE), was used to estimate the noise exposure level for each subject, which was calculated using the following formula (20):

$$\mathrm{CNE} = L_{\mathrm{Aeq,8h}} + 10\log T,$$

where $L_{\mathrm{Aeq,8h}}$ is the equivalent sound pressure level in A weight of 8 continuous hours of a work-day, which was measured and analyzed using the personal exposure dosimeter (Aihua, ASV5910 type, Hangzhou, China). Subjects were required to wear the dosimeter on the shoulder for five work-days to calculate the average $L_{\mathrm{Aeq,8h}}$. T is the working time-length in years obtained from the questionnaire.

Data-Driven Cluster Analysis

Seven variables including standardized values of thresholds at frequencies of 0.5, 1, 2, 3, 4, 6, and 8 kHz were input for k-means cluster analysis performed using R software (version 4.0.3) (21). The optimal number of clusters was selected according to the within cluster sum of squares (WSS) (22), the number of clusters from 2 to 15 was tried, and the last one that significantly reduced the WSS (at the inflection point of the curve) was selected as the optimal number of clusters (**Figures 2A–E**). Data-driven cluster analysis was performed in data from two factories (dataset 1 and dataset 2) separately, and then repeated in the total data.

Statistical Analysis

Data analysis was performed by using IBM SPSS version 24.0 software (SPSS Inc., Chicago, IL, USA) except for cluster analysis. Continuous variables are expressed as the mean (standard deviation, SD), and categorical variables are presented as percentages (n [%]). Statistical significance for differences in continuous variables was examined using Student's t test (between dataset 1 and dataset 2) or ANOVA (between cluster subtypes with Dunn-Bonferroni tests for *post-hoc* analyses), and categorical variables were compared by the chi-square test. Multinomial logistic regression models were used to analyze relevant factors of different clusters of audiometric subtypes. For the hierarchical regression, age was categorized into 3 groups (<30, 30–45, and >45 years). For

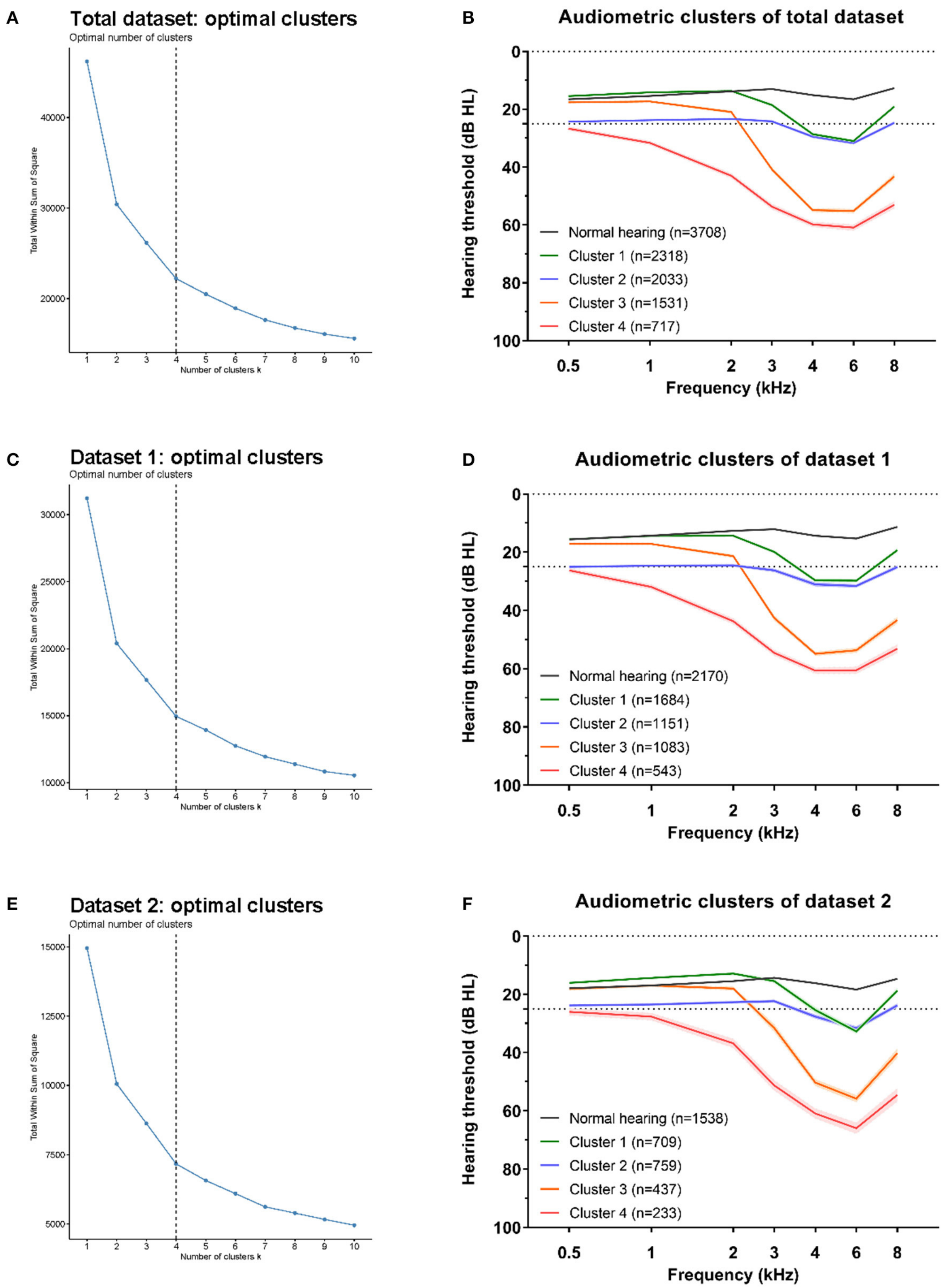


FIGURE 2 | Optimal clusters for three datasets. The within cluster sum of squares (WSS) decrease with the increment of clusters number, and the optimal number of clustering was selected at the last one significantly reduced the WSS (at the inflection point of the curve). The optimal clustering were all at number of four (black dotted line) for total dataset **(A)**, dataset 1 **(C)**, and dataset 2 **(E)**. The average hearing thresholds over 0.5–8 kHz frequencies of normal hearing subjects and those four clusters were shown for total dataset **(B)**, dataset 1 **(D),** and dataset 2 **(F)**.

TABLE 1 | Characteristics of subjects in different datasets.

Variables	Total (n = 10,307)	Dataset 1 (n = 6,631)	Dataset 2 (n = 3,676)	P value*
Age (years), mean (SD)	34.5 (8.8)	36.2 (8.6)	31.4 (8.3)	<0.001
CNE (dB[A]), mean (SD)	91.2 (22.7)	92.0 (22.0)	89.8 (23.9)	<0.001
Sex, n (%)				0.007
Males	9,165 (88.9)	5,855 (88.3)	3,310 (90.0)	
Females	1,142 (11.1)	776 (11.7)	366 (10.0)	
BMI, n (%)				<0.001
Non-obese	9,391 (90.9)	6,000 (90.5)	3,371 (91.7)	
Obese	936 (9.1)	631 (9.5)	305 (8.3)	
Hearing difficulty, n (%)				<0.001
No	7,955 (77.2)	4,952 (74.7)	3,003 (81.7)	
Yes	2,352 (22.8)	1,679 (25.3)	673 (18.3)	
Tinnitus, n (%)				<0.001
No	6,971 (67.6)	4,327 (65.3)	2,644 (71.9)	
Yes	3,336 (32.4)	2,304 (34.7)	1,032 (28.1)	
HPD use, n (%)				<0.001
<4 h/work-day	7,384 (71.6)	4,936 (74.4)	2,448 (66.6)	
≥4 h/work-day	2,923 (28.4)	1,695 (25.6)	1,228 (33.4)	
Earphone use, n (%)				<0.001
<1 h/day	5,844 (56.7)	3,418 (51.5)	2,426 (66.0)	
≥1 h/day	4,463 (43.3)	3,213 (48.5)	1,250 (34.0)	
Tobacco consumption, n (%)				0.156
<10 cigarettes/day	6,436 (62.4)	4,174 (62.9)	2,262 (61.5)	
≥10 cigarettes/day	3,871 (37.6)	2,457 (37.1)	1,414 (38.5)	
Alcohol consumption, n (%)				<0.001
<50 g/day	7,558 (73.3)	5,076 (76.5)	2,482 (67.5)	
≥50 g/day	2,749 (26.7)	1,555 (23.5)	1,194 (32.5)	
Hearing loss, n (%)				<0.001
No	3,708 (36.0)	2,170 (32.7)	1,538 (41.8)	
Yes	6,599 (64.0)	4,461 (67.3)	2,138 (58.2)	

SD, standard deviation; BMI, body mass index; CNE, cumulative noise exposure; HPD, hearing protective device.
**Comparisons were between dataset 1 and dataset 2.*

all models, odds ratios (ORs) and 95% confidence intervals (CIs) are presented. A 2-tailed $P < 0.05$ was considered statistically significant.

RESULTS

Basic Characteristics of Subjects

A total of 10,307 Chinese Han subjects (9,165 males [88.9%], mean age 34.5 [SD 8.8] years, mean CNE 91.2 [SD 22.7] dB[A]) were included. Among all subjects, 3,708 (36.0%) had completely normal hearing over 0.5–8 kHz frequencies. The total subjects were recruited from two independent factories in a shipyard, who had similar types of occupational tasks, despite significantly different distributions of sex, age, CNE, hearing loss, and other characteristics. The distributions of age, CNE, sex, BMI, hearing difficulty, tinnitus, HPD use, earphone use, tobacco consumption, and alcohol consumption are shown in **Table 1**.

Clusters of Audiometric Phenotypes

To classify NIHL into novel audiometric phenotypes, we used the k-means clustering method in audiograms with hearing loss. We repeated the cluster process, respectively, in total dataset (all the hearing loss audiograms, $n = 6{,}599$), dataset 1 (hearing loss audiograms from factory 1, $n = 4{,}461$) and dataset 2 (hearing loss audiograms from factory 2, $n = 2{,}138$) to verify that the cluster structure described for each dataset was reproducible.

For all three datasets, the optimal number of clusters was four according to the WSS decreasing curve (**Figures 2A–E**), and the audiometric phenotypes of four clusters identified from different datasets were qualitatively similar. In total, 6,239 /6,599 (94.5%) audiograms in total dataset clusters were classified into the same subtype according to the distinct clusters in dataset 1 (4,286 /4,461, 96.1%) and dataset 2 (1,953 /2,138, 91.3%), the consistency of subtypes by cluster analysis in two distinct datasets and total dataset showed in **Table 2**. The average hearing thresholds of normal hearing subjects and those four clusters are shown for each dataset (**Figures 2B–F**).

TABLE 2 | The consistency of subtypes by cluster analysis in two independent datasets and total dataset.

Consistency, *n* (%)	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Total
Dataset 1	1,549 (100.0)	1,123 (89.9)	1,075 (97.3)	539 (96.6)	4,286 (96.1)
Dataset 2	703 (91.4)	737 (94.0)	354 (83.1)	159 (100.0)	1,953 (91.3)
Total	2,252 (97.2)	1,860 (91.5)	1,492 (93.3)	698 (97.4)	6,239 (94.5)

Relevant Characteristics of Audiometric Phenotypes

Audiograms with hearing loss were then classified into 4 subtypes for cluster analysis of the total dataset, which were named 4–6 kHz sharp-notched (original cluster 1, **Figure 3A**), 4–6 kHz flat-notched (original cluster 2, **Figure 4A**), 3–8 kHz notched (original cluster 3, **Figure 5A**), and 1–8 kHz notched (original cluster 4, **Figure 6A**) phenotypes, referring to the frequency range, and shape of their audiometric notches. Hearing thresholds at frequencies of 0.5–8 kHz of the four subtypes were significantly different from each other (all the P values < 0.001). In comparison with the 4–6 kHz sharp- and flat-notched subtypes, subjects manifested as the 3–8 kHz and 1–8 kHz notched subtypes were significantly older, with higher noise exposure, as well as higher proportions of males, hearing difficulties and tinnitus. In *post-hoc* analyses, for the 4–6 kHz flat-notched audiogram, the average age of subjects of this subtype was similar ($P = 0.293$) to that of the 4–6 kHz sharp-notched audiogram, while the mean CNE was slightly smaller ($P = 0.008$) than that of the 4–6 kHz sharp-notched audiogram, but significantly larger ($P < 0.001$) than that of the normal-hearing audiogram. The proportions of females, hearing difficulties, tinnitus, and earphone uses were higher in subjects with the 4–6 kHz flat-notched audiogram than that in the 4–6 kHz sharp-notched audiogram. Moreover, the average hearing thresholds of the 4–6 kHz flat-notched audiogram at frequencies of 0.5–3 kHz were obviously higher than that of the 4–6 kHz sharp-notched audiogram (all the P value < 0.001). The detailed distribution of characteristics in subjects with different audiometric phenotypes is shown in **Table 3**.

Variables that showed significant differences between audiometric phenotypes were included in the multinomial logistic regression analysis (**Table 4**). Age, male sex, tobacco consumption, and alcohol consumption were risk factors for all subtypes, while the HPD use was a protective factor. CNE was associated with three of all subtypes except for the 4–6 kHz flat-notched phenotype. Tinnitus was associated with three of all subtypes except for the 4–6 kHz sharp-notched phenotype. Self-reported hearing difficulty was only related to the 1–8 kHz notched phenotype, which reflected the most severe NIHL subtype.

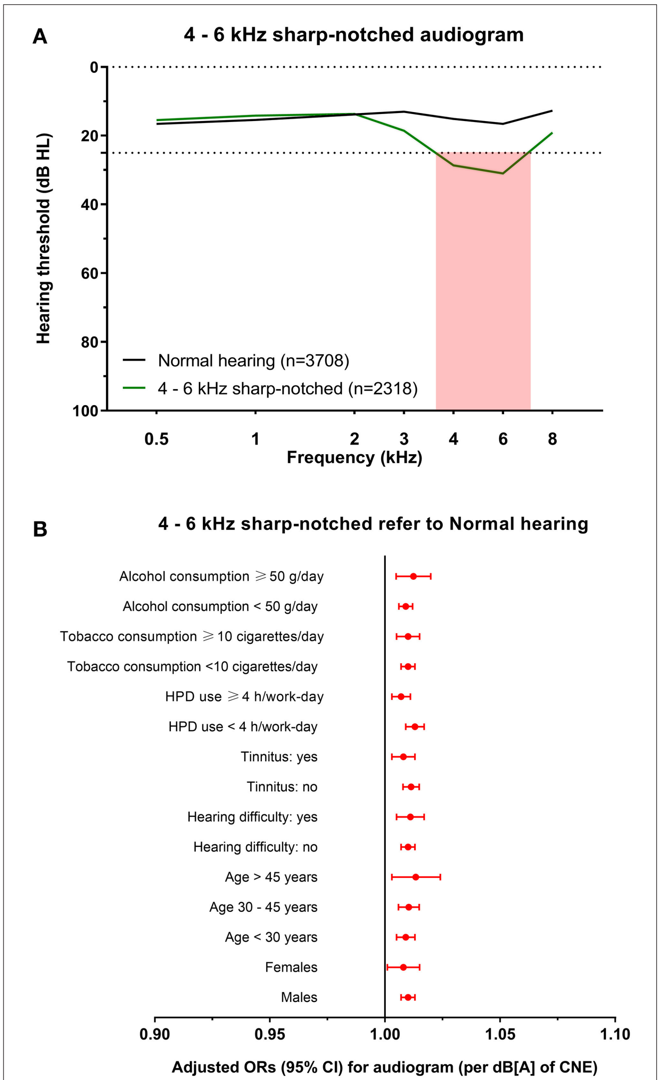


FIGURE 3 | The 4–6 kHz sharp-notched audiogram and its association with noise exposure. **(A)** The average hearing thresholds over 0.5–8 kHz frequencies of subjects with normal hearing and the cluster of 4–6 kHz sharp-notched audiogram. The pink shade includes the range of notched frequencies. **(B)** Adjusted ORs with 95% CI of CNE increment (per dB[A]) for the 4–6 kHz sharp-notched audiogram refer to normal hearing audiogram after stratification of sex, age, hearing difficulty, tinnitus, HPD use, tobacco consumption, and alcohol consumption.

Specific Influence of Noise Exposure on Audiometric Phenotypes

To explore the specific influence of noise exposure dose on audiometric phenotypes among populations with different characteristics, we performed hierarchical regression analysis of audiometric phenotypes stratified by confounding factors (sex, age, CNE, HPD use, hearing difficulty, tinnitus, tobacco consumption, and alcohol consumption). According to

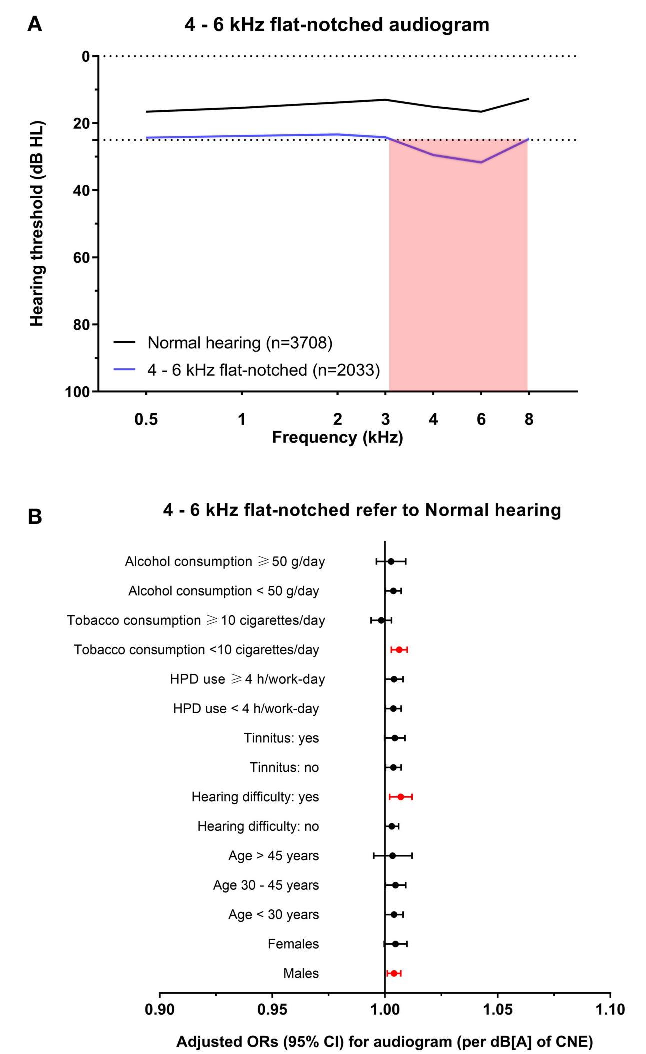


FIGURE 4 | The 4–6 kHz flat-notched audiogram and its association with noise exposure. **(A)** The average hearing thresholds over 0.5–8 kHz frequencies of subjects with normal hearing and the cluster of 4–6 kHz flat-notched audiogram. The pink shade includes the range of notched frequencies. **(B)** Adjusted ORs with 95% CI of CNE increment (per dB[A]) for the 4–6 kHz flat-notched audiogram refer to normal hearing audiogram after stratification of sex, age, hearing difficulty, tinnitus, HPD use, tobacco consumption, and alcohol consumption.

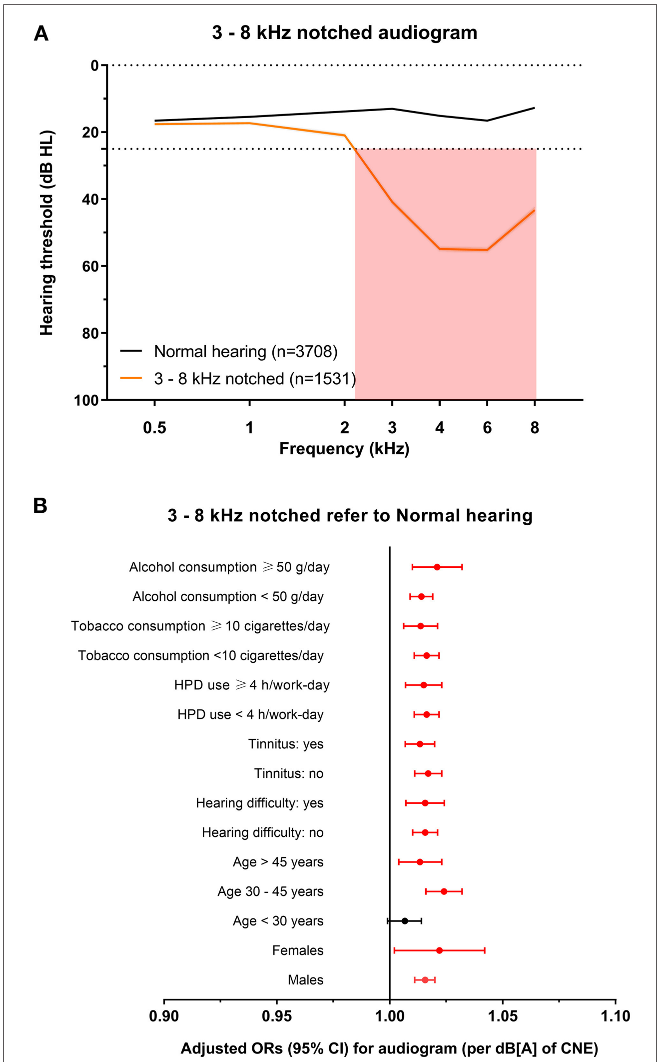


FIGURE 5 | The 3–8 kHz notched audiogram and its association with noise exposure. **(A)** The average hearing thresholds over 0.5–8 kHz frequencies of subjects with normal hearing and the cluster of 3–8 kHz notched audiogram. The pink shade includes the range of notched frequencies. **(B)** Adjusted ORs with 95% CI of CNE increment (per dB[A]) for the 3–8 kHz notched audiogram refer to normal hearing audiogram after stratification of sex, age, hearing difficulty, tinnitus, HPD use, tobacco consumption, and alcohol consumption.

the adjusted ORs of noise exposure dose for different phenotypes after stratification, the increment of CNE was stably associated with the 4–6 kHz sharp-notched phenotype (**Figure 3B**), as well as associated with the 3–8 kHz notched phenotype except among younger subjects (<30 years old) (**Figure 5B**) and the 1–8 kHz notched phenotypes except for females and younger population (**Figure 6B**). In contrast, CNE was almost unrelated to the 4–6 kHz flat-notched phenotype (**Figure 4B**), except for population who were males, with hearing difficulty and little tobacco consumption.

DISCUSSION

In this study we performed a cluster analysis of noise-exposed population who had some degree of hearing loss. By using the audiometric thresholds over 0.5–8 kHz of the total hearing loss dataset (n = 6,599), we developed the cluster model and identified four phenotypes with distinct audiogram subtypes of hearing loss. We repeated the cluster analysis in two independent parts of the total dataset, dataset 1 (n = 4,461) and dataset 2 (n = 2,138) where we were able to replicate the clusters into four similar phenotypes. The relevant demographic and behavioral characteristics of population with different hearing

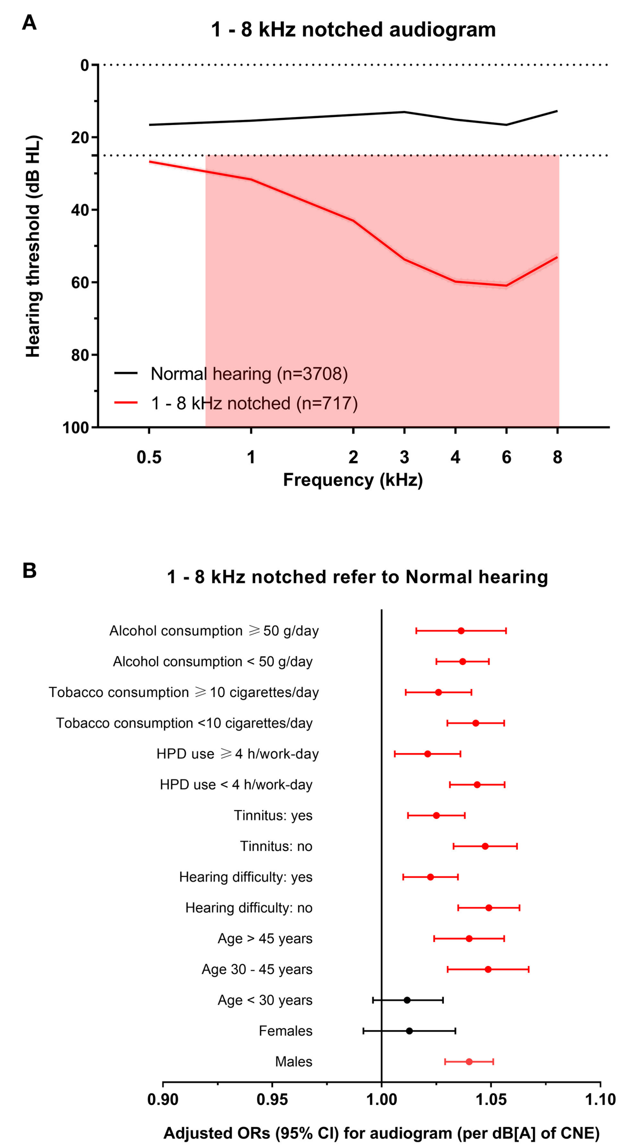


FIGURE 6 | The 1–8 kHz notched audiogram and its association with noise exposure. **(A)** The average hearing thresholds over 0.5–8 kHz frequencies of subjects with normal hearing and the cluster of 1–8 kHz notched audiogram. The pink shade includes the range of notched frequencies. **(B)** Adjusted ORs with 95% CI of CNE increment (per dB[A]) for the 1–8 kHz notched audiogram refer to normal hearing audiogram after stratification of sex, age, hearing difficulty, tinnitus, HPD use, tobacco consumption, and alcohol consumption.

loss phenotypes were analyzed in comparison with the normal hearing population ($n = 3,708$).

Our main finding was that hearing loss in noise-exposed population consisted of four audiogram subtypes that had different characteristics and associations with noise exposure levels. In line with previous studies, we found the presence of a "notch" at high frequencies of 3, 4, and 6 kHz with recovery at 8 kHz in most hearing loss audiograms, some of which extended to involve even 1 kHz and 2 kHz (14–16). Therefore, we named the phenotypes 4–6 kHz sharp-notched, 4–6 kHz flat-notched, 3–8 kHz notched, and 1–8 kHz notched audiograms.

In the present study, the 4–6 kHz sharp-notched audiogram, 3–8 kHz notched audiogram, and 1–8 kHz notched audiogram were strongly related to noise exposure, which represented three distinct subtypes of NIHL. This result supported the conventional description of noise-induced high-frequency audiometric notches (8, 9, 23) based on data-driven evidence. The occurrence of 4–6 kHz sharp-notched audiogram was highest among all subtypes with constant correlation to the noise exposure, which could be regarded as a typical subtype of NIHL. While the 3–8 kHz notched audiogram and 1–8 kHz notched audiogram that manifested as more severe subtypes of NIHL involved wider ranges of frequency, which were less likely to appear among younger populations and even females. This finding agreed with several previous studies suggesting that the risk of NIHL in males was significantly higher than that in females (12, 15, 24), as well as the effects of aging may extend the hearing loss frequencies to 8 kHz and even low frequencies, which reduces the prominence of the typical "notch" in audiograms of individuals with excess noise exposure (8, 9).

In particular, the 4–6 kHz flat-notched audiogram was the second most common subtype of hearing loss after the 4–6 kHz sharp-notched audiogram, however, it seemed to be unrelated to noise exposure, but associated with age, sex, and some behavioral factors according to the logistic regression (**Table 4**). Although the mean CNE of subjects with the 4–6 kHz flat-notched audiogram was significantly larger than that of the normal-hearing audiogram, it might due to the longer working-length of subjects with the 4–6 kHz flat-notched audiogram, who were also older than those with the normal-hearing audiogram. In addition, the average hearing thresholds of the 4–6 kHz flat-notched audiogram at lower frequencies were higher than that of the 4–6 kHz sharp-notched audiogram, despite of the similar mean age, and CNE. However, in consideration of the obvious differences in sex, hearing difficulty, tinnitus, and earphone use between the two subtypes, we speculated that there should be other factors (such as individual behaviors and genetic heterogeneity) influencing the audiometric phenotypes, which should be further explored in future studies. This finding may provide an explanation for some previous studies reporting that audiometric notches also commonly occur in individuals without any previous noise exposure and have been associated with other factors (14, 15, 25). The Nord-Trøndelag Hearing Loss Study analyzed the various definitions of notched audiograms in the 3–6 kHz range [defined by Coles et al. (9), Hoffman et al. (26), Wilson and Mcardle (27)] in 49 774 subjects aged 20–101 years. The prevalence of those notches varied from 60 to 70% in the most noise-exposed men, but was also common in men without any occupational noise exposure. Another study using the Hoffmann notch to analyze audiograms of US adults from the NHANES (16) showed that though 8.2% of 1,223 self-reported occupational noise-exposed individuals had bilateral high-frequency audiometric notches, 5.2% of 2,360 individuals without noise exposure also had bilateral notches. Those artificial definitions of notches probably included

TABLE 3 | Characteristics of subjects in different audiometric phenotypes.

Variables	Normal hearing (n = 3,708)	4–6 kHz sharp-notched (n = 2,318)	4–6 kHz flat-notched (n = 2,033)	3–8 kHz notched (n = 1,531)	1–8 kHz notched (n = 717)	P value*
Age (years), mean (SD)	29.9 (7.0)[b,c,d,e]	34.9 (7.7)[a,d,e]	35.4 (8.9)[a,d,e]	39.6 (8.1)[a,b,c,e]	43.5.4 (7.8)[a,b,c,d]	<0.001
CNE (dB[A]), mean (SD)	85.4 (27.8)[b,c,d,e]	93.8 (17.6)[a,c,d,e]	91.5 (22.6)[a,b,d,e]	96.9 (14.3)[a,b,c,e]	100.0 (13.3)[a,b,c,d]	<0.001
Sex, *n* (%)						<0.001
Males	3,065 (82.7)[b,c,d,e]	2,144 (92.5)[a,c,d,e]	1,775 (87.3)[a,b,d,e]	1,492 (97.5)[a,b,c]	689 (96.1)[a,b,c]	
Females	643 (17.3)[b,c,d,e]	174 (7.5)	258 (12.7)	39 (2.5)	28 (3.9)	
BMI, *n* (%)						0.209
non-obese	3,356 (90.5)	2,118 (91.4)	1,831 (90.1)	1,409 (92.0)	657 (91.6)	
Obese	352 (9.5)	200 (8.6)	202 (9.9)	122 (8.0)	60 (8.4)	
Hearing difficulty, *n* (%)						<0.001
No	2,914 (78.6)[c,d,e]	1,844 (79.6)[c,d,e]	1,555 (76.5)[a,b,e]	1,168 (76.3)[a,b,e]	474 (66.1)[a,b,c,d]	
Yes	794 (21.4)	474 (20.4)	478 (23.5)	363 (23.7)	243 (33.9)	
Tinnitus, *n* (%)						<0.001
No	2,600 (70.1)[c,d,e]	1,640 (70.8)[c,d,e]	1,370 (67.4)[a,b,d,e]	957 (62.5)[a,b,c,e]	404 (56.3)[a,b,c,d]	
Yes	1,108 (29.9)	678 (29.2)	663 (32.6)	574 (37.5)	313 (43.7)	
HPD use, *n* (%)						<0.001
<4 h/work-day	2,290 (61.8)[b,c,d,e]	1,719 (74.2)[a,d,e]	1,516 (74.6)[a,d,e]	1,256 (81.4)[a,b,c,e]	613 (85.5)[a,b,c,d]	
≥4 h/work-day	1,418 (38.2)	599 (25.8)	517 (25.4)	285 (18.6)	104 (14.5)	
Earphone use, *n* (%)						0.004
<1 h/day	2,151 (58.0)[c,d,e]	1,361 (58.7)[c,d,e]	1,112 (54.7)[a,b]	836 (54.6)[a,b]	384 (53.6)[a,b]	
≥1 h/day	1,557 (42.0)	957 (41.3)	921 (45.3)	695 (45.4)	333 (46.4)	
Tobacco consumption, *n* (%)						<0.001
<10 cigarettes/day	2,451 (66.1)[b,c,d,e]	1,371 (59.1)[a,c,d]	1,300 (63.9)[a,b,d,e]	886 (57.9)[a,c]	428 (59.7)[a,c]	
≥10 cigarettes/day	1,257 (33.9)	947 (40.9)	733 (36.1)	645 (42.1)	289 (40.3)	
Alcohol consumption, *n* (%)						<0.001
<50 g/day	3,023 (81.5)[b,c,d,e]	1,577 (68.0)[a,c,d]	1,498 (73.7)[a,b,d,e]	982 (64.1)[a,b,c,e]	478 (66.7)[a,c,d]	
≥50 g/day	685 (18.5)	741 (32.0)	535 (26.3)	549 (35.9)	239 (33.3)	

SD, standard deviation; BMI, body mass index; CNE, cumulative noise exposure; HPD, hearing protective device.
**Comparisons were between different audiometric phenotypes.*
Significantly different from normal hearing[a], 4–6 kHz sharp-notched[b], 4–6 kHz flat-notched[c], 3–8 kHz notched[d], and 1–8 kHz notched[e] audiograms in post-hoc analyses of ANOVA or chi-square test ($P < 0.05$).

TABLE 4 | Multinomial logistic regression models of audiometric phenotypes.

Variables	Refer to normal hearing			
(OR [95% CI])	4–6 kHz sharp-notched	4–6 kHz flat-notched	3–8 kHz notched	1–8 kHz notched
Age (per years)	**1.09** (1.08–1.09)	**1.10** (1.09–1.10)	**1.16** (1.15–1.17)	**1.23** (1.22–1.25)
CNE (per dB[A])	**1.01** (1.01–1.01)	1.00 (1.00–1.01)	**1.02** (1.01–1.02)	**1.04** (1.03–1.05)
Male sex	**2.76** (2.27–3.35)	**1.72** (1.44–2.05)	**9.63** (6.81–13.63)	**6.44** (4.25–9.75)
Hearing difficulty (self-reported yes)	0.92 (0.80–1.06)	1.06 (0.92–1.22)	0.91 (0.77–1.08)	**1.36** (1.10–1.67)
Tinnitus (self-reported yes)	1.03 (0.91–1.17)	**1.19** (1.04–1.35)	**1.53** (1.32–1.78)	**1.84** (1.51–2.24)
HPD use ≥4 h/work-day	**0.84** (0.74–0.95)	**0.86** (0.75–0.98)	**0.77** (0.65–0.90)	**0.76** (0.59–0.97)
Earphone use ≥1 h/day	0.9 (0.81–1.01)	1.10 (0.97–1.23)	1.07 (0.93–1.22)	1.01 (0.84–1.22)
Tobacco consumption ≥10 cigarettes/day	**1.17** (1.03–1.32)	**1.10** (0.97–1.25)	**1.26** (1.09–1.45)	**1.33** (1.09–1.62)
Alcohol consumption ≥50 g/day	**1.51** (1.32–1.72)	**1.27** (1.11–1.47)	**1.54** (1.32–1.8)	**1.25** (1.02–1.54)

Reference variables: Sex, females; Hearing difficulty, self-reported no; Tinnitus, self-reported no; HPD use, <4 h/work day; Earphone use, <1 h/day; Tobacco consumption, <10 cigarettes/day; Alcohol consumption, <50 g/day.
Bold type: $P < 0.05$.
OR, odds ratio; CI, confidence interval; BMI, body mass index; CNE, cumulative noise exposure; HPD, hearing protective device.

this 4–6 kHz flat-notched audiogram, which may limit the specificity of using high-frequency audiometric notch for the diagnosis of NIHL.

As many previous studies reported (11, 28, 29), we found that age, sex, tobacco, and alcohol consumption were confounding influencing factors of hearing loss other than noise exposure. Using HPDs in an environment with loud noise exposure for hours every work-day likely protected individuals from NIHL, despite audiometric subtypes. In addition, we found that tinnitus was associated with the degree of hearing loss rather than the most typical NIHL subtype, while self-reported hearing difficulty was only closely related to the most severe subtype of hearing loss with speech frequencies impairment. These findings are approximately consistent with previous studies that reported that tinnitus is usually accompanied by hearing loss (30), and self-reported hearing status could not sensitively reflect high-frequency hearing loss (31).

It is widely accepted that audiometric phenotypes are based on presumed underlying auditory histopathology, which suggests the causes and degree of auditory organ damage (32, 33). A few previous studies have performed cluster analysis in clinical audiograms. Interestingly, the notched audiometric phenotype was always distinguished out as a separate cluster (18, 19), and we assumed that it should indicate the NIHL phenotype, although the noise exposure history of those patients was not reported. Here we propose to use this cluster classification to identify audiometric phenotypes for the evaluation of NIHL, since the typical NIHL in a specific population may manifest as different subtypes of notched audiograms, and suggest different management approaches. For instance, the presence of 4–6 kHz sharp-notched audiogram in younger females might be a strong signal indicating NIHL, in contrast, the 4–6 kHz flat-notched audiogram should not be evidence of NIHL. This would facilitate optimal assessment of NIHL.

The main strength of our study is that it first provides various reproducible audiometric subtypes of NIHL by data-driven analysis in a relatively large-scale noise-exposed population. Another strength was that our study was based on consideration of detailed noise exposure history, questionnaire information and audiometric data from standardized protocols, which can give a more nuanced picture than clinical data. Previously Zhao et al. developed machine learning models for the prediction of NIHL (34), which were based on hypothesis-driven or supervised analysis. Instead, for the first time to our knowledge, we performed an unsupervised data-driven cluster analysis to identify the unknown audiometric phenotypes associated with noise exposure, and to describe the relevant characteristics of distinct subtypes of NIHL. However, there are also some limitations. First, this cross-sectional study did not allow robust causal inference, although the employees were supposed to have a pre-work health examination to ensure normal hearing at baseline. Second, all subjects in this study were collected in the same region of China and they may not represent the whole noise-exposed population. Furthermore, we cannot at this stage claim that the new subtypes represent different etiologies of NIHL, or that this clustering is the optimal classification of NIHL phenotypes.

In conclusion, we were able to repeat and identify distinct audiometric phenotypes of NIHL in large-scale noise-exposed populations with different relevant characteristics, by using cluster analysis. Moreover, given the technological advances in machine learning, our study provides a sight into the prospect of involving data-driven audiogram mining for the precise diagnosis and treatment of NIHL in future studies.

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 authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the Ninth People's Hospital affiliated to Shanghai Jiao Tong University School of Medicine. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

QW, MQ, ZH, and HW: contributed to the study design. QW and MQ: writing the original draft. MQ, LY, JS, YH, and KH: contributed to the acquisition of data. QW, CL, and JL contributed to the analysis of data, ZH and HW contributed to the supervision and funding acquisition of the work, review and revision of the draft. All authors contributed to the interpretation of data and critical revision of the draft. All authors gave final approval of the version to be published.

FUNDING

This study was supported by grant 81730028 to HW from the National Natural Science Foundation of China, grant 2017YFC1001804 to HW from the National Key Research and Development Program of China. This work was also supported by grant 20ZR1431200 to ZH from the Natural Science Foundation of Shanghai, grant YG2021ZD14 to ZH from the Medical Engineering Cross Research Funding of Shanghai Jiao Tong University.

ACKNOWLEDGMENTS

We thank all the subjects for their participation and the security department of shipyard for the cooperation.

REFERENCES

1. Masterson EA, Bushnell PT, Themann CL, Morata TC. Hearing impairment among noise-exposed workers - United States, 2003-2012. *MMWR Morb Mortal Wkly Rep*. (2016) 65:389–94. doi: 10.15585/mmwr.mm6515a2
2. Eichwald J, Scinicariello F. Survey of teen noise exposure and efforts to protect hearing at school United States, 2020. *MMWR Morb Mortal Wkly Rep*. (2020) 69:1822–6. doi: 10.15585/mmwr.mm6948a5
3. Yao C, Ma AK, Cushing SL, Lin VYW. Noise exposure while commuting in Toronto - a study of personal and public transportation in Toronto. *J Otolaryngol Head Neck Surg*. (2017) 46:62. doi: 10.1186/s40463-017-0239-6
4. Eichwald J, Scinicariello F, Telfer JL, Carroll YI. Use of personal hearing protection devices at loud athletic or entertainment events among adults - United States, 2018. *MMWR Morb Mortal Wkly Rep*. (2018) 67:1151–5. doi: 10.15585/mmwr.mm6741a4
5. Basner M, Babisch W, Davis A, Brink M, Clark C, Janssen S, et al. Auditory and non-auditory effects of noise on health. *Lancet*. (2014) 383:1325–32. doi: 10.1016/S0140-6736(13)61613-X
6. Wilson BS, Tucci DL, Merson MH, O'donoghue GM. Global hearing health care: new findings and perspectives. *Lancet*. (2017) 390:2503–15. doi: 10.1016/S0140-6736(17)31073-5
7. Lutman ME, Coles RR, Buffin JT. Guidelines for quantification of noise-induced hearing loss in a medicolegal context. *Clin Otolaryngol*. (2016) 41:347–57. doi: 10.1111/coa.12569
8. Mirza R, Kirchner DB, Dobie RA, Crawford J. Occupational noise-induced hearing loss. *J Occup Environ Med*. (2018) 60:e498–e501. doi: 10.1097/JOM.0000000000001423
9. Coles RR, Lutman ME, Buffin JT. Guidelines on the diagnosis of noise-induced hearing loss for medicolegal purposes. *Clin Otolaryngol Allied Sci*. (2000) 25:264–73. doi: 10.1046/j.1365-2273.2000.00368.x
10. Ali S, Morgan M, Ali UI. Is it reasonable to use 1 and 8 kHz anchor points in the medico-legal diagnosis and estimation of noise-induced hearing loss? *Clin Otolaryngol*. (2015) 40:255–9. doi: 10.1111/coa.12362
11. Morris RJ. On the need for an "Industry-Wide Standard" definition of an audiometric notch and the lack of specificity of the audiometric notch in the diagnosis of noise-induced hearing loss. *J Occup Environ Med*. (2019) 61:e226. doi: 10.1097/JOM.0000000000001550
12. Lie A, Skogstad M, Johnsen TS, Engdahl B, Tambs K. A cross-sectional study of hearing thresholds among 4627 Norwegian train and track maintenance workers. *BMJ Open*. (2014) 4:e005529. doi: 10.1136/bmjopen-2014-005529
13. Peng Y, Fan C, Hu L, Peng S, Xie P, Wu F, et al. Tunnel driving occupational environment and hearing loss in train drivers in China. *Occup Environ Med*. (2019) 76:97–104. doi: 10.1136/oemed-2018-105269
14. Nondahl DM, Shi X, Cruickshanks KJ, Dalton DS, Tweed TS, Wiley TL, et al. Notched audiograms and noise exposure history in older adults. *Ear Hear*. (2009) 30:696–703. doi: 10.1097/AUD.0b013e3181b1d418
15. Lie A, Skogstad M, Johnsen TS, Engdahl B, Tambs K. The prevalence of notched audiograms in a cross-sectional study of 12,055 railway workers. *Ear Hear*. (2015) 36:e86–92. doi: 10.1097/AUD.0000000000000129
16. Carroll YI, Eichwald J, Scinicariello F, Hoffman HJ, Deitchman S, Radke MS, et al. Vital signs: noise-induced hearing loss among adults - United States 2011-2012. *MMWR Morb Mortal Wkly Rep*. (2017) 66:139–44. doi: 10.15585/mmwr.mm6605e3
17. Lisboa PJ, Etchells TA, Jarman IH, Chambers SJ. Finding reproducible cluster partitions for the k-means algorithm. *BMC Bioinformatics. (*2013) 14(Suppl 1):S8. doi: 10.1186/1471-2105-14-S1-S8
18. Lee CY, Hwang JH, Hou SJ, Liu TC. Using cluster analysis to classify audiogram shapes. *Int J Audiol*. (2010) 49:628–33. doi: 10.3109/14992021003796887
19. Parthasarathy A, Romero Pinto S, Lewis RM, Goedicke W, Polley DB. Data-driven segmentation of audiometric phenotypes across a large clinical cohort. *Sci Rep*. (2020) 10:6704. doi: 10.1038/s41598-020-63515-5
20. Zhao YM, Qiu W, Zeng L, Chen SS, Cheng XR, Davis RI, et al. Application of the kurtosis statistic to the evaluation of the risk of hearing loss in workers exposed to high-level complex noise. *Ear Hear*. (2010) 31:527–32. doi: 10.1097/AUD.0b013e3181d94e68
21. R Foundation for Statistical Computing. *R: A Language and Environment for Statistical Computing*. Austria (2020). Available online at: https://www.R-project.org/ (accessed December 20, 2020).
22. Steinley D, Brusco MJ. Choosing the number of clusters in κ-means clustering. *Psychol Methods*. (2011) 16:285–97. doi: 10.1037/a0023346
23. Mcbride DI, Williams S. Audiometric notch as a sign of noise induced hearing loss. *Occup Environ Med*. (2001) 58:46–51. doi: 10.1136/oem.58.1.46
24. Bhatt IS, Guthrie O. Analysis of audiometric notch as a noise-induced hearing loss phenotype in US youth: data from the National Health And Nutrition Examination Survey, 2005-2010. *Int J Audiol*. (2017) 56:392–9. doi: 10.1080/14992027.2017.1278799
25. Osei-Lah V, Yeoh LH. High frequency audiometric notch: an outpatient clinic survey. *Int J Audiol*. (2010) 49:95–8. doi: 10.3109/14992020903300423
26. Hoffman HJ, Ko CW, Themann CL, Dillon CF, Franks JR. Reducing Noise-induced Hearing Loss (NIHL) to achieve us healthy people 2010 goals. *Am J Epidemiol*. (2006) 163:S122. doi: 10.1093/aje/163.suppl_11.S122-c
27. Wilson RH, Mcardle R. Characteristics of the audiometric 4,000 Hz notch (744,553 veterans) and the 3,000, 4,000, and 6,000 Hz notches. (539,932 veterans). *J Rehabil Res Dev*. (2013) 50:111–32. doi: 10.1682/JRRD.2011.11.0225
28. Tikka C, Verbeek JH, Kateman E, Morata TC, Dreschler WA, Ferrite S. Interventions to prevent occupational noise-induced hearing loss. *Cochrane Database Syst Rev*. (2017) 7:Cd006396. doi: 10.1002/14651858.CD006396.pub4
29. Golmohammadi R, Darvishi E. The combined effects of occupational exposure to noise and other risk factors - a systematic review. *Noise Health*. (2019) 21:125–41. doi: 10.4103/nah.NAH_4_18
30. Baguley D, Mcferran D, Hall D. Tinnitus. *Lancet*. (2013) 382:1600–7. doi: 10.1016/S0140-6736(13)60142-7
31. Tsimpida D, Kontopantelis E, Ashcroft D, Panagioti M. Comparison of self-reported measures of hearing with an objective audiometric measure in adults in the english longitudinal study of ageing. *JAMA Netw Open*. (2020) 3:e2015009. doi: 10.1001/jamanetworkopen.2020.15009
32. Schmiedt RA, Lang H, Okamura HO, Schulte BA. Effects of furosemide applied chronically to the round window: a model of metabolic presbyacusis. *J Neurosci*. (2002) 22:9643–50. doi: 10.1523/JNEUROSCI.22-21-09643.2002
33. Demeester K, Van Wieringen A, Hendrickx JJ, Topsakal V, Fransen E, Van Laer L, et al. Audiometric shape and presbycusis. *Int J Audiol*. (2009) 48:222–32. doi: 10.1080/14992020802441799
34. Zhao Y, Li J, Zhang M, Lu Y, Xie H, Tian Y, et al. Machine learning models for the hearing impairment prediction in workers exposed to complex industrial noise: a pilot study. *Ear Hear*. (2019) 40:690–9. doi: 10.1097/AUD.0000000000000649

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.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 12 April 2021
doi: 10.3389/fnins.2021.643205

Effects of Wireless Remote Microphone on Speech Recognition in Noise for Hearing Aid Users in China

*Jing Chen, Zhe Wang, Ruijuan Dong, Xinxing Fu, Yuan Wang and Shuo Wang**

Beijing Institute of Otolaryngology, Otolaryngology—Head and Neck Surgery, Beijing Tongren Hospital, Capital Medical University, Beijing, China

Objective: This study was aimed at evaluating improvements in speech-in-noise recognition ability as measured by signal-to-noise ratio (SNR) with the use of wireless remote microphone technology. These microphones transmit digital signals via radio frequency directly to hearing aids and may be a valuable assistive listening device for the hearing-impaired population of Mandarin speakers in China.

Methods: Twenty-three adults (aged 19–80 years old) and fourteen children (aged 8–17 years old) with bilateral sensorineural hearing loss were recruited. The Mandarin Hearing in Noise Test was used to test speech recognition ability in adult subjects, and the Mandarin Hearing in Noise Test for Children was used for children. The subjects' perceived SNR was measured using sentence recognition ability at three different listening distances of 1.5, 3, and 6 m. At each distance, SNR was obtained under three device settings: hearing aid microphone alone, wireless remote microphone alone, and hearing aid microphone and wireless remote microphone simultaneously.

Results: At each test distance, for both adult and pediatric groups, speech-in-noise recognition thresholds were significantly lower with the use of the wireless remote microphone in comparison with the hearing aid microphones alone ($P < 0.05$), indicating better SNR performance with the wireless remote microphone. Moreover, when the wireless remote microphone was used, test distance had no effect on speech-in-noise recognition for either adults or children.

Conclusion: Wireless remote microphone technology can significantly improve speech recognition performance in challenging listening environments for Mandarin speaking hearing aid users in China.

Keywords: hearing aids, sensorineural hearing loss, signal-to-noise ratio, wireless remote microphone, speech-in-noise recognition

OPEN ACCESS

Edited by:
Fan-Gang Zeng,
University of California, Irvine,
United States

Reviewed by:
Yi Yuan,
University of Florida, United States
Jan-Willem A. Wasmann,
Radboud University Nijmegen Medical Centre, Netherlands

****Correspondence:***
Shuo Wang
shannonwsh@aliyun.com

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *19 December 2020*
Accepted: *18 March 2021*
Published: *12 April 2021*

Citation:
Chen J, Wang Z, Dong R, Fu X, Wang Y and Wang S (2021) Effects of Wireless Remote Microphone on Speech Recognition in Noise for Hearing Aid Users in China. Front. Neurosci. 15:643205. doi: 10.3389/fnins.2021.643205

INTRODUCTION

With the advancement and evolution of current hearing technology, a variety of digital audio electronic devices have become more prevalent in the general population. These accessories which allow wireless transfer of high quality audio, such as smartphone compatible wireless earphones, have enhanced listening experiences for normal-hearing listeners, with the potential of similar accessories improving listening experiences for the hearing-impaired as well (Picou, 2020).

Sensorineural hearing loss not only reduces the sensitivity to sound and dynamic range of auditory perception for hearing-impaired listeners, but also reduces their frequency and temporal resolution that can lead to difficulty in speech comprehension (Moore, 2013). Hearing aids have been proven to be an effective solution in compensating for hearing loss in the loudness domain, but cannot compensate sufficiently for issues with frequency or temporal resolution. These issues become more pronounced in listening environments where the target sounds are masked by competing sounds. An effective way to help people with sensorineural hearing loss in more challenging listening environments is to improve the audibility of the target signal. The signal-to-noise ratio (SNR) is defined as the ratio of speech signals to noise and is frequently used to indicate the quality of the target signal in challenging communication environments. Research has shown that the speech-in-noise recognition ability in people with sensorineural hearing loss is significantly reduced when the SNR is at or below 5 dB. In contrast, the speech-in-noise recognition performance for normal-hearing listeners are minimally impacted at a SNR of 0 dB (Dong et al., 2015). Wilson et al. (2007) reported that individuals with a moderate hearing loss required an increased SNR of up to 10 dB to achieve the same speech understanding as individuals with normal hearing. Generally speaking, SNR depends on three key determinants: the presence of background noise, the distance between the listener and the target signal, and the reverberation in the listening environment. If the competing noise level is stable and the distance between the listener and target speech increases, the effective SNR for the listener will decrease. Studies have shown that to achieve better speech signal clarity, the distance between the listener and the signal source should be no greater than 1.8 m (Fickes, 2003). Blazer (2007) reported that students with hearing loss were able to achieve 95% on speech recognition tasks when they were 1.8 m apart from the source of interest, and only 60% when they were 7.3 m apart from the source of interest. In addition to the target distance, Reverberation Time (RT) plays an important role as well: the longer the RT in the communication environment, the more difficult it is for people with sensorineural hearing loss to communicate. Studies have shown that reducing reverberation time from 1.2 to 0.3 s can lead to 11 dB improvement in SNR (David and William, 1984). Furthermore, some studies have shown that optimal SNR for speech perception is dependent on a child's age, with younger children requiring a more favorable SNR to obtain similar speech recognition scores as adults. Adult-like performance is reached at the age of 10–12 years in stationary noise conditions (Koopmans et al., 2018).

Hearing aids today can provide listeners with a clear, high-quality sound experience in a quiet environment, but they deteriorate in the presence of noise (Bentler et al., 2016). Kochkin (2002) found that nearly 45% of hearing aid users were dissatisfied with their hearing aid performance in a noisy environment. One of the main goals of the development of current hearing aid technology is to improve speech recognition in complex listening environments and improve SNR in conditions where noise, distance, and reverberation can interfere. One of the technologies, directional microphones, can significantly improve speech recognition in noise for specific listening environments. Directional microphones perform best when the speech is presented from the front, the noise is in the rear, and the target speaker is in close proximity. However, in a real-world communication environment, directional microphones may fall short as these conditions are often not met (Kreikemeier et al., 2013; Picou et al., 2014).

When hearing aids and their microphones alone do not provide adequate assistance, some of the most beneficial wearable technology comes in the form of assistive listening devices which can also effectively improve the SNR for hearing aid users. For example, remote microphone hearing assistance technology (HAT) is widely used for hearing-impaired children. Amongst various HAT devices, wireless frequency modulation (FM) system is an early development that is still widely used. A typical FM unit consists of a small transmitter and microphone, which picks up the voice of a speaker and sends the clean speech to a radio-frequency (RF) receiver plugged into the hearing aid of a listener. Using a FM system to transmit the teacher's voice directly to the student is equivalent to reducing the communication distance to within 3–6 inches. Boothroyd showed that using the FM system in a noisy environment resulted in the same speech recognition as in a quiet environment (Boothroyd and Guerrero, 2004). For FM systems, American Speech-Language-Hearing Association (ASHA) recommends, ". . . the basic goal is that the FM system should increase the level of the perceived speech, in the listener's ear, by approximately 10 dB" [American Speech-Language-Hearing Association (ASHA), 2002]. Lewis et al. (2004) reported that on average, subjects improved by 14.2–16.7 dB with the use of one FM receiver over the use of two hearing aids alone in the directional microphone mode. Current hearing aids often utilize digital radio frequency technology, such as Bluetooth Remote Microphones, which wirelessly connect to hearing aids. These devices function similarly to FM systems and can wirelessly transmit audio signals to hearing aids over long distances (up to 10 m). In most hearing aid applications, this technology does not require an extra receiving device like a traditional FM system, as the digital wireless antenna is built into the hearing aids. This makes them more convenient to carry and operate than traditional FM systems. Research has clearly indicated that the use of remote microphone systems statistically improved speech recognition in noise, relative to unaided and hearing aid-only conditions for adults with hearing loss (Jace et al., 2015; Rodemerk and Galster, 2015).

Mandarin Chinese is a tonal language with four phonologically distinctive tones characterized by syllable-level fundamental frequency (F0) contour patterns. These pitch contours are commonly described as high–level (tone 1), high–rising (tone 2), falling–rising (tone 3), and high–falling (tone 4) (Lin, 1988). According to a study of the hearing disabled population from four provinces in 2016, about 5% of the population in mainland China have hearing impairment (Hu et al., 2016). However, it was speculated that only about 7–10% of those hearing impaired listeners have been fitted with hearing aids (Zhang, 2009), suggesting a large unreached population of hearing-impaired Chinese listeners that could benefit from the use of hearing aids and the assistive listening devices.

Previous studies (Jace et al., 2015; Rodemerk and Galster, 2015; Bentler et al., 2016) have shown that the remote microphone HAT can significantly improve the speech recognition ability of the hearing-impaired people who communicate in English in noise. However, there are few reports on the application of this technology in the hearing impaired population who speak Mandarin Chinese. It is our interest to investigate how much improvement Chinese hearing-impaired listeners may benefit from the current wireless remote microphone technology. This study was aimed at evaluating improvements in speech-in-noise recognition ability as measured by signal-to-noise ratio (SNR) with the use of wireless remote microphone technology.

MATERIALS AND METHODS

Subjects

Two groups of subjects were recruited in this study. All subjects had a history of digital hearing aids use for more than 1 year but no experience with HAT in combination with their hearing devices. Twenty-three native Mandarin Chinese-speaking adult subjects (6 females and 17 males) participated in the adult group. The subjects were between 19 and 80 years old (Mean = 63.4, *SD* = 18.7) and had relatively symmetric sensorineural hearing loss in both ears. The mean pure-tone hearing thresholds at frequencies of 500, 1,000, 2,000, and 4,000 Hz ($PTA_{0.5 to\ 4\ kHz}$) across the two ears for the groups of subjects ranged from 40 to 75 dB HL, as shown in **Figure 1**.

Fourteen native Mandarin Chinese-speaking children (7 females and 7 males) were recruited for the children's group. These subjects were between 8 and 17 years old (Mean = 13.1, *SD* = 3.2), and had bilateral sensorineural hearing loss with a $PTA_{0.5 to\ 4\ kHz}$ ranging from approximately 35 to 80 dB HL, as shown in **Figure 1**.

Hearing Aid Fitting Equipment

In this study, adult participants were fitted with two ReSound LiNX2 962 Receiver-In-The-Ear (RIE) hearing aids, and pediatric participants were fitted with two ReSound UP 988 Behind-the-Ear (BTE) hearing aids. Pediatric participants utilized their own earmolds during the test, which were coupled to hearing aids. Pediatric subjects with good low frequency hearing had earmolds with small vents., which would have negligible effect on the gain of the amplified sound path (Dillon, 2012). A ReSound Mini Microphone was paired to the test hearing aids in all cases. The ReSound Mini Microphone is a small personal streaming device which utilizes 2.4 GHz digital wireless technology to transmit sounds from the remote microphone or the output signal from any external audio source, directly to a Resound hearing aid. The remote microphone can be clipped onto the speaker's clothing or placed on a surface to transmit the voices of multiple speakers. It provides a wireless link between the speaker and the listener with no additional hardware or connections required. The audio frequency range of the Mini Microphone is from 100 to 8,000 Hz. A remote control was used to change the hearing aid program during the test. The ReSound Aventa 3.10 software was used to fit hearing aids for subjects.

Test Equipment and Materials

The test equipment included five Audioengine2 + active speakers, four of which were used to present noise signals and one was used to present the target speech. The Mandarin Hearing in Noise Test (MHINT) was used to test the speech recognition ability for adults. The HINT is a standardized and recorded test that can be used to estimate the signal-to-noise ratio (SNR) at which the sentences embedded in background noise can be repeated correctly 50% of the time. MHINT is the Mandarin version of HINT. The MHINT materials consist of 12 lists, each containing 20 sentences. Each sentence contains 10 Chinese characters. The presentation level is response dependent; lowered or raised according to a participant's correct or incorrect response of the test material. Presentation levels were decreased by 2 dB after a correct response and increased by 2 dB after an incorrect response. The reception threshold for sentences (RTSs) was calculated using this adaptive procedure (Wong et al., 2007). The Mandarin Hearing in Noise Test for Children (MHINT-C) was used to test the children. The MHINT-C materials consist of 12 lists, each containing 10 sentences (Chen and Wong, 2020). For each test condition, a list of 10 MHINT sentences were presented in a randomized order. Speech-shaped noise (SSN) was used as the masker noise in the present study.

Test Environment

The test location was a quiet meeting room measured at 5.5 × 8 × 2.5 m, with ambient background noise levels below 45 dB A. The testing room resembles a typical classroom setting with a reverberation time of 0.46 s. As shown in **Figure 2**, four speakers were utilized for the noise located at the four corners of the room, 0.75 m away from the walls in the corners (N1–N4) of the test room, 45 degrees away from the two vertical walls, facing the center of the room. The participants were seated at the S0, 1.5 m away from the back wall. The speakers presenting speech signals were located at 0°azimuth distanced at 1.5, 3, and 6 m directly in front of the participant (S1, S2, and S3 conditions, respectively), resulting in a critical distance of 1.41 m. The wireless remote microphone was set to directional and clipped to the participant's collar. In the test, the wireless remote microphone was suspended 25 cm below the speaker to simulate the distance and orientation between the speaker's mouth and the Mini Microphone in practical applications. The speaker was set at ear level for each participant. Four speakers labeled N1, N2, N3, and N4 were used to deliver speech-shaped noise simultaneously at a calibrated noise level of 65 dB A at S0.

Test Procedures

Adult subjects were fitted with ReSound LiNX2 962 RIE hearing aids on both ears according to their audiograms. The proprietary fitting prescription of Audiogram + in the ReSound Aventa 3 software was used. The hearing aid microphones were set to a fixed directional response, and all other advanced signal processing features (e.g., directional processing, digital noise reduction, wind noise reduction, reverberation processing, frequency lowering) were disabled. The ReSound Mini Microphone was paired with the hearing aids and set at

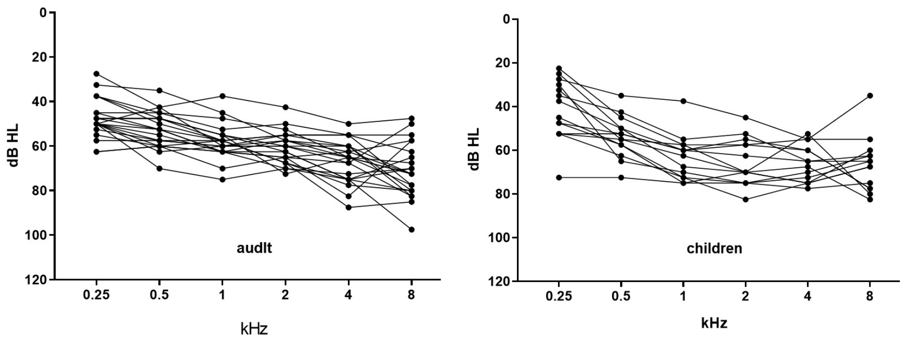


FIGURE 1 | The average hearing thresholds between the two ears of each individual adult and child. The horizontal axis represents the frequency. The vertical axis represents the hearing thresholds in dB HL. Each dot represents the average threshold of the left and right ears at this frequency of one individual.

a hearing aid to Mini Microphone ratio of 1:1. (i.e., remote microphone and hearing aid microphone were set such that each contributed equally to the output audio signal). There were nine test conditions consisting of three test distances and three program settings. Each program setting was tested at each distance condition. The program settings included directional microphones active (HA_D), the Mini Microphone active only (MM), and both hearing aid microphones and MM active (HA_D + MM). The nine test conditions were carried out in a randomized order. The remote control was used by the investigator to switch the hearing aid program settings. For each test condition, a list of MHINT materials were presented to obtain the speech recognition threshold in SNR. During the test, the SSN noise level was fixed at 65 dB A.

Pediatric subjects were fitted with bilateral Resound UPS988-DLW BTE hearing aids based on their audiograms with a DSL v5 fitting prescription. The programming of the hearing aids were the same as those utilized for the adult subjects, with the addition of an omnidirectional microphone mode (HA_O) program. Hence, for the pediatric subjects, there were twelve test conditions (four test program settings at three distances each) carried out in a randomized order. For each test condition, a list of MHINT-C materials were used to obtain recognition thresholds in SNR.

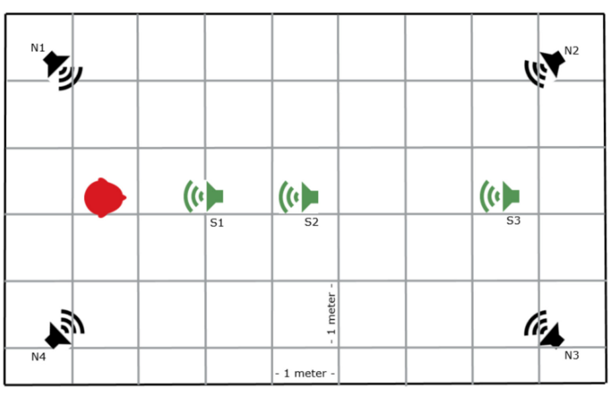


FIGURE 2 | Five speaker set up in the testing room. Four speakers playing the noise placed in the four corners (N1–N4) of the test room, and 0.75 m from the wall corner. The speaker playing speech signals were located at S1, S2, and S3, respectively. S0 is the location of the subject (Red mark).

Statistical Methods

Statistical analysis was carried out using Statistics Package for Social Science (SPSS) 16.0. A repeated-measures analysis of variance (RM-ANOVA) was conducted to determine if there was any overall statistical significance among the outcome SNRs across the three or four device settings at the three test distances for both adult and pediatric groups. The test distances and device settings were considered the independent variables. SNR was considered the dependent variable.

RESULTS

Speech Recognition for Adults

The SNRs obtained under nine test conditions for adults are detailed in **Table 1**. The RM-ANOVA revealed a statistically significant difference among each device setting [$F(2, 65) = 267.91$, $p < 0.05$]. There was a significant interaction effect between the distance and device settings [$F(4, 62) = 7.77$, $p < 0.05$]. This indicates that SNR is affected by the interaction between the distance and device settings.

Figure 3 illustrates the speech-in-noise recognition threshold (measured in SNR) of adult subjects with different test distances and different test device settings. The results showed that at the same test distance, the SNR thresholds under three device settings were significantly different from each other ($p < 0.05$). Performance was significantly better when the MM was active compared to the hearing aid microphone alone ($p < 0.05$). The performance with the MM alone was significantly better than the performance with HA_D + MM (all $p < 0.05$).

Moreover, the results showed that with the hearing aid microphone alone, the SNR for the 1.5 m condition was significantly better than those for the 3–6 m conditions ($p < 0.05$), with no significant differences in SNR between 3 and 6 m conditions. When the MM was active, the test distance had no effect on SNRs ($p > 0.05$).

TABLE 1 | The speech in noise recognition thresholds in SNRs (dB) obtained under each test condition for both adults and children.

	Adult			Children			
	HA__D	HA + MM	MM	HA__0	HA__D	HA + MM	MM
1.5 m	5.55 ± 5.90	−2.23 ± 6.87	−5.46 ± 7.50	4.09 ± 5.56	1.66 ± 1 ± 6.19	−4.37 ± 5.66	−9.26 ± 1 ± 5.11
3 m	11.77 ± 6.90	−3.30 ± 6.99	−5.88 ± 1 ± 6.71	8.16 ± 1 ± 5.28	6.77 ± 6.12	−3.66 ± 7.23	−9.65 ± 4.55
6m	13.11 ± 7.08	−4.87 ± 5.96	−7.53 ± 7.08	10.16 ± 8.92	8.47 ± 8.14	-4.47 ± 8.10	−10.38 ± 5.00

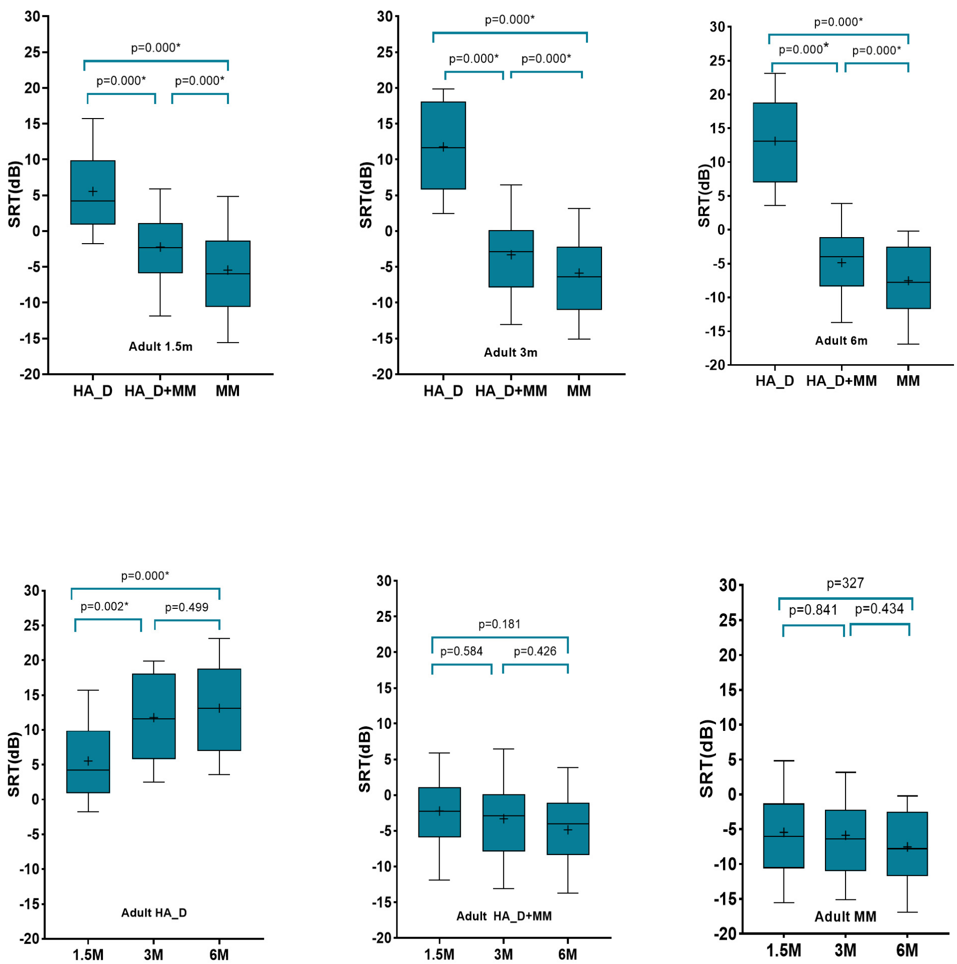


FIGURE 3 | Comparison of speech in noise recognition threshold (measured in SNR) of adult subjects with different test distances and different device settings. Boundaries of the boxes indicate the 25th and 75th percentile. Whiskers indicate the 10th and 90th percentiles. Solid lines denote the median. Plus denotes the mean.

Speech Recognition for Children

The SNRs obtained under twelve test conditions for children are detailed in **Table 1**. The RM-ANOVA revealed a statistically significant difference among each device setting [F (2, 65) = 267.91, $p < 0.05$]. There was a significant interaction effect between test distance and device settings ($p < 0.05$). This indicates that the SNR is affected by the interaction between the distance and device settings.

Figure 4 illustrates the speech-in-noise recognition threshold (measured in SNR) of children subjects with varying test distances and test device settings. The results showed that the SNR thresholds under 1.5 m conditions for four device settings were significantly different from each other ($p < 0.05$). Performance was significantly better for the two conditions with the MM active, in comparison with the hearing aid microphone alone ($p < 0.05$) regardless of microphone directionality. The performance for the MM alone was significantly better than the performance with MM + HA (all $p < 0.05$). When the test distance was set at 3 and 6 m, there were no differences in performance between HA_O and HA_D.

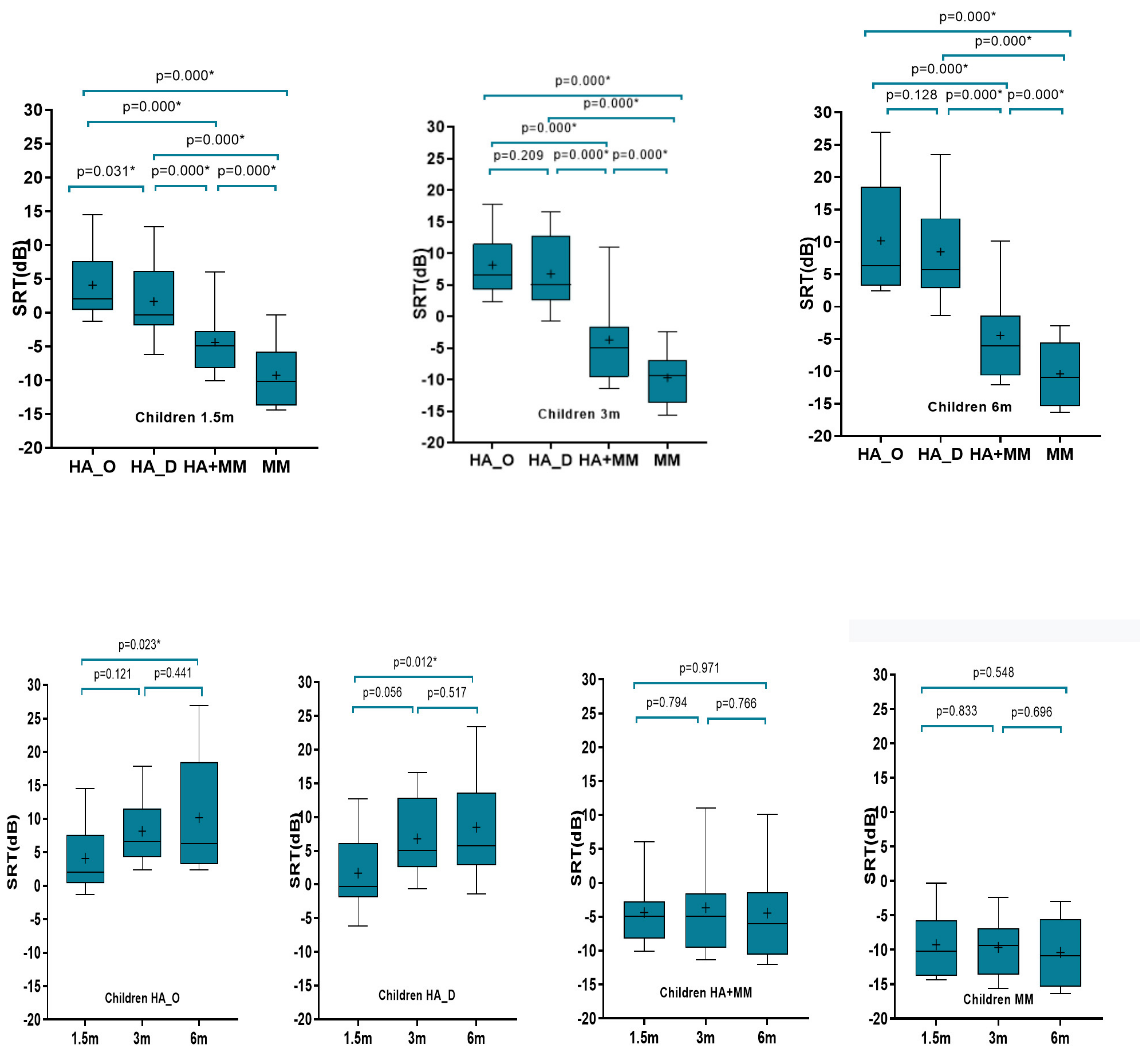


FIGURE 4 | Comparison of speech in noise recognition threshold (measured in SNR) of pediatric subjects at different test distances and different test device settings. Boundaries of the boxes indicate the 25th and 75th percentile. Whiskers indicate the 10th and 90th percentiles. Solid lines denote the median. Plus (+) denotes the mean.

The results show that using the hearing aid microphone alone, regardless if directional or omnidirectional, the SNR in the 1.5 m condition was significantly better than that for the 6 m condition, with no significant difference in SNR between 1.5 and 3 m, 3 and 6 m. When the MM was active, the test distance had no effect on speech in noise recognition thresholds ($p > 0.05$).

DISCUSSION

A remote microphone can be connected with a hearing aid and/or a cochlear implant to improve the speech recognition ability for patients with sensorineural hearing loss in noise. In a study of children with moderate to severe sensorineural hearing loss wearing bilateral hearing aids, Lewis et al. investigated the effects of remote microphone technology on speech perception in noise relative to hearing aid only conditions. Results revealed that the use of bilateral FM audio streaming significantly improved SNR relative to the omnidirectional hearing aids alone by 16–22.7 dB, confirming that adult listeners with hearing loss benefit from the use of a personal remote microphone system (Lewis et al., 2004). Research by Jantien et al. (2017) showed that using a wireless remote microphone in a noisy environment improved Speech Reception Thresholds (SRTs)in adults with bilateral cochlear implants by 5.4 dB. In a study with preschool children using remote microphones as personal wireless systems with cochlear implants, Clare and Gill (2018) showed that with professional guidance and training at home, this technology has the potential to improve education and communication environments for preschool children. In the present study, the benefit of using the remote microphone was consistent in all three test distances for both adults and children. Speech in noise recognition thresholds, measured in SNR, at all test distances decreased significantly, indicating a significant improvement in the speech recognition performance in noise. Compared with the HA_D setting, the results with the MM improved by 11–19.5 dB for adults and by 10–18.9 dB for children. The benefits of adding the MM compared to the HA alone increased as the test distance increased. Regardless of whether the HA was set to omnidirectional or fixed directional, increases in distance resulted in a rapid decrease in SNR for the hearing aid microphones only condition. In the conditions where MM was used, the distance between the MM and sound source remained constant even though the listener wearing hearing aids moved further away. Thus, the SNR at the location where the target speech was detected remained the same. The use of a wireless remote microphone could very well improve the problem of reduced signal-to-noise ratios due to greater distances by increasing the desired sound in communication environments.

In this study, we found that for both adult and pediatric subjects, speech in noise recognition thresholds using the MM alone were significantly better than using HA_D + MM. This finding is similar to Linda et al., who reported that when using a FM-only microphone setting, the SNR is determined primarily by the SNR of the FM microphone; when both HA and FM microphones are active, the SNR is determined by the highest level of the speech, which is typically at the FM microphone, and the highest level of noise at either the FM or HA microphone. Linda showed that better performance was observed in the FM-only compared to FM + HA condition. The amount of improvement in the SNR is determined by the levels of noise at the FM and HA microphones. When the noise levels are similar at the two microphones, an improvement in the SNR of 2 dB is expected (Linda and Kristen, 2016). In the present study, the use of MM alone could provide 3 dB improvement in SNR compared to HA_D + MM settings for adult subjects, and 6.5 dB improvement for pediatric subjects. This phenomenon was more distinct in children. Compared with adults, it is difficult for children to concentrate on listening tasks in low SNR conditions (Ryan and Patricia, 2011). Therefore, the negative impact of low SNR listening environments on children is greater than that seen in adults. The results of this study also showed that children are more likely to experience "floor effects" than adults in hearing aid microphone only conditions.

It has been established that the use of hearing aids with directional microphones can improve speech communication in noise for people with sensorineural hearing loss; however, varying degrees of improvement have been reported. Early research showed that directional microphones can improve speech in noise by 6–8 dB compared to omnidirectional microphones (Gravel et al., 1999; Kuk et al., 1999). Ricketts et al. performed HINT tests on 47 adults using five different hearing aids to evaluate the advantages of directional microphones. Speech was presented from a 0° azimuth with simulated cafeteria noise presented from 30°, 105°, 180°, 225°, and 330° azimuths (Ricketts et al., 2001). An average directional benefit of 2.2–2.9 dB compared to omnidirectional microphones was reported (Ryan and Patricia, 2011). In the present study, when the test distance was 1.5 m, the directional microphone responses were significantly better than omnidirectional responses for children. When the test distance increased to 3 and 6 m, there were no significant differences among the directional and the omnidirectional microphones. The directional microphone advantage disappeared, the possible reason being that the increase in test distance led to the rapid decline of signal-to-noise ratio, resulting in the "floor" effect. The directional microphone loses its advantage under the condition of low signal-to-noise ratio, which leads to no significant differences among the directional and the omnidirectional microphones.

In this study, when listening via the hearing aid microphones only, speech in noise recognition performance in adult subjects decreased as the test distance increased from 1.5 to 3 m. However, in children subjects, the same trend as that of adults was observed, but did not reach a standard of statistical significance. The consideration may be related to the small number of children subjects. Whether adult or child subject, no further decrease was observed when the test distance increased from 3 to 6 m. This was not a surprising finding as performance often decreases with increasing distance (Wilson et al., 2007). The lack of a further decrease between the 3 and 6 m test distances could be due to a "floor effect" where decreased signal levels at further distance did not result in an even poorer speech in noise recognition performance, perhaps due to reverberation and distortion of the speech signal caused by the test environment.

Compared with English, most initial consonants in Mandarin are voiceless. This results in initial consonants with low sound intensities as voiceless signals do not entail the vibration of the vocal folds and makes Mandarin comparatively more difficult to recognize in noise. This study showed that the Mini Microphone can effectively improve speech communication in Mandarin-speaking patients with sensorineural hearing loss.

Currently, multiple hearing aid manufacturers have introduced digital wireless remote microphones compatible with their range of hearing aid and cochlear implant technologies. Operation of these devices is simple, could be easily adopted by hearing aid users, especially older adult users, and adds relatively little cost to the hearing aid purchase. For future studies, a comparative study of speech intelligibility, speech delay, and cumulative power consumption of multiple digital wireless devices used in conjunction with hearing aids may be considered. Lastly, the Bluetooth SIG (Special Interest Group) has introduced a digital wireless standard for manufacturers of hearing aids and wireless accessories, as well as consumer devices. With the increase of such technology, Bluetooth digital signal coverage in public spaces such as theaters and cinemas may increase, improving accessibility for hearing aid users who may be able to use remote microphones in public spaces to improve communication and listening.

In the current study, loudspeakers were used as the source of speech signals. The role of lip-reading and facial expressions in communication was not fully examined. However, in daily communication, lip reading and facial expressions play a vital role in understanding speech, especially for hearing-impaired individuals and children (Chodosh et al., 2020). The results of this study showed that MM alone provided the best speech recognition ability in a noisy environment for both adults and children, but this result should not be interpreted as a basis to deactivate a hearing aid microphone in noisy environments. For hearing-impaired children, hearing aid microphones can increase the chances of incidental learning (Vermeulen et al., 2012; Klein et al., 2018). HA + MM may be considered as a part of a more comprehensive program, where both target speech and incidental learning are desired.

Lastly, one of the limitations of the current study is that,only native speakers of Mandarin Chinese were selected. For future studies, bilingual (e.g., Mandarin Chinese and English) adults and children could be recruited to evaluate the effect of remote microphone and assistive listening devices in both tonal and non-tonal languages. In addition to the aforementioned wearable assistive listening devices, speech-to-text conversion apps for smart phones have been designed specifically to provide communication redundancy for individuals with hearing loss. These apps have been shown to improve communication for those with hearing loss, especially the profoundly hearing impaired population in certain listening situations (Pragt et al., 2020).

CONCLUSION

The addition of a wireless remote microphone to bilaterally worn hearing aids compensates for increased distance from the sound source. The use of a wireless remote microphone can significantly improve speech in noise communication performance in Chinese hearing-impaired listeners.

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/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board at the Beijing Institute of Otolaryngology and Beijing Tongren Hospital. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

SW, JC, and RD contributed conception and design of the study. JC, RD, ZW, and YW performed the experiments. JC, SW, XF, and RD performed the statistical analyses. JC and SW wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work was funded in part by grants from the Natural Science Foundation of China (Grant Nos. 81870715 and 81200754), the Promotion Grant for High-Level Scientific and Technological Elites in Medical Science from the Beijing Municipal Health Bureau (Grant No. 2015-3-012), the 2017 Excellent Talents in Medical Science from the Dongcheng District of Beijing.

ACKNOWLEDGMENTS

We thank all the subjects participating in this study and our colleagues at the Clinical Audiology Center in Beijing Tongren Hospital for recruiting the hearing-impaired subjects. We also thank Dr. Jiong Hu and Dr. Esther Pugh for their help with editing this manuscript.

REFERENCES

American Speech-Language-Hearing Association (ASHA) (2002). *Guidelines for Fitting and Monitoring of FM Systems*. Rockville, MD: ASHA.

Bentler, R., Mueller, H. G., and Ricketts, T. A. (2016). *Modern Hearing Aids*. New York, NY: Thieme, 171–236.

Blazer, C. (2007). *Improving the Classroom Environment: Classroom Audio Technology (Miami-Dade County Public Schools Information Capsule, Research Services, 607).Wagener et al*. Miami, FL: Miami-Dade County Public Schools Information, 11.

Boothroyd, A., and Guerrero, N. (2004). Hearing aid accessories for adults: the remote FM microphone. *Ear Hear*. 25, 22–33. doi: 10.1097/01.aud.0000111260.46595.ec

Chen, Y., and Wong, L. L. N. (2020). Development of the mandarin hearing in noise test for children. *Int. J. Audiol.* 59, 707–712. doi: 10.1080/14992027.2020.1750717

Chodosh, J., Weinstein, B. W., and Blustein, J. (2020). Face masks can be devastating for people with hearing loss. *BMJ* 370:m2683. doi: 10.1136/bmj.m2683

Clare, B., and Gill, D. (2018). The use of the Cochlear Mini Microphone(MM) as a personal radio system (FM) with young children who are deaf. *Cochlear Implants Int.* 19, 330–337. doi: 10.1080/14670100.2018.1505324

David, B., and William, S. (1984). Signal-to-noise ratio advantage of binaural hearing aids and directional microphones under different levels of reverberation. *J. Speech Hear. Disord.* 49, 278–286. doi: 10.1044/jshd.4903.278

Dillon, H. (2012). *Hearing Aids*. New York, NY: Thieme, 136.

Dong, R. J., Wang, S., and Liu, D. X. (2015). Speech perception in noise in adults with auditory neuropathy spectrum disorders. *Chin. J. Otolog.* 13, 604–607.

Fickes, M. (2003). *The Sounds of a Sound Education. School Planning and Management*. https://spaces4learning.com/articles/2003/06/01/the-sounds-of-a-sound-education.aspx (accessed March 27, 2021).

Gravel, J., Fausel, N., and Liskow, C. (1999). Children's speech recognition in noise using omnidirection al and dual-microphone hearing aid technology. *Ear Hear.* 20, 1–11. doi: 10.1097/00003446-199902000-00001

Hu, X. Y., Zheng, X. Y., and Ma, F. R. (2016). Prevalence of hearing disorders in China:a population-based survey in four provinces of China. *Chin. J. Otorhinolaryngol. Head Neck Surg.* 51, 819–825.

Jace, W., Mila, M. D., and Erin, S. (2015). Research article evaluation of performance with an adaptive digital remote microphone. *Am. J. Audiol.* 24, 440–450. doi: 10.1044/2015_aja-15-0018

Jantien, L., Vroegop, J., and Gertjan, D. (2017). Evaluation of a wireless remote microphone in bimodal cochlear implant recipients. *Int. J. Audiol.* 56, 643–649. doi: 10.1080/14992027.2017.1308565

Klein, K. E., Wu, Y. H., Stangl, El, and Bentler, R. A. (2018). Using a digital language processor to quantify the auditory environment and the effect of hearing aids for adults with hearing loss. *J. Am. Acad. Audiol.* 29, 279–291. doi: 10.3766/jaaa.16114

Kochkin, S. (2002). Consumers rate improvements sought in hearing instruments. *Hear. Rev.* 9, 18–22.

Koopmans, W., Goverts, S. T., and Smits, C. (2018). Speech recognition abilities in normal-hearing children 4 to 12 years of age in stationary and interrupted noise. *Ear Hear.* 39, 1091–1103. doi: 10.1097/AUD.0000000000000569

Kreikemeier, S., Margolf-Hackl, S., and Raether, J. (2013). Comparison of different directional microphone technologies for moderate-to-severe hearing loss. *Hear. Rev.* 20, 44–45.

Kuk, F., Kollofski, C., and Brown, S. (1999). Use of a digital hearing aid with directional microphones in school-aged children. *J. Am. Acad. Audiol.* 10, 535–548.

Lewis, M. S., Crandell, C. C., and Valente, M. (2004). Speech perception in noise: directional microphones versus frequency modulation (FM) systems. *J. Am. Acad.* 15, 426–439. doi: 10.3766/jaaa.15.6.4

Lin, M. C. (1988). *The Acoustical Properties and Perceptual Characteristics of Mandarin Tones*. Beijing: Zhongguo Yuwen, 182–193.

Linda, W. N., and Kristen, C. (2016). Frances PH. the effects of FM and hearing aid microphone settings, FM gain, and ambient noise levels on SNR at the tympanic membrane. *J. Am. Acad. Audiol.* 27, 117–125. doi: 10.3766/jaaa.15012

Moore, B. C. J. (2013). *An Introduction of the Psychology of Hearing*, 6th Edn. New York, NY: Thieme, 67–123.

Picou, E. M. (2020). MarkeTrak 10 (MT10) survey results demonstrate high satisfaction with and benefits from hearing aids. *Semin. Hear.* 41, 21–36.

Picou, E. M., Aspell, E., and Ricketts, T. A. (2014). Potential benefits and limitations of three types of directional processing in hearing aids. *Ear Hear.* 35, 339–352. doi: 10.1097/aud.0000000000000004

Pragt, L., Hengel, P. V., Grob, D., and Wasmann, J. W. (2020). *Speech Recognition Apps for the Hearing Impaired and Deaf. VCCA2020 Conference*. https://computationalaudiology.com/ai-speech-recognition-apps-for-hearing-impaired-and-deaf/ (accessed February 17, 2021).

Ricketts, T., Lindey, G., and Henry, P. (2001). Impact of compression and hearing aid style on directional hearing aid benefit and performance. *Ear Hear.* 22, 348–361. doi: 10.1097/00003446-200108000-00009

Rodemerk, K. S., and Galster, J. A. (2015). The benefit of remote microphones using four wireless protocols. *J. Am. Acad. Audiol.* 26, 724–731. doi: 10.3766/jaaa.15008

Ryan, W. M., and Patricia, G. S. (2011). Audibility-based predictions of speech recognition for childrenand adults with normal hearing. *J. Acoust. Soc. Am.* 130, 4070–4081. doi: 10.1121/1.3658476

Vermeulen, A., Raeve, L. D., Langereis, M., and Snik, A. (2012). Changing realities in the classroom for hearing-impaired children with cochlear implant. *Deafness Educ. Int.* 14, 36–47. doi: 10.1179/1557069x12y.0000000004

Wilson, R. H., McArdle, R. A., and Smith, S. L. (2007). An evaluation of the BKB SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss. *J. Speech Lang. Hear. Res.* 50, 844–856. doi: 10.1044/1092-4388(2007/059)

Wong, L. L. N., Sigfrid, D. S., and Liu, S. (2007). Development of the Mandarin Hearing in Noise Test ‘(MHINT). *Ear Hear.* 28(Suppl. 2), 70S–74S.

Zhang, H. (2009). Consideration of the development of hearing aids. *News Rev.* 24, 70–72.

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.

frontiers in Psychology

BRIEF RESEARCH REPORT
published: 02 June 2021
doi: 10.3389/fpsyg.2021.656052

Check for updates

Investigation of an MAA Test With Virtual Sound Synthesis

Ruijie Meng[1,2†], Jingpeng Xiang[1,2†], Jinqiu Sang[1,2], Chengshi Zheng[1,2], Xiaodong Li[1,2], Stefan Bleeck[3], Juanjuan Cai[4*] and Jie Wang[5*]*

[1] Key Laboratory of Noise and Vibration Research, Institute of Acoustics, Chinese Academy of Sciences, Beijing, China, [2] University of Chinese Academy of Sciences, Beijing, China, [3] Institute of Sound and Vibration Research, University of Southampton, Hampshire, United Kingdom, [4] State Key Laboratory of Media Convergence and Communication, Communication University of China, Beijing, China, [5] School of Electronics and Communication Engineering, Guangzhou University, Guangzhou, China

OPEN ACCESS

Edited by:
Qinglin Meng,
South China University of Technology,
China

Reviewed by:
Ellen Peng,
University of Wisconsin-Madison,
United States
Liping Zhang,
Shandong University ENT Hospital,
China

***Correspondence:**
Jinqiu Sang
sangjinqiu@mail.ioa.ac.cn
Juanjuan Cai
caijuanjuan@cuc.edu.cn
Jie Wang
wangjie@gzhu.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology

Received: *20 January 2021*
Accepted: *06 May 2021*
Published: *02 June 2021*

Citation:
Meng R, Xiang J, Sang J, Zheng C, Li X, Bleeck S, Cai J and Wang J (2021) Investigation of an MAA Test With Virtual Sound Synthesis. Front. Psychol. 12:656052. doi: 10.3389/fpsyg.2021.656052

The ability to localize a sound source is very important in our daily life, specifically to analyze auditory scenes in complex acoustic environments. The concept of minimum audible angle (MAA), which is defined as the smallest detectable difference between the incident directions of two sound sources, has been widely used in the research fields of auditory perception to measure localization ability. Measuring MAAs usually involves a reference sound source and either a large number of loudspeakers or a movable sound source in order to reproduce sound sources at a large number of predefined incident directions. However, existing MAA test systems are often cumbersome because they require a large number of loudspeakers or a mechanical rail slide and thus are expensive and inconvenient to use. This study investigates a novel MAA test method using virtual sound source synthesis and avoiding the problems with traditional methods. We compare the perceptual localization acuity of sound sources in two experimental designs: using the virtual presentation and real sound sources. The virtual sound source is reproduced through a pair of loudspeakers weighted by vector-based amplitude panning (VBAP). Results show that the average measured MAA at 0° azimuth is 1.1° and the average measured MAA at 90° azimuth is 3.1° in a virtual acoustic system, meanwhile the average measured MAA at 0° azimuth is about 1.2° and the average measured MAA at 90° azimuth is 3.3° when using the real sound sources. The measurements of the two methods have no significant difference. We conclude that the proposed MAA test system is a suitable alternative to more complicated and expensive setups.

Keywords: localization acuity, the frontal MAA, the lateral MAA, virtual sound synthesis, VBAP

1. INTRODUCTION

The smallest perceptually detectable difference between the azimuths of two sound sources is called the minimum audible angle (MAA) (Mills, 1958). In 1958, Mills proposed the concept of MAA to measure perceptional auditory spatial acuity and since then, the MAA has been used in many studies on sound localization and auditory perception. For example, the MAA test was used to investigate the precedence effect in sound localization (Litovsky and Macmillan, 1994) or to measure the sound localization acuity of children with cochlear implants (Saberi et al., 1991; Litovsky et al., 2006; Tyler et al., 2010).

Sound source localization is important for auditory scene analysis (McAdams, 1984, 1993; Tyler et al., 2002; Grieco-Calub and Litovsky, 2010; Kerber and Seeber, 2012). There is an increasing demand for affordable and convenient assessment of sound localization ability especially for the hearing impaired and the early identification of hearing loss in children. Often in experimental designs, researchers are restricted to loudspeakers with fixed positions, often with 10° or more separation. It would therefore be preferable to have a controlled method to render virtual stimuli at any angle when measuring the MAA at any desired incident direction.

The MAA measurement method has been conducted in previous researches to measure sound localization acuity with real sound sources (Mills, 1958; Perrott, 1969, 1993; Harris and Sergeant, 1971; Perrott et al., 1989; Saberi et al., 1991; Grantham et al., 2003; Van Deun et al., 2009; Tyler et al., 2010). Various such techniques were developed in the past: Mills (Mills, 1958) used rotating poles to change the incident direction of stimuli in the horizontal plane and the MAA value is about 1°. This apparatus was also popular later in related studies. For example Saberietal (Saberi et al., 1991) used a system of counter-balanced speakers on the pole to measure MAAs in the lateral and dorsal planes. Van Deun et al. (2009) used nine loudspeakers positioned in the frontal horizontal field to measure sound localization, sound lateralization, and binaural masking level differences in young children. Tyler et al. (2010) set up an auditory training system with eight loudspeakers to improve binaural hearing in noise and localization. Perrott (1969, 1993) used 13 loudspeakers in the MAA study with different signal onsets in the horizontal plane and another array with 14 loudspeakers. Harris and Sergeant (1971) set up a track upon which a loudspeaker rode on a little cart, and MAA was computed from the stimulus of Gaussian white noise moving left and right. In Litovsky and Macmillan's experiment (Litovsky and Macmillan, 1994), MAAs were estimated for single noise bursts, and for burst pairs that satisfied the conditions of the precedence effect, but the loudspeakers had to be moved manually between trials. All of these experimental designs based on the real source reproduction are complex, a better design is expected to be applied to clinical utility with easier experiments.

Using a rotating boom method, Mills (1958) measured the MAAs in various directions in the horizontal plane using a two-alternative forced choice procedure. He reported MAAs of about 1°. Similar results were later found by Perrott et al. (1989). For a broadband 0.9 kHz high pass noise, the measured MAA at 0° azimuth is about 1.2° (Perrott, 1993). For broadband noise the measured MAA at 0° azimuth is about 1.6° (Grantham et al., 2003).

Virtual sound synthesis methods were used in studies of virtual reality and artificial sound field generation (McAdams, 2000; Daniel et al., 2010). Existing virtual sound synthesis methods mainly include wave-field synthesis (WFS), Ambisonics, vector-based amplitude panning (VBAP) and binaural synthesis. Wave-field synthesis (WFS) developed by Berkhout et al. (1993) enables the synthesis of sound fields within a rather large listening area. Localization accuracy with wave-field synthesis (WFS) was evaluated using an MAA listening test paradigm (Völk et al., 2012b; Völk, 2016). Ambisonics was firstly proposed by Michael Gerzon as a point source solution for a small listening area and was extended to higher orders of spherical harmonics so that the listening area can be extended significantly (Gerzon, 1977). However, sound reproduction systems through WFS or Ambisonics require tens of loudspeakers. Binaural synthesis (BS) is widely used as a tool aiming at eliciting specific auditory perceptions by means of headphones. An evaluation method was proposed, addressing the binaural synthesis quality by comparing the MAAs measured in the synthesized situation versus the corresponding real situation (Völk et al., 2012a). Völk argued for the use of virtual acoustic methods in psychoacoustics and auditory studies because of their relatively simple application (Völk, 2013). Hohmann discussed the current state and the perspectives of virtual reality technology used in the lab for designing complex audiovisual communication environments for hearing assessment and hearing device design and evaluation, the result showed that the virtual reality lab in its current state marks a step toward more ecological validity in lab-based hearing and hearing device research (Hohmann et al., 2020). Ahrens investigated source localization accuracy with the head mounted displays (HMD) in virtual reality providing a varying amount of visual information, which showed that the lateral localization error induced by wearing HMD was due to alterations of HRTF (Ahrens et al., 2019). However, BS requires individualized head related transfer functions (HRTFs) which are difficult to measure. Berger proposed auditory source localization could be improved for users of generic HRTFs via cross-modal learning (Berger et al., 2018). Pausch employed perceptual tests to evaluate a recently proposed binaural real-time auralization system for hearing aid (HA) users (Pausch and Fels, 2020). But, problems like virtual sound images perceived internalized with binaural synthesis still need to be overcome (Kulkarni and Colburn, 1998). The vector-based amplitude panning (VBAP) was proposed by Pulkki (1997) as stereophonic principles aiming to synthesize an arbitrary sound source between selected pair or triplet of loudspeakers in a plane or in the three-dimension space (Pulkki, 2001a,b; Pulkki and Karjalainen, 2008). Pulkki investigated the localization accuracy of the VBAP method, it was shown that the high-frequency interaural level difference (ILD) cues roughly propose the same directions as low-frequency interaural time difference (ITD) (Pulkki and Karjalainen, 2001). Gröhn (Pulkki, 2001a) conducted a localization accuracy test with VBAP reproduction and non-individualized HRTF reproduction, finding the median value of median azimuth error were 5.6° and 8.3°, the VBAP in this experiment showed the same accuracy as the direct loudspeaker reproduction. The setup of VBAP is relatively simple, however, whether VBAP can be an alternative to conventional methods in hearing research has not been established yet, and some basic perceptual effects such as the MAAs at different reproduction angles should be validated.

In this study, we investigate the feasibility to use the VBAP method to measure the MAAs at 0 ° azimuth and 90° azimuth. This method could reproduce source positions for a single listener at a sweet spot regardless of head rotation. However, the result in sound localization acuity through VBAP is not known yet. This paper first introduces the setup of experiments including the VBAP method and a baseline method. Experiment results are given in section 3, followed by discussions in section 4. Finally, the conclusions are drawn in section 5.

2. MATERIALS AND METHODS

Setup

The process of producing the stimuli using VBAP is explained in detail in Pulkki (2001a) and summarized here. For a desired azimuthal incident direction ϕ the signal amplitudes of the selected pair of loudspeakers located at θ_1, θ_2 are controlled with gain factors g_1, g_2. The amplitude gains g_1, g_2 are calculated based on Equations (1, 2). Equation (1) calculates the sound amplitudes as a function of incident direction

$$\begin{pmatrix} \cos\theta_1 & \cos\theta_2 \\ \sin\theta_1 & \sin\theta_2 \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \cos\phi \\ \sin\phi \end{pmatrix} \tag{1}$$

and Equation (2) shows how to calculate the normalized amplitude gains:

$$g_1 = \frac{a_1}{a_1^2 + a_2^2}; g_2 = \frac{a_2}{a_1^2 + a_2^2} \tag{2}$$

In the measurement of the MAA at 0° azimuth (the frontal MAA) with the virtual sound synthesis system, two loudspeakers are located symmetrically with ϕ_0 =30° at each side of the reference as shown in **Figure 1A**. In the measurement of the MAA at 90° azimuth (the lateral MAA) with the virtual sound synthesis system, two loudspeakers are located symmetrically with ϕ_0 =15° at each side of the reference as shown in **Figure 1B**. Generally, when the aperture between loudspeakers is wider, the localization accuracy is worse (Pulkki and Karjalainen, 2001). The head rotation can be corrected using the tangent law (Pulkki, 1997):

$$\frac{\tan\phi}{\tan\phi_0} = \frac{g_1 - g_2}{g_1 + g_2} \tag{3}$$

Theoretically, an accurate synthesis is possible in the horizontal plane by weighting the amplitude gains of the pair of loudspeakers. Therefore, the VBAP method is a promising candidate to provide a simple method of measuring MAA using just two fixed loudspeakers. Sounds were presented at 65 dB(A) with a background level of 28 dB(A), measured with a sound level meter.

Participants

Nineteen normal-hearing (thresholds < 20 dB HL, measured in a hearing screening test) volunteers participated in the study (aged between 22 and 29). All participants had participated in psychoacoustic localization experiments before and were considered as experienced listeners. A listening room with dimensions of 12.92 m × 6.94 m × 2.67 m (Length × Width × Height) and the reverberation time T60 of 0.1 s was used as the test environment.

VBAP Measurement Procedure

Prior to test trials, participants received training to familiarize them with the procedure. The participants were instructed and positioned in a seat 1.8 m away from the loudspeakers in those two experiments. Broadband white noises (0.1–8 kHz) were used as stimuli. The noise stimulus were a train of three 100-ms bursts of Gaussian noise, with a 500-ms silence between bursts. A pair of Bose MusicMonitors were used for sound reproduction through Realtek(R) Audio sound card, in addition, a silent speaker was placed in the middle as a visual reference. **Figure 1** showed the frontal and the lateral MAA measurement configuration. In the frontal MAA measurement experiment, the two speakers were placed symmetrically on the left and right of the participants at a fixed angle of 30° in reference to 0° azimuth. In each trial, the stimuli were presented from the right or the left randomly. Participants were instructed to indicate the perceived side of the stimuli in each trial by pointing with their hand toward the right or the left side. The results were recorded by the experimenter seated behind the participant. In the very first trial, the stimuli were presented from 30° (right or left). The initial 30° shift was chosen to ensure that it comfortably exceeded the expected MAAs of all participants. In the lateral MAA measurement experiment, the two speakers were placed symmetrically on the front and back of the participants at a fixed angle of 15° in reference to 90° azimuth. In each trial, the stimuli were presented from the front or the back randomly. Participants were instructed to indicate the perceived side of the stimuli in each trial by pointing with their hand toward the front or the back side. The results were recorded by the experimenter seated behind the participant. In the very first trial, the stimuli were presented from 15° (front or back). This initial 15° shift was chosen to ensure that it comfortably exceeded the expected MAAs of all participants. A 3-down/1-up adaptive procedure (Levitt, 1971) was used to determine the reproduction angle for the next trial, which could be smaller or larger than the previous separation, so as to find the 79.4% correct point on a psychometric function (Schütt et al., 2016). The angular step sizes in the frontal MAA measurement were determined by Parameter Estimation by Sequential Testing (PEST) (Litovsky and Macmillan, 1994), and were: 30°, 15°, 8°, 4°, 2°, 1°, 0.5°. And the angular step sizes in the lateral MAA measurement were: 15°, 8°, 4°, 2°, 1°. The presentation side (left/front or right/back) in each trial was chosen randomly. The experiment ended after six reversals (a reversal is an increase in angle following a decrease, or vice versa), this procedure is converging toward the 79.4% point of the psychometric function. After discarding the first 2 reversals the MAA is defined here as the angular threshold where about 79.4% of all judgments of the relative positions of the sound sources are correct. The average experiment duration for each individual was around 30 min.

Baseline Measurement Procedure

In order to verify the accuracy of the VBAP system, an MAA experiment system with the real sources shown in **Figure 2** was used as the baseline comparison. In the frontal MAA measurement, this baseline system consisted of a pair of loudspeakers: one was located on the left and the other was located on the right side at either 1°, 2°, 4°, 8° symmetrically

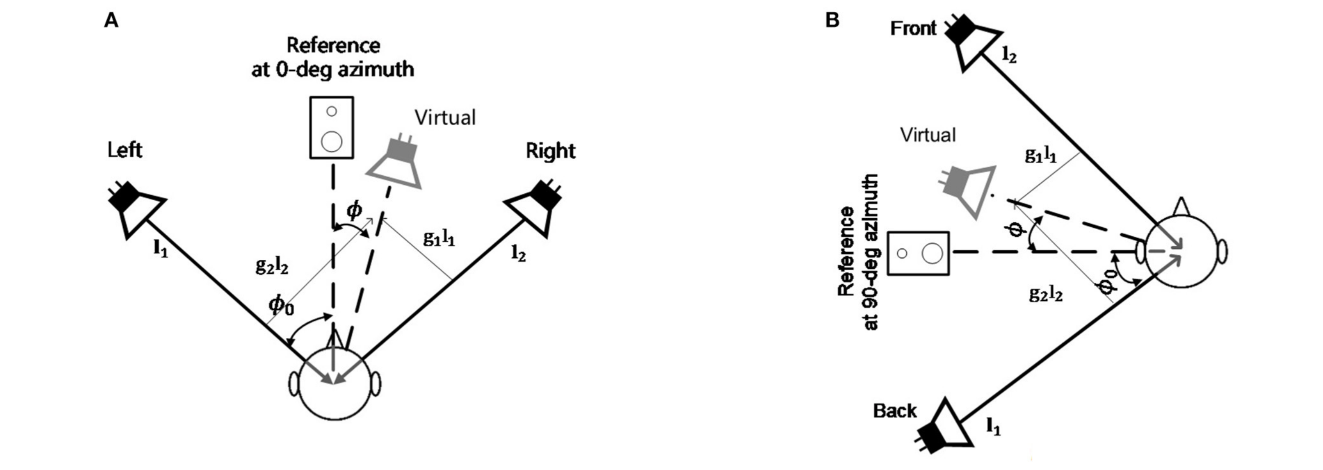


FIGURE 1 | The VBAP system configuration. **(A)** The configuration of the frontal MAA measurement experiment; **(B)** The configuration of the lateral MAA measurement experiment.

in reference to 0° azimuth. In the lateral MAA measurement, this baseline system consisted of a pair of loudspeakers: one was located at the front and the other was located on the back side at either 1°, 2°, 4°, 8°,15° symmetrically in reference to 90° azimuth. The stimuli and the test room were the same as those in the previous VBAP measurement procedure. The participants were instructed and positioned in a seat 1.8 m away from the loudspeakers in those two experiments. When playing the left and right (or the front and back) sound randomly, the participants were answered whether the sound is on the left or the right (front or back). The procedure was conducted twenty times at each angular separation of loudspeakers. The results were averaged and provided a percent correct indicating how often the participants correctly identified the localization. The average experiment duration for each individual was around 30 min.

3. RESULTS

The average MAAs at 0° azimuth measured by the VBAP method is 1.1° with a range from 0.8° to 1.7° and a standard deviation of 0.3°. The average MAAs at 90° azimuth measured by the VBAP method is 3.1° with a range from 0.8° to 5.7° and a standard deviation of 1.3°. Using the described adaptive method, the MAA is the angle where the psychometric function is 79.4% correct. To establish an equivalent threshold from the baseline method, we employed the following method: percent correct rates were calculated for each angle, and the resulting data were fitted with a psignifit function (Schütt et al., 2016). The percent correct at each angular separation in the baseline method was extracted from subjects' answer data, based on which the fitting curves were used to estimate corresponding MAAs with judgments 79.4% correct. The MAA at 0° azimuth is 1.2° with a range from 0.6° to 1.7° and a standard deviation of 0.3°. The MAA at 90° azimuth is 3.3° with a range from 1.8° to 5.6° and a standard deviation of 1.1°, which is taken as the MAAs from the baseline method. This result is consistent with previous findings that MAA at 0° azimuth is about 1° with a range from 0.7° to 2.5° (Mills, 1958; Perrott, 1969; Harris and Sergeant, 1971; Tyler et al., 2010). In the frontal MAA measurement experiment, we assume that participants have an average of about 50% correct at 0° and almost achieve 100% accuracy at the angular separation of 8°, and in the lateral MAA measurement experiment, we assume that participants have an average of about 50% correct at 0° and achieve 100% accuracy at the angular separation of 15°. The MAA results of the two experiments are shown in **Figure 3**. Paired t-tests of participants' MAAs in both methods were performed to test if there is a significant difference between the baseline data and the VBAP data. As the calculated p-values ($t = 0.43$, $p > 0.05$, Cohen's $d = 0.10$ in the frontal experiment; $t = 1.30$, $p > 0.05$, Cohen's $d = 0.30$ in the lateral experiment) are bigger than the p critical value ($p = 0.05$ with 95% confidence), the null hypothesis is accepted, meaning that there is no statistical difference in the same participant's MAA between the VBAP method and the baseline method. For the same participant, we also found that the performance in the frontal MAA measurement experiment is not statistically correlated with the performance in the lateral experiment ($r = 0.02$, $p > 0.05$ in baseline method, $r = 0.01$, $p > 0.05$ in VBAP method). This may mean that people who perform well in the frontal MAA measurement experiment do not necessarily perform well in the lateral MAA measurement experiment.

To further illustrate the similarity between the baseline method and the VBAP method, we calculated the average percent correct of different angular separation in both experiments (marked star and circle in **Figure 4**) and fitted curve of the average percent correct of the group at each angle (see **Figure 4** dash and dash-dot line). The error bar means variance of the correct percent at each angle in **Figure 4**. For the frontal experiment, the variance of the deviation between the individual measurement accuracy of each angle is 18.65, 13.11, 9.18, 4.03, and 0%, respectively at 0.5°, 1°, 2°, 4°, 8° in the VBAP method.

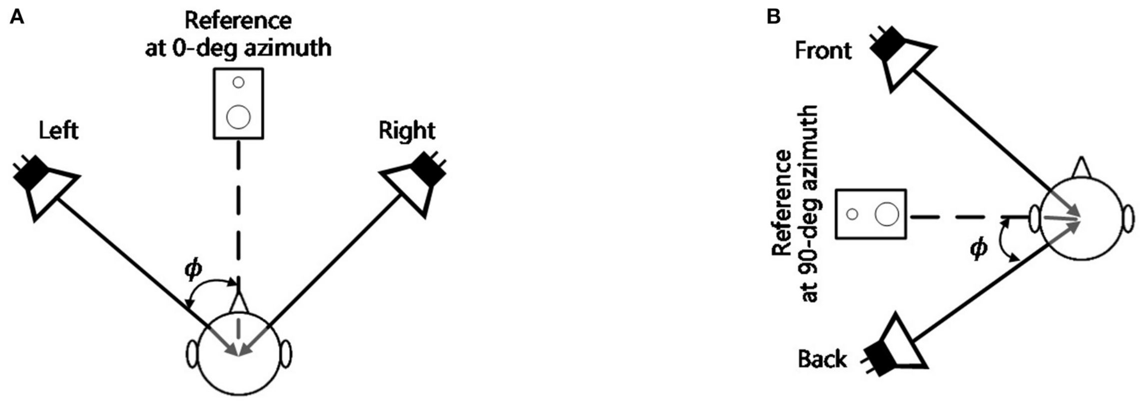


FIGURE 2 | The verification system configuration. **(A)** The configuration of the frontal MAA measurement experiment (0° azimuth), the angle $\phi = \pm 1°, \pm 2°, \pm 4°, \pm 8°$; **(B)** The configuration of the lateral MAA measurement experiment (90° azimuth), the angle $\phi = \pm 1°, \pm 2°, \pm 4°, \pm 8°, \pm 15°$.

For the frontal experiment, the variance of the deviation between the individual measurement accuracy of each angle is 12.08, 6.86, 1.91, and 0%, respectively at 1°, 2°, 4°, 8° in the baseline method. For the lateral experiment, the variance of the deviation between the individual measurement accuracy of each angle is 14.01, 16.05, 15.70, 3.48, 2.90%, respectively at 1°, 2°, 4°, 8°, 15° in the VBAP method and 21.74, 15.72, 12.89, 5.84, 2.17%, respectively at 1°, 2°, 4°, 8°, 15° in the basline method. We compared the percent correct of each angle in VBAP method and the baseline method for each individual participant. Paired t-tests of participants' results in both methods were performed ($\phi = 1°$, $t = 1.14$, $p > 0.05$, Cohen's $d = 0.26$; $\phi = 2°$, $t = 2.25$, $p = 0.04 < 0.05$, Cohen's $d = 0.51$; $\phi = 4°$, $t = 1.84$, $p > 0.05$, Cohen's $d = 0.42$ in the frontal experiment. $\phi = 1°$, $t = 0.33$, $p > 0.05$, Cohen's $d = 0.08$; $\phi = 2°$, $t = 0.34$, $p > 0.05$, Cohen's d $= 0.08$; $\phi = 4°$, $t = 1.44$, $p > 0.05$, Cohen's $d = 0.33$; $\phi = 8°$, $t = 1.31$, $p > 0.05$, Cohen's $d = 0.30$; $\phi = 15°$, $t = 0.44$, $p > 0.05$, Cohen's $d = 0.10$ in the lateral experiment). The above analysis shows that the calculated p-values are bigger than the p critical value (0.05 with 95% confidence) except when ϕ is 2° (Cohen's $d = 0.51$, medium effect), which indicates that the two methods are not significantly different to some extent. However, more samples are needed to strongly support the non-significant difference hypothesis.

4. DISCUSSION

We compared the MAAs at 0° and 90° azimuth determined in the VBAP method and the measured baseline results, and showed that there is no significant difference in the results obtained by the two methods. To further verify the substitutability of the VBAP method, we conducted acoustic simulations to analyze the binaural cues (ITD and ILD) of the stimuli delivered via VBAP. By convolving the generic non-individualized HRTF of the KEMAR mannequin with the stimuli, the left and right signals are obtained. We divided the stimuli into 16 critical bands with a gammatone filterbank and estimated corresponding ITDs and ILDs through the Binaural Cue Selection Toolbox (Faller and Merimaa, 2004). The simulation results of the ILDs and ITDs in the VBAP method and the baseline method are compared in **Figure 5**. The left and the right columns show the results for 0-degree azimuth and 90-degree azimuth of incidences, respectively. The top and bottom rows show the results for the ITDs and ILDs, respectively. The VBAP delivers ITDs and ILDs closely consistent with those delivered by the real sound source. Therefore, we conclude that the virtual sound synthesis system is a valid alternative to the conventional apparatus, e.g., a cart runs in the track (Harris and Sergeant, 1971), large scale loudspeaker array (Harris and Sergeant, 1971; Perrott and Saberi, 1990), or a sound boom balanced by weights (Saberi et al., 1991) and can provide a compact and affordable listening test system for measuring MAAs.

This could be useful in the future as an additional tool to diagnose hearing impairment in a clinical setting, and could also be used for the hearing aid fitting process. Due to the principle of rendering virtual sound sources within the angular range between two loudspeakers, a slight misplacement would introduce a large deviation of the incident direction. This position-sensitive attribute is particularly obvious in the hearing tests where small angular differences are required. Improving the localization accuracy of the apparatus as well as the participants' localization accuracy would be beneficial. To reduce the uncertainty of participants' localization, a head-tracking system monitoring the participants' head position would also be useful. However, an appropriate head fixation limiting the head motions is a cheaper option. Moreover, the present sample size is small, and the feasibility of the VBAP method needs to be further verified. These limitations are important issues for our future research, and they are also inevitable problems in clinical applications. Finally, reducing the interval angle between each pair of loudspeakers is likely to provide higher localization acuity in sound source reproduction. However, a large interval angle can provide more virtual sound source locations flexibly without having to move or add loudspeakers. We need

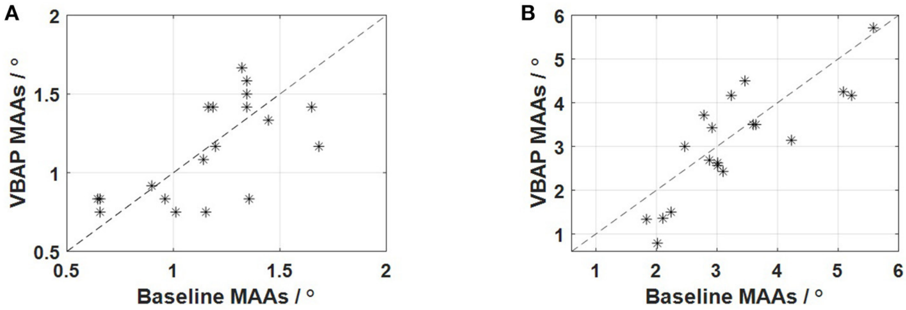


FIGURE 3 | Comparison of the results of the two methods. **(A)** The results of the frontal MAA measurement experiment; **(B)** The results of the lateral MAA measurement experiment.

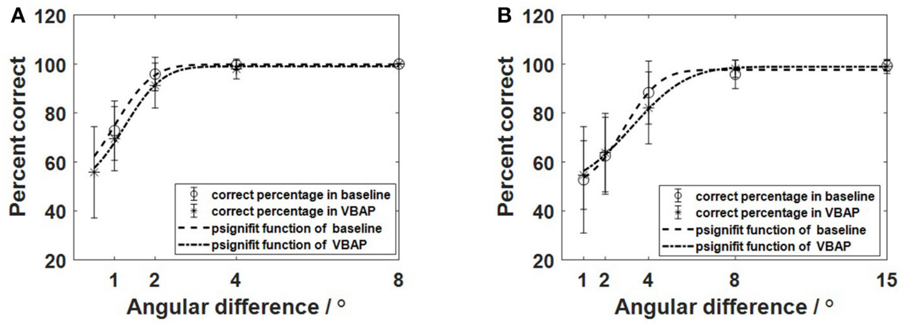


FIGURE 4 | The percentage of correct answers fitted by psignifit function in the baseline (dash line) and the VBAP method (dash-dot line). **(A)** The results of the front MAA measurement experiment; **(B)** The results of the lateral MAA measurement experiment.

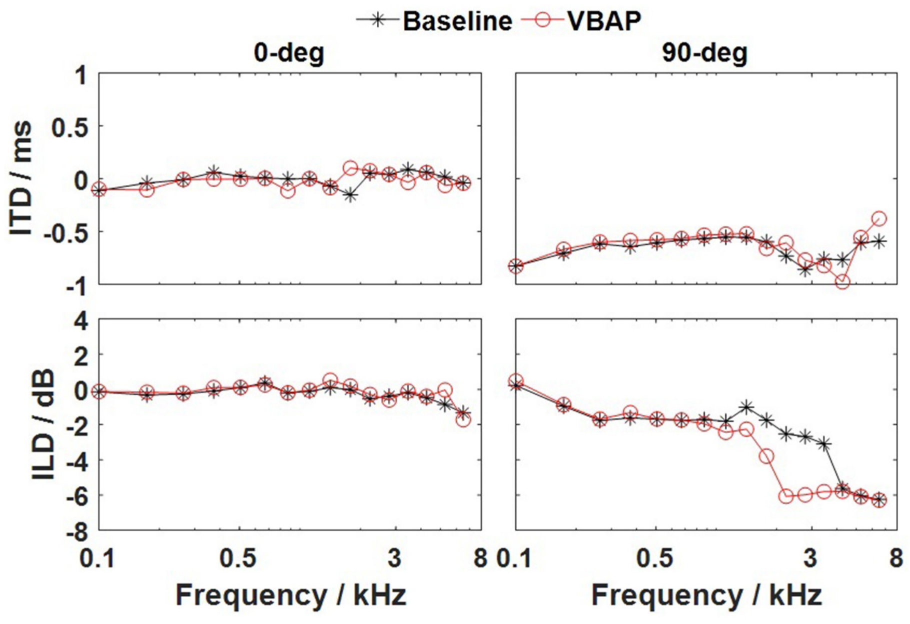


FIGURE 5 | ITDs and ILDs as a function of the critical bands for the baseline method (solid line marked with the black star) and the VBAP method (solid line marked with the red circle). Left column, 0° azimuth of incidence; and the right column, 90° azimuth of incidence. The top and bottom rows show the results for the ITDs and ILDs, respectively.

to balance between the speaker arrangement flexibility and localization acuity.

5. CONCLUSIONS

We evaluated the feasibility of a virtual acoustic method to measure MAAs, because conventional apparatuses are usually complicated to use. We used a setup with two loudspeakers driven by sounds based on the vector-based amplitude panning (VBAP) principle. Results show that a resolution around 1° at 0° azimuth and around 3° at 90° azimuth can be achieved by the virtual acoustic test system. To validate the results of MAA test, a baseline measure with real loudspeakers was established with the same participants. The results of "real MAAs" and "virtual MAAs" are not significantly different and thus provide validation of the proposed MAA measurement method.

The virtual acoustic methods provide a convenient and affordable alternative to implement experiments in hearing research and they have the potential for a wider range of applications. For example, assessment of localization skill in hearing-aid fitting and children's localization training in the critical period of auditory development (Harrison et al., 2005). Since the loudspeakers are fixed during the experiment, such methods can be quite convenient for studies involved moving sound sources such as moving minimum audible angle (Hughes and Kearney, 2016).

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

Written informed consent was obtained from each participant in accordance with the local legislation and institutional requirements.

AUTHOR CONTRIBUTIONS

RM: methodology, data analysis, writing the second draft based on the review comments. JW and JC: methodology and modification. SB and CZ: writing-review and editing. XL: writing-review and editing, supervision. JS: methodology, writing-review and editing. JX: methodology, software, formal analysis, and writing-original draft. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Science Fund of China (Nos. 12074403 and 11974086), the Open Research Project of the State Key Laboratory of Media Convergence and Communication, Communication University of China, China (No. SKLMCC2020KF005), the Guangzhou Science and Technology Plan Project (No. 201904010468), and Guangzhou University Science Research Project (No. YJ2021008).

ACKNOWLEDGMENTS

We would like to thank those subjects who participated in the listening tests. We would also like to express our great appreciation to the editor and the reviewers.

REFERENCES

Ahrens, A., Lund, K. D., Marschall, M., and Dau, T. (2019). Sound source localization with varying amount of visual information in virtual reality. *PLoS ONE* 14:e0214603. doi: 10.1371/journal.pone.0214603

Berger, C. C., Gonzalez-Franco, M., Tajadura-Jiménez, A., Florencio, D., and Zhang, Z. (2018). Generic hrtfs may be good enough in virtual reality. improving source localization through cross-modal plasticity. *Front. Neurosci.* 12:21. doi: 10.3389/fnins.2018.00021

Berkhout, A. J., de Vries, D., and Vogel, P. (1993). Acoustic control by wave field synthesis. *J. Acoust. Soc. Am.* 93, 2764–2778. doi: 10.1121/1.405852

Daniel, A., Nicol, R., and McAdams, S. (2010). "Multichannel audio coding based on minimum audible angles," in *Audio Engineering Society Conference: 40th International Conference: Spatial Audio: Sense the Sound of Space*, Audio Engineering Society.

Faller, C., and Merimaa, J. (2004). Source localization in complex listening situations: Selection of binaural cues based on interaural coherence. *J. Acoust. Soc. Am.* 116, 3075–3089. doi: 10.1121/1.1791872

Gerzon, M. A. (1977). *Multi-system ambisonic.* Wireless World 83, 43–47.

Grantham, D. W., Hornsby, B. W., and Erpenbeck, E. A. (2003). Auditory spatial resolution in horizontal, vertical, and diagonal planes. *J. Acoust. Soc. Am.* 114, 1009–1022. doi: 10.1121/1.1590970

Grieco-Calub, T. M., and Litovsky, R. Y. (2010). Sound localization skills in children who use bilateral cochlear implants and in children with normal acoustic hearing. *Ear Hear.* 31:645. doi: 10.1097/AUD.0b013e3181e50a1d

Harris, J. D., and Sergeant, R. L. (1971). Monaural/binaural minimum audible angles for a moving sound source. *J. Speech Hear. Res.* 14, 618–629. doi: 10.1044/jshr.1403.618

Harrison, R. V., Gordon, K. A., and Mount, R. J. (2005). Is there a critical period for cochlear implantation in congenitally deaf children? analyses of hearing and speech perception performance after implantation. *Dev. Psychobiol.* 46, 252–261. doi: 10.1002/dev.20052

Hohmann, V., Paluch, R., Krueger, M., Meis, M., and Grimm, G. (2020). The virtual reality lab: Realization and application of virtual sound environments. *Ear Hear.* 41(Suppl. 1):31S. doi: 10.1097/AUD.0000000000000945

Hughes, S., and Kearney, G. (2016). "Moving virtual source perception in 2d space," in *Audio Engineering Society Conference: 2016 AES International Conference on Audio for Virtual and Augmented Reality*, Audio Engineering Society.

Kerber, I. S., and Seeber, I. B. U. (2012). Sound localization in noise by normal-hearing listeners and cochlear implant users. *Ear Hear.* 33:445. doi: 10.1097/AUD.0b013e318257607b

Kulkarni, A., and Colburn, H. S. (1998). Role of spectral detail in sound-source localization. *Nature* 396, 747–749. doi: 10.1038/25526

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J. Acoust. soc. Am.* 49, 467–477. doi: 10.1121/1.1912375

Litovsky, R. Y., Johnstone, P. M., Godar, S., Agrawal, S., Parkinson, A., Peters, R., et al. (2006). Bilateral cochlear implants in children: localization acuity measured with minimum audible angle. *Ear Hear.* 27:43. doi: 10.1097/01.aud.0000194515.28023.4b

Litovsky, R. Y., and Macmillan, N. A. (1994). Sound localization precision under conditions of the precedence effect: effects of azimuth and standard stimuli. *J. Acoust. Soc. Am.* 96, 752–758. doi: 10.1121/1.411390

McAdams, S. (1984). The auditory image: a metaphor for musical and psychological research on auditory organization. *Adv. Psychol.* 19, 289–323. doi: 10.1016/S0166-4115(08)62356-0

McAdams, S. (1993). "Recognition of sound sources and events," in *Thinking in Sound: The Cognitive Psychology of Human Audition*, 146–198. doi: 10.1093/ACPROF:OSO/9780198522577.003.0006

McAdams, S. (2000). The psychomechanics of real and simulated sound sources. *J. Acoust. Soc. Am.* 107, 2792–2792. doi: 10.1121/1.428981

Mills, A. W. (1958). On the minimum audible angle. *J. Acoust. Soc. Am.* 30, 237–246. doi: 10.1121/1.1909553

Pausch, F., and Fels, J. (2020). Localization performance in a binaural real-time auralization system extended to research hearing aids. *Trends Hear.* 24:2331216520908704. doi: 10.1177/2331216520908704

Perrott, D. R. (1969). Role of signal onset in sound localization. *J. Acoust. Soc. Am.* 45, 436–445. doi: 10.1121/1.1911392

Perrott, D. R. (1993). "Auditory and visual localization: Two modalities, one world," in *Audio Engineering Society Conference: 12th International Conference: The Perception of Reproduced Sound*, Audio Engineering Society.

Perrott, D. R., Marlborough, K., Merrill, P., and Strybel, T. Z. (1989). Minimum audible angle thresholds obtained under conditions in which the precedence effect is assumed to operate. *J. Acoust. Soc. Am.* 85, 282–288. doi: 10.1121/1.397735

Perrott, D. R., and Saberi, K. (1990). Minimum audible angle thresholds for sources varying in both elevation and azimuth. *J. Acoust. Soc. Am.* 87, 1728–1731. doi: 10.1121/1.399421

Pulkki, V. (1997). Virtual sound source positioning using vector base amplitude panning. *J. Audio Eng. Soc.* 45, 456–466.

Pulkki, V. (2001a). Localization of amplitude-panned virtual sources ii: two-and three-dimensional panning. *J. Audio Eng. Soc.* 49, 753–767. doi: 10.1016/S0920-4105(01)00137-1

Pulkki, V. (2001b). *Spatial sound generation and perception by amplitude panning techniques (Ph.D. dissertation).* Helsinki University of Technology, Espoo, Finland.

Pulkki, V., and Karjalainen, M. (2001). Localization of amplitude-panned virtual sources i: stereophonic panning. *J. Audio Eng. Soc.* 49, 739–752. doi: 10.1121/1.1377288

Pulkki, V., and Karjalainen, M. (2008). Multichannel audio rendering using amplitude panning [dsp applications]. *IEEE Sig. Proc. Mag.* 25, 118–122. doi: 10.1109/MSP.2008.918025

Saberi, K., Dostal, L., Sadralodabai, T., and Perrott, D. (1991). Minimum audible angles for horizontal, vertical, and oblique orientations: lateral and dorsal planes. *Acta Acust. United Acust.* 75, 57–61.

Schütt, H. H., Harmeling, S., Macke, J. H., and Wichmann, F. A. (2016). Painfree and accurate bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vis. Res.* 122, 105–123. doi: 10.1016/j.visres.2016.02.002

Tyler, R. S., Parkinson, A. J., Wilson, B. S., Witt, S., Preece, J. P., and Noble, W. (2002). Patients utilizing a hearing aid and a cochlear implant: speech perception and localization. *Ear Hear.* 23, 98–105. doi: 10.1097/00003446-200204000-00003

Tyler, R. S., Witt, S. A., Dunn, C. C., and Wang, W. (2010). Initial development of a spatially separated speech-in-noise and localization training program. *J. Am. Acad. Audiol.* 21, 390–403. doi: 10.3766/jaaa.21.6.4

Van Deun, L., Van Wieringen, A., Van den Bogaert, T., Scherf, F., Offeciers, F. E., Van de Heyning, P. H., et al. (2009). Sound localization, sound lateralization, and binaural masking level differences in young children with normal hearing. *Ear Hear.* 30, 178–190. doi: 10.1097/AUD.0b013e318194256b

Völk, F. (2013). Interrelations of virtual acoustics and hearing research by the example of binaural synthesis.

Völk, F. (2016). "Minimum-audible angles in wave-field synthesis: a case study," in *Audio Engineering Society Convention 141*, Audio Engineering Society.

Völk, F., Gottschalk, A., and Fastl, H. (2012a). "Evaluation of binaural synthesis by minimum audible angles," in *Tagungsband Fortschritte der Akustik-DAGA, Darmstadt (2012)*, Darmstadt.

Völk, F., Mühlbauer, U., and Fastl, H. (2012b). "Minimum audible distance (mad) by the example of wave field synthesis," in *Tagungsband Fortschritte der Akustik-DAGA, Darmstadt (2012)*, Darmstadt.

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.

frontiers in Digital Health

REVIEW
published: 10 August 2021
doi: 10.3389/fdgth.2021.724370

A State-of-Art Review of Digital Technologies for the Next Generation of Tinnitus Therapeutics

Grant D. Searchfield[1,2,3], Philip J. Sanders[1,2,3], Zohreh Doborjeh[1,2,3], Maryam Doborjeh[4], Roger Boldu[5], Kevin Sun[1] and Amit Barde[6]*

[1] Section of Audiology, The University of Auckland, Auckland, New Zealand, [2] Eisdell Moore Centre, The University of Auckland, Auckland, New Zealand, [3] Centre for Brain Research, The University of Auckland, Auckland, New Zealand, [4] School of Engineering, Computer and Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand, [5] Augmented Human Laboratory, Auckland Bioengineering Institute, The University of Auckland, Auckland, New Zealand, [6] Empathic Computing Laboratory, Auckland Bioengineering Institute, The University of Auckland, Auckland, New Zealand

OPEN ACCESS

Edited by:
Fan-Gang Zeng,
University of California, Irvine,
United States

Reviewed by:
Jameel Muzaffar,
University Hospitals Birmingham NHS Foundation Trust, United Kingdom
Magdalena Sereda,
University of Nottingham,
United Kingdom

****Correspondence:***
Grant D. Searchfield
g.searchfield@auckland.ac.nz

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: *13 June 2021*
Accepted: *12 July 2021*
Published: *10 August 2021*

Citation:
Searchfield GD, Sanders PJ, Doborjeh Z, Doborjeh M, Boldu R, Sun K and Barde A (2021) A State-of-Art Review of Digital Technologies for the Next Generation of Tinnitus Therapeutics. Front. Digit. Health 3:724370. doi: 10.3389/fdgth.2021.724370

Background: Digital processing has enabled the development of several generations of technology for tinnitus therapy. The first digital generation was comprised of digital Hearing Aids (HAs) and personal digital music players implementing already established sound-based therapies, as well as text based information on the internet. In the second generation Smart-phone applications (apps) alone or in conjunction with HAs resulted in more therapy options for users to select from. The 3rd generation of digital tinnitus technologies began with the emergence of many novel, largely neurophysiologically-inspired, treatment theories that drove development of processing; enabled through HAs, apps, the internet and stand-alone devices. We are now of the cusp of a 4th generation that will incorporate physiological sensors, multiple transducers and AI to personalize therapies.

Aim: To review technologies that will enable the next generations of digital therapies for tinnitus.

Methods: A "state-of-the-art" review was undertaken to answer the question: what digital technology could be applied to tinnitus therapy in the next 10 years? Google Scholar and PubMed were searched for the 10-year period 2011–2021. The search strategy used the following key words: "tinnitus" and ["HA," "personalized therapy," "AI" (and "methods" or "applications"), "Virtual reality," "Games," "Sensors" and "Transducers"], and "Hearables." Snowballing was used to expand the search from the identified papers. The results of the review were cataloged and organized into themes.

Results: This paper identified digital technologies and research on the development of smart therapies for tinnitus. AI methods that could have tinnitus applications are identified and discussed. The potential of personalized treatments and the benefits of being able to gather data in ecologically valid settings are outlined.

Conclusions: There is a huge scope for the application of digital technology to tinnitus therapy, but the uncertain mechanisms underpinning tinnitus present a challenge and many posited therapeutic approaches may not be successful. Personalized AI modeling based on biometric measures obtained through various sensor types, and assessments of individual psychology and lifestyles should result in the development of smart therapy platforms for tinnitus.

Keywords: review, digital, therapy, tinnitus, treatment, technology, biometrics

INTRODUCTION

Tinnitus is commonly referred to as "ringing in the ears;" it is the perception of a sound in the absence of a sound source. Tinnitus is the result of a complex cascade of changes within the auditory and emotional networks of the brain that occur following ear or head injury (1). Tinnitus can have mild through to catastrophic effect on life-quality; it can disrupt hearing, attention and sleep, result in anxiety, and depression (2). The incidence of significant tinnitus is highest amongst older populations (3). There is currently no cure for tinnitus, largely due to its heterogeneity (4). Much of the severity of tinnitus relates to the sufferer's psychological response to abnormal auditory and emotional inputs. Stress, anxiety and depression have been shown to occur with tinnitus or contribute to a greater, negative and sustaining reaction to tinnitus (5). Cognitive processes are suspected to affect how dominating tinnitus may or may not become, how easy or difficult it may be to ignore, and whether attention can easily be diverted away from, or is captured by, this endogenous signal (6). In the absence of effective medication for tinnitus a combination of sound-based therapy, to disrupt auditory processing of the tinnitus signal, and counseling therapy to reduce the potentiation of the tinnitus signal by emotional neural networks, has become common (7, 8). This therapy is usually provided by audiologists and is demanding of both clinician and patient time. Its benefits are often only apparent after months and may require the expense of HAs (9). Such sound therapy is very beneficial for some patients but is of limited or no benefit for others (10). This means that considerable cost in health delivery is incurred without any certainty of benefit (8).

Digital processing has enabled the development of several generations of technology for treating tinnitus. Computers have been used for tinnitus assessment since the early 1980s (11) and the internet ushered in online information (and misinformation) becoming readily available. But it was the first commercially successful digital HAs in the mid-1990s that marked the beginning of the first generation of tinnitus treatments based around digital technology (12). The first-generations of digital HAs appeared to be more effective tinnitus therapy tools than their analog predecessors (13). The beginning of the Apple iTunes store in 2001 facilitated the availability and knowledge surrounding downloading of sounds to personal music players (e.g., MP3 players, iPods). MP3 players enabled earlier self-help tinnitus masking strategies developed using the SONY Walkman (14) to be replicated, but with a greater playtime, more listening options and longer battery life. Digitization of sound files and MP3 players has continued to enable the rapid prototyping of sound therapies (15). As of April 2021, a search on the Apple iTunes store identified over 100 digital albums identified as "tinnitus relief." In the early 2000s a stand-alone tinnitus treatment device, resembling an MP3 player, called the Neuromonics was released based on the concept of systematic acoustic desensitization (16). It implemented a two-stage therapy of noise combined with a music calibrated to have a flat spectrum and weighted to the individuals audiogram profile (17).

We identify the second generation of digital tinnitus devices as offering user-selectable therapy options and, in the case of wearable devices, wireless connectivity. With increases in memory and processing capacity HA manufacturers were able to include tinnitus treatment sounds in many of their models, usually consisting of broadband noise derivatives, some synthesizing surf like sounds (18), others fractal sounds (digital chimes) (19). The emergence of "made for iPhone" HAs saw Smart-phone apps able to connect with HAs via Bluetooth (20). At first Bluetooth power demands required an intermediary device receiving the Bluetooth transmission and retransmitting as a lower drain signal to HAs (21). This inconvenience was addressed when direct connection between phone and HAs become possible. The first universal (Android and Apple) connective HA became available in 2017–2018 (22). As Smartphones became popular their onboard MP3 software negated the need for a separate dedicated MP3 player. The ubiquitous nature of the Smartphone enabled tinnitus apps to be developed by manufacturers and third party developers, these first generation apps consisted primarily of sound libraries that extended the range of sounds available to the listener. Online tinnitus clearinghouses (e.g., www.tinnitus.org.uk) and clinic websites (www.tinnitustunes.com) made a wider range of resources available, including podcasts and sound files to download.

The 3rd generation of digital tinnitus technologies emerged from a plethora of new therapy concepts based on specific, putative, neurophysiological mechanisms (1). Third generation therapies attempt to personalize to a characteristic of the users' tinnitus (7). Digital processing algorithms enabled these new therapies to be implemented through HAs, apps, the internet and stand-alone devices. Sound therapies may be based on tinnitus pitch and use notched sound in an attempt to achieve lateral inhibition (23, 24), patterned tones to create desynchronization (25) or be tinnitus replicas for nocturnal habituation (26). Digital processing has also facilitated multimodal stimulation pairing sound with vagus nerve or trigeminal nerve stimulation. Apps and online counseling services have become more interactive (for example chatbots). Despite technological improvements 3rd generation therapies appear to offer little in population-based benefits above 2nd generation (and possibly 1st generation) approaches (27). Almost without exception tinnitus technology has been made available commercially before clinical trials have shown efficacy. It is possible that some of these 3rd generation devices and therapeutics are very effective, but for whom and when is unclear. Certain subsets of the population may be more responsive to one or other therapies, and this may not be static but change with chronification or dynamically through the day to circadian rhythms. To address this variability, we need to be able to ascertain individual differences that predict the most appropriate therapy, and potentially adjust in real-time to a marker of some tinnitus property that can be modified. Recent AI technologies have presented the opportunity to realize such goals (28). In recent years, AI methods have been implemented in a variety of health settings for the purpose of early diagnosis, developing smart therapy platforms and prediction of response to treatment. The rapid growth of AI-driven technologies has allowed tinnitus researchers to consider AI methods and applications to address open questions in tinnitus data analytics, tinnitus management and accelerating decision-making for choosing the best course of treatment, more specifically, toward development of smart therapy. It is our ambit that such smart tinnitus devices represent the 4th generation of the evolution of tinnitus therapeutic technology.

We are on the cusp of a 4th generation of digital tinnitus therapy that we believe will be defined by the incorporation of physiological sensors, multiple therapy options and AI to personalize therapies. This new generation of tinnitus technology coincides with maturing wearable technology, and in particular Hearables. Hearables are ear level wearable computers or computer interfaces (29, 30). While several start-up companies have come and gone, the release of Hearables by large consumer electronics companies and the inclusion of Hearable features into HAs indicate a pathway for tinnitus technology innovation. As awareness for the heterogeneity of tinnitus continues to grow within the scientific community, research is beginning to move toward precise treatment of tinnitus which is tailored for an individual (31, 32). For tinnitus treatments to be truly individualized, one must understand the physiological and psychological biomarkers of tinnitus and how they influence treatment outcome and selection (7). Progress toward precision, data-driven, treatment of tinnitus requires either large datasets to better understand tinnitus heterogeneity or in-depth repeated measures in individuals in which the technology adapts to, or learns, personal preferences and effective decisions (33). AI can be applied to these datasets (34, 35).

In this state-of-the-art review we will catalog and describe technology that has the potential to deliver real-time customized tinnitus treatment that extends beyond the decision making capability and capacity of clinicians alone. Advancements in mobile computing technology enable ecologically valid technology-based interventions tailored to individual needs. The aim of this review is to identify and discuss potential sensors, transducers and algorithms that may comprise the next generations of digital therapies for tinnitus. To capture information in peer reviewed journals, industry whitepapers and forums, a "state-of-art" review was implemented. State-of-the-art reviews are a specific form of review that focus on current issues and new perspectives, often in areas with a need of further research (36). This is an inclusive form of review, that captures information from a wide variety of sources; it does not exclude material based on a quality criterion in the way systematic reviews do. A trade off in sourcing a wide range of information is the inclusion of some material that be of low evidence quality, for example expert opinion.

METHODS

A state-of-art review (36) was undertaken in March 2021 with cataloging of results in April 2021. The aim of the review was: To review proposed and potential sensors, transducers and algorithms that may comprise the next generations of digital therapies for tinnitus. Google Scholar and PubMed were searched for the 10-year period 2011–2021. The search strategy used the following key words: "tinnitus" and ["Hearing aids," "personalized therapy," "Artificial intelligence" (and "methods" or "applications"), "Virtual reality," "Games," "Sensors," and "Transducers"] after an initial search an additional search term "Hearables" was added. The reference lists of these articles were searched for additional pertinent articles including in Gray literature (e.g., public domain consultancy documents, consumer electronics magazines and blogs), older articles were included if they provided context. The authors' knowledge of expert topic areas were used to identify gaps in the search outputs and fill these with appropriate source material. Studies were charted according to technology types and purposes: Hearing aids and cochlear implants, Hearables, Internet-based therapies, Dedicated sound and multimodal therapy devices, Apps, Virtual and Augmented reality, EMA, Sensors, and AI. The articles were cross-references if content lay across categories. A narrative was constructed from the chosen material, based around a pragmatic worldview.

RESULTS

Hearing Aids

Hearing Aids (HAs) have been used as tinnitus management tools since at least the late 1940s (37) and this continued with the arrival of fully digital aids in the mid 1990s. HA are used in tinnitus management to reduce accompanying hearing handicap, reduce the levels of attention paid to tinnitus, compensate for deafferentation, and by raising the audibility of environmental sounds so that tinnitus can be masked (38). In the early 2000s the United Kingdom National Health Service modernized their hearing aid program, and digital aids became available to NHS patients. It was found that tinnitus outcomes improved with a shift from analog to digital aids (13). The digital benefit was attributed to greater high frequency amplification (13). New digital processing strategies within 2nd generation tinnitus HA and the emergence of made for iPhone HAs (39) may have resulted in some increment improvements in outcomes, although evidence for this is poor. Sound introduced in this next generation of HAs included synthesized natural sounds, such as ocean surf sounds (18). Streamed nature sounds and Broad Band Noise (BBN) have been found to be equally effective over 6 months (40). Digitally-rendered, fractal sounds resembling musical chimes have also been used (41). Most trials of fractal sounds have shown benefit but have been open label, with large individual variability (19, 42).

A scoping review found HAs were beneficial for tinnitus management (43). A Cochrane review of amplification for tinnitus and hearing loss (44) identified only 1 study (45) meeting their quality criteria which compared digital HA to sound generator (maskers), both groups improved on the Tinnitus Handicap Inventory but there was no statistical difference between groups. A follow-up review (46) included eight studies (with a total of 590 participants). Seven of the studies investigated HAs, four combination HAs and three sound generators. There was insufficient evidence to differentiate between outcomes of the sound therapy options and the level of evidence of an overall benefit was low (46). So, while widely adopted for tinnitus control there is little evidence for 2nd generation hearing aid tinnitus efficiency above 1st generation approaches. The current volume of evidence suggests that hearing aid amplification is an effective way to treat tinnitus, but the research that supports this evidence is not of a high quality (18, 43).

Third generation digital hearing aid tinnitus solutions are translations from body-worn dedicated sound therapy devices based on novel neurophysiological theories (described in a subsequent section). In a case example, applied through HAs, Acoustic Coordinated Reset Neuromodulation, a treatment using patterned tones, seemed feasible (47). Notched noise and music have been used in attempts to inhibit tinnitus (23, 24). Notched amplification takes a similar approach to the use of notched music, but in this case amplification of sound is not applied surrounding tinnitus pitch. There is currently no compelling evidence that novel sound processing adds benefit to conventional amplification for tinnitus (48, 49).

The next generations of HAs are going to apply increasing levels of AI and incorporate biosensors (50). Already HAs are beginning to feature better fidelity (51) and have begun to incorporate sensors such as for fall detection (52) mirroring similar developments in fitness trackers and Hearables described later. As hearing loss often accompanies tinnitus HAs are a logical platform for data-driven wearable tinnitus therapy. Efforts to develop cognitively controlled HAs (53) and the development of ear-based EEG (54) could be extended to tinnitus treatment solutions with real-time adjustment based on AI.

Cochlear Implants

Cochlear implants are surgical acoustic-electrical transducers that comprise a digital signal processor similar to a hearing aid and a multiple electrode array that is inserted into the cochlear (55). The array provides direct stimulation of the auditory nerve through the tonotopic electrode array when hearing aid amplification would be insufficient to improve hearing. Cochlear implants replace the role of HAs when hearing loss is severe (55). Cochlear implants provide electrical stimulation based on sound patterns, but can be considered a Sound Therapy device because they activate auditory pathways. Interest in cochlear implants for tinnitus therapy began in the 1980s (56), but early experience with implantation for tinnitus showed limited success (57). Cochlear implants are becoming more common therapy options when tinnitus accompanies a severe unilateral hearing loss (58) Tinnitus Sound Therapy strategies developed for HAs, for example apps, are also being trialed with implants (59). A systematic review indicated low-level evidence for the benefits of cochlear implants on tinnitus; they appear to help people who had severe tinnitus prior to the implant, but do carry some risk that the implant surgery may exacerbate or initiate tinnitus (60). As knowledge of tinnitus mechanisms advance there are great opportunities to introduce novel stimulation paradigms through cochlear implants that target tinnitus in addition to, or alternatively, to the focus on speech understanding.

Hearables

Crum (30) likened the ear to a biological "USB" port, presumably with the brain as the "CPU." A range of sensors and transducers worn in the ear can measure or affect physiology. These ear level devices have become generically known as "Hearables" (29, 30). A Hearable can be defined as a device that fits in the ear that contains a wireless link (29). We prefer to define the Hearable as an ear level wearable computer or computer interface. Both definitions encompass many HAs as well as ear level Bluetooth headphones and fitness trackers. HAs have been reviewed in a separate section primarily because of their established long history in tinnitus therapy. Although HAs and Hearables appear to be converging technologies they can, at present, be separated by their primary consumers, the music listening, fitness focused, younger public (Hearables) and those persons with greater than mild hearing loss (HAs). It is with this in mind that after setting the initial review criteria it was evident that AI and sensors are being, or could be, incorporated into Hearables for tinnitus therapy, consequently "Hearables" was added as a separate search term. The ambit of the search for this section of the review is to scope the development of Hearable design and their current or potential application to tinnitus.

In the mid 2010s several startup companies emerged in the Hearable space to considerable consumer electronics and media attention. Biometric sensors saw the Bragi Dash emerge as a trend setter amongst these first generation Hearables (61). Some of these companies are no longer operating [e.g., Soundhawk (39); Doppler (62) or have moved from manufacturing to focus on AI development as a 3rd party for other manufacturers Bragi (63)]. The 2016 introduction of Apple Air Pods, for which "Live Listen" allows functioning as basic HAs, signaled an important juncture for access to augmented communication (39) and digital tinnitus therapies. Recent Apple accessibility updates to Air Pods include the availability of background sounds, that could be used for tinnitus masking. The recent purchase of the consumer headphone brand Sennheiser by the hearing aid manufacturer and retailer Sonova is a further indication of market convergence and opportunities for capturing market share across the spectrum of hearing, and potentially tinnitus, needs (64).

Hearables could be used as an alternative to HAs as device for 2nd generation therapy, for example wireless access to sound libraries [e.g., Bose Sleepbuds (65)]. Onboard access to internet of voice (29) may enable virtual counselors beyond simple chatbots. The real promise of this technology is the potential to combine biometrics (e.g., EEG, heart rate, temperature, skin resistance, blood oxygen, and stress hormone levels) with auditory or other sensory stimulation (30). These measurements and response could inform status of mental effort, stress, engagement and attention, direction of vision and physical health (30). The types of sensors that would enable Hearables to be used in 4th generation digital therapies are described in greater depth in a later section of this review. In an opinion piece Crum (30) specifically mentions tinnitus as a health application for Hearables. Bragi began to explore the potential of incorporating tinnitus treatments in future generations of their Hearable (66) before moving away from manufacturing. At the time of writing Nuheara was one of the few Hearable companies highlighting hearing loss (67), and had identified tinnitus as a market (https://www.nuheara.com/how-it-helps/tinnitus-relief/).

Internet-Based Therapies

Internet-based therapies include Internet Cognitive Behavioral Therapy (iCBT), online counseling, peer support and tailored sound. The 1st generation internet was used for tinnitus in a manner similar to a written self-help book, as a repository for information. ICBT programs developed greater content as a form of 2nd generation digital tinnitus treatment. ICBT typically consists of text-based modules for tinnitus patients to work through. An example is the Tinnitus E-program, a 10–12 week self-directed approach consisting of education and information about tinnitus, management, resources available, training for psychological strategies, social support, and monitoring of tinnitus (68). In addition to written information, behavior change techniques such as relaxation methods are available as downloadable MP3 files (68). ICBT in various forms has a good level of evidence to support its use (69). Tinnitus Tunes (www.tinnitustunes.com; established in 2016) is another 2nd generation internet-based digital tinnitus therapy. Subscribers undertake self-directed activities to complement their clinician's advice. It offers a 12-week structured program consisting five separate steps: (1) Education and information so that members can remove any false beliefs they have about tinnitus. (2) Information about role of different clinicians. (3) Managing stress associated with tinnitus (e.g., relaxation sound files, audio podcasts on visualization and progressive relaxation). (4) Training the patient's brain to ignore tinnitus through attention refocusing and adaptation. (5) Prevention of relapse through lifestyle tips. Weekly emails are sent to users that include case studies and lifestyle tips to suit user needs.

Notched sound/music is a 3rd generation online therapy; based on a specific tinnitus mechanism and customized to the individual. Notched sound involves customizing sound (usually music) by removing of sound energy in a band around the patient's reported tinnitus pitch. Two examples are https://www.audionotch.com and https://www.tinnitracks.com/en. The concept is based on experimental work that notched sound may result in lateral inhibition (23). The user chooses the audio signal they wish to have notched (e.g., music or white noise) an online processing algorithm applies the notch, then sounds are played over the computer or downloaded to be played on a portable device. The use of notched sound from the internet or app has only a low level evidence base (70). A double-blinded controlled trial of notched sound found no significant change in the primary outcome of the Tinnitus Questionnaire, although there was a change in a tinnitus loudness rating (24).

Ecological Momentary Assessments and Mobile Crowdsensing

EMA and mobile crowdsensing have been gathering momentum in tinnitus research in recent years [see recent reviews (71, 72)]. EMA describes the collection of data from participants as they go about their everyday lives, outside of research laboratories or clinical appointments. The ubiquity of smartphones in modern society has made implementation of this approach much more feasible and cost-effective than in the past. Smartphones have in-built sensors, processing, and data transfer capabilities. Participants can be notified via their phone at any time to complete surveys (also delivered on their phone) or data can be collected from the in-built sensors and connected devices such as smart watches (73). The EMA approach stands in stark contrast to the highly controlled environments used in traditional clinical research. These two styles of data collection each have their own strengths and weaknesses, and can serve to inform and compliment the other. With a focus on group effects, traditional clinical research aims to remove as much variability as possible from the testing environment between appointments (within and between participants) in order to try and reduce confounding factors (74). However, these research settings lack ecological validity as they tend not to reflect the participants' everyday environment. EMAs offer ecologically valid measurements at the expense of control over the environment. Interaction between effects and different environments is often an interesting and relevant avenue for research (72, 75). Furthermore, while group studies are undeniably useful for assessing treatments, their methods inherently lack the nuance required for understanding individual differences in conditions and responses to treatments. The EMA approach has been useful in fields investigating conditions of high heterogeneity of symptoms and treatment responses such as psychology and psychiatry (76). One of the reasons that successful tinnitus treatments have remained elusive and of variable benefit is because of the highly heterogeneous nature of the condition in terms of etiology, experience, reaction, and response to treatment (4).

Smartphone-delivered EMA tinnitus studies have shown high engagement from participants, especially from people with more severe symptoms (77–81). A large, longitudinal study (81) found that there was a high drop-off rate after the first few days of initial engagement, but that predictors of continued engagement were evident at these early stages. They suggested that personalized motivators should be considered to increase adherence. This suggests that people will engage with tinnitus-focused EMAs but future apps could increase adherence through more engaging and interactive content, such as gamification of data collection methods (82). The incorporation of sensor data in such apps could lead to more targeted timing of EMAs and perhaps less reliance on subjective survey data (though this data is still highly relevant in tinnitus research at this stage). Software is already being developed to support integration of sensor data with

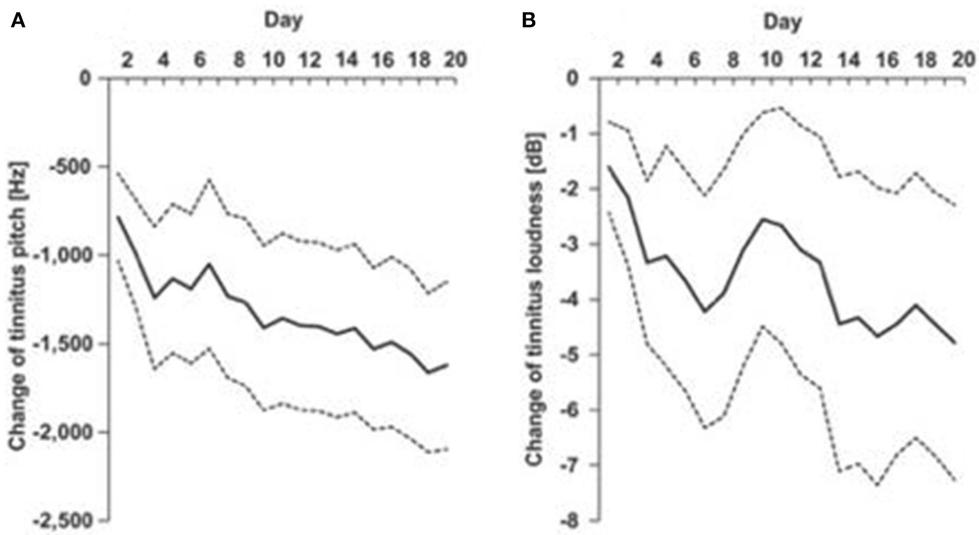


FIGURE 1 | EMA of tinnitus pitch **(A)** and loudness **(B)** recorded each day of a 20 day multisensory training program [Spiegel et al. (89), used with permission of authors].

treatment plans and allow information flow between patients, clinicians, and researchers (83).

EMA studies have revealed various factors accounting for within- and between-subject differences in tinnitus perception, including time of day, emotional state and arousal, stress level, concentration, and environmental sound (71, 84, 85). Measures of tinnitus impact (such as annoyance and perceived loudness) can fluctuate over a day, and between days (85). EMAs collected over long periods could be used to examine and identify patterns within patients and potentially enable the development of precision treatments in terms of the timing and targets of treatment.

Data collected from the combination of sensor and EMA technologies has the potential to advance individualized tinnitus treatment and research for patients, clinicians and researchers. Patients could be given more control over their tinnitus by being informed about situations and environments that are likely to increase or decrease their tinnitus symptoms (71). This information could be delivered and interventions suggested in real-time if AI algorithms were developed to predict these situations accurately and produce feedback based on case history, EMA and biosensor data (86). Researchers could also use biosensor data to automate the delivery of EMAs to gain subjective data during high or low stress situations (72). There is potential for data mining opportunities with large anonymized EMA datasets that could detect patterns in data to elucidate tinnitus subtypes, environmental impacts and circadian patterns, and inform treatment.

Mobile crowdsensing and EMA tend to be more cost-effective to implement than clinical trials, and require less labor (especially with increasing automation capabilities likely in the future), have the added benefits of ecological validity, the ability to examine heterogeneous populations, and the elimination of recall bias (84). It has also been suggested that these methods could inform clinical trials. Selection of participants (and therefore sensitivity) could be improved with machine learning used to identify participants most likely to respond to therapies (78). Time of day has been identified as a key consideration for when measurements are taken (84, 85). Accounting for this will improve the accuracy of trial results. EMA and mobile crowdsensing tinnitus research supports the development of interventions that reduce stress and improve emotional stability to improve tinnitus symptoms (71). The research has also revealed the need to further investigate circadian factors in tinnitus, and the underlying chronobiological mechanisms (85, 87).

EMA has been employed in tinnitus using: Palm pilots (77) text messaging linked to surveys (72) a TrackYourTinnitus app on Smartphones (85) and Smartwatches (73), and as part of a serious games (82, 88) and multisensory Training (89, 90) (**Figure 1**). All these methods for EMA are feasible with response rates to text reminder EMA for tinnitus being high (79–88%) (72, 78). Average time to complete questionnaires on smartwatch were 7.8% longer than on a smartphone (215.4 s & 198.6 s respectively) (85).

Common outcomes among studies include: tinnitus fluctuates within and between days, there is a link between emotional dynamics, stress and tinnitus loudness and distress, and some correlation between increased severity of tinnitus and different times of the day (71) although not all studies showed such variation (77). Tinnitus loudness and distress have been found to be most severe at night and early morning, while stress was most severe in the afternoon (85). The ability to track tinnitus changes over a day may enable individual tailoring in timing of therapeutic interventions, possibly in tune with circadian rhythms (85). In addition to question-based EMA tracking treatment progress through pitch and loudness matching is achievable (82, 88). The addition of biosensors would increase the information being obtained, and with AI, enable real-time adjustments.

Biosensors

Physiological assessment of function has been part of medical diagnosis for decades. Measures such as blood pressure (BP), heart beat rate (HR), electrocardiograms (ECG/EKG), functional magnetic resonance imaging (MRI) magnetoencephalograms (MEG), and electroencephalograms (EEG) are routinely used in clinics, hospitals and research labs to assess cardiac (BP, ECG), and brain (EEG/MRI/MEG) function. BP, ECG, and EEG along with many other objective physiological, or psychological, function tests can be undertaken in the field, in ecologically valid settings (91, 92). Although there is no single physiological objective measure of tinnitus itself, years of lab based search has identified many related markers of tinnitus related activity, for example neural networks associated with tinnitus (93), measures of emotion, and stress (94). New miniaturized wearable technology is now available to make longitudinal measures of physiological function (95–104) that can be related to behavioral indices of tinnitus.

A strong link between tinnitus and stress has been found epidemiologically, and is seen in clinical practice; patients often report tinnitus onset or increased tinnitus severity in response to stressful events (94). People with tinnitus display dysfunction in both the short-term stress response mediated by the sympathetic nervous system, and the long-term stress response mediated by the hypothalamic pituitary adrenal axis (105). Imaging studies have found increased activity in autonomic brain regions in those with tinnitus, which may act to generate and maintain the tinnitus percept and the related emotional distress (106). If tinnitus related distress is maintained long-term, the sympathetic response may become blunted (105). It is feasible to continuously monitor various physiological variables, including the inter-beat interval (from which heart rate and heart rate variability can be derived), blood volume pulse and skin conductance and relate these to tinnitus (107). Skin conductivity can be used to measure stress, it may be possible to measure cortisol and volatile organic components from the skin (103) possibly using flexible electronic patches (99, 101).

EEG is a non-invasive electrophysiological measure of brain function with high time resolution. In laboratories it is normally recorded from multiple wired electrodes across the skull, often held in position by a cap and electrical conductance enhanced with a conducting gel (97). We are now able to move measures of EEG outside of the lab and record frequently across hours or days, this information can be stored and or relayed back to the researchers. A key feature of wearable EEG is the integration of the electrodes and connecting leads into headsets or ear-level devices (102, 104). Various forms of wearable EEG have been developed and trialed, silver electrodes in a custom earmold (108) cloth electrodes with non-custom earbud (96, 109) conductive silicone electrodes in the ear and around the ear (104) and fingered electrodes to improve contact (97). At present there is a tradeoff between the convenience and usability of these devices and resolution and/or range of applications. Fewer electrodes means brain regions removed from the recording site contribute less to the recorded signal and there is less data. The form of the EEG headset should match the most important spatial region(s) generating the activity of interest, otherwise spatial smearing of activity various brain regions may occur.

Disruption of sleep is a common complaint amongst tinnitus sufferers. Polysomnography (PSG) is the gold standard for objective sleep monitoring (110). PSG uses EEG, electromyography, electro-oculography, electrocardiography, pulse oximetry, and numerous other measures. PSG is undertaken in a laboratory and is expensive and labor intensive. Actigraphy is less expensive method, that can be used at home, and estimates periods of wakefulness and sleep from timing, intensity, and duration of movements using inertial sensors (98). Incorporation of heart rate and variability, skin conductance and temperature along with movement should, depending on algorithms used, improve accuracy (98).

An example of the use of sensors in assessing therapy outcome was a small 12 participant randomized crossover study comparing visualization therapy with visualization paired with self-selected nature sounds. Sleep was assessed using wrist-based actigraphy (Actiwatch 2®) (111). Sleep quality was measured by actigraphy estimates of total sleep time, sleep onset latency, sleep fragmentation, and wake after sleep onset. Sleep onset latency significantly improved following both treatment conditions (111).

Medical monitoring and consumer electronics have converged with the development of smart fitness tracking watches. Much like the convergence of HAs and Hearables this blending of wearable technology is a potent catalyst to apply to the Tinnitus field. Consumer fitness and activity Smartwatches are popular, and provide increasingly accurate measures of function (98). Biosensors, alongside smartphone acquired context such as geospatial information (112), weather conditions and EMA may cue real-time treatment modification (100). Monitoring emotion may enable identification in changes in mood that may necessitate different therapy (96), sensors may be used to provide personalized coaching for health behaviors (95).

Apps

Several recent reviews have previously identified tinnitus-related apps, although scope, inclusion/exclusion criteria, and definitions differed between them (83, 113–118). Tinnitus patients have been surveyed about their preferences in apps (113), but aside from a small trials (119) evidence of benefit from apps for tinnitus is absent. Sereda et al. (113) found the five most commonly used apps in descending order from highest were: (1) White Noise Free, (2) Oticon Tinnitus Sound, (3) Relax Melodies–Sleep Sounds, (4) myNoise, and (5) Tinnitus Therapy Lite. Due to the fast development and release of new apps on various platforms, the list of apps is ever-growing and changing. Therefore, the number of tinnitus-related apps available at any time is difficult to report but recent reviews have suggested over 200 (116, 118). However, a common observation was that very few apps had been scientifically validated or tested for efficacy, and there was a high risk of bias in many of the studies that were available (83, 117). Furthermore, Sereda et al. (113) found that most of the apps that they reviewed were self-help apps that did involve clinicians. App services identified included CBT, assessments and self-measurement of hearing and tinnitus symptoms, EMA, serious games, and sound therapies (115). App-based tinnitus sound therapies take the form of masking, notched music, and hearing aid control. While these therapies are well-established, there is no guarantee that a given app is implementing them correctly. Similarly, without expert guidance app users may not use the therapies correctly or as intended, and some therapies may not be appropriate for some individuals.

Mobile delivery of therapies has several advantages over traditional clinic-based therapies. Barriers to accessing care such as distance or hesitancy to engage in face-to-face therapy can be overcome. Therapy, questionnaires, and compliance measures can all be implemented on the same device (120). Mobile therapy programs are also generally cheaper to implement, can reach a wider audience, and require less clinician time. Hauptmann et al. (120) found that tinnitus pitch matches measured in clinic vs. in an app did not differ, but the app measures showed less variability than the clinical measures. This is an example of how technology can save time for clinicians and even produce more reliable results. However, clinician involvement is still desirable in order to direct patients to valid and appropriate apps, and instruct them on correct use of therapies (116, 118). Smart technology opens up the possibility of collecting objective data regarding the user's physiology and environment in real-life settings. This can be achieved through the in-built sensors present in modern smartphones (e.g., GPS, microphone, camera, gyroscopes, and accelerometers) (115) as well as peripheral wearable devices such as smart-watches and hearables. However, limitations of hardware and differences between devices must be considered when interpreting the data (121). Other limitations to mobile therapy can include lack of validation, lack of expert supervision, and incorrect use. In research, samples can differ between platforms; newer platforms tend to attract younger participants than traditional advertisements through clinics (122). Drop off in usage after initial sign up is common in internet and mobile based therapy and research (119), although the reasons why are not well-understood (123). Inconsistencies in terminology have been identified as a problem in internet/digital psychological interventions that can lead to miscommunication and can hinder systematic reviews and research (124). Mitigation of inconsistent terminology should be considered in the tinnitus field to enable evidence-based validation of digital therapies. One review even suggested an app review board, similar to journal editorial boards to develop and ensure standards are met before apps can be recommended by clinicians (114).

Auditory Training and Serious Games

Auditory training is a learning method in which listeners are taught to make perceptual distinctions about sounds being presented. A review undertaken at the beginning of our decade of interest found improvement in outcome measures in nine of ten studies after auditory training, but with low-moderate levels of evidence (125). Frequency discrimination tasks have been the primary training mode (125, 126), but frequency categorization training has been suggested as an alternative method (127). The training tasks may have had benefit through attention mechanisms as opposed to pure sensory improvements in discrimination (6, 125, 128).

Gamification of training may be an important consideration to maintain motivation and compliance with auditory training (128). "Serious games" do not have entertainment as their primary goal but, instead, are intended to change behavior or teach new skills while being engaging and enjoyable (129, 130). A systematic review of application to psychotherapy and meta-analysis of serious games for mental health have indicated, with some caveats to research quality, that serious games are effective (129, 130). A game based on ignoring distractor sounds resembling tinnitus (on that day) while receiving points reward in identifying non-tinnitus target sounds has been developed and tested (82, 88). Users advanced through different levels of an increasing number of distractors across 20 days of 30 min game-play. A feasibility study demonstrated the tasks were achievable and the system was useable; preliminary data indicated significant reduction in tinnitus handicap (88). A controlled trial demonstrated Tinnitus Functional Index scores improved as did performances on audio and visual attention tasks (82). The N1 auditory evoked potential latency was also reduced for sounds remote from tinnitus pitch (82). The concepts of training to focus on target sounds while suppressing background sounds used by Wise and colleagues (6, 82, 88) have been mirrored by other developers (131).

Dedicated Sound Therapy Devices

Dedicated sound therapy devices are desktop, hand-held or body worn tinnitus devices that have a single purpose: tinnitus treatment. Neuromonics® therapy is a habituation-based, 1st generation passive music sound therapy. The Neuromonics® "Oasis" became available in the mid-2000s as a hand held digital sound player with Bose headphones. It is now available as a download for Smartphones, the "Oasis Pro" (https://neuromonics.com). It uses music that has been spectrally flattened (to reduce bass dominance) and adjusted to the individuals hearing (usually a treble increase). It consists of two stages: stage 1 is noise with modified music, stage 2 is modified music alone. Multiple trials of the treatment have been undertaken indicating success in reducing negative psychological aspects of tinnitus (17, 132). Neuromonics® treatment outcomes have been found to be equivalent to ear-level maskers (133).

Acoustic CR® (Coordinated Reset) Neuromodulation was developed from an electrical stimulation paradigm to treat Parkinson's disease (134). It is available as a handheld processor and wired air conduction headphones: the Desyncra™. The tinnitus treatment consists of temporally patterned tones of frequencies that span a tonal pitch match. The treatment aims to desynchronize aberrant neural ensembles. We consider the Desyncra a 3rd generation treatment, being based on a plausible neurophysiological mechanism and with personalization (tinnitus pitch); however evidence for its benefits are limited (135). An unpublished controlled trial suggested no benefit of the Desyncra treatment over an active control (134). Another 3rd generation device was the Otoharmonics® Levo System.

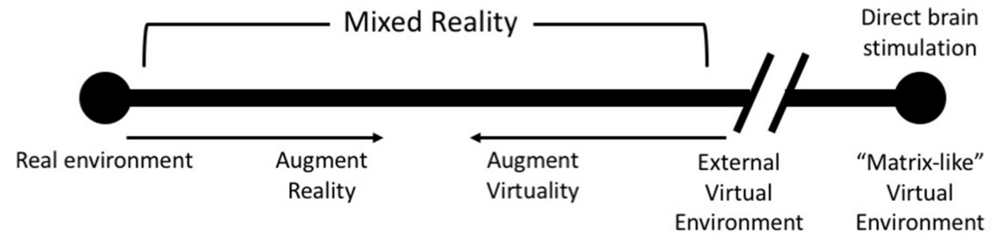


FIGURE 2 | An update of Milgram's Reality-Virtuality Continuum (150) proposed by Skarbez et al. (151) [image redrawn and modified from Skarbez et al. (151), Figure 2, page 3].

This consisted of tinnitus synthesis software Apple iPod and headphones. It used a tinnitus replica sound that was played during sleep and was based on the hypothesis that tinnitus emerges to replace an input deficit, matched sound should interrupt or reverse this (136). There is limited clinical evidence available demonstrating its efficacy, one trial showed clinically meaningful change in a questionnaire after 3 months (26); the Otoharmonics company appears to have ceased operating.

Multimodal Therapies

As the name suggests, therapies using multiple modes don't just use sound, but couple it with some other sensory stimulation or nerve modulation. Evidence of multimodal stimulation benefit comes from animal models of tinnitus (137, 138). The MicroTransponder, Serenity® pairs sounds with Vagal Nerve stimulation as an implanted device (139), The Neuromod Lenire couples sound stimulation with tongue tip™ trigeminal stimulation (140). The Neosensory "Duo" combines sound with wrist haptic stimulation (141). These three systems are available clinically in some countries. Computer-based perceptual training has also trialed combining sound, tactile and visual stimuli (89, 90). Clinical outcomes appear variable with questions as to what combination of sound and other stimulation is optimal (142). Further evidence is need to confirm that these treatments offer clinically meaningful benefits above auditory stimulation alone.

Virtual and Augmented Reality for Tinnitus Therapy

AR and VR have been used for the purpose of entertainment (143, 144), enterprise (145) and health care (146, 147). Another popular use for both these forms of technology has been for collaboration in virtual spaces (148, 149). The last decade has seen the emergence of Augmented Reality (AR) and Virtual Reality (VR) as healthcare tools. These emerging technologies are capable of generating immersive environments that leverage the body's perceptual capabilities. While VR provides a completely computer generated environment which isolates a user from the real world, AR incorporates the real world by augmenting the environment, visually and aurally, the user inhabits (150). VR and AR lie on opposite spectrums of the Reality-Virtuality continuum described by Milgram et al. (150) (**Figure 2**).

Researchers have suggested the use of VR for tinnitus; stating that the ability to expose patients to ecologically valid environments in a safe and regulated manner may help deliver more effective therapies for the treatment of tinnitus (152). Furthermore, at least one study demonstrates the use of VR for the treatment of tinnitus (153). While the results from that study show no differences between Cognitive Behavioral Therapy and VR (153), there are many parameters to explore, particularly the association of simultaneously occurring sensory cues. Most of the technological focus in virtual and augmented reality has been on visual stimuli. But virtual auditory scenes are important for immersive environments. Efforts have been made to create realistic auditory avatars of tinnitus (154) and HRTFs used for spatial rendering of sound to manipulate tinnitus perception through training (82) or masking (155). There is the possibility that by harnessing such auditory signal processing, coupled to vision and haptics, the perceived reality of tinnitus might be changed from an unreal phantom sound in the head to something resembling an ecological valid sound (156).

Familiarity with one's surroundings and other factors can significantly affect the efficacy of the treatment that is provided (75). Another important factor to consider is which of AR and/or VR technology is best for the purpose. Both AR and VR have their advantages and disadvantages. For example, with AR we can use the patient's natural surroundings in combination with computer generated auditory and visual cues to deliver therapy. While such an approach offers the best ecological validity, it does not afford complete control of the environment. In cases where control of the virtual environment is desirable in order to manipulate delivery of the therapy, the inherent flexibility of VR can help achieve this with relative ease. More work needs to be done, both with AR and VR, to explore how these technologies can be used in an effective manner for tinnitus. Some of this work could possibly involve exploring how "Matrix-like" direct brain stimulation (**Figure 3**) could be used to manipulate the interoceptive sense to deliver effective treatment (151).

Artificial Intelligence: Methods and Applications in Tinnitus

The 4th generation of digital tinnitus therapies will almost certainly require AI to automate functions and make use of multimodal sets of data acquired through biosensors. Developments of new algorithms in AI, and its adoption by healthcare providers, have been transforming the tinnitus field in many ways, with impacts on areas including personalized diagnosis, prognosis and smart therapeutics (157). This section

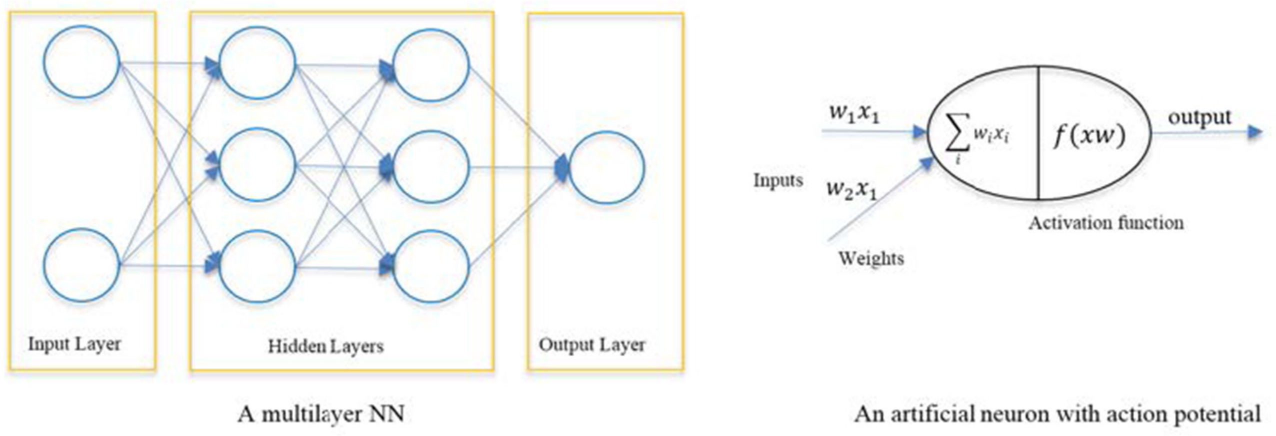


FIGURE 3 | Block diagram showing components of an artificial neuron.

focuses on the state-of-the-art methodological developments in AI in tinnitus studies.

Analytical AI Methods for Tinnitus

AI can be used to develop intelligent systems and devices. Smart algorithms can learn from multimodal sets of data to extract meaningful patterns that can indicate certain health outcome (diagnosis and prognosis), more accurately and faster than traditional approaches (157). The application of AI techniques to tinnitus started relatively recently and is so far limited. Thus far, AI algorithms have mostly been used in operational aspects of tinnitus such as comparative analysis (tinnitus vs. control), evaluating tinnitus-related distress, and individualizing tinnitus treatments through feature selection, classification, and prediction tasks (34, 83, 158).

AI has been applied to develop advanced systems (machines) that can learn from data, so-called Machine Learning (ML) in which a variety of computational architectures and learning algorithms have been emerged to increase the accuracy of decision making and decision support (159). The most commonly developed ML systems are based on Artificial Neural Networks (ANN) which loosely model the information processing mechanism observed in neurons in a biological brain (160). ANNs are organized in three main layers of neurons (input layer, hidden layer, and output) (**Figure 3**). The input layer nodes pass the input data (e.g., biosensors) to the ANN, the hidden layer neurons are computational units that learn from the input data while applying non-linear functions which link the input samples to the proper output neurons (diagnosis labels) (160, 161). State-of-the art in ANN is Deep Learning (DL) in which several hidden layers of neurons are used, each performing automatic extraction informative features from the input data and then pass them to the next layer (162). This is a modern variation of ANN that permits practical application and optimized implementation of non-linear classifications.

The main AI methods that are being used in tinnitus research alongside current and potential AI-driven applications are described in subsequent sections (**Figure 4**).

Machine Learning in Tinnitus

ML is a field of study that applies the principles of computer and mathematical science and statistics to create computational models, which are used for future predictions (based on past data) and identifying patterns in data (159). The use of ML has increased in healthcare applications and has been applied to behavioral, EEG, functional Magnetic resonance imaging fMRI, and Functional near-infrared spectroscopy (fNIRS) data (163–165). The learning algorithms in ML can be divided into two main groups: supervised and unsupervised (166). Supervised methods are based on learning from labeled datasets to create a predictive statistical model based on mapping input data to an output decision (prognosis or diagnosis outcomes). In unsupervised learning, data is unlabeled and the algorithms learn from the input data to detect differences and categorized the data into different groups (166, 167).

ML algorithms solve several tasks including clustering, classification and prediction. Clustering is an approach in data mining, pattern recognition, and knowledge discovery (168). This aims to objectively organize data samples into homogeneous groups, where the data samples within a group are similar (168). Classification is used for decision making by categorizing the data samples into different classes (diagnosis labels), while prediction is to provide an early detection of an outcome (e.g., response to treatment) (169). In the tinnitus literature, the classification of brain data has often been done using different ML methods. One of the commonly used method is the Support Vector Machine (SVM) which is based on a supervised learning to detect the relationship between the data samples and their class label information (163, 170, 171). SVM learns from data to assign a hyperplane in an optimal position in the data space such that the samples are best separated with respect to their classes (159). Multilayer Perceptron (MLP) is a another ML approach which is a well-known architecture of ANN with supervised

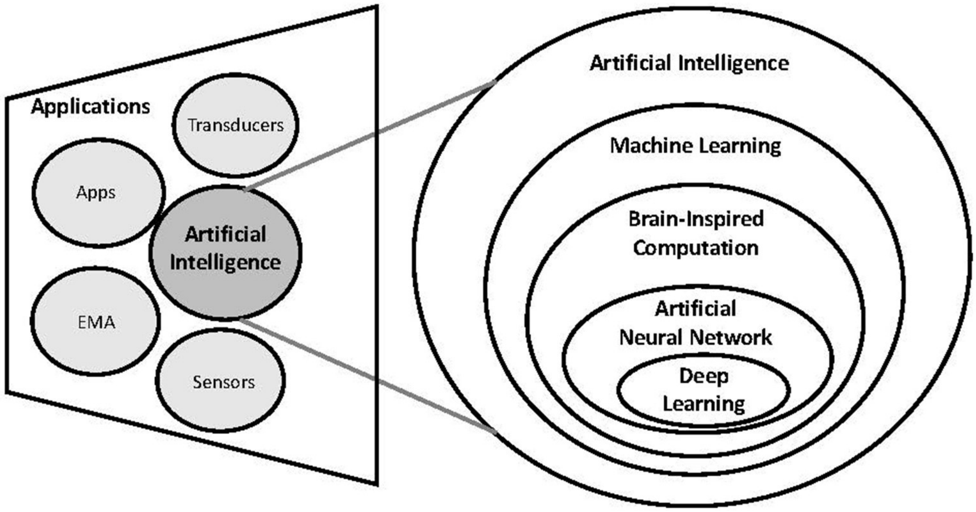


FIGURE 4 | Analytical Artificial Intelligence methods and applications in Tinnitus.

learning algorithms that can perform non-linear classification and prediction (160).

Although EEG is widely applied to tinnitus research, few ML methods have been developed to classify tinnitus patients from healthy people using EEG. Sun et al. (171) proposed a multi-view intact space learning method to distinguish EEG signals and classify the tinnitus patients from healthy people using a SVM classifier; with accuracy of 99%. Monaghan et al. (172) applied SVM techniques to classify (at the individual level) tinnitus from healthy people, based on their Auditory Brainstem Responses. Their findings showed the existence of objective features in neural activity generated by the inner ear and early auditory brain that vary between individuals with/without subjective tinnitus with quite high accuracy (80%) (172). This approach shows potential to be developed into a diagnostic tool.

In order to increase the classification accuracy, researchers have tried to use ML feature selection methods to extract the most important variables from EEG data. For example, Liu et al. (163) studied cortical/subcortical morphological neuroimaging biomarkers that may characterize idiopathic tinnitus using ML methods. They used a hybrid feature selection algorithm combining the F-score and sequential forward floating selection (SFFS). SFFS is a search algorithm that is used to reduce the dimensions of feature space to improve the computational efficiency; it removes irrelevant features or noise, without losing the informative patterns in the data. The results suggested a combination of 13 cortical/subcortical brain regions had the highest classification accuracy for effectively differentiating patients with tinnitus from healthy subjects (163). In addition to EEG data, Shoushtarian et al. (165) collected fNIRS data to differentiate tinnitus patients from control participants and to identify fNIRS features associated with tinnitus severity. The Naïve Bayes classifiers (a mathematical formula for determining probability of an outcome occurring, based on a previous outcomes) were used to classify patients with tinnitus from controls. An accuracy of 87.32% was obtained to distinguish patients with slight/ mild vs. moderate/ severe tinnitus (165). These findings show the feasibility of using fNIRS and ML to develop an objective measure of tinnitus that might enable clinicians to provide new treatment plans.

No treatment is currently able to eliminate the perception of tinnitus, but reducing its impact is possible (70). However, treatment is complicated by the large variability in tinnitus, and response to treatments, amongst sufferers (4). Research developing machine learning methods for early prediction of the effectiveness of tinnitus interventions based on the response of tinnitus individuals have been undertaken (165, 173, 174). For example, Schecklmann et al. (173), used a new cluster analysis based on the multimodal datasets including Positron-Emission Tomography (PET) and clinical variables, and extracted the most important predictor variables to improve accuracy. Their findings showed that clustering according to patients imaging data (PET data) is feasible and might provide a new approach for identifying tinnitus sub-types. Niemann et al. (174) developed a model to predict depression severity after outpatient therapy based on variables obtained before therapy among tinnitus sufferers. In this study, a decision tree classifier, which is a supervised ML model was used to split the data samples into different outcomes (tree leaves) by passing them through several decision nodes (tree nodes) and assigning them to proper branches in the tree (174). The results indicated an accuracy of 89% for detection of depression severity after treatment using data extracted from questionnaire answered before treatment. By incrementally reducing the number of features on predictive performance the set of predictive features (the number of questions) required may be minimized. Therefore, determinants of tinnitus-related distress provide valuable information about tinnitus categorization and desired therapy planning. Niemann et al. (175) also identified that gender-associated differences may facilitate a more detailed identification of symptom profiles. AI may heighten treatment response rates, and help to create access for vulnerable tinnitus populations that are potentially less visible in clinical settings (35). Niemann et al. (35) generated different regression models in the dataset and finally classified the samples with respect to various regressions.

Tinnitus patients' psychological symptom-based phenotypes comparison with tinnitus have been explored in a Gaussian mixture model (176). It was found that specific symptom profiles (e.g., anxiety) were significantly correlated with cochlear implant users' tinnitus characteristics. The Gaussian mixture model was found as a promising ML tool for identifying psychological symptom-based phenotypes.

Artificial Neural Networks in Tinnitus

To gain a mechanistic understanding of how tinnitus develops in the brain, we need to design a biologically plausible computational model that mimic both tinnitus formation and perception, then evaluate the preliminary models using brain and behavioral experiments (177). ANNs are computational models directly inspired by, and partially modeled on biological neural networks (160). They are capable of modeling and processing non-linear relationships between inputs and outputs in parallel (161). The brain is a highly interactive and deep learning network, but nearly all multivariate models employed in brain data analysis are linear and do not model interactions. Understanding the dynamic patterns of spatiotemporal brain data through traditional machine learning methods is limited because temporal features of the data manifest complex interactions that change dynamically over time. Therefore, it is crucial to develop new computational models that are capable of learning spatiotemporal interactions between multivariate data streams. Durai et al. (34) used a behavioral case series, alongside EEG and a brain-inspired artificial neural network model, to evaluate the effect of three masking sounds therapy on tinnitus and associated symptoms across 12 months. The method was able to predict sound therapy responders (93% accuracy) from non-responders (100% accuracy) using baseline EEG recordings. The authors further used ANN model to examine the effects of Acoustic Residual Inhibition on EEG function, as well as the predictive ability of the model (93%) (158). This approach may aid in the development of predictive models for treatment selection.

Despite advances in AI, relatively little is known on how best to incorporate it into health service delivery. Personalized modeling of tinnitus could enable classification/prediction of an individual patients profile (178). In contrast to global modeling (the conventional AI systems the create a computational model for the entire dataset), personalized modeling learns from the most relevant datasets for the individual (personalized subset of features and samples). It increases the model efficiency as a smaller block of data features can be selected and used for AI learning algorithms. Personalized AI models can be also used in tracking the effectiveness of a treatment over time and evaluate the treatment success at an individual level. This could, as an example, lead to application of an AI clinical decision tool to direct care toward higher probability of success treatments improving: tinnitus therapy outcomes, more efficient in in time and cost, in turn reducing burden to the wider community.

DISCUSSION

The results of this review indicate that there are many digital technologies in use for tinnitus management, with an even greater number of technologies that demonstrate the potential to address the issue of tinnitus. A potential key to unlock success in tinnitus therapy is to address the heterogeneity of tinnitus and dispense targeted therapies (7). An individual's susceptibility to, and experience of, tinnitus is not divorced from the environment that surrounds them. Therefore, addressing multiple biopsychosocial factors (**Figure 5**) may be necessary to holistically treat tinnitus. Environment factors including circadian rhythms (87) and stress appear to interact with individual sensitivities (5, 75) to modulate tinnitus. A treatment method that could account for such modifiers would seem invaluable. The resolution in understanding an individual's tinnitus experience with and without treatment may be essential for very effective treatments, due to tinnitus heterogeneity. Physiological predictors for treatment effectiveness, and potentially adjustments for environment, may lead to highly personalized real-time adjustments to the individual and their environment. Sensor technology coupled with EMA and AI offers the promise of real-time delivery of personalized therapies.

State-Of-the-Art Review Strengths and Weaknesses

This "state-of-the-art" review was undertaken to answer the question: what digital technology could be applied to tinnitus therapy in the next 10 years? State-of-the-art reviews are a subtype of narrative review that focus on current recent knowledge and highlight how research may advance this further. All review types have strengths and weaknesses (36). By focussing on the last decade this review has captured developments in the rapidly developing field of digital technology, that are, or could reasonably, be applied to tinnitus. The authors are knowledgeable in fields of behavioral science, audiology, artificial intelligence and engineering so are familiar with the topic and have been able to identify and fill gaps in the literature search by referring to missed peer reviewed publications and by use of gray literature. Gray literature includes manufacturer publications and consumer electronics publications. This literature does not provide high quality evidence, but it is current, and addresses commercial questions not commonly discussed in scientific papers. This review covers a breadth of material that a systematic review would reject as not meeting apriori quality criteria. The value of expertise from the authors must be considered in light of risk for bias:

> "*a subject expert may simply provide a particularly idiosyncratic and personal perspective on current and future priorities*" [(36), p. 102].

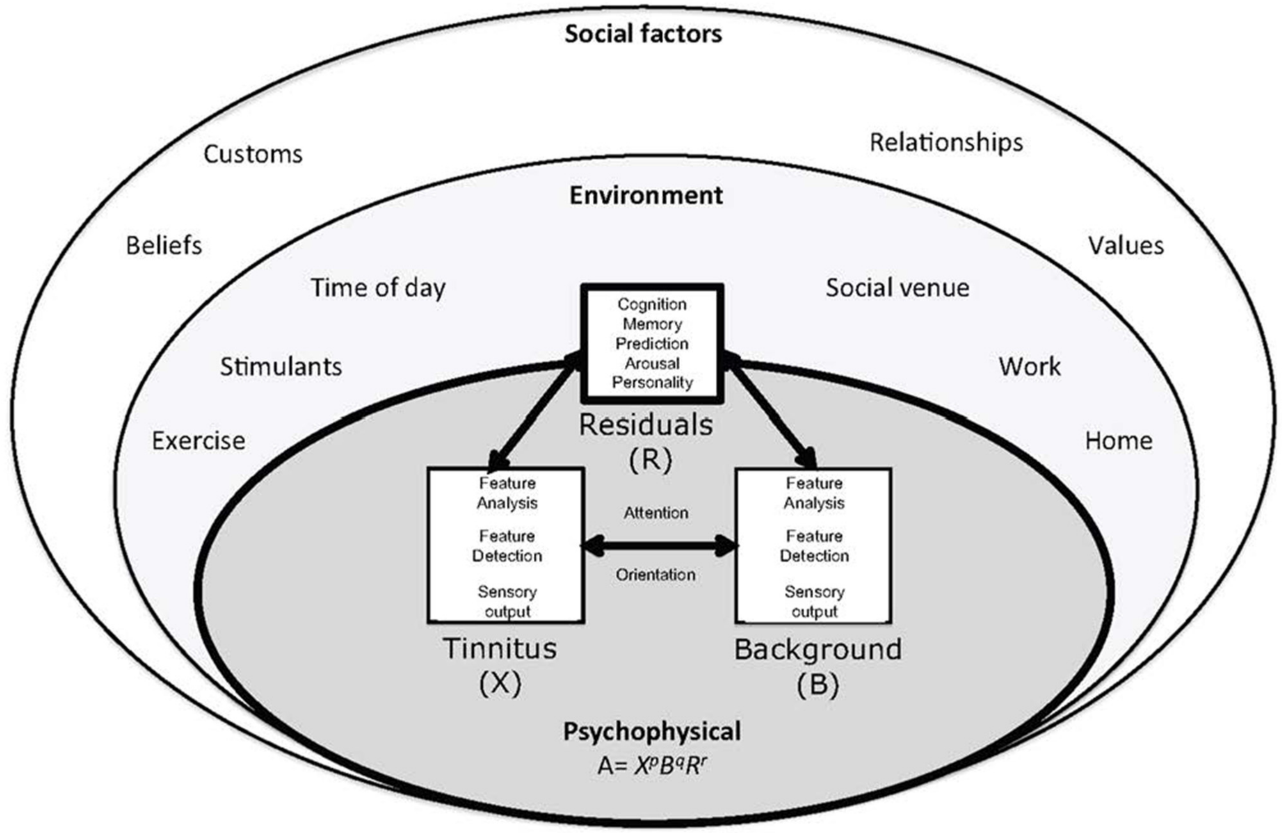


FIGURE 5 | The ecological model of tinnitus. It consists of a psychophysical core described by adaptation level theory. In adaptation level theory tinnitus and background sound perception are under influence of individual psychology factors classified as "residuals." These factors are influenced by the environment and social context. Helson (179) expressed this relationship mathematically: $A = X^pB^qR^r$. The adaptation level (A) is the weighted product of: X, the intensity of tinnitus signal, *B*, intensity of background neural activity, and *R*, intensity of residual components (e.g., memory, arousal, and personality). The weighting coefficients *p, q*, and *r* determine the relative contributions of components to adaptation level. These factors are under the influence of environmental and psychosocial factors [(75) with permission of the authors].

With this is mind further research using scoping or systematic review methods, perhaps with a narrower focus, are recommended.

Future Digital Therapeutics for Tinnitus and Research Priorities

Multifactorial treatments may be needed to address the diversity in tinnitus neurophysiology and patient goals. Recent development in smart mobile apps offers a large variety of functions that can be used for the clinical interventions and diagnosis in the chronic tinnitus. AI and machine learning tools could be used to learn from the trend of data and extract meaningful patterns for the purpose of precise prediction and classification of tinnitus, this information could inform counseling, hearing assistance and sound therapy. These concepts could be further developed toward making a smart therapy system for tinnitus, that may lend itself to use of AI decision tools and real-time treatment selection based on physiological markers. This will necessitate development of personalized AI models based on a group of individuals' data with similar characteristics. Our research group is working toward a Precision Sound Therapy™ that examines individual differences and treatment goals, and will employ AI to aid therapy selection (**Figure 6**).

The convergence of consumer and clinical devices is happening quickly in the hearing space; the release of the first Bose HA (180) the release of Jabra branded hearing aids by GN ReSound (181) and the purchase of Sennheiser by Sonova (64) being prime examples. Changes in the hearing aid space is likely to migrate to tinnitus therapy as well. The technical similarities between Hearables and HA are obvious (microphone, Bluetooth, signal processing, speaker/receiver) but the differences are still significant. Hearing access is used to promote Hearable devices, but their primary market is not, with a few exceptions, hearing impaired or tinnitus sufferers. Most Hearables are designed for entertainment and/or fitness tracking first while hearing impairment is a secondary concern, and potentially a marketing strategy (29). HAs are worn near continuously requiring low battery consumption and high comfort, need to be

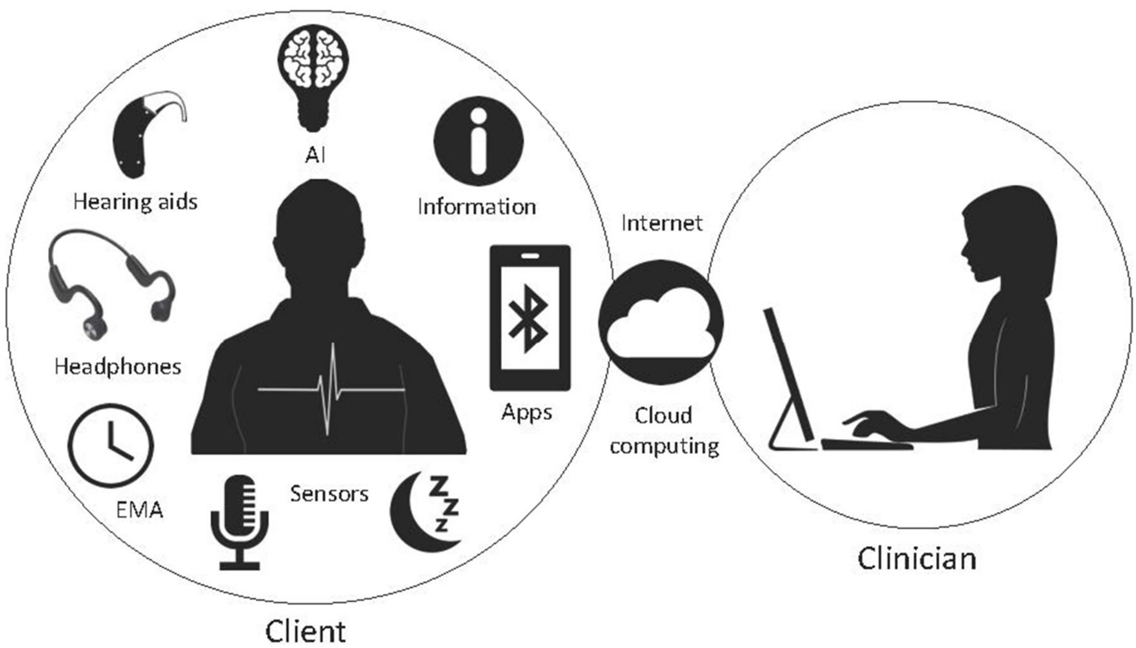


FIGURE 6 | A potential future digital therapeutic system consisting of an app-based therapy that uses AI to configure transducers, counseling, EMA, biosensors and connect to clinicians via cloud computing.

free of occlusion for voice quality and have to manage acoustic feedback, they also require different types of support than consumer electronics. Although universal devices may benefit from volume and mass production the majority of end-users do not have disabilities. This runs some risk that highly focused technology development undertaken by HA manufacturers could be compromised by the more generic solutions offered by consumer electronics companies. Importantly HAs are often accompanied by chronic medical conditions requiring clinical management, it is not clear how consumer-driven models will mitigate risk of undiagnosed pathology, especially when that consumer technology may mask the true problem, potentially delaying diagnosis. Debate as to the best model(s) for delivery of tinnitus management, self-help, self-directed and clinician led services also needs to occur for tinnitus.

This review has shown that clinicians and researchers in the tinnitus field do not lack imagination and innovation in their use of digital technology. However, many ideas appear to have been translated into commercial products before concepts are proven. To advance the field and develop effective digital therapeutics we suggest 6 key priorities for tinnitus technology research.

1. Tinnitus researchers should explore new and emerging technologies through appropriate proof-of-concept trials before the expense of randomized clinical trials and the lure of commercialization is entertained. Innovation is important, but not at the expense of evidence.
2. Physiological measures of tinnitus (or in their absence known associated measures) need to be included in trials as often as possible alongside behavioral measures so as to develop a reliable compendium of biomarkers for tinnitus therapy.
3. Wearable biosensors need to be applied with EMA to establish real-time patterns in tinnitus related physiology. The meaning and value of such measures need to be ascertained.
4. AI methods to adjust therapies to physiological-EMA measures need to be developed and tested to ascertain whether personalized tinnitus therapies can benefit from modifying response according to the patient's physiology and environment in real-time.
5. New health-delivery models should be developed with end-user communities. Data driven approaches need to ensure data privacy. Patient concerns regarding data use and data sovereignty need to be studied across cultures.
6. The role of the clinician in providing tinnitus digital therapies needs to researched from an efficacy, cost and consumer perspective. The CoVID-19 pandemic has illustrated the value of remote care and access to services outside clinic walls. The value proposition of new technology relative to established patterns of clinical care should be explored.

CONCLUSIONS

The bourgeoning industry for digital tinnitus services is an exciting area but current opinion is that it should be used as an adjunct to, rather than a replacement of, clinical care. The uncertain mechanisms underpinning tinnitus present a challenge and many posited therapeutic approaches may not be successful. Some current therapies appear to be driven by technology innovation capability and theory rather than evidence. Due to the heterogeneity of tinnitus, response to various treatments differs between individuals. Holistic programs that offer multiple therapeutic facets such as sound therapy, multisensory stimulation, information, guided meditation, and counseling may address this heterogeneity. Personalized AI modeling based on biometric measures obtained through various sensor types, and assessments of individual psychology and lifestyles should result in the development of smart therapy platforms for tinnitus.

AUTHOR CONTRIBUTIONS

PS undertook the initial database search and initial consideration of title relevance to the study, he was primarily responsible for sections related to EMA and apps. ZD and MD reviewed articles related to AI. KS provided the review of stress and tinnitus. AB was responsible for the AR/VR sections. RB reviewed the technical accuracy of the manuscript. GS was primarily responsible for the introduction, reviews of hearing aids, hearables, cochlear implants, internet-based therapies, perceptual training, dedicated sound therapy devices, multimodal therapies, sensors, and the discussion, had overall oversight, and editorial responsibility for the review. All authors contributed to the synthesis of concepts and editing of the manuscript.

FUNDING

GS research was currently funded by the Auckland Medical Research Foundation and Return on Science.

REFERENCES

1. Searchfield GD, Zhang J, Biswas R, De Ridder D, Deutsch B, Hall D, et al. Emerging topics in the behavioral neuroscience of tinnitus. *Curr Top Behav Neurosci*. (2021) doi: 10.1007/7854_2020_217. [Epub ahead of print].
2. Park HM, Jung J, Kim JK, Lee YJ. Tinnitus and its association with mental health and health-related quality of life in an older population: a nationwide cross-sectional study. *J Appl Gerontol*. (2020) 733464820966512. doi: 10.1177/0733464820966512
3. Wu B, Searchfield G, Exeter D, Lee A. Tinnitus prevalence in New Zealand. *N Z Med J*. (2015) 128:24–34.
4. Cederroth CR, Gallus S, Hall DA, Kleinjung T, Langguth B, Maruotti A, et al. Towards an understanding of tinnitus heterogeneity. *Front Aging Neurosci*. (2019) 11:53. doi: 10.3389/fnagi.2019.00053
5. Kleinstäuber M, Weise C. Psychosocial variables that predict chronic and disabling tinnitus: a systematic review. *Curr Top Behav Neurosci*. (2021). doi: 10.1007/7854_2020_213. [Epub ahead of print].
6. Searchfield GD, Morrison-Low J, Wise K. Object identification and attention training for treating tinnitus. *Prog Brain Res*. (2007) 166:441–60. doi: 10.1016/S0079-6123(07)66043-9
7. Searchfield GD, Durai M, Linford T. A state-of-the-art review: personalization of tinnitus sound therapy. *Front Psychol*. (2017) 8:1599. doi: 10.3389/fpsyg.2017.01599
8. Stockdale D, McFerran D, Brazier P, Pritchard C, Kay T, Dowrick C, et al. An economic evaluation of the healthcare cost of tinnitus management in the UK. *BMC Health Serv Res*. (2017) 17:577. doi: 10.1186/s12913-017-2527-2
9. Searchfield GD, Kaur M, Martin WH. Hearing aids as an adjunct to counseling: tinnitus patients who choose amplification do better than those that don't. *Int J Audiol*. (2010) 49:574–9. doi: 10.3109/14992021003777267
10. Brennan-Jones CG, Thomas A, Hoare DJ, Sereda M. Cochrane corner: sound therapy (using amplification devices and/or sound generators) for tinnitus. *Int J Audiol*. (2020) 59:161–5. doi: 10.1080/14992027.2019.1643503
11. Mitchell P, Moffat D, Fallside F. Computer-aided tinnitus characterization. *Clin Otolaryngol All Sci*. (1984) 9:35–42. doi: 10.1111/j.1365-2273.1984.tb01469.x
12. Knebel SBH, Bentler RA. Comparison of two digital hearing aids. *Ear Hear*. (1998) 19:280–9. doi: 10.1097/00003446-199808000-00003
13. Trotter M, Donaldson I. Hearing aids and tinnitus therapy: a 25-year experience. *J Laryngol Otol*. (2008) 122:1052–6. doi: 10.1017/S002221510800203X
14. Al-Jassim A. The use of the walkman mini-stereo system in the management of tinnitus. *J Laryngol Otol*. (1987) 101:663–5. doi: 10.1017/S0022215100102488
15. Durai M, Searchfield GD. A mixed-methods trial of broad band noise and nature sounds for tinnitus therapy: group and individual responses modeled under the adaptation level theory of tinnitus. *Front Aging Neurosci*. (2017) 9:44. doi: 10.3389/fnagi.2017.00116
16. Davis P, Wilde R, Steed L. Habituation-Based rehabilitation technique using acoustic desensitisation protocol. In: *Proceedings of the Seventh International Tinnitus Seminar*. (2002). p. 188–90.
17. Davis PB, Paki B, Hanley PJ. Neuromonics tinnitus treatment: third clinical trial. *Ear Hear*. (2007) 28:242–59. doi: 10.1097/AUD.0b013e3180312619
18. Sereda M, Davies J, Hall DA. Pre-market version of a commercially available hearing instrument with a tinnitus sound generator: feasibility of evaluation in a clinical trial. *Int J Audiol*. (2017) 56:286–94. doi: 10.1080/14992027.2016.1254822
19. Simonetti P, Vasconcelos LG, Oiticica J. Effect of fractal tones on the improvement of tinnitus handicap inventory functional scores among chronic tinnitus patients: an open-label pilot study. *Int Arch Otorhinolaryngol*. (2018) 22:387–94. doi: 10.1055/s-0038-1635575
20. Weaver J. Made-for-iPhone hearing aid has broad appeal, early adopters report. *Hear J*. (2014) 67:28–30. doi: 10.1097/01.HJ.0000449902.49313.f1
21. Kerckhoff J, Listenberger J, Valente M. Advances in hearing aid technology. *Contemp Issues Commun Sci Disord*. (2008) 35:102–12. doi: 10.1044/cicsd_35_F_102
22. Hearing R. *Sonova Announces New SWORD "Made for All"*. (2017). Available online at: https://www.hearingreview.com/hearing-products/sonova-announces-new-sword-made-platform-phonak-audeo-b-direct-2
23. Okamoto H, Stracke H, Stoll W, Pantev C. Listening to tailor-made notched music reduces tinnitus loudness and tinnitus-related auditory cortex activity. *Proc Natl Acad Sci USA*. (2010) 107:1207–10. doi: 10.1073/pnas.0911268107
24. Stein A, Wunderlich R, Lau P, Engell A, Wollbrink A, Shaykevich A, et al. Clinical trial on tonal tinnitus with tailor-made notched music training. *BMC Neurol*. (2016) 16:38. doi: 10.1186/s12883-016-0558-7
25. Wegger M, Ovesen T, Larsen DG. Acoustic coordinated reset neuromodulation: a systematic review of a novel therapy for tinnitus. *Front Neurol*. (2017) 8:36. doi: 10.3389/fneur.2017.00036
26. Theodoroff SM, McMillan GP, Zaugg TL, Cheslock M, Roberts C, Henry JA. Randomized controlled trial of a novel device for tinnitus sound therapy during sleep. *Am J Audiol*. (2017) 26:543–54. doi: 10.1044/2017_AJA-17-0022
27. Hesse G. Evidence and evidence gaps in tinnitus therapy. *GMS Curr Top Otorhinolaryngol Head Neck Surg*. (2016) 15:Doc04. doi: 10.3205/cto000131
28. Reddy S, Fox J, Purohit MP. Artificial intelligence-enabled healthcare delivery. *J R Soc Med*. (2019) 112:22–8. doi: 10.1177/0141076818815510
29. Hunn N. *The Market for Hearable Devices 2016–2020*. WiFore Wireless Consulting. (2016). Available online at: http://bit.ly/hearables2020
30. Crum P. Hearables: here come the: technology tucked inside your ears will augment your daily life. *IEEE Spectrum*. (2019) 56:38–43. doi: 10.1109/MSPEC.2019.8701198
31. Tutton R. Personalizing medicine: futures present and past. *Soc Sci Med*. (2012) 75:1721–8. doi: 10.1016/j.socscimed.2012.07.031

32. Tzounopoulos T, Balaban C, Zitelli L, Palmer C. Towards a mechanistic-driven precision medicine approach for tinnitus. *J Assoc Res Otolaryngol.* (2019) 20:115–31. doi: 10.1007/s10162-018-00709-9
33. Schlee W, Hall DA, Canlon B, Cima RF, de Kleine E, Hauck F, et al. Innovations in doctoral training and research on tinnitus: the European school on interdisciplinary tinnitus research (ESIT) perspective. *Front Aging Neurosci.* (2018) 9:447. doi: 10.3389/fnagi.2017.00447
34. Durai M, Sanders P, Doborjeh Z, Doborjeh M, Wendt A, Kasabov N, et al. Prediction of tinnitus masking benefit within a case series using a spiking neural network model. *Prog Brain Res.* (2021) 260:129–65. doi: 10.1016/bs.pbr.2020.08.003
35. Niemann U, Boecking B, Brueggemann P, Mebus W, Mazurek B, Spiliopoulou M. Tinnitus-related distress after multimodal treatment can be characterized using a key subset of baseline variables. *PLoS ONE.* (2020) 15:e0228037. doi: 10.1371/journal.pone.0228037
36. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Inform Libr J.* (2009) 26:91–108. doi: 10.1111/j.1471-1842.2009.00848.x
37. Saltzman M, Ersner MS. A hearing aid for the relief of tinnitus aurium. *Laryngoscope.* (1947) 57:358–66. doi: 10.1288/00005537-194705000-00005
38. Searchfield GD. Sense and sensibility: a review of the behavioral neuroscience of tinnitus sound therapy and a new typology. *Curr Top Behav Neurosci.* (2020). doi: 10.1007/7854_2020_183. [Epub ahead of print].
39. Gilmore JN. Design for everyone: apple airpods and the mediation of accessibility. *Crit Stud Media Commun.* (2019) 36:482–94. doi: 10.1080/15295036.2019.1658885
40. Barozzi S, Del Bo L, Crocetti A, Dyrlund O, Passoni S, Zolin A, et al. A comparison of nature and technical sounds for tinnitus therapy. *Acta Acust United Acust.* (2016) 102:540–6. doi: 10.3813/AAA.918971
41. Sweetow RW, Sabes JH. Effects of acoustical stimuli delivered through hearing aids on tinnitus. *J Am Acad Audiol.* (2010) 21:461–73. doi: 10.3766/jaaa.21.7.5
42. Tyler RS, Deshpande AK, Lau CC, Kuk F. The effectiveness of the progression of widex zen tinnitus therapy: a pilot study. *Am J Audiol.* (2017) 26:283–92. doi: 10.1044/2017_AJA-16-0009
43. Shekhawat GS, Searchfield GD, Stinear CM. Role of hearing aids in tinnitus intervention: a scoping review. *J Am Acad Audiol.* (2013) 24:747–62. doi: 10.3766/jaaa.24.8.11
44. Hoare DJ, Edmondson-Jones M, Sereda M, Akeroyd MA, Hall D. Amplification with hearing aids for patients with tinnitus and co-existing hearing loss. *Cochrane Database Syst Rev.* (2014) 31:CD010151. doi: 10.1002/14651858.CD010151.pub2
45. Parazzini M, Del Bo L, Jastreboff M, Tognola G, Ravazzani P. Open ear hearing aids in tinnitus therapy: an efficacy comparison with sound generators. *Int J Audiol.* (2011) 50:548–53. doi: 10.3109/14992027.2011.572263
46. Sereda M, Xia J, El Refaie A, Hall D, Hoare D. Sound therapy (using amplification devices and/or sound generators) for tinnitus. *Cochrane Database Syst Rev.* (2018) 12:CD013094. doi: 10.1002/14651858.CD013094.pub2
47. Hauptmann C, Williams M, Vinciati F, Haller M. Technical feasibility of acoustic coordinated reset therapy for tinnitus delivered via hearing aids: a case study. *Case Rep Otolaryngol.* (2017) 2017:5304242. doi: 10.1155/2017/5304242
48. Haab L, Lehser C, Corona-Strauss FI, Bernarding C, Seidler H, Hannemann R, et al. Implementation and long-term evaluation of a hearing aid supported tinnitus treatment using notched environmental sounds. *IEEE J. Transl. Eng. Health Med.* (2019) 7:1600109. doi: 10.1109/JTEHM.2019.2897570
49. Marcrum SC, Picou EM, Steffens T, Hannemann R, Vielsmeier V, Schecklmann M, et al. Conventional versus notch filter amplification for the treatment of tinnitus in adults with mild-to-moderate hearing loss. *Prog Brain Res.* (2020) 260:235–52. doi: 10.1016/bs.pbr.2020.06.020
50. Popelka GR, Moore BCJ. Future directions for hearing aid development. In: Popelka G, Moore B, Fay R, Popper A, editors. *Hearing Aids. Springer Handbook of Auditory Research*, Vol. 56. Cham: Springer (2016). doi: 10.1007/978-3-319-33036-5_11
51. Groth J, Ruggles D, Ellison J. Sizing up hearing aids in the 21st century: is there still room for improvement. *Retriev. Jan.* (2021) 6:2020.
52. Burwinkel JR, Xu B, Crukley J. Preliminary examination of the accuracy of a fall detection device embedded into hearing instruments. *J Am Acad Audiol.* (2020) 31:393–430. doi: 10.3766/jaaa.19056
53. Borgström BJ, Brandstein MS, Ciccarelli GA, Quatieri TF, Smalt CJ. Speaker separation in realistic noise environments with applications to a cognitively-controlled hearing aid. *Neural Netw.* (2021) 140:136–47. doi: 10.1016/j.neunet.2021.02.020
54. Fiedler L, Obleser J, Lunner T, Graversen C. Ear-EEG allows extraction of neural responses in challenging listening scenarios—a future technology for hearing aids? In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* Orlando, FL: IEEE (2016). p. 5697–700. doi: 10.1109/EMBC.2016.7592020
55. Boisvert I, Reis M, Au A, Cowan R, Dowell RC. Cochlear implantation outcomes in adults: a scoping review. *PLoS ONE.* (2020) 15:e0232421. doi: 10.1371/journal.pone.0232421
56. Brackmann D. Reduction of tinnitus in cochlear-implant patients. *J Laryngol Otol*. (1981) 163–5.
57. Thedinger B, House W, Edgerton B. Cochlear implant for tinnitus. *Ann Otol Rhinol Laryngol.* (1985) 94:10–3. doi: 10.1177/000348948509400102
58. Arts RA, George EL, Stokroos RJ, Vermeire K. cochlear implants as a treatment of tinnitus in single-sided deafness. *Curr Opin Otolaryngol Head Neck Surg.* (2012) 20:398–403. doi: 10.1097/MOO.0b013e3283577b66
59. Tyler RS, Owen RL, Bridges J, Gander PE, Perreau A, Mancini PC. Tinnitus suppression in cochlear implant patients using a sound therapy app. *Am J Audiol.* (2018) 27:316–23. doi: 10.1044/2018_AJA-17-0105
60. Ramakers GG, van Zon A, Stegeman I, Grolman W. The effect of cochlear implantation on tinnitus in patients with bilateral hearing loss: A systematic review. *Laryngoscope.* (2015) 125:2584–92. doi: 10.1002/lary.25370
61. Painter L. *Bragi Dash Pro Review*. Tech Advisor (2018). Available online at: https://www.techadvisor.com/review/bragi-dash-pro-3683343/
62. Pierce D. *Inside the Downfall of Doppler Labs.* Wired (2017). Available online at: https://www.wired.com/story/inside-the-downfall-of-doppler-labs/
63. Hollister S. *Bragi Confirms It's Pivoting Out of the Wireless Earbuds Business.* The Verge (2019). Available online at: https://www.theverge.com/2019/4/2/18292401/bragi-dash-pro-wireless-earbuds-hardware-software-pivot-exit
64. Prosoundnews. *Sonova Acquires Sennheiser's Consumer Division.* (2021). Available online at: https://www.radioworld.com/news-and-business/business-and-law/sonova-acquires-sennheisers-consumer-division
65. Krasnoff B. *Bose Sleepbuds II Review: Get Your Eight Hours.* The Verge (2020). Available online at: https://www.theverge.com/22176950/bose-sleepbuds-ii-earbuds-review
66. Taylor B. Bragi announces tinnitus relief feature in new earbuds. *Hear Health Technol Matter*. (2018). Available online at: https://hearinghealthmatters.org/hearingnewswatch/2018/bragi-tinnitus-relief-feature-new-hearable-earbuds/
67. Dacus M, Rawool V. *Audibility, cost-acceptibility and cosmetic appearance of IQ buds BOOST in adults over the age of 50* (Honors thesis). Communication Sciences and Disorders, University of Missisipi (2020). p. 1388. Available online at: https://egrove.olemiss.edu/hon_thesis/1388
68. Greenwell K, Featherstone D, Hoare DJ. The application of intervention coding methodology to describe the tinnitus e-programme, an internet-delivered self-help intervention for tinnitus. *Am J Audiol.* (2015) 24:311–5. doi: 10.1044/2015_AJA-14-0089
69. Mehta S, Peynenburg VA, Hadjistavropoulos HD. Internet-delivered cognitive behaviour therapy for chronic health conditions: a systematic review and meta-analysis. *J Behav Med.* (2019) 42:169–87. doi: 10.1007/s10865-018-9984-x
70. Zenner, H.-P., Delb W, Kröner-Herwig B, Jäger B, Peroz I, et al. A multidisciplinary systematic review of the treatment for chronic idiopathic tinnitus. *Euro Arch Oto Rhino Laryngol.* (2017) 274:2079–91. doi: 10.1007/s00405-016-4401-y
71. Schlee W, Kraft R, Schobel J, Langguth B, Probst T, Neff P, et al. Momentary assessment of tinnitus-how smart mobile applications advance our understanding of tinnitus. In: Baumeister H, Montag C, editors. *Digital Phenotyping and Mobile Sensing. Studies in*

Neuroscience, Psychology and Behavioral Economics. Cham: Springer (2019). doi: 10.1007/978-3-030-31620-4_13

72. Deutsch B, Piccirillo J. Momentary analysis of tinnitus: considering the patient. *Curr Top Behav Neurosci*. (2020). doi: 10.1007/7854_2020_176. [Epub ahead of print].
73. Schickler M, Pryss R, Reichert M, Heinzelmann M, Schobel J, Langguth B, et al. Using wearables in the context of chronic disorders: results of a pre-study. In: *2016 IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE (2016). p. 68–9. doi: 10.1109/CBMS.2016.40
74. Landgrebe M, Azevedo A, Baguley D, Bauer C, Cacace A, Coelho C, et al. Methodological aspects of clinical trials in tinnitus: a proposal for an international standard. *J Psychosom Res*. (2012) 73:112–21. doi: 10.1016/j.jpsychores.2012.05.002
75. Searchfield GD. Tinnitus what and where: an ecological framework. *Front Neurol*. (2014) 5:271. doi: 10.3389/fneur.2014.00271
76. Gromatsky M, Sullivan SR, Spears AP, Mitchell E, Walsh S, Kimbrel NA, et al. Ecological momentary assessment (EMA) of mental health outcomes in veterans and servicemembers: a scoping review. *Psychiatry Res*. (2020) 292:113359. doi: 10.1016/j.psychres.2020.113359
77. Henry JA, Galvez G, Turbin MB, Thielman EJ, McMillan GP, Istvan JA. Pilot study to evaluate ecological momentary assessment of tinnitus. *Ear Hear*. (2012) 32:179. doi: 10.1097/AUD.0b013e31822f6740
78. Wilson MB, Kallogjeri D, Joplin CN, Gorman MD, Krings JG, Lenze EJ, et al. Ecological momentary assessment of tinnitus using smartphone technology: a pilot study. *Otolaryngol Head Neck Surg*. (2015) 152:897–903. doi: 10.1177/0194599815569692
79. Goldberg RL, Piccirillo ML, Nicklaus J, Skillington A, Lenze E, Rodebaugh TL, et al. Evaluation of ecological momentary assessment for tinnitus severity. *JAMA Otolaryngol Head Neck Surg*. (2017) 143:700–6. doi: 10.1001/jamaoto.2017.0020
80. Pryss R, Probst T, Schlee W, Schobel J, Langguth B, Neff P, et al. Prospective crowdsensing versus retrospective ratings of tinnitus variability and tinnitus–stress associations based on the TrackYourTinnitus mobile platform. *Int J Data Sci Analyt*. (2019) 8:327–38. doi: 10.1007/s41060-018-0111-4
81. Schleicher M, Unnikrishnan V, Neff P, Simoes J, Probst T, Pryss R, et al. Understanding adherence to the recording of ecological momentary assessments in the example of tinnitus monitoring. *Sci Rep*. (2020) 10:22459. doi: 10.1038/s41598-020-79527-0
82. Wise K, Kobayashi K, Magnusson J, Welch D, Searchfield GD. Randomized controlled trial of a perceptual training game for tinnitus therapy. *Games Health J*. (2016) 5:141–9. doi: 10.1089/g4h.2015.0068
83. Mehdi M, Riha C, Neff P, Dode A, Pryss R, Schlee W, et al. Smartphone apps in the context of tinnitus: systematic review. *Sensors*. (2020) 20:1725. doi: 10.3390/s20061725
84. Pryss R, Reichert M, Langguth B, Schlee W. Mobile crowd sensing services for tinnitus assessment, therapy, and research. In: *2015 IEEE International Conference on Mobile Services*. IEEE (2015). p. 352–9. doi: 10.1109/MobServ.2015.55
85. Probst T, Pryss RC, Langguth B, Rauschecker JP, Schobel J, Reichert M, et al. Does tinnitus depend on time-of-day? An ecological momentary assessment study with the "TrackYourTinnitus" application. *Front Aging Neurosci*. (2017) 9:253. doi: 10.3389/fnagi.2017.00253
86. Muniandi LP, Schlee W, Pryss R, Reichert M, Schobel J, Kraft R, et al. Finding tinnitus patients with similar evolution of their ecological momentary assessments. In: *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE (2018). p. 112–7. doi: 10.1109/CBMS.2018.00027
87. Zacharia T, James J, Prakash H, Mohan RT, Rajashekhar B. The effect of circadian rhythm on the perceived tinnitus severity: a preliminary study. *Int Tinnitus J*. (2014) 19:41–5. doi: 10.5935/0946-5448.20140006
88. Wise K, Kobayashi K, Searchfield G. Feasibility study of a game integrating assessment and therapy of tinnitus. *J Neurosci Methods*. (2015) 249:1–7. doi: 10.1016/j.jneumeth.2015.04.002
89. Spiegel DP, Linford T, Thompson B, Petoe MA, Kobayashi K, Stinear CM, et al. Multisensory attention training for treatment of tinnitus. *Sci Rep*. (2015) 5:10802. doi: 10.1038/srep10802
90. Searchfield G, Spiegel D, Poppe T, Durai M, Jensen M, Kobayashi K, et al. A proof-of-concept study comparing tinnitus and neural connectivity changes following multisensory perceptual training with and without a low-dose of fluoxetine. *Int J Neurosci*. (2020) 31:433–44. doi: 10.1080/00207454.2020.1746310
91. Adjamian P. The application of electro-and magneto-encephalography in tinnitus research–methods and interpretations. *Front Neurol*. (2014) 5:228. doi: 10.3389/fneur.2014.00228
92. Guzik P, Malik M. ECG by mobile technologies. *J Electrocardiol*. (2016) 49:894–901. doi: 10.1016/j.jelectrocard.2016.07.030
93. Mohagheghian F, Makkiabadi B, Jalilvand H, Khajehpoor H, Samadzadehaghdam N, Eqlimi E, et al. Computer-aided tinnitus detection based on brain network analysis of EEG functional connectivity. *J Biomed Phys Eng*. (2019) 9:687–98. doi: 10.31661/JBPE.V0I0.937
94. Elarbed A, Fackrell K, Baguley DM, Hoare DJ. Tinnitus and stress in adults: a scoping review. *Int J Audiol*. (2020) 60:171–82. doi: 10.1080/14992027.2020.1827306
95. Hermens H, op den Akker H, Tabak M, Wijsman J, Vollenbroek M. Personalized coaching systems to support healthy behavior in people with chronic conditions. *J Electromyogr Kinesiol*. (2014) 24:815–26. doi: 10.1016/j.jelekin.2014.10.003
96. Athavipach C, Pan-Ngum S, Israsena P. A wearable in-ear EEG device for emotion monitoring. *Sensors*. (2019) 19:4014. doi: 10.3390/s19184014
97. Casson AJ. Wearable EEG and beyond. *Biomed Eng Lett*. (2019) 9:53–71. doi: 10.1007/s13534-018-00093-6
98. De Zambotti M, Cellini N, Goldstone A, Colrain IM, Baker FC. Wearable sleep technology in clinical and research settings. *Med Sci Sports Exerc*. (2019) 51:1538–57. doi: 10.1249/MSS.0000000000001947
99. Gao W, Ota H, Kiriya D, Takei K, Javey A. Flexible electronics toward wearable sensing. *Acc Chem Res*. (2019) 52:523–33. doi: 10.1021/acs.accounts.8b00500
100. Johnson KT, Picard RW. Advancing neuroscience through wearable devices. *Neuron*. (2020) 108:8–12. doi: 10.1016/j.neuron.2020.09.030
101. Ling Y, An T, Yap LW, Zhu B, Gong S, Cheng W. Disruptive, soft, wearable sensors. *Adv Mater*. (2020) 32:1904664. doi: 10.1002/adma.201904664
102. Park S, Han CH, Im CH. Design of wearable eeg devices specialized for passive brain–computer interface applications. *Sensors*. (2020) 20:4572. doi: 10.3390/s20164572
103. Zamkah A, Hui T, Andrews S, Dey N, Shi F, Sherratt RS. Identification of suitable biomarkers for stress and emotion detection for future personal affective wearable sensors. *Biosensors*. (2020) 10:40. doi: 10.3390/bios10040040
104. Valentin O, Viallet G, Delnavaz A, Cretot-Richert G, Ducharme M, Monsarat-Chanon H, et al. Custom-fitted in-and around-the-ear sensors for unobtrusive and on-the-go EEG acquisitions: development and validation. *Sensors*. (2021) 21:2953. doi: 10.3390/s21092953
105. Hébert S, Lupien SJ. The sound of stress: blunted cortisol reactivity to psychosocial stress in tinnitus sufferers. *Neurosci Lett*. (2007) 411:138–42. doi: 10.1016/j.neulet.2006.10.028
106. Elgoyhen AB, Langguth B, De Ridder D, Vanneste S. Tinnitus: perspectives from human neuroimaging. *Nat Rev Neurosci*. (2015) 16:632–42. doi: 10.1038/nrn4003
107. Sun K. *Investigating the Use of Heart Rate Entertainment for Reducing Tinnitus Using Biosensors*. (Master of Audiology), Auckland: The University of Auckland (2021).
108. Looney D, Kidmose P, Mandic DP. Ear-EEG: user-centered and wearable BCI. In: Guger C, Allison B, Leuthardt E, editors. *Brain-Computer Interface Research. Biosystems & Biorobotics*, Vol. 6. Berlin; Heidelberg: Springer (2014). doi: 10.1007/978-3-642-54707-2_5
109. Kidmose P, Looney D, Ungstrup M, Rank ML, Mandic DP. A study of evoked potentials from ear-EEG. *IEEE Trans Biomed Eng*. (2013) 60:2824–30. doi: 10.1109/TBME.2013.2264956
110. Blackwell T, Redline S, Ancoli-Israel S, Schneider JL, Surovec S, Johnson NL, et al. Comparison of sleep parameters from actigraphy and polysomnography in older women: the SOF study. *Sleep*. (2008) 31:283–91. doi: 10.1093/sleep/31.2.283

111. Spittal J. *Restarting the Sleeping Brain With Targeted Tinnitus Memory Reactivation.* (Master of Audiology), Auckland: The University of Auckland (2019).
112. Kraft R, Birk F, Reichert M, Deshpande A, Schlee W, Langguth B, et al. Design and implementation of a scalable crowdsensing platform for geospatial data of tinnitus patients. In: *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS).* IEEE (2019). p. 294–9. doi: 10.1109/CBMS.2019.00068
113. Sereda M, Smith S, Newton K, Stockdale D. Mobile apps for management of tinnitus: users' survey, quality assessment, and content analysis. *JMIR mHealth uHealth.* (2019) 7:e10353. doi: 10.2196/10353
114. Casale M, Costantino A, Rinaldi V, Forte A, Grimaldi M, Sabatino L, et al. Mobile applications in otolaryngology for patients: an update. *Laryngosc Invest Otolaryngol.* (2018) 3:434–8. doi: 10.1002/lio2.201
115. Kalle S, Schlee W, Pryss RC, Probst T, Reichert M, Langguth B, et al. Review of smart services for tinnitus self-help, diagnostics and treatments. *Front Neurosci.* (2018) 12:541. doi: 10.3389/fnins.2018.00541
116. Deshpande AK, Shimunova T. A comprehensive evaluation of tinnitus apps. *Am J Audiol.* (2019) 28:605–16. doi: 10.1044/2019_AJA-18-0135
117. Mehdi M, Dode A, Pryss R, Schlee W, Reichert M, Hauck FJ. Contemporary review of smartphone apps for tinnitus management and treatment. *Brain Sci.* (2020) 10:867. doi: 10.3390/brainsci10110867
118. Nagaraj MK, Prabhu P. Internet/smartphone-based applications for the treatment of tinnitus: a systematic review. *Euro Arch Oto Rhino Laryngol.* (2020) 277:649–57. doi: 10.1007/s00405-019-05743-8
119. Henry JA, Thielman E, Zaugg T, Kaelin C, Choma C, Chang B, et al. Development and field testing of a smartphone "app" for tinnitus management. *Int J Audiol.* (2017) 56:784–92. doi: 10.1080/14992027.2017.1338762
120. Hauptmann C, Wegener A, Poppe H, Williams M, Popelka G, Tass PA. Validation of a mobile device for acoustic coordinated reset neuromodulation tinnitus therapy. *J Am Acad Audiol.* (2016) 27:720–31. doi: 10.3766/jaaa.15082
121. Rizzi L, Bertolasi N, Ghelfi G. Sound masking on ios devices. masking everyday noises and tinnitus. In: *Audio Engineering Society Convention 144.* Milan: Audio Engineering Society (2018).
122. Probst T, Pryss RC, Langguth B, Spiliopoulou M, Landgrebe M, Vesala M, et al. Outpatient tinnitus clinic, self-help web platform, or mobile application to recruit tinnitus study samples? *Front Aging Neurosci.* (2017) 9:113. doi: 10.3389/fnagi.2017.00113
123. Melville KM, Casey LM, Kavanagh DJ. Dropout from internet-based treatment for psychological disorders. *Br J Clin Psychol.* (2010) 49:455–71. doi: 10.1348/014466509X472138
124. Smoktunowicz E, Barak A, Andersson G, Banos RM, Berger T, Botella C, et al. Consensus statement on the problem of terminology in psychological interventions using the internet or digital components. *Internet Intervent.* (2020) 21:100331. doi: 10.1016/j.invent.2020.100331
125. Hoare DJ, Stacey PC, Hall DA. The efficacy of auditory perceptual training for tinnitus: a systematic review. *Ann Behav Med.* (2010) 40:313–24. doi: 10.1007/s12160-010-9213-5
126. Hemanth N, Ghosh PV. Effect of frequency discrimination training on tinnitus in individuals with flat sensorineural hearing loss. *Ind J Otolaryngol Head Neck Surg.* (2021) 1–8. doi: 10.1007/s12070-021-02433-y
127. Jepsen K, Sanders M, Searchfield G, Kobayashi K. Perceptual training of tinnitus. In: *Proceedings of 'Tinnitus Discovery': Asia-Pacific Tinnitus Symposium 11-12th September 2009.* Auckland: The New Zealand Medical Journal (2010).
128. Hoare DJ, Van Labeke N, McCormack A, Sereda M, Smith S, Al Taher H, et al. Gameplay as a source of intrinsic motivation in a randomized controlled trial of auditory training for tinnitus. *PLoS ONE.* (2014) 9:e107430. doi: 10.1371/journal.pone.0107430
129. Eichenberg C, Schott M. Serious games for psychotherapy: a systematic review. *Games Health J.* (2017) 6:127–35. doi: 10.1089/g4h.2016.0068
130. Lau HM, Smit JH, Fleming TM, Riper H. Serious games for mental health: are they accessible, feasible, and effective? A systematic review and meta-analysis. *Front Psychiatry.* (2017) 7:209. doi: 10.3389/fpsyt.2016.00209
131. Schickler M, Pryss R, Reichert M, Schobel J, Langguth B, Schlee W. Using mobile serious games in the context of chronic disorders: a mobile game concept for the treatment of tinnitus. In: *2016 IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS).* IEEE (2016). p. 343–8. doi: 10.1109/CBMS.2016.9
132. Tavora Vieira D, Eikelboom R, Ivey G, Miller S. A multi-centre study on the long-term benefits of tinnitus management using neuromonics tinnitus treatment. *Int Tinnitus J.* (2011) 16:111–7. doi: 10.1007/978-1-60761-145-5_75
133. Newman CW, Sandridge SA. A comparison of benefit and economic value between two sound therapy tinnitus management options. *J Am Acad Audiol.* (2012) 23:126–38. doi: 10.3766/jaaa.23.2.7
134. Adamchic I, Hauptmann C, Barnikol UB, Pawelczyk N, Popovych O, Barnikol TT, et al. Coordinated reset neuromodulation for Parkinson's disease: proof-of-concept study. *Move Disord.* (2014) 29:1679–84. doi: 10.1002/mds.25923
135. Williams M, Hauptmann C, Patel N. Acoustic CR neuromodulation therapy for subjective tonal tinnitus: a review of clinical outcomes in an independent audiology practice setting. *Front Neurol.* (2015) 6:54. doi: 10.3389/fneur.2015.00054
136. Pedemonte M, Drexler D, Rodio S, Geisinger D, Bianco A, Pol-Fernandes D, et al. Tinnitus treatment with sound stimulation during sleep. *Int Tinnitus J.* (2010) 16:37–43.
137. Engineer ND, Riley JR, Seale JD, Vrana WA, Shetake JA, Sudanagunta SP, et al. Reversing pathological neural activity using targeted plasticity. *Nature.* (2011) 470:101–4. doi: 10.1038/nature09656
138. Riffle TL, Martel DT, Jones GR, Shore SE. Bimodal auditory electrical stimulation for the treatment of tinnitus: preclinical and clinical studies. *Curr Top Behav Neurosci.* (2020). doi: 10.1007/7854_2020_180. [Epub ahead of print].
139. Tyler R, Cacace A, Stocking C, Tarver B, Engineer N, Martin J, et al. Vagus nerve stimulation paired with tones for the treatment of tinnitus: a prospective randomized double-blind controlled pilot study in humans. *Sci Rep.* (2017) 7:11960. doi: 10.1038/s41598-017-12178-w
140. Hamilton C, D'Arcy S, Pearlmutter BA, Crispino G, Lalor EC, Conlon BJ. An investigation of feasibility and safety of bi-modal stimulation for the treatment of tinnitus: an open-label pilot study. *Neuromodul Technol Neural Interface.* (2016) 19:832–7. doi: 10.1111/ner.12452
141. Classon D. *New Program Uses Sound and Touch to Help Ease Tinnitus.* Health Hearing (2021). Available online at: https://www.healthyhearing.com/report/53207-Neosensory-duo-wearable-for-tinnitus
142. Conlon B, Langguth B, Hamilton C, Hughes S, Meade E, Connor CO, et al. Bimodal neuromodulation combining sound and tongue stimulation reduces tinnitus symptoms in a large randomized clinical study. *Sci Transl Med.* (2020) 12:eabb2830. doi: 10.1126/scitranslmed.abb2830
143. Tan CT, Soh D. Augmented reality games: A review. In: *Proceedings of Gameon-Arabia*, Alexandria: Eurosis (2010).
144. Plecher D, Ludl M, Klinker G. Designing an AR-escape-room with competitive and cooperative mode. In: Weyers B, Lürig C, Zielasko D, editors. *GI VR / AR Workshop*. Gesellschaft für Informatik e.V (2020). doi: 10.18420/vrar2020_30
145. Gavish N, Gutiérrez T, Webel S, Rodríguez J, Peveri M, Bockholt U, et al. Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. *Interact Learn Environ.* (2015) 23:778–98. doi: 10.1080/10494820.2013.815221
146. Emmelkamp PM, Meyerbröker K, Morina N. Virtual reality therapy in social anxiety disorder. *Curr Psychiatry Rep.* (2020) 22:32. doi: 10.1007/s11920-020-01156-1
147. Mekbib DB, Han J, Zhang L, Fang S, Jiang H, Zhu J, et al. Virtual reality for upper limb rehabilitation in patients with stroke: a meta-analysis of randomized clinical trials. *Brain Injury.* (2020) 34:456–65. doi: 10.1080/02699052.2020.1725126
148. Piumsomboon T, Dey A, Ens B, Lee G, Billinghurst M. the effects of sharing awareness cues in collaborative mixed reality. *Front Robot AI.* (2019) 6:5. doi: 10.3389/frobt.2019.00005

149. Yang J, Sasikumar P, Bai H, Barde A, Sörös G, Billinghurst M. The effects of spatial auditory and visual cues on mixed reality remote collaboration. *J Multi User Interfaces.* (2020) 14:337–52. doi: 10.1007/s12193-020-00331-1
150. Milgram P, Takemura H, Utsumi A, Kishino F. Augmented reality: a class of displays on the reality-virtuality continuum. In: *Telemanipulator and Telepresence Technologies*. Vol. 2351. International Society for Optics and Photonics (1995). p. 282–92.
151. Skarbez R, Smith M, Whitton MC. Revisiting Milgram and Kishino's Reality-Virtuality Continuum. *Front Virtu Real.* (2021) 2:27. doi: 10.3389/frvir.2021.647997
152. Guitton MJ. Telemedicine in tinnitus: feasibility, advantages, limitations, and perspectives. *ISRN Otolaryngol.* (2013) 2013:218265. doi: 10.1155/2013/218265
153. Malinvaud D, Londero A, Niarra R, Peignard P, Warusfel O, Viaud-Delmon I, et al. Auditory and visual 3D virtual reality therapy as a new treatment for chronic subjective tinnitus: results of a randomized controlled trial. *Hear Res.* (2016) 333:127–35. doi: 10.1016/j.heares.2015.12.023
154. Vajsakovic D, Maslin M, Searchfield GD. Principles and methods for psychoacoustic evaluation of tinnitus. *Curr Top Behav Neurosci.* (2021). doi: 10.1007/7854_2020_211. [Epub ahead of print].
155. Searchfield GD, Kobayashi K, Hodgson, S.-A., Hodgson C, Tevoitdale H, et al. Spatial masking: Development and testing of a new tinnitus assistive technology. *Assist Technol.* (2016) 28:115–25. doi: 10.1080/10400435.2015.1110214
156. Durai M, Doborjeh Z, Sanders PJ, Vajsakovic D, Wendt A, Searchfield GD. Behavioral outcomes and neural network modeling of a novel, putative, recategorization sound therapy. *Brain Sci.* (2021) 11:554. doi: 10.3390/brainsci11050554
157. Moawad GN, Elkhalil J, Klebanoff JS, Rahman S, Habib N, Alkatout I. Augmented realities, artificial intelligence, and machine learning: clinical implications and how technology is shaping the future of medicine. *J Clin Med.* (2020) 9:3811. doi: 10.3390/jcm9123811
158. Sanders PJ, Doborjeh ZG, Doborjeh MG, Kasabov NK, Searchfield GD. Prediction of acoustic residual inhibition of tinnitus using a brain-inspired spiking neural network model. *Brain Sci.* (2021) 11:52. doi: 10.3390/brainsci11010052
159. Mohri M, Rostamizadeh A, Talwalkar A. *Foundations of Machine Learning.* Cambridge: MIT press (2018).
160. Koprinkova-Hristova P, Mladenov V, Kasabov NK. Artificial neural networks. *Eur Urol.* (2015) 40:245. doi: 10.1007/978-3-319-09903-3
161. Kasabov N. *Springer Handbook of Bio-/Neuro-Informatics*. Berlin: Springer Science & Business Media. (2013). doi: 10.1007/978-3-642-30574-0
162. Goodfellow I, Bengio Y, Courville A, Bengio Y. *Deep Learning.* Cambridge: MIT press (2016).
163. Liu Y, Niu H, Zhu J, Zhao P, Yin H, Ding H, et al. Morphological neuroimaging biomarkers for tinnitus: evidence obtained by applying machine learning. *Neural Plast.* (2019) 2019:1712342. doi: 10.1155/2019/1712342
164. Haugg A, Renz FM, Nicholson AA, Lor C, Götzendorfer SJ, Sladky R, et al. Determinants of real-time fMRI neurofeedback performance and improvement - a machine learning mega-analysis. *bioRxiv*. (2020). doi: 10.1101/2020.10.21.349118
165. Shoushtarian M, Alizadehsani R, Khosravi A, Acevedo N, McKay CM, Nahavandi S, et al. Objective measurement of tinnitus using functional near-infrared spectroscopy and machine learning. *PLoS ONE.* (2020) 15:e0241695. doi: 10.1371/journal.pone.0241695
166. Lison P. An introduction to machine learning. *Lang Technol Group.* (2015) 1:1–35. doi: 10.1142/9789814335461_0001
167. Ayodele TO. Types of machine learning algorithms. In: Zhang Y, editor. *New Advances in Machine Learning*, Vol. 3. IntechOpen (2010). p. 19–48.
168. Al-Omary AY, Jamil MS. A new approach of clustering based machine-learning algorithm. *Knowl Based Syst.* (2006) 19:248–58. doi: 10.1016/j.knosys.2005.10.011
169. Kotsiantis SB. Supervised machine learning: a review of classification techniques. *Informatica*. (2007) 31:249–68.
170. Song J, Lee D, Choi I, Lee K. Objective diagnosis of tinnitus using resting-state EEC big data-based support vector machine learning. *J Hear Sci.* (2017) 7:133.
171. Sun ZR, Cai YX, Wang SJ, Wang CD, Zheng YQ, Chen YH, et al. Multi-view intact space learning for tinnitus classification in resting state EEG. *Neural Process Lett.* (2019) 49:611–24. doi: 10.1007/s11063-018-9845-1
172. Monaghan JJM, Liu Z, Yao L, McAlpine D. Redicting tinnitus from auditory brainstem responses using machine learning. In: Goehring T, editor. *Virtual Conference on Computational Audiology.* (2020). Available online at: https://computationalaudiology.com/
173. Schecklmann M, Lehner A, Poeppl TB, Kreuzer PM, Hajak G, Landgrebe M, et al. Cluster analysis for identifying sub-types of tinnitus: a positron emission tomography and voxel-based morphometry study. *Brain Res.* (2012) 1485:3–9. doi: 10.1016/j.brainres.2012.05.013
174. Niemann U, Brueggemann P, Boecking B, Mazurek B, Spiliopoulou M. Development and internal validation of a depression severity prediction model for tinnitus patients based on questionnaire responses and socio-demographics. *Sci Rep.* (2020) 10:4664. doi: 10.1038/s41598-020-61593-z
175. Niemann U, Boecking B, Brueggemann P, Mazurek B, Spiliopoulou M. Gender-Specific differences in patients with chronic tinnitus—baseline characteristics and treatment effects. *Front Neurosci.* (2020) 14:487. doi: 10.3389/fnins.2020.00487
176. Smith S, Kitterick P, Scutt P, Baguley D, Pierzycki R. An exploration of psychological symptom-based phenotyping of adult cochlear implant users with and without tinnitus using a machine learning approach. *Prog Brain Res*. (2020) 260:283–300. doi: 10.1016/bs.pbr.2020.10.002
177. Krauss P, Schilling A. Towards a cognitive computational neuroscience of auditory phantom perceptions. *arXiv*. (2020) arXiv:2010.01914
178. Doborjeh M, Kasabov N, Doborjeh Z, Enayatollahi R, Tu E, Gandomi AH. Personalised modelling with spiking neural networks integrating temporal and static information. *Neural Netw.* (2019) 119:162–77. doi: 10.1016/j.neunet.2019.07.021
179. Helson H. *Adaptation-Level Theory*. New York, NY: Harper and Row (1964).
180. Sanchez K. *Bose Gets Into Hearing Aid Business With New FDA-Cleared SoundControl Hearing Aids*. The Verge (2021). Available online at: https://www.theverge.com/2021/5/11/22430485/bose-soundcontrol-hearing-aid-fda-cleared
181. Hearing R. *GN Launches Jabra Enhance Pro Hearing Aids at Costco.* The Hearing Review (2021). Available online at: https://www.hearingreview.com/hearing-products/hearing-aids/otc/jabra-enhance-pro-hearing-aids

Conflict of Interest: GS has commercial interests in tinnitus therapy, he is a director of Tinnitus Tunes a subscription based tinnitus therapy website and has received research funding from Hearing Aid manufacturers.

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.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 13 August 2021
doi: 10.3389/fnins.2021.722970

Effects of Adaptive Non-linear Frequency Compression in Hearing Aids on Mandarin Speech and Sound-Quality Perception

*Shuang Qi[1,2], Xueqing Chen[1,2], Jing Yang[3], Xianhui Wang[4], Xin Tian[5], Hsuanyun Huang[5], Julia Rehmann[6], Volker Kuehnel[6], Jingjing Guan[5] and Li Xu[4]**

[1] Beijing Tongren Hospital, Capital Medical University, Beijing, China, [2] Key Laboratory of Otolaryngology—Head and Neck Surgery, Beijing Institute of Otolaryngology, Capital Medical University, Ministry of Education, Beijing, China, [3] Department of Communication Sciences and Disorders, University of Wisconsin—Milwaukee, Milwaukee, WI, United States, [4] Division of Communication Sciences and Disorders, Ohio University, Athens, OH, United States, [5] Sonova China, Shanghai, China, [6] Sonova AG, Stäfa, Switzerland

OPEN ACCESS

Edited by:
Changxin Zhang,
East China Normal University, China

Reviewed by:
Chenghui Jiang,
Nanjing Medical University, China
Joshua Alexander,
Purdue University, United States

Correspondence:
Li Xu
xul@ohio.edu

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Neuroscience

Received: *09 June 2021*
Accepted: *26 July 2021*
Published: *13 August 2021*

Citation:
Qi S, Chen X, Yang J, Wang X, Tian X, Huang H, Rehmann J, Kuehnel V, Guan J and Xu L (2021) Effects of Adaptive Non-linear Frequency Compression in Hearing Aids on Mandarin Speech and Sound-Quality Perception. Front. Neurosci. 15:722970. doi: 10.3389/fnins.2021.722970

Objective: This study was aimed at examining the effects of an adaptive non-linear frequency compression algorithm implemented in hearing aids (i.e., SoundRecover2, or SR2) at different parameter settings and auditory acclimatization on speech and sound-quality perception in native Mandarin-speaking adult listeners with sensorineural hearing loss.

Design: Data consisted of participants' unaided and aided hearing thresholds, Mandarin consonant and vowel recognition in quiet, and sentence recognition in noise, as well as sound-quality ratings through five sessions in a 12-week period with three SR2 settings (i.e., SR2 off, SR2 default, and SR2 strong).

Study Sample: Twenty-nine native Mandarin-speaking adults aged 37–76 years old with symmetric sloping moderate-to-profound sensorineural hearing loss were recruited. They were all fitted bilaterally with Phonak Naida V90-SP BTE hearing aids with hard ear-molds.

Results: The participants demonstrated a significant improvement of aided hearing in detecting high frequency sounds at 8 kHz. For consonant recognition and overall sound-quality rating, the participants performed significantly better with the SR2 default setting than the other two settings. No significant differences were found in vowel and sentence recognition among the three SR2 settings. Test session was a significant factor that contributed to the participants' performance in all speech and sound-quality perception tests. Specifically, the participants benefited from a longer duration of hearing aid use.

Conclusion: Findings from this study suggested possible perceptual benefit from the adaptive non-linear frequency compression algorithm for native Mandarin-speaking adults with moderate-to-profound hearing loss. Periods of acclimatization should be taken for better performance in novel technologies in hearing aids.

Keywords: hearing aids, non-linear frequency compression, speech recognition, Mandarin Chinese, sound quality, acclimatization, adult

INTRODUCTION

High-frequency components of acoustic signals convey useful information in speech and music. They play an important role in sound-quality perception, sound localization, speech perception in noise, and language development in children (Stelmachowicz et al., 2002, 2004; Monson et al., 2014; Moore, 2016). Many patients with sensorineural hearing loss have difficulty accessing high-frequency information. For this population, the most common intervention is to wear hearing aids. However, due to the limitation of audible bandwidth for speech information above 5 kHz in the conventional processing hearing aids (Boothroyd and Medwetsky, 1992; Moeller et al., 2007) and the presence of cochlear dead regions (Moore, 2001, 2004; Zhang et al., 2014), the aided performance in many hearing-aid users is not satisfactory. The frequency-lowering technique provides a practical solution because it shifts inaudible high frequencies to audible low-frequency regions (Simpson, 2009; Alexander, 2013; Mao et al., 2017). Among many different frequency lowering algorithms, non-linear frequency compression (NLFC) has been implemented in modern commercial hearing aids, such as Phonak Naida hearing aids. The key concept of NLFC is to disproportionally compress high frequencies into lower-frequency regions. In the first-generation of NLFC (known as SoundRecover or SR), two parameters, cut-off frequency (CT) and compression ratio (CR), determine the start point and strength of compression, respectively. Sound with frequencies below the CT remains unchanged but sound above the CT is compressed.

For people with severe-to-profound hearing loss, more aggressive settings with a lower CT and a higher CR are required because the patients have a narrower audible frequency bandwidth and the inaudible frequencies start at a lower frequency point in comparison to people with mild or moderate hearing loss. While the use of a lower CT ensures that a wider range of high frequencies can be shifted down so that they become audible to hearing aid users, it may also introduce unwanted detrimental effects to consonant and vowel perception (Alexander, 2016; Yang et al., 2018) and sound-quality perception (McDermott, 2011; Parsa et al., 2013; Souza et al., 2013). Therefore, to achieve a balance between audibility (lower CT) and fidelity (higher CT), Phonak introduced a new adaptive NLFC algorithm (known as SoundRecover 2 or SR2) in which the CT is switched between a low cut-off (CT1) and a high cut-off (CT2) based on the short-term energy distribution of the input signal (Rehmann et al., 2016). When the system detects sound energy at a relatively low-frequency region (e.g., vowels), CT2 is used so that the formants are not disturbed. When the incoming signal is a high-frequency sound (e.g., consonants), the system uses CT1. Technically, the adaptive NLFC preserves the spectral structure of vowel sounds and other low-frequency speech information and allows the accessibility of high-frequency information that is compressed and shifted to the lower-frequency region.

So far, there has been a number of studies examining the efficacy of NLFC on various aspects of speech perception including phoneme and word recognition, sentence perception, and sound-quality perception (Glista et al., 2009; Wolfe et al., 2010, 2011, 2017; Ching et al., 2013; Parsa et al., 2013; Brennan et al., 2014, 2017; Hopkins et al., 2014; McCreery et al., 2014; Picou et al., 2015; Alexander and Rallapalli, 2017; Chen et al., 2020; Xu et al., 2020). While many studies reported lower detection thresholds and improved perceptual accuracies with NLFC-fitted hearing aids in comparison to hearing devices fitted with conventional processing (CP) (Ching et al., 2013; Alexander et al., 2014; Zhang et al., 2014; Ching and Rattanasone, 2015), some studies reported no additional benefit in phoneme audibility, or sentence recognition with the NLFC algorithm (Perreau et al., 2013; Bentler et al., 2014; Picou et al., 2015). In addition, within those studies that found improved perceptual performance with NLFC, some reported that the benefit of NLFC was not ubiquitously shown in all tested subjects (Simpson et al., 2005; Glista et al., 2009; McCreery et al., 2014). For example, Simpson et al. (2005) tested the recognition of monosyllabic words with NLFC vs. CP in 17 participants with sloping moderate-to-severe sensorineural hearing loss. Only eight of them showed improved recognition accuracy and one participant demonstrated decreased accuracy with NLFC compared to CP. As summarized in Akinseye et al. (2018), there still lacks convincing evidence supporting the superiority of NLFC over CP in all hearing-related tasks.

As adaptive NLFC is a newly developed algorithm, only a few studies tested the use of this algorithm in hearing-impaired listeners (e.g., Glista et al., 2017, 2019; Xu et al., 2020). In these studies, the researchers compared the perceptual performance with CP, NLFC, and adaptive NLFC of the tasks including phoneme perception, word recognition, and sound-quality ratings in hearing-impaired children and/or adults. Wolfe et al. (2017) reported a lower threshold for phoneme detection and higher accuracy for phoneme and word recognition in the tested children. Glista et al. (2017) found that both NLFC and adaptive NLFC provided greater benefits on phonetic recognition than CP. However, there was no significant difference between static NLFC and adaptive NLFC on phoneme perception. In a recent study, Xu et al. (2020) evaluated the efficacy of the adaptive NLFC (i.e., SR2) on phoneme detection, speech detection threshold, and sound-quality ratings in Mandarin-speaking hearing-impaired adults. In that study, five SR2 settings (SR2-off, SR2-default, SR2-weak, SR2-strong 1, and SR2-strong 2) with various fitting parameters for CT1, CT2, and CR were compared. The results revealed that the hearing-impaired listeners showed improved (lowered) phoneme detection and speech detection thresholds with the two strong settings than weak setting or off condition. However, different settings did not exert a significant influence on sound-quality ratings. While Xu et al. (2020) study focused on detection ability through one-time tests, the current study aimed to provide a more comprehensive evaluation to test the impact of adaptive NLFC on different aspects of perception ability including Mandarin consonant, vowel, and sentence recognition. Additionally, because each participant was tested multiple times spanning a 3-month period, the design of this study enabled us to examine how the perceptual performance would change as a function of increased experience with adaptive NLFC-fitted hearing aids.

MATERIALS AND METHODS

Participants

The participants included 29 Mandarin-speaking adults (13 females and 16 males) aged between 37 and 76 years old ($M = 66.7$, $SD = 8.8$). All participants were diagnosed with sloping moderate-to-profound sensorineural hearing loss. The average pure-tone audiometric thresholds between 500 and 4,000 Hz for both ears were between 40 and 90 dB HL. The individual and group average pure-tone thresholds are shown in **Figure 1**. The duration of hearing loss ranged from 2 to 40 years, with a mean of 13.1 years. All participants met the following recruitment criteria: (1) symmetric sloping sensorineural hearing loss (i.e., interaural difference $\leq$ 15 dB at all octave frequencies from 250 to 8,000 Hz) with air-bone gaps at any frequency $\leq$ 15 dB; (2) normal middle ear function as indicated by tympanometry and otoscopy examinations; (3) no diagnosed cognitive or mental impairments (able to communicate effectively with their families and the investigators); (4) no experience of hearing aids with frequency lowering schemes prior to participating in the present study; and (5) native Mandarin speakers in daily life. Twenty-two of the participants had used hearing aids before participating in the study whereas seven had no hearing aid experience. All 29 participants completed all five test sessions. This study protocol was reviewed and approved by the Institutional Review Boards of Ohio University and Beijing Tongren Hospital.

Hearing Aid Fitting

All participants were bilaterally fitted with experimental hearing aids (Phonak Nadia V90-SP) programmed in Phonak's Target fitting software (v. 5.1). To ensure proper amplification of sound across the entire speech spectrum and to limit acoustic feedback, occluding hard ear-molds made of acrylic materials were used with different vent sizes based on the recommendation of the fitting software. After the feedback and real-ear test using the estimated RECD (real-ear-to-coupler difference) and recommended vents, the APDT (Adaptive Phonak Digital Tonal) gain algorithm was chosen as the prescriptive target. The output gain level was initially set to 100%, decreasing in a 10% step size in case the participant reported that the hearing aids were too loud. Three SR2 settings were tested in the study: SR2 off, SR2 default, and SR2 strong (i.e., moving three steps toward Audibility relative to default on the upper slider). As a group, the parameters CT1 and CT2 changed from 3.77 $\pm$ 1.35 (mean $\pm$ SD) and 5.12 $\pm$ 1.21 kHz in SR2 default to 2.17 $\pm$ 0.55 and 3.78 $\pm$ 0.67 kHz in SR2 strong. The group average of the parameter CR remained unchanged in SR2 default and SR2 strong (1.22 $\pm$ 0.10 vs. 1.22 $\pm$ 0.06). Note that the use of lower CT1 and CT2 in SR2 strong could potentially move more high-frequency energy to the audible range but create greater disruption of low-frequency information. A schematic diagram of the signal processing for SR2 is available in our previous acoustic study of non-linear frequency compression (See Figure 1 of Yang et al., 2018). The other advanced functions (such as noise reduction, directionality, etc.) were all set as default. All the adjustments were performed by experienced audiologists. Moreover, all the settings embedded in the hearing aids remained the same throughout the process of the entire study. All participants wore the same experimental hearing aids with only setting varied (see Procedures below) throughout the study period.

Perceptual Tasks and Outcome Measures

The perceptual performance of each participant was evaluated through speech perception tests and sound-quality rating tasks. The speech perception tests included Mandarin-Chinese consonant, vowel, and sentence recognition tests.

Consonant Recognition

The consonant recognition test included five Mandarin fricatives (i.e., f /f/, s /s/, x /ɕ/, sh /ʂ/, and h /x/) and six Mandarin affricates (i.e., z /ts/, c /tsʰ/, j /tɕ/, q /tɕʰ/, zh /tʂ/, ch /tʂʰ/) embedded in a /Ca/ syllable in tone 1. The 11 words containing the target consonants are 发 (fā), 撒 (sā), 扎 (zā), 擦 (cā), 虾 (xiā), 家 (jiā), 掐 (qiā), 沙 (shā), 渣 (zhā), 插 (chā), and 哈 (hā). The tokens were recorded from 6 adult Mandarin speakers (3 males and 3 females). Thus, the consonant recognition test comprised 66 tokens (11 words × 6 speakers) that were randomly presented to the participants. The intensity of the stimuli was set at 65 dB SPL.

Vowel Recognition

The Mandarin vowel list included 12 Mandarin vowels (i.e., /a/, /ai/, /ao/, /ɤ/, /i/, /iao/, /ie/, /iou/, /ou/, /u/, /uei/, /uo/) embedded in a /dV/ syllable structure in tone 1. The 12 monosyllabic words are 搭 (dā), 呆 (dāi), 刀 (dāo), 嘚 (dē), 低 (dī), 雕 (diāo), 跌 (diē), 丢 (diū), 兜 (dōu), 督 (dū), 堆 (duī), and 多 (duô). The tokens were recorded from the same 6 adult Mandarin speakers. In total, there were 72 tokens for the vowel recognition test (12 words × 6 speakers). The intensity of the stimuli was also set at 65 dB SPL.

Sentence Recognition

The material used for sentence recognition was Mandarin Hearing in Noise Test (MHINT) (Wong et al., 2007). MHINT contains 12 sentence lists. Each list is composed of 20 sentences each of which is 10 Chinese characters long. The intensity of the sentence stimuli was fixed at 65 dB SPL. In order to reduce the ceiling effect for sentence recognition, a speech-spectrum-shaped noise (Xu et al., 2021) was mixed with the sentences at a signal-to-noise ratio (SNR) of +5 dB. For each SR2 setting, participants were tested with one different list randomly chosen from the 12 sentence lists. A total of 11 lists were used throughout the five sessions in which two lists were used for sessions 1, 2, 4, and 5 and three lists were used for session 3 (see below in Procedures section for details on test sessions and SR2 conditions). The final score was calculated based on the percent of correct characters in each test list.

Sound-Quality Rating

All participants were asked to rate the loudness, clarity, naturalness, and overall sound quality of different types of sounds including own voice, male voice, female voice, bird chirp, and music. Own voice was referred to as the participant's natural and spontaneous vocal production in daily life after wearing the

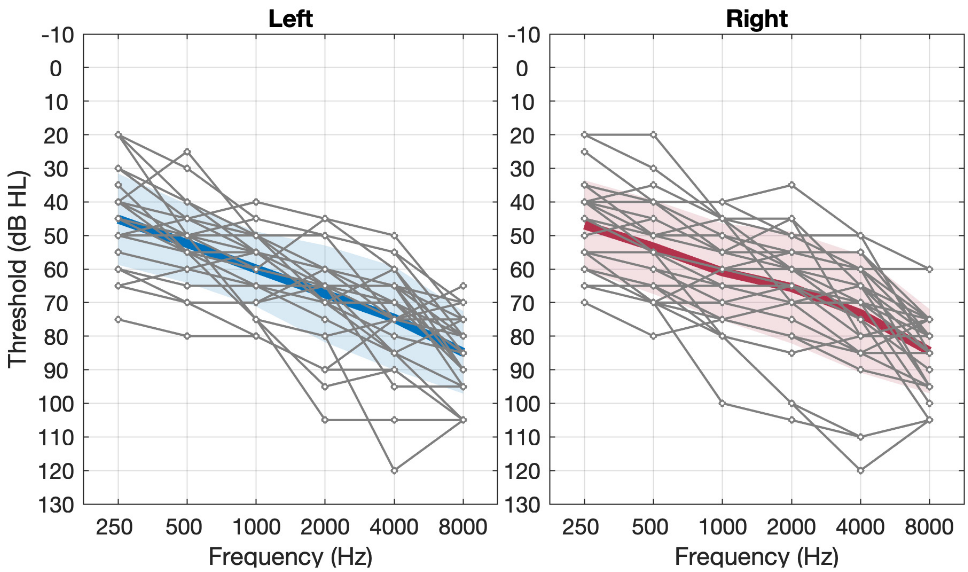


FIGURE 1 | Unaided pure-tone thresholds of **left** and **right** ears in the 29 subjects. The thin gray lines represent individual thresholds and the thick black line represents the group mean threshold (n = 29). The shaded area represents ±1 S.D. of the mean.

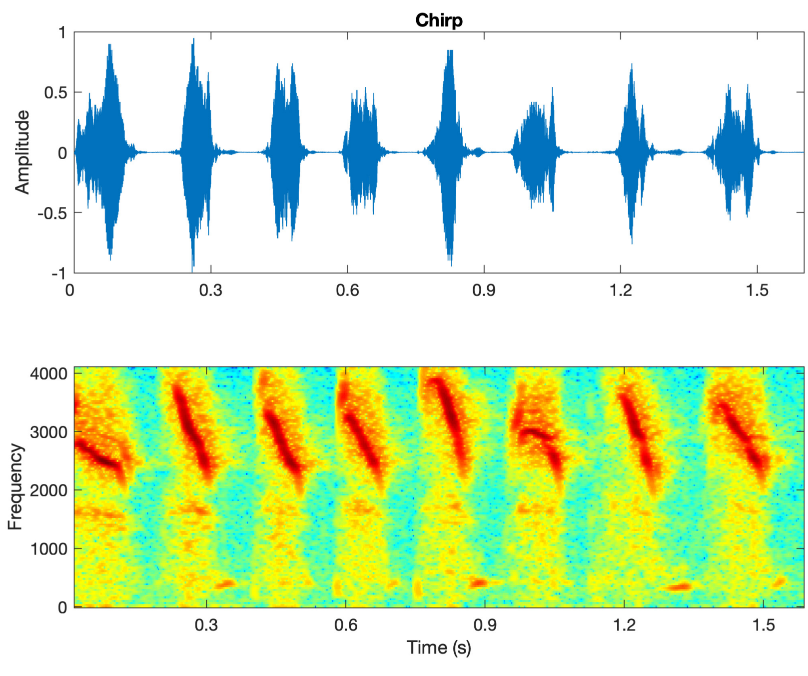


FIGURE 2 | The waveform and spectrogram of the bird chirps. The spectrogram shows that the chirps are rapid downward frequency sweeps from 4000 to 2,000 Hz approximately.

experimental hearing aids. The male voice and female voice were reading text composed of 127 Chinese characters by one male and one female talker. The lengths of the recordings were 34 and 36 s for the male and female voices, respectively. They were presented at 65 dB(A) in quiet. The bird chirps were provided by MATLAB software (MathWorks, Natick, MA). The original 8 chirps (**Figure 2**) were repeated twice to form 24 chirps with a total duration of 5.5 s. The music was a piece of recorded piano music excerpted from a classic and well-known Chinese folk music entitled "Liang Zhu (The Butterfly Lovers)." The bird chirps and music were presented at 65 and 70 dB(A), respectively.

The five types of stimuli were presented in a fixed order as above with no repetition being allowed. After listening to each stimulus, participants were asked to rate four aspects of

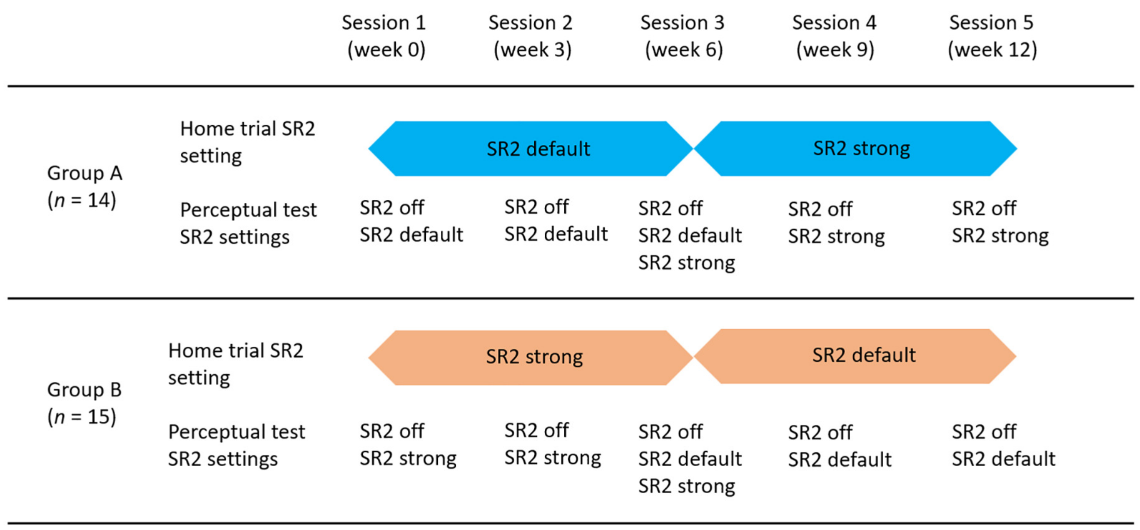


FIGURE 3 | Diagram of test procedures and group assignments. There were 5 sessions equally distributed in the 12-week period. Participants were randomly assigned into two groups. Group A (n = 14) used SR2 default as the home trial setting for the first 6 weeks and switched to SR2 strong for the home trial in the second 6 weeks. Group B (n = 15) used SR2 strong for the first 6 weeks and SR2 default for the second 6 weeks in their home trial.

sound quality including loudness, clarity, naturalness, and overall quality using a continuous bar with two ends being 0 (extremely poor) and 10 (perfect). No practice was provided for this task.

Procedures

All participants were tested at five different sessions separated by 3 weeks between each two consecutive sessions. In the first session, the participants were fitted with bilateral hearing aids (Phonak Nadia V90-SP BTE) with individualized hard ear-molds. The sound-field aided thresholds with SR2 off, SR2 default, and SR2 strong settings were measured using warble tones at 250, 500, 1,000, 2,000, 3,000, 4,000, 6,000, and 8,000 Hz. Perceptual tests including speech perception (i.e., consonant, vowel, and sentence recognition) and sound-quality rating with SR2 off and SR2 default or SR2 strong settings were then conducted. After the first session, the participants were sent home with the hearing aids on either SR2 default or SR2 strong. To count-balance the order of SR2 settings, 14 of the 29 participants wore SR2 default for the first 6 weeks (Group A) and the remaining 15 of the 29 participants wore SR2 strong for the first 6 weeks (Group B). After 6 weeks, the SR2 settings for both groups were switched. **Figure 3** illustrates the SR2 settings used in the home trial and those used in the perceptual tests in the lab. The order of SR2 setting used in the perceptual tests was randomized across participants and test sessions. All tests were conducted in a sound booth, with the background noise below 30 dB A. The stimuli were presented through a loudspeaker located at 1.45 m in front of the participants at 0° azimuth.

Statistical Analyses

Statistical analyses were conducted using R software (version 3.63). The percent-correct data of the recognition tests were treated as binomial data (Thornton and Raffin, 1978). A generalized linear mixed-effect model (GLMM) (Warton and Hui, 2011) was used to investigate the impacts of (1) SR2 settings (off, default, and strong) and (2) test sessions on the percent-correct scores. Furthermore, we analyzed the potential interactions between SR2 setting and test session. For the sound-quality rating data, linear mixed models (LMM) were performed separately for each category of sound-quality percept (i.e., loudness, clarity, naturalness, and overall preference). The three main factors were (1) SR2 settings, (2) test sessions, and (3) sound types (i.e., own voice, male voice, female voice, bird chirp, and music).

RESULTS

Group Aided Thresholds

The group mean unaided and aided hearing thresholds under the three SR2 settings (**Figure 4**) were analyzed using repeated-measures analysis of variance (ANOVA) by frequency separately. Compared with unaided thresholds, aided hearing thresholds were better at all frequencies (with F-values ranging from 8.87 to 77.92, all $p < 0.001$) except 250 Hz [$F_{(3,\ 118)} = 0.07$, $p = 0.98$]. Tukey-adjusted pairwise comparisons revealed no significant differences in aided hearing thresholds among SR2 off, SR2 default, and SR2 strong settings from 250 Hz to 6 kHz (all adjusted $p > 0.05$). At 8 kHz, the hearing thresholds for SR2 off setting and unaided condition were comparable (adjusted $p = 0.34$), and those for SR2 default and SR2 strong settings were not significantly different (adjusted $p = 0.29$). However, the hearing threshold was significantly higher at for unaided or SR2 off setting than for SR2 default setting (unaided vs. SR2 default: adjusted $p < 0.0001$; SR2 off vs. SR2 default: adjusted $p < 0.0001$) and SR2 strong setting (unaided vs. SR2 strong: adjusted $p < 0.0001$; SR2 off vs. SR2 strong: adjusted $p < 0.0001$).

Speech Perception Tests

Individual and group average performance of consonant, vowel, and sentence recognition with the three SR2 settings in five sessions is shown in **Figure 5**. All participants underwent a 12-week trial of the two SR2-enabled settings (i.e., SR2 default and SR2 strong), each of which for 6 weeks, respectively. As explained in **Figure 3** and associated text, 14 of the 29 participants (Group A) used SR2 default setting in the first 6 weeks, while the other 15 participants (Group B) used SR2 strong setting first. In the sixth week, the two SR2 settings were switched. This counterbalance design minimized potential order effects. As shown in **Figure 5**, no apparent order effects were observed in the speech recognition data. An independent *t*-test comparing the recognition performance between Groups A and B revealed no order effect [consonant recognition: $t(27) = -0.57$, $p = 0.57$; vowel recognition: $t(27) = -1.17$, $p = 0.24$; sentence recognition: $t(27) = -0.95$, $p = 0.35$]. Thus, in the following presentations of results, data from Groups A and B were pooled together and were not treated separately.

Large individual variability in speech recognition, especially in consonant recognition in quiet and sentence recognition in noise (+5 dB SNR), was evident (**Figure 5**). Pearson correlation analyses showed that consonant recognition scores in the three SR2 settings were correlated with the participants' aided thresholds (with corresponding SR2 settings) at high frequencies (i.e., 4,000, 6,000, or 8,000 Hz) as well as averaged thresholds across 500, 1,000, 2,000, and 4,000 Hz ($PTA_{250-4{,}000\ Hz}$) (correlation coefficients ranging from −0.411 to −0.741, *z*-test, all $p < 0.05$) with exception of one condition [i.e., 8,000 Hz with the SR2 strong setting ($r = -0.311$, $p = 0.1$)]. In addition, we used the difference scores in speech recognition between SR2 default and SR2 off settings or those between SR2 strong and SR2 off settings as potential NLFC benefit scores. However, Pearson correlation analyses revealed no significant correlation of the latter and the participants' aided hearing thresholds at high frequencies or $PTA_{250-4{,}000\ Hz}$. We also found that the participants' age was not correlated with any of the speech recognition performance nor the potential NLFC benefit scores.

Figure 6 plots the group mean speech-recognition results of the three SR2 settings as a function of the test session. For consonant recognition, the GLMM analysis revealed that both SR2 settings (i.e., SR2 off, SR2 default, and SR2 strong) and test sessions (i.e., 1, 2, 3, 4, and 5) were significant factors for the recognition performance [SR2 settings: $\chi^2(2, N = 29) = 21.90$, $p < 0.\ 0001$; sessions: $\chi^2(4, N = 29) = 68.95$, $p < 0.0001$]. *Post-hoc* multiple comparisons revealed a modestly better performance for SR2 default setting than for SR2 off setting (adjusted $p = 0.001$) or SR2 strong setting (adjusted $p < 0.001$). No significant differences between SR2 off and SR2 strong settings were observed (adjusted $p = 0.41$). The participants also showed improved performance from Session 1 to Session 5. Multiple comparisons revealed that the participants performed similarly at Sessions 4 and 5 (adjusted $p = 0.98$) which were both significantly better than at Sessions 1, 2, and 3 (all adjusted $p < 0.001$). The interaction between test session and SR2 setting was not significant [$\chi^2(8, N = 29) = 4.31$, $p = 0.83$].

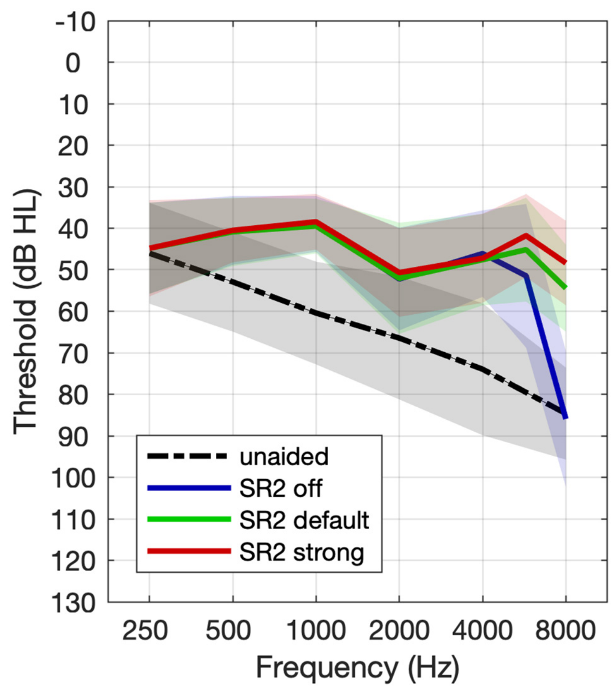


FIGURE 4 | Mean group unaided and aided hearing thresholds. The black line represents the unaided thresholds. The dashed line, the solid line, and the dotted lines represent free-field thresholds for the aided thresholds with SR2 off, SR2 default, and SR2 strong settings, respectively. The shaded area represents ±1 S.D. of the mean.

For vowel recognition, unlike consonant recognition, only test session contributed to the improved performance [$\chi^2(4, N = 29) = 61.73$, $p < 0.00$]. Specifically, the vowel recognition at Session 5 was significantly better than that at Sessions 1, 2, and 3 (all adjusted $p < 0.001$). Between Sessions 4 and 5, however, no significant differences were observed (adjusted $p = 0.006$). This was also true for the pairwise comparisons among Sessions 1, 2, and 3 (all adjusted $p > 0.1$).

For sentence recognition in noise, similar to consonant perception, both test sessions and SR2 settings contributed to the improved performance [session: $\chi^2(2, N = 29) = 19.96$, $p < 0.0001$; SR2 setting: $\chi^2(4, N = 29) = 302.22$, $p < 0.0001$] and no interaction was found between the two factors [$\chi^2(8, N = 29) = 8.27$, $p = 0.41$]. The sentence-recognition performance for SR2 default setting was the highest as compared to SR2 off setting (adjusted $p = 0.001$) or SR2 default setting (adjusted $p = 0.003$). No difference between SR2 off and strong was found (adjusted $p = 0.99$).

Consonant confusion analyses were then conducted to illustrate potential confusion patterns that might help to explain the modest improvement with the frequency-lowering technique. **Figure 7** shows the consonant confusion matrices averaged across all 5 sessions. Of all tested consonants, the alveolar sounds /s, ts, ts^h/ (s, z, c) showed the lowest recognition accuracy and alveolopalatal sounds /ɕ, tɕ, tɕ^h/ (x, j, q) showed the highest accuracy. Among the five places of

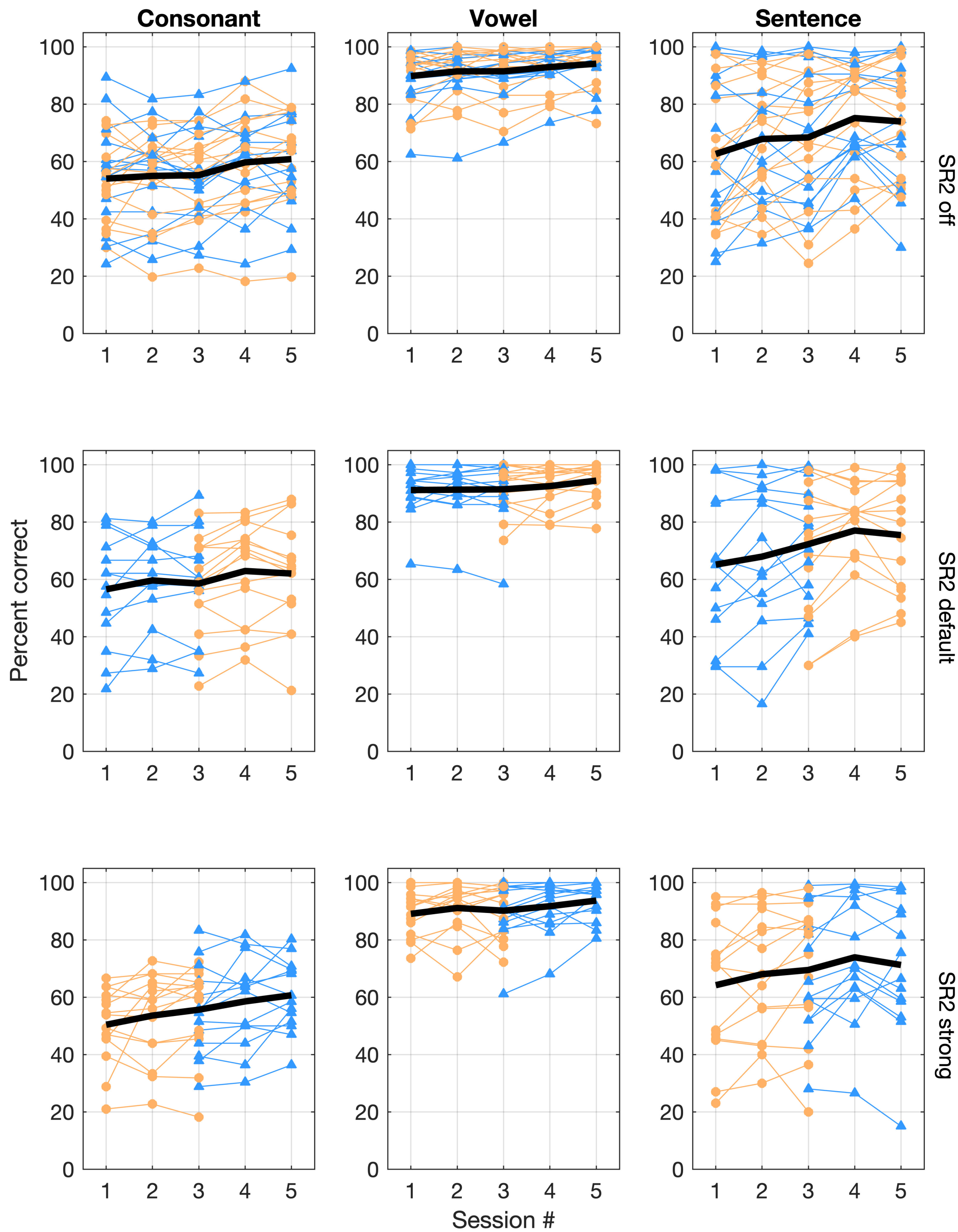


FIGURE 5 | Individual and group mean speech-recognition performance. Consonant and vowel recognition in quiet, and sentence recognition in noise (+5 dB SNR) are represented in the three columns whereas performance for the three SR2 settings is represented in the three rows of panels. In each panel, performance scores (% correct) are plotted as a function of test sessions. Each thin line represents one participant. Those in Group A are plotted with triangles in blue whereas those in Group B are plotted with circles in orange. The thick black line represents the overall group mean performance.

articulation, confusions mainly occurred between the alveolar sounds /s, ts, ts^h/ (s, z, c) and the retroflex postalveolar /ʂ, tʂ, $tʂ^h$/ (sh, zh, ch), which showed an asymmetrical pattern. That is, the alveolar sounds were more likely to be recognized as the retroflex sounds, rather than the other way around. Additionally, as the SR setting became stronger,

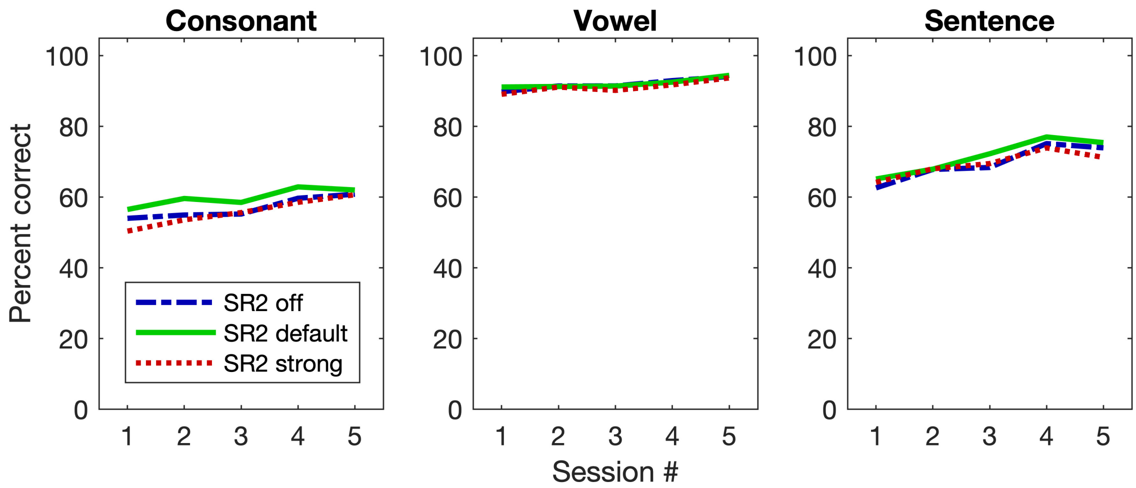


FIGURE 6 | Group mean performance of speech-recognition performance. The three panels are for consonant and vowel recognition in quiet, and sentence recognition in noise (+5 dB SNR), respectively. The three different lines represent the group mean results for the three SR2 settings.

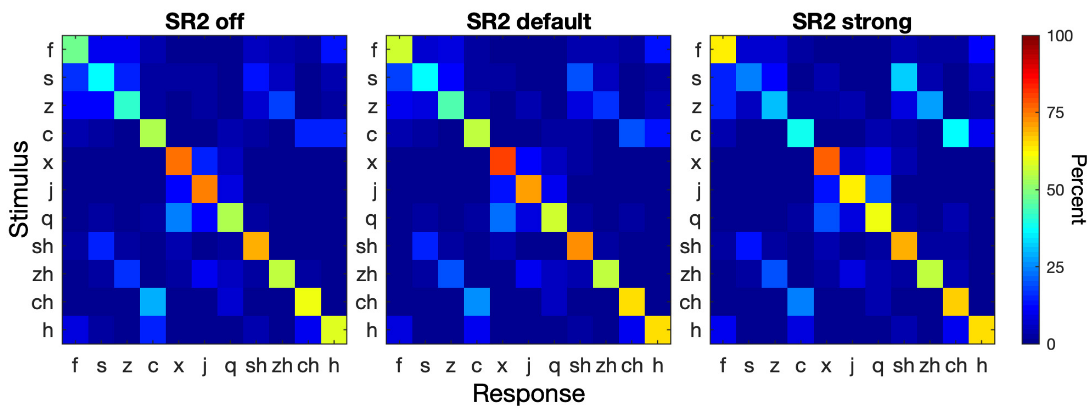


FIGURE 7 | Consonant confusion matrices. Consonant recognition data were collapsed from all 5 sessions. The three panels show the confusion matrices for SR2 off, SR2 default, and SR2 strong settings. In each panel, the stimulus is represented by the ordinate and the response by the abscissa. The color of each cell in a matrix represents the percent of a stimulus being identified as a particular consonant (see color bar on the right).

the degree of confusion of alveolar sounds as retroflex sounds increased.

Sound-Quality Rating

Figure 8 plots the average sound-quality ratings from the 29 participants in the three SR2 settings and five test sessions. The LMM analyses revealed different patterns in each category of sound-quality percept (loudness, clarity, naturalness, and overall preference). For the rating of loudness, the analysis yielded significant main effects of test session [$F_{(4,1555.6)} = 106.44$, $p < 0.0001$] and type of stimulus [e.g., own voice, male, female, bird chirps, and music, $F_{(4,1553.8)} = 25.7$, $p < 0.0001$]. SR2 settings [$F_{(2,1554.4)} = 1.01$, $p = 0.36$] as well as the interaction between test session and SR2 setting [$F_{(8,1558.4)} = 0.38$, $p = 0.93$] were not significant. The participants' satisfaction with loudness improved progressively from Session 1 to Session 5 (all adjusted $p < 0.001$). Among the five different types of stimuli, the rating score for own voice was significantly lower than all the other four types of stimuli (all adjusted $p < 0.001$). In addition, the overall quality rating for the female voice was found to be higher than for the male voice (adjusted $p = 0.028$). No other significant differences were observed (all adjusted $p > 0.05$).

In terms of clarity rating, all the three main factors were significant [test session: $F_{(4,1558.9)} = 4.10$, $p = 0.016$; SR2 setting: $F_{(2,1557.6)} = 71.16$, $p < 0.0001$; type of stimulus: $F_{(4,1557)} = 18.59$, $p < 0.0001$] but there was no interaction between test session and SR2 setting [$F_{(8,1561.7)} = 0.72$, $p = 0.68$]. The participants rated clarity significantly higher with SR2 default setting than with SR2 strong setting (adjusted $p = 0.018$), while no difference was detected between SR2 off and SR2 default settings (adjusted $p = 0.17$) or between SR2 off and SR2 strong settings (adjusted $p = 0.041$). The clarity rating at Sessions 4 and 5 was the highest compared to that at Sessions 1, 2, and 3 (all adjusted $p < 0.001$), yet no significant difference was observed at Session 5 compared

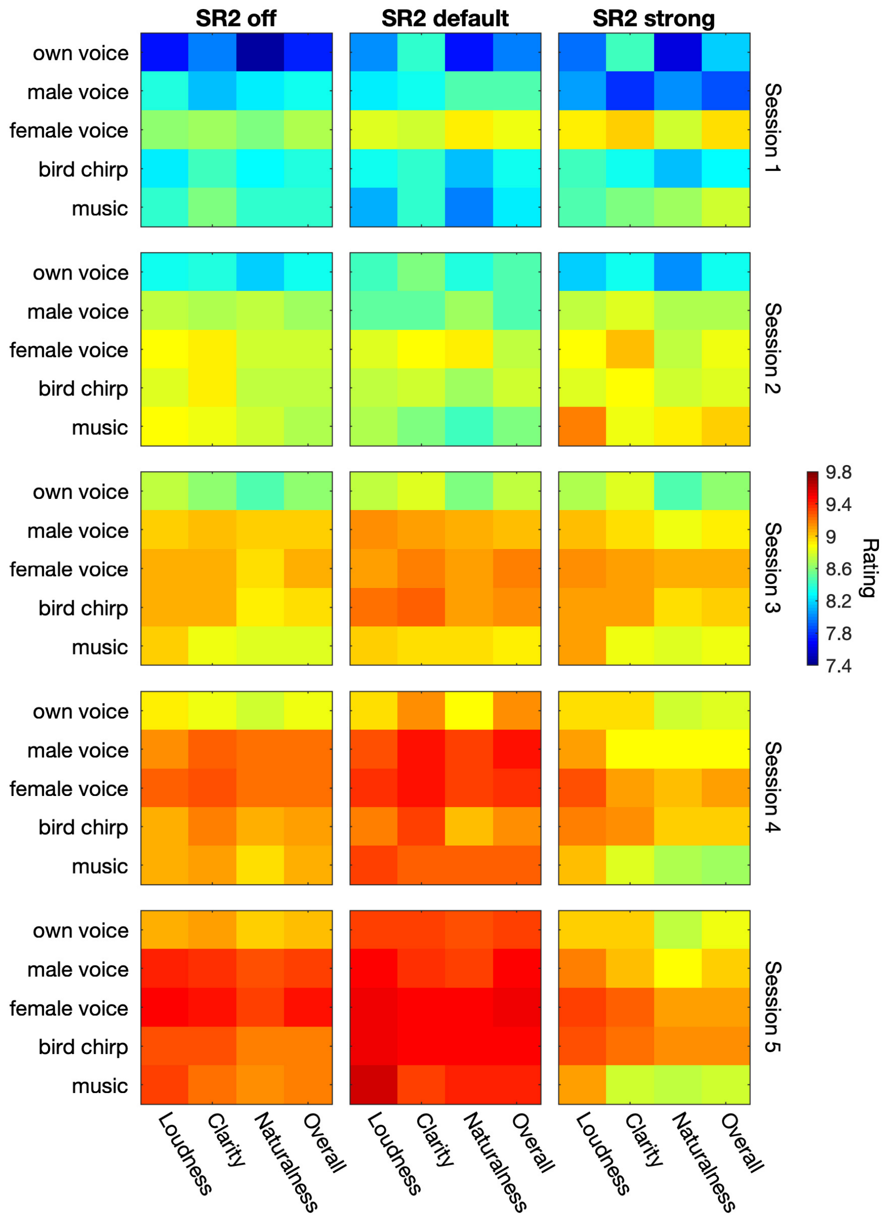


FIGURE 8 | The average sound-quality ratings from the 29 participants. The three columns represent SR2 off, SR2 default, and SR2 strong setting, respectively. From top to down, the 5 rows indicate the 5 sessions. In each panel, the five types of stimuli (i.e., own voice, male voice, female voice, bird chirps, and music) are represented in rows whereas the category of percepts (i.e., loudness, clearness, naturalness, and overall quality) are represented in columns. The color in each cell represents the rating score as indicated by the color bar on the right.

with Session 4 (adjusted $p = 0.59$). Finally, own voice clarity scores were found to be the lowest among all types of stimuli (all adjusted $p < 0.001$). The female voice, on the other hand, received rating scores significantly higher than the male voice (adjusted $p < 0.001$) or music (adjusted $p = 0.001$) but similar scores to the chirp (adjusted $p = 0.39$). All other pairwise comparisons were not significant otherwise (all adjusted $p > 0.1$).

For the naturalness and overall rating, all the three main factors (i.e., SR2 setting, test session, and type of stimulus) played significant roles in the outcome measures [Naturalness: $F_{(2,1557.7)} = 6.14$, $p = 0.002$; $F_{(4,1558.9)} = 78.00$, $p < 0.0001$; $F_{(4,1557.1)} = 34.53$, $p < 0.0001$; Overall: $F_{(2,1557.7)} = 6.39$, $p = 0.0017$; $F_{(4,1558.9)} = 78.15$, $p < 0.0001$; $F_{(4,1557.1)} = 24.80$, $p < 0.0001$] but not the interaction terms [$F_{(8,1561.6)} = 0.86$, $p = 0.55$; $F_{(8,1561.7)} = 0.60$, $p = 0.78$]. Both naturalness and overall ratings improved from Session 1 to 4 (naturalness: all adjusted $p < 0.0001$; overall quality: all adjusted $p < 0.0001$) but not from Session 4 to 5 (naturalness: adjusted $p = 0.27$; overall quality: adjusted $p = 0.11$). Both ratings were found to be better for SR2 default setting than for SR2 strong setting (both adjusted

$p = 0.002$), but no difference was found for the other comparisons among SR2 settings. Among different types of stimuli, own voice was rated with the lowest scores for both naturalness (all adjusted $p < 0.0001$) and overall quality scores (all adjusted $p < 0.0001$). In addition, the female voice was rated with significantly higher naturalness than music (adjusted $p = 0.009$) and with higher overall quality scores than the other types of stimuli (all adjusted $p < 0.05$).

DISCUSSION

In the present study, we evaluated the efficacy of the new adaptive NLFC scheme on Mandarin speech perception and sound-quality ratings in adult hearing aid users. A series of tests were conducted over a 3 month period to evaluate aided hearing thresholds, consonant and vowel recognition in quiet, sentence recognition in noise, and subjective sound-quality ratings in 29 participants with severe-to-profound hearing impairment.

For sound detection, our results demonstrated an improvement of more than 30 dB in detecting high-frequency sounds at 8 kHz with the application of adaptive NLFC scheme (**Figure 4**). A similar tendency was found for 6 kHz but the change of the detection threshold with the adaptive NLFC on vs. off was only 5 dB. This change was not statistically nor clinically significant. The reduced effect of SR2 on the detection of the 6-kHz tone in the present study may be due to the fact that the hearing loss at 6 kHz for most of our participants was not too severe (**Figure 1**) and that the hearing aids provided adequate amplification at that frequency even with the conventional processing scheme (**Figure 4**). Nonetheless, the detection results were consistent with findings by Xu et al. (2020) in which a group of 15 adult hearing-aid users with sloping severe-to-profound sensorineural hearing loss showed substantial improvement in detecting high-frequency Mandarin phonemes such as /s/ (centered at both 6 and 9 kHz) and /ɕ/ ("x"), especially with the stronger SR2 settings.

For vowel recognition, most of the participants had recognition scores above 90% correct (**Figure 5**). The adaptive NLFC scheme did not show a significant impact on the recognition performance in this experiment. This result was similar to previous findings in vowel recognition using the static NLFC scheme (i.e., SR) (Yang et al., 2018; Chen et al., 2020). The first three formants that characterize vowel identity reside in relatively low frequency regions, usually lower than 4 kHz. In contrast to high-frequency sounds, vowel sounds are typically accessible in patients with sensorineural hearing impairments. SR2, even with a strong setting, enables well-preserved formant patterns due to the application of two cutoffs. In the current study, even under the SR2 strong setting, the average CT2 (3.78 kHz) is out of the range of the second and third formants for most vowels (Alexander, 2016). With the application of adaptive cutoffs, the formant structure of Mandarin vowel sounds was preserved well and vowel recognition accuracy remained very high. Another possible reason could be that Mandarin has a small inventory of monophthongal vowel phonemes and less crowded vowel space in comparison to many languages such as English. Even though the compression process might modify the spectral features of certain vowels [e.g., F2 of Mandarin high front vowel /i/ as shown in Yang et al. (2018)], the distorted spectral structure introduced a limited detrimental effect on recognition. Therefore, Mandarin vowel recognition was not negatively affected by the adaptive NLFC scheme, even in the strong setting.

Unlike vowel recognition, the use of adaptive NLFC significantly improved consonant recognition performance. Chen et al. (2020) reported improved Mandarin consonant recognition with SR default setting. In the present study, we observed improved accuracy for consonant recognition with SR2 default in comparison to SR2 off. It is noteworthy that the participants' average hearing loss in the present study was more severe at 4 and 8 kHz than that in the previous SR study by Chen et al. (2020). The unaided thresholds at 4 and 8 kHz were approximately 10 dB higher in the present study than those in Chen et al. (2020). The significant improvement of consonant recognition with SR2 default setting in patients with more severe hearing loss suggests a possible application of this setting in patients with a wider range of hearing loss. While the SR2 default setting significantly improved the recognition accuracy, a stronger setting did not improve the consonant recognition performance further. This finding is similar to the outcome reported in Xu et al. (2020) that a stronger compression does not ensure better recognition performance. A possible explanation is that strong compressions might cause more confusion for high-frequency consonants and thus offset the benefit of better detection. When the incoming signals contain predominantly high-frequency energy, as in the consonants, CT1 (with higher compression) will be applied. In this study, the average CT1s were 3.77 kHz (SR2 default) and 2.17 kHz (SR2 strong), suggesting that a lower cut-off frequency for SR2 could limit recognition performance. The better performance with SR2 default than SR2 strong suggested that there might be an optimal range of frequency compression for individuals with high-frequency hearing loss, as suggested by several previous studies (Johnson and Light, 2015; Scollie et al., 2016; Glista and Scollie, 2018).

Our consonant confusion analyses (**Figure 7**) showed that not all consonants were equally affected by NLFC. The positive effects of NLFC on recognition of certain Mandarin consonants was negated by deterioration of other consonants, resulting in a small overall effect of NLFC on consonant recognition. The asymmetrical confusion pattern between alveolar sounds and retroflex sounds as well as greater confusion as the compression setting changed from SR2 default to SR2 strong was consistent with the acoustic change caused by frequency compression (Yang et al., 2018). The acoustic energy of retroflex sounds /ʂ, tʂ, tʂʰ/ (sh, zh, ch) concentrates at a lower frequency region that was less affected by frequency compression. By contrast, the alveolar sounds /s, ts, tsʰ/ (s, z, c) have spectral energy distributed at a higher-frequency region. Frequency compression shifted the spectral energy of alveolar sounds to a lower-frequency region similar to retroflex sounds. The more aggressive compression setting caused greater distortion of spectral energy distribution and thus greater confusion. While both alveolar and alveolopalatal sounds are high-frequency sounds that present accessibility challenges in people with severe hearing impairments, the higher accuracy and less confusion of alveolopalatal sounds were likely associated with the phonotactic constraints in Mandarin Chinese. Mandarin /ɕ, tɕ, tɕ^{h}/ (x, j, q) is always followed by /i/, /y/, or vowels starting with these two. In the current study, the alveolopalatal sounds were followed by /ia/ which differed from other tested consonants that were followed by /a/. The distinct vowel environment likely assisted in the recognition by the hearing aid users.

Although adaptive NLFC technology helped improve patients' consonant recognition and audibility of high-frequency information in quiet, it did not provide significant perceptual benefits to sentence recognition in noise. Previous research revealed mixed findings on sentence recognition in noise with the NLFC technology (Glista et al., 2009; Perreau et al., 2013; Bentler et al., 2014; McCreery et al., 2014; Picou et al., 2015). Chen et al. (2020) reported improved sentence recognition accuracy in patients with SR on than with SR off. In the present study, sentence recognition performance did not show significant change among different SR2 settings. However, similar to Chen et al. (2020), the participants in the present study demonstrated substantial individual differences in sentence recognition. While some participants (8/29) showed a benefit of more than 5 percentage points with SR2 default or SR2 strong setting than SR2 off across all five sessions; some (8/29) had a decreased accuracy of more than 5 percentage points; the others (13/29) demonstrated a small change less than 5 percentage points. Varying degrees of hearing loss (from moderate to profound) in the 29 participants could result in different optimal compression situations. However, our correlational analyses did not find a correlation between the aided thresholds and the amount of NLFC benefits in speech recognition. Cognitive ability and level of linguistic knowledge could also potentially affect sentence recognition in noise. Although age is not the only factor related to cognitive ability and level of linguistic knowledge and we found no correlation between age and speech recognition performance in our sample, other studies have indicated that those factors might be especially important for elderly patients with profound hearing loss (Best et al., 2018; Nuesse et al., 2018; Vermeire et al., 2019). Another note is that the sentence materials used in this study were recorded by a male talker. Our data suggested that female voice tended to receive higher sound-quality ratings than the male voice. Wang and Xu (2020) also showed that Mandarin tone recognition in noise using a female voice yielded higher scores than using a male voice. It would be interesting to examine whether female voice would lead to a greater improvement for sentence recognition in NLFC conditions.

Sound quality is another core criterion in evaluating patients' willingness to accept technology and degree of comfort after using it. Some researchers reported no obvious changes in sound-quality ratings with or without NLFC (Uys et al., 2012; Parsa et al., 2013; Kirchberger and Russo, 2016; Tseng et al., 2018; Chen et al., 2020). Some researchers found that patients with moderate to severe sensorineural hearing loss showed a preference for listening to music with NLFC scheme (Uys et al., 2016). As reported in Xu et al. (2020) who used the same test materials for sound-quality ratings as the present study, no deterioration of sound quality was found between stronger settings of SR2 and SR2 off setting, which suggested a great tolerance to the adaptive NLFC. Consistent results were found in the present study. In general, participants favored SR2 default setting over SR2 off or SR2 strong settings (**Figure 8**). Furthermore, our study demonstrated that patients generally favored female voice over the other four types of materials (i.e., own voice, male voice, chirp, music) in the four perceptual categories of loudness, clarity, naturalness, and overall quality rating. By contrast, own voice had the least favored sound quality, at least for the first three sessions. Such poor rating in "own voice" is usually attributed to the occlusion effects caused by the hard ear-molds used in the present study. However, we have no direct evidence of perceived occlusion effect in our participants. It was worth noting that through periods of adaption, our participants rated their own voice significantly better in the fifth session than in their first three sessions, which indicated the importance of acclimatization for the SR2 function as further elaborated below.

Researchers have found that a period of acclimatization is necessary for the human brain to gradually learn and obtain benefits from any newly applied technologies including the NLFC scheme. Giroud et al. (2017) found that it took several weeks for hearing aid users to acclimate themselves to the new hearing aid algorithm of NLFC. Glista et al. (2009) reported that the average time for patients to adapt to the NLFC processor was approximately 10 weeks. In the present study, all participants had a period of 12 weeks to adapt to the NLFC scheme. Test session was found to be the main factor for vowel, consonant, sentence recognition, and sound-quality ratings. It should be noted that no interaction of test session by SR2 setting was evident. This suggested that the improvement might be the outcome of perceptual learning or training effects (i.e., participants' increased familiarity with the test materials and procedure) instead of auditory acclimatization, *per se*. This finding was consistent with the previous SR study (Chen et al., 2020). The present study only examined speech-recognition performance and subjective ratings for 3 months. On one hand, we observed continuous improvement in the first few weeks, indicating that the participants adapted to the new hearing aids in a very short period. On the other hand, had we extended the observation to a longer period, we might see further continuous improvement in both speech recognition and subjective quality ratings. Future studies will be necessary to demonstrate the long-term benefits of using the NLFC technology in hearing-impaired listeners. Further, one limitation of the present study is the lack of a control condition in which the SR2 off setting is used in the hearing aids for the same amount of time as the SR2 default and SR2 strong settings. With such a control condition, we would be in a better position to evaluate whether the perceptual benefits (such as in consonant recognition task and overall sound-quality ratings) resulted from a training effect or from the application of the SR2 technology itself.

CONCLUSION

In summary, the adaptive NLFC technology implemented the hearing aids provided benefit to high-frequency sound detection, as well as Mandarin consonant recognition, and sound-quality ratings in listeners with moderate-to-profound sensorineural hearing loss. As expected, vowel recognition performance with or without the adaptive NLFC algorithm was consistently good, indicating no deterioration of vowel perception using the adaptive NLFC. However, for sentence recognition in noise, the adaptive NLFC algorithm showed limited effects. Among different settings, SR2 default provided more benefits than SR2 strong setting in our group of participants with moderate to profound hearing loss. Furthermore, participants showed continuous improvement of recognition performance with increased length of SR2 use, which indicated a potential effect of auditory acclimatization.

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 the Institutional Review Boards of Ohio University and Beijing Tongren Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

XC, JR, VK, and LX conceived and designed the study. SQ, XT, HH, and JG collected the data. XW performed the statistical analyses. SQ, JY, and LX wrote the manuscript. All authors discussed the results and implications and commented on the article at all stages.

FUNDING

The study was supported, in part, by a research contract from Sonova AG.

REFERENCES

Akinseye, G. A., Dickinson, A.-M., and Munro, K. J. (2018). Is non-linear frequency compression amplification beneficial to adults and children with hearing loss? A systematic review. *Int. J. Audiol.* 57, 262–273. doi: 10.1080/14992027.2017.1420255

Alexander, J. (2013). Individual variability in recognition of frequency-lowered speech. *Semin. Hear.* 34, 086–109. doi: 10.1055/s-0033-1341346

Alexander, J. M. (2016). Nonlinear frequency compression: influence of start frequency and input bandwidth on consonant and vowel recognition. *J. Acoust. Soc. Am.* 139, 938–957. doi: 10.1121/1.4941916

Alexander, J. M., Kopun, J. G., and Stelmachowicz, P. G. (2014). Effects of frequency compression and frequency transposition on fricative and affricate perception in listeners with normal hearing and mild to moderate hearing loss. *Ear Hear.* 35, 519–532. doi: 10.1097/AUD.0000000000000040

Alexander, J. M., and Rallapalli, V. (2017). Acoustic and perceptual effects of amplitude and frequency compression on high-frequency speech. *J. Acoust. Soc. Am.* 142:908. doi: 10.1121/1.4997938

Bentler, R., Walker, E., McCreery, R., Arenas, R. M., and Roush, P. (2014). Nonlinear frequency compression in hearing aids: impact on speech and language development. *Ear Hear.* 35, e143–e152. doi: 10.1097/AUD.0000000000000030

Best, V., Ahlstrom, J. B., Mason, C. R., Roverud, E., Perrachione, T. K., Kidd, G., et al. (2018). Talker identification: effects of masking, hearing loss, and age. *J. Acoust. Soc. Am.* 143:1085. doi: 10.1121/1.5024333

Boothroyd, A., and Medwetsky, L. (1992). Spectral distribution of /s/ and the frequency response of hearing aids. *Ear Hear.* 13, 150–157. doi: 10.1097/00003446-199206000-00003

Brennan, M. A., Lewis, D., McCreery, R., Kopun, J., and Alexander, J. M. (2017). Listening effort and speech recognition with frequency compression amplification for children and adults with hearing loss. *J. Am. Acad. Audiol.* 28, 823–837. doi: 10.3766/jaaa.16158

Brennan, M. A., McCreery, R., Kopun, J., Hoover, B., Alexander, J., Lewis, D., et al. (2014). Paired comparisons of nonlinear frequency compression, extended bandwidth, and restricted bandwidth hearing aid processing for children and adults with hearing loss. *J. Am. Acad. Audiol.* 25, 983–998. doi: 10.3766/jaaa.25.10.7

Chen, X., You, Y., Yang, J., Qian, J., Lu, Q., Kuehnel, V., et al. (2020). Effects of nonlinear frequency compression on Mandarin speech and sound-quality perception in hearing-aid users. *Int. J. Audiol.* 59, 524–533. doi: 10.1080/14992027.2020.1761035

Ching, T. Y., Day, J., Zhang, V., Dillon, H., Van Buynder, P., Seeto, M., et al. (2013). A randomized controlled trial of nonlinear frequency compression versus conventional processing in hearing aids: speech and language of children at three years of age. *Int. J. Audiol.* 52 Suppl 2, S46–S54. doi: 10.3109/14992027.2013.858281

Ching, T. Y. C., and Rattanasone, N. X. (2015). Intelligibility of speech produced by children with hearing loss: Conventional amplification versus nonlinear frequency compression in hearing aids. *Commun. Disord. Deaf. Study Hear. Aids* 3, 135-1–131-8.

Giroud, N., Lemke, U., Reich, P., Matthes, K. L., and Meyer, M. (2017). The impact of hearing aids and age-related hearing loss on auditory plasticity across three months – An electrical neuroimaging study. *Hear Res.* 353, 162–175. doi: 10.1016/j.heares.2017.06.012

Glista, D., Hawkins, M., Bohnert, A., Rehmann, J., Wolfe, J., and Scollie, S. (2017). The effect of adaptive nonlinear frequency compression on phoneme perception. *Am. J. Audiol.* 26, 531–542. doi: 10.1044/2017_AJA-17-0023

Glista, D., Hawkins, M., Vaisberg, J. M., Pourmand, N., Parsa, V., and Scollie, S. (2019). Sound quality effects of an adaptive nonlinear frequency compression processor with normal-hearing and hearing-impaired listeners. *J. Am. Acad. Audiol.* 30, 552–563. doi: 10.3766/jaaa.16179

Glista, D., and Scollie, S. (2018). The Use of frequency lowering technology in the treatment of severe-to-profound hearing loss: a review of the literature and candidacy considerations for clinical application. *Semin. Hear.* 39, 377–389. doi: 10.1055/s-0038-1670700

Glista, D., Scollie, S., Bagatto, M., Seewald, R., Parsa, V., and Johnson, A. (2009). Evaluation of nonlinear frequency compression: clinical outcomes. *Int. J. Audiol.* 48, 632–644. doi: 10.1080/14992020902971349

Hopkins, K., Khanom, M., Dickinson, A. M., and Munro, K. J. (2014). Benefit from non-linear frequency compression hearing aids in a clinical setting: the effects of duration of experience and severity of high-frequency hearing loss. *Int. J. Audiol.* 53, 219–228. doi: 10.3109/14992027.2013.873956

Johnson, E. E., and Light, K. C. (2015). A Patient-centered, provider-facilitated approach to the refinement of nonlinear frequency compression parameters based on subjective preference ratings of amplified sound quality. *J. Am. Acad. Audiol.* 26, 689–702. doi: 10.3766/jaaa.14053

Kirchberger, M., and Russo, F. A. (2016). Harmonic frequency lowering: effects on the perception of music detail and sound quality. *Trends Hear.* 20, 1–12. doi: 10.1177/2331216515626131

Mao, Y., Yang, J., Hahn, E., and Xu, L. (2017). Auditory perceptual efficacy of nonlinear frequency compression used in hearing aids: a review. *J. Otol.* 12, 97–111. doi: 10.1016/j.joto.2017.06.003

McCreery, R. W., Alexander, J., Brennan, M. A., Hoover, B., Kopun, J., and Stelmachowicz, P. G. (2014). The influence of audibility on speech recognition with nonlinear frequency compression for children and adults with hearing loss. *Ear Hear.* 35, 440–447. doi: 10.1097/AUD.0000000000000027

McDermott, H. J. (2011). A technical comparison of digital frequency-lowering algorithms available in two current hearing aids. *PLoS One* 6:e22358. doi: 10.1371/journal.pone.0022358

Moeller, M. P., Hoover, B., Putman, C., Arbataitis, K., Bohnenkamp, G., Peterson, B., et al. (2007). Vocalizations of infants with hearing loss compared with infants with normal hearing: part I–phonetic development. *Ear Hear.* 28, 605–627. doi: 10.1097/aud.0b013e31812564ab

Monson, B. B., Hunter, E. J., Lotto, A. J., and Story, B. H. (2014). The perceptual significance of high-frequency energy in the human voice. *Front. Psychol.* 5:587. doi: 10.3389/fpsyg.2014.00587

Moore, B. C. (2001). Dead regions in the cochlea: diagnosis, perceptual consequences, and implications for the fitting of hearing AIDS. *Trends Amplif.* 5, 1–34. doi: 10.1177/108471380100500102

Moore, B. C. (2016). A review of the perceptual effects of hearing loss for frequencies above 3 kHz. *Int. J. Audiol.* 55, 707–714. doi: 10.1080/14992027.2016.1204565

Moore, B. C. J. (2004). Dead regions in the cochlea: conceptual foundations, diagnosis, and clinical applications. *Ear Hear.* 25, 98–116. doi: 10.1097/01.aud.0000120359.49711.d7

Nuesse, T., Steenken, R., Neher, T., and Holube, I. (2018). Exploring the link between cognitive abilities and speech recognition in the elderly under different listening conditions. *Front. Psychol.* 9:678. doi: 10.3389/fpsyg.2018.00678

Parsa, V., Scollie, S., Glista, D., and Seelisch, A. (2013). Nonlinear frequency compression: effects on sound quality ratings of speech and music. *Trends Amplif.* 17, 54–68. doi: 10.1177/1084713813480856

Perreau, A. E., Bentler, R. A., and Tyler, R. S. (2013). The contribution of a frequency-compression hearing aid to contralateral cochlear implant performance. *J. Am. Acad. Audiol.* 24, 105–120. doi: 10.3766/jaaa.24.2.4

Picou, E. M., Marcrum, S. C., and Ricketts, T. A. (2015). Evaluation of the effects of nonlinear frequency compression on speech recognition and sound quality for adults with mild to moderate hearing loss. *Int. J. Audiol.* 54, 162–169. doi: 10.3109/14992027.2014.961662

Rehmann, J., Jha, S., and Allegro Baumann, S. (2016). *Sound Recover2—The first Adaptive Frequency Compression Algorithm.* Phonak Insight Paper. Stäfa: Phonak.

Scollie, S., Glista, D., Seto, J., Dunn, A., Schuett, B., Hawkins, M., et al. (2016). Fitting frequency-lowering signal processing applying the american academy of audiology pediatric amplification guideline: updates and protocols. *J. Am. Acad. Audiol.* 27, 219–236. doi: 10.3766/jaaa.15059

Simpson, A. (2009). Frequency-lowering devices for managing high-frequency hearing loss: a review. *Trends Amplif.* 13, 87–106. doi: 10.1177/1084713809336421

Simpson, A., Hersbach, A. A., and McDermott, H. J. (2005). Improvements in speech perception with an experimental nonlinear frequency compression hearing device. *Int. J. Audiol.* 44, 281–292. doi: 10.1080/14992020500060636

Souza, P. E., Arehart, K. H., Kates, J. M., Croghan, N. B. H., and Gehani, N. (2013). Exploring the limits of frequency lowering. *J. Speech Lang Hear. Res.* 56:1349. doi: 10.1044/1092-4388(2013/12-0151)

Stelmachowicz, P. G., Pittman, A. L., Hoover, B. M., and Lewis, D. E. (2002). Aided perception of /s/ and /z/ by hearing-impaired children. *Ear Hear.* 23, 316–324. doi: 10.1097/00003446-200208000-00007

Stelmachowicz, P. G., Pittman, A. L., Hoover, B. M., Lewis, D. E., and Moeller, M. P. (2004). The importance of high-frequency audibility in the speech and language development of children with hearing loss. *Arch. Otolaryngol. Head Neck Surg.* 130, 556–562. doi: 10.1001/archotol.130.5.556

Thornton, A. R., and Raffin, M. J. (1978). Speech-discrimination scores modeled as a binomial variable. *J. Speech Hear. Res.* 21, 507–518. doi: 10.1044/jshr.2103.507

Tseng, W. H., Hsieh, D. L., Shih, W. T., and Liu, T. C. (2018). Extended bandwidth nonlinear frequency compression in Mandarin-speaking hearing-aid users. *J. Formos Med. Assoc.* 117, 109–116. doi: 10.1016/j.jfma.2017.01.013

Uys, M., Latzel, M., and Van Dyk, M. (2016). The Performance of music and speech with different frequency lowering devices. *Commun. Disord. Deaf. Study Hear. Aids* 4:163. doi: 10.4172/2375-4427.1000163

Uys, M., Pottas, L., Vinck, B., and van Dijk, C. (2012). The influence of non-linear frequency compression on the perception of music by adults with a moderate to sever hearing loss: subjective impressions. *S. Afr. J. Commun. Disord.* 59, 53–67.

Vermeire, K., Knoop, A., De Sloovere, M., Bosch, P., and van den Noort, M. (2019). Relationship between working memory and speech-in-noise recognition in young and older adult listeners with age-appropriate hearing. *J. Speech Lang. Hear. Res.* 62, 3545–3553. doi: 10.1044/2019_JSLHR-H-18-0307

Wang, X., and Xu, L. (2020). Mandarin tone perception in multiple-talker babbles and speech-shaped noise. *J. Acoust. Soc. Am.* 147, EL307–EL313. doi: 10.1121/10.0001002

Warton, D. I., and Hui, F. K. C. (2011). The arcsine is asinine: the analysis of proportions in ecology. *Ecology* 92, 3–10. doi: 10.1890/10-0340.1

Wolfe, J., Duke, M., Schafer, E. C., Rehmann, J., Jha, S., Allegro Baumann, S., et al. (2017). Preliminary evaluation of a novel non-linear frequency compression scheme for use in children. *Int. J. Audiol.* 56, 976–988. doi: 10.1080/14992027.2017.1358467

Wolfe, J., John, A., Schafer, E., Nyffeler, M., Boretzki, M., and Caraway, T. (2010). Evaluation of nonlinear frequency compression for school-age children with moderate to moderately severe hearing loss. *J. Am. Acad. Audiol.* 21, 618–628. doi: 10.3766/jaaa.21.10.2

Wolfe, J., John, A., Schafer, E., Nyffeler, M., Boretzki, M., Caraway, T., et al. (2011). Long-term effects of non-linear frequency compression for children with moderate hearing loss. *Int. J. Audiol.* 50, 396–404. doi: 10.3109/14992027.2010.551788

Wong, L. L., Soli, S. D., Liu, S., Han, N., and Huang, M. W. (2007). Development of the mandarin hearing in noise test (MHINT). *Ear. Hear.* 28, 70S–74S. doi: 10.1097/AUD.0b013e31803154d0

Xu, L., Voss, S. C., Yang, J., Wang, X., Lu, Q., Rehmann, J., et al. (2020). Speech perception and sound-quality rating with an adaptive nonlinear frequency compression algorithm in Mandarin-speaking hearing aid users. *J. Am. Acad. Audiol.* 31, 590–598. doi: 10.1055/s-0040-1709450

Xu, L., Xi, X., Patton, A., Wang, X., Qi, B., and Johnson, L. (2021). A cross-language comparison of sentence recognition using American English and Mandarin Chinese HINT and AzBio sentences. *Ear Hear.* 42, 405–413. doi: 10.1097/AUD.0000000000000938

Yang, J., Qian, J., Chen, X., Kuehnel, V., Rehmann, J., von Buol, A., et al. (2018). Effects of nonlinear frequency compression on the acoustic properties and recognition of speech sounds in Mandarin Chinese. *J. Acoust. Soc. Am.* 143:1578. doi: 10.1121/1.5027404

Zhang, T., Dorman, M. F., Gifford, R., and Moore, B. C. (2014). Cochlear dead regions constrain the benefit of combining acoustic stimulation with electric stimulation. *Ear Hear* 35, 410–417. doi: 10.1097/AUD.0000000000000032

Zhang, V. W., Ching, T. Y., Van Buynder, P., Hou, S., Flynn, C., Burns, L., et al. (2014). Aided cortical response, speech intelligibility, consonant perception and functional performance of young children using conventional amplification or nonlinear frequency compression. *Int J Pediatr Otorhinolaryngol* 78, 1692–1700. doi: 10.1016/j.ijporl.2014.07.022

Conflict of Interest: XT, HH, and JG were employed by the company Sonova China, Shanghai. JR and VK were employed by the company Sonova AG Switzerland.

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.

frontiers in Digital Health

METHODS
published: 13 August 2021
doi: 10.3389/fdgth.2021.723348

Bayesian Pure-Tone Audiometry Through Active Learning Under Informed Priors

Marco Cox[1]* and *Bert de Vries*[1,2]

[1] Signal Processing Systems Group, Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, Netherlands, [2] GN Hearing, Eindhoven, Netherlands

Pure-tone audiometry—the process of estimating a person's hearing threshold from "audible" and "inaudible" responses to tones of varying frequency and intensity—is the basis for diagnosing and quantifying hearing loss. By taking a probabilistic modeling approach, both optimal tone selection (in terms of expected information gain) and hearing threshold estimation can be derived through Bayesian inference methods. The performance of probabilistic model-based audiometry methods is directly linked to the quality of the underlying model. In recent years, Gaussian process (GP) models have been shown to provide good results in this context. We present methods to improve the efficiency of GP-based audiometry procedures by improving the underlying model. Instead of a single GP, we propose to use a GP mixture model that can be conditioned on side-information about the subject. The underlying idea is that one can typically distinguish between different types of hearing thresholds, enabling a mixture model to better capture the statistical properties of hearing thresholds among a population. Instead of modeling all hearing thresholds by a single GP, a mixture model allows specific types of hearing thresholds to be modeled by independent GP models. Moreover, the mixing coefficients can be conditioned on side-information such as age and gender, capturing the correlations between age, gender, and hearing threshold. We show how a GP mixture model can be optimized for a specific target population by learning the parameters from a data set containing annotated audiograms. We also derive an optimal tone selection method based on greedy information gain maximization, as well as hearing threshold estimation through Bayesian inference. The proposed models are fitted to a data set containing roughly 176 thousand annotated audiograms collected in the Nordic countries. We compare the predictive accuracies of optimized mixture models of varying sizes with that of an optimized single-GP model. The usefulness of the optimized models is tested in audiometry simulations. Simulation results indicate that an optimized GP mixture model can significantly outperform an optimized single-GP model in terms of predictive accuracy, and leads to significant increases the efficiency of the resulting Bayesian audiometry procedure.

Keywords: active learning, audiometry, Bayesian inference, Gaussian process, machine learning, probabilistic modeling

OPEN ACCESS

Edited by:
Dennis L. Barbour, Washington University in St. Louis, United States

Reviewed by:
Ming Huang, Nara Institute of Science and Technology (NAIST), Japan
Herman Myburgh, University of Pretoria, South Africa

***Correspondence:**
Marco Cox
m.g.h.cox@tue.nl

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: *10 June 2021*
Accepted: *23 July 2021*
Published: *13 August 2021*

Citation:
Cox M and de Vries B (2021) Bayesian Pure-Tone Audiometry Through Active Learning Under Informed Priors. Front. Digit. Health 3:723348. doi: 10.3389/fdgth.2021.723348

1. INTRODUCTION

Hearing loss is typically represented by an *audiogram*, which depicts the *hearing threshold* (HT) (i.e., the lowest sound intensity level that can still be perceived) at a set of standard frequencies ranging from 125 Hz to 8 kHz. The hearing threshold levels are usually measured through a process called *pure-tone audiometry* (PTA) (1). In PTA, the subject provides a series of "audible" and "inaudible" responses to pure-tones of various frequencies and intensities, and those responses are used to estimate the HT at the required frequencies. Multiple protocols for selecting tone frequencies and intensities have been developed, the most common of which is a staircase "up 5 dB–down 10 dB" approach known as the Hughson-Westlake protocol (2).

We present a method for performing hearing threshold estimation with optimal efficiency in terms of number of interactions required to achieve a given accuracy level. The method is based on a full probabilistic treatment of the estimation problem. At its core is a probabilistic model which captures a range of statistical properties of the hearing threshold. In short, this probabilistic model encodes a probability distribution over hearing thresholds, and describes in a probabilistic way how the subject's response to a stimulus ("audible" or "inaudible") is generated. Specifically, we propose a model based on a weighted mixture of Gaussian processes (3). Based on this model, our method works in the following way:

1. Optimize the model parameters with respect to a data set of annotated audiometric records from a large number of people. The goal of this step is to train the model to match the statistical distribution of hearing thresholds in the data set as well as possible. This optimization procedure is independent of the subject, and yields a non-personalized prior model.
2. Optionally condition the prior model on side-information from the subject, such as age an gender, to improve the accuracy of the predictive distribution.
3. Determine which stimulus to present such that the "audible" or "inaudible" response to it provides the maximum amount of information about the hearing threshold. Given the model, this stimulus can be derived theoretically using information-theoretic criteria.
4. Present the optimal stimulus to the subject and collect the response.
5. Update the probabilistic model based on the response. This step involves Bayesian inference, and combines the non-personalized model with the data (i.e., the subject's responses to stimuli) to obtain the posterior distribution of the model. The result includes a revised estimate of the hearing threshold including uncertainty bands.
6. Go to step 3 and repeat until the hearing threshold estimate is sufficiently accurate.

The process of repeatedly selecting the most informative next stimulus and updating the probabilistic model based on the response is called an *active learning loop* (4), and it has been shown to significantly reduce the total number of required test tones to reach a certain accuracy level (5). However, the success of the active learning approach hinges on the quality of the probabilistic model at hand. If the model is flexible and accurate enough, it should theoretically outperform any empirical method based on heuristics. If on the other hand the model fails to capture important aspects of the underlying dynamics, the quality of the method will suffer. It is important to note that once the model has been developed and fitted to a data set, the remaining steps 2–6 basically have unique optimal solutions that can be derived theoretically and just have to be translated correctly into an algorithm.

In prior work, the Gaussian process (GP) model has been used to model the hearing threshold as a continuous function of frequency (5–7). The parameters of the GP were typically chosen empirically. In this work we propose the use of a more flexible class of models, namely weighted mixtures of GPs. The rationale behind this is that a lot of hearing thresholds have one of several typical shapes (8). By capturing the statistics of these distinct shapes by separate GP models, the quality of model might be improved significantly. Additionally, the proposed model can be conditioned on side-information such as age and gender of the subject, further increasing the predictive accuracy. The trained and possibly conditioned model can be viewed as an *informed prior* since it is based on information about the target population (through the training set) and the available side-information about the subject (age and/or gender). We show that it is possible to derive theoretical solutions for optimal tone selection and Bayesian model inference under the more flexible model.

The application of Bayesian methods and information-theoretic criteria to obtain information-efficient audiometry procedures has a long history (9–12). Most of the early methods relied on probabilistic models of the HT at individual frequencies, or captured dependencies among a discrete set of frequencies. More recently, GP-based methods with a continuous frequency scale have been introduced (5–7, 13) and validated experimentally (14, 15). The aim of this work is to improve upon those methods by increasing the quality of the underlying model, both by increasing the complexity of the model and by fitting the model to data.

In the remainder, we mathematically specify the proposed model and derive an algorithm to fit its parameters to a set of annotated audiograms. Next, we outline the (approximate) Bayesian inference algorithm required to update the model, as well as the algorithm for selecting the optimal next stimulus. The proposed model is trained on a large data set containing $\sim$176 thousand audiograms annotated with age en gender. The resulting Bayesian PTA method based on an informed prior is tested through various simulations.

2. MATERIALS AND EQUIPMENT

All methods and models have been implemented in the Julia programming language (16). TensorFlow (16) is used as the computational back-end for fitting the hearing threshold models.

2.1. Data Source

The results reported in this paper related to model learning and simulations are based on a proprietary data set. This anonymized data set contains the ages, genders, and audiograms of both ears of 88,237 people from the Nordic countries who visited an audiologist. In total, the data set contains 176,474 audiograms annotated with age and gender. The audiograms specify the hearing thresholds with a resolution of 5 dB on (subsets of) the following frequencies: 125, 250, 500, 750, 1,000, 1,500, 2,000, 3,000, 4,000, 6,000, 8,000 Hz. The presented methods are independent of the specific data set that is used for model learning and simulations.

3. METHODS

We first introduce the probabilistic model on which our method is based. Next, we describe how the model parameters can be "learned" from a data set of annotated audiograms. Subsequently, we show how the model is used to estimate the HT from a set of responses to stimuli by performing Bayesian inference. Finally, we illustrate how the model enables the identification the most informative next stimulus given the responses so far.

3.1. Probabilistic Hearing Loss Model

The complete probabilistic model consists of two parts: a user response model and a hearing threshold model. We introduce these components separately, and then combine them to obtain the complete model.

3.1.1. User Response Model

A PTA procedure is assumed to involve a sequence of *trials*. A trial consists of a single pure-tone stimulus of a certain frequency f in Hertz and intensity level h, together with a binary response label y indicating whether the stimulus was audible or inaudible to the subject. The intensity level h is expressed in dB *hearing level* (dB-HL), which is a relative sound pressure level in which 0 dB-HL corresponds to the hearing threshold of the average person with no hearing impairment. The subject's response depends on the presented stimulus, and is encoded in the following way:

$$y(f,h) = \begin{cases} +1, & \text{if } (f,h) \text{ is audible,} \\ -1, & \text{otherwise.} \end{cases} \tag{1}$$

A trial is represented by a tuple containing all relevant quantities: (f, h, y).

By definition, stimuli near the subject's hearing threshold will not yield consistent responses. To capture the uncertainty in the response generating process, a probabilistic user response model is required. This model describes how a user determines their response to a stimulus if their "true" HT were known. In our model, the "true" HT is assumed to be evaluated under white Gaussian noise $\mathcal{N}(0, \sigma_p^2)$. In other words, the model assumes that a stimulus is audible if and only if its intensity exceeds the subject's "true" HT at the corresponding frequency by some random margin. This leads to the following formal response model:

$$\begin{aligned} P(y \mid f, h) &= \Pr\{y \cdot (h - \mathcal{HT}(f)) > \mathcal{N}(0, \sigma_p^2)\} \\ &= \int_{-\infty}^{y\cdot(h-\mathcal{HT}(f))} \mathcal{N}(h' \mid 0, \sigma_p^2)\mathrm{d}h' \\ &= \Phi\left(\frac{y \cdot (h - \mathcal{HT}(f))}{\sigma_p}\right), \end{aligned} \tag{2}$$

where $\mathcal{HT}$ denotes the unknown "true" hearing threshold as a function of frequency. Φ is the cumulative density function of the standard normal distribution. Since $y \in \{-1, +1\}$, probability distribution $P(y \mid f, h)$ is a Bernoulli distribution. For simplicity, the perceptual noise variance parameter σ_p^2 is independent of frequency, but this can easily be relaxed. The value of σ_p^2 can either be learned from a data set of actual pure-tone responses, or it can be tuned empirically.

3.1.2. Hearing Threshold Model

The hearing threshold model specifies a probability distribution over hearing thresholds. Instead of treating hearing thresholds at distinct frequencies as independent quantities, we assume the HT to be a smooth function of frequency. Since the human auditory perception of frequency shifts is non-linear, it makes sense to model the HT in a psycho-acoustical space that resembles the human perception better than the linear frequency domain. Technically, the psycho-acoustical space is a warped frequency space in which the distance between frequencies better resembles the human perception of frequency shifts. This is useful since our model aims to exploit properties of the HT that are more natural to interpret on a psycho-acoustical scale. Various psycho-acoustical scales are being used in the field of audiology, such as the "Mel" scale, the "semitone" scale, and the "Bark" scale (1). All of these scales roughly match the semi-logarithmic frequency axis in typical audiogram plots.

For the HT model it is not very important which specific frequency transformation is used, as long as the transformation is invertible. For the Bark scale, multiple transformations with varying degrees of complexity and accuracy have been proposed (17, 18). In our model we adopt the simple Bark transformation described in (18):

$$\text{bark}(f) \triangleq 6 * \sinh^{-1}\left(\frac{f}{600}\right), \tag{3}$$

with f in Hz. For notational convenience, we will use x to denote a transformed frequency and in the remainder of this paper we use $\mathcal{HT}(x)$ to denote the HT as a function of the transformed frequency.

We obtain a probabilistic model for hearing thresholds by assuming that a HT is a smooth function of transformed frequency, drawn from a Gaussian process (GP). A GP is a probability distribution over continuous functions, and it is fully characterized by a covariance function and a mean function (3). The approach of modeling the HT by a GP has already been proposed before, for example in (6) and (7). However, instead of assuming the HT to be generated by a single GP, we assume the HT to be generated by one of C independently parameterized GPs. The main idea behind this choice is that HTs tend to be of one of several distinct types in terms of location, slope, and smoothness. Mixing multiple GPs has two important benefits. Firstly, it should enable the model to capture the statistical properties of distinct typical HT types with a higher resolution than a single GP, leading to a more accurate model. Secondly, it allows the selection of the individual GP to be dependent on side-information such as age and gender. Intuitively, this means that the individual GP components in the model could capture different HT types corresponding (for example) to mild hearing loss, typical old-age HTs, and "cookie-bite hearing loss." For example, if the subject's age is available, the model could exploit it by adjusting the a-priori probabilities of the different HT types. For age 80 we would expect the GP corresponding to "typical old-age HTs" to get a higher relative probability compared to age 40. By having a more complex model, we hope to leverage large data sets of annotated audiometric records to be able to *learn* accurate models that capture as many statistical properties as possible.

3.1.3. Complete Probabilistic Model

Our model assumes responses to stimuli to be generated in the following way:

1. Randomly select a GP component $c \in [1, \ldots, C]$ from a categorical distribution whose parameters depend on any available side-information $\mathcal{I}$ about the subject: $c \sim \text{Categorical}(\alpha(\mathcal{I}))$.
2. Randomly generate an HT curve from the selected GP: $t \sim \mathcal{GP}(m_c, k_c)$. Parameters m_c and k_c respectively denote the mean function and covariance function or *kernel* of the GP. The mean function is assumed to be a third-order polynomial, and the covariance function is a squared exponential kernel (3). Note that t is a continuous, real-valued function of transformed frequency. The choice for the squared exponential kernel follows prior work (6, 7), and is based on the idea of modeling the HT as a smooth, continuous function of frequency.
3. For each stimulus in the procedure, randomly generate the response based on t according to the response model from Equation (2).

$\mathcal{I}$ is a set of discrete features, in this case corresponding to the subject's age and gender. The age can be unspecified or an integer between 0 and 120, and gender can be unspecified, female or male.

Formally, this leads to the following generative process for a PTA procedure involving N trials on the same subject:

$$c \sim \text{Categorical}(\alpha(\mathcal{I})), \tag{4a}$$

$$t \mid c \sim \mathcal{GP}(m_c, k_c), \tag{4b}$$

$$\forall i \in [1, \ldots, N] : y_i \mid x_i, h_i, t \sim \text{Bernoulli}\left[\Phi\left(\frac{h_i - t(x_i)}{\sigma_p}\right)\right]. \tag{4c}$$

3.2. Model Learning

The model from Equation (4) includes a significant number of parameters, specifically:

- C: the number of individual GPs in the mixture model.
- $\alpha(\mathcal{I})$: the probabilities of individual GPs as a function on side-information $\mathcal{I}$. Note that $\alpha(\mathcal{I})$ is a vector whose elements sum to 1, and can be interpreted as the conditional mixing weights of the individual GPs.
- $m_1, \ldots, m_C$: the mean functions of the GPs. The mean functions are constrained to be third-order polynomials in transformed frequency space.
- $k_1, \ldots, k_C$: the covariance functions of the GPs, constrained to be squared exponential kernels. A squared exponential kernel is parameterized by a variance parameter and a length-scale parameter. The variance parameter regulates the width of the GP's uncertainty bands around the mean function, and the length-scale parameter regulates the smoothness of the GP. For notational convenience, we use $\theta_1, \ldots, \theta_C$ to denote the parameters of the covariance functions.
- σ_p: the standard deviation of the perceptual noise under which the HT is evaluated.

In general, using a complex model only makes sense if its parameters can be "learned" from a data set. In this context, "learning" means optimizing the parameters such that the model captures the statistics of the data in the data set as well as possible. In other words: "learning" the model implicitly means extracting as much relevant information as possible from the data set, and storing it in the model parameters. We provide an outline for how the model parameters can be "learned" through maximum likelihood estimation.

Since increasing the number of mixture components is guaranteed to lead to a more accurate model (at the cost of more complexity), we propose to optimize C empirically. Additionally, σ_p will also be chosen empirically since it is not possible to learn σ_p from a data set that consists of (annotated) audiograms without actual responses to individual stimuli. The remaining parameters can be optimized to a data set containing audiograms annotated with optional side-information, for example by using maximum likelihood estimation (MLE). In MLE, the parameters are tuned by maximizing the likelihood that the model assigns to the data.

Assume we have a data set containing audiograms that are optionally annotated with side information. Each audiogram defines the hearing thresholds of a subject's ear at a fixed set of standard audiometric frequencies $\mathcal{F}$, for example $\mathcal{F} = \{250, 500, 750, 1,000, 1,500, 2,000, 3,000, 4,000, 6,000, 8,000\}$ Hz. Since the audiograms are only defined at a discrete set of frequencies, the "GP mixture" distribution of infinite dimensionality reduces to a "Gaussian mixture" distribution of finite dimensionality $\|\mathcal{F}\|$. By exploiting this property, MLE can be performed in three steps:

1. Perform MLE of a Gaussian mixture model (GMM), for example using the well-known expectation-maximization (EM) algorithm (19).
2. Optimize $m_1, \ldots, m_C$ and $\theta_1, \ldots, \theta_C$ by minimizing the Kullback-Leibler divergence between the predictive distributions of the GP mixture and the GMM from step 1. In this step, the parameters of the GP mixture are tuned such that its predictive distribution matches that of the GMM at

the discrete set of frequencies $\mathcal{F}$. The optimization can be implemented by writing the Kullback-Leibler divergence in a framework that supports automatic differentiation, and then using the automatically calculated gradients to perform a gradient-based optimization. We implemented this step in TensorFlow (20), and used standard gradient descent to perform the optimization.
3. Implement $\alpha(\mathcal{I})$ by nearest-neighbor regression. For every audiogram that contains side-information, the posterior mixing coefficients are calculated using Bayesian inference, and stored in a lookup table indexed by $\mathcal{I}$, averaging over any duplicate entries. The value of $\alpha(\mathcal{I})$ is then obtained by performing a nearest-neighbor lookup. This approach works well if the data set is large relative to the cardinality of $\mathcal{I}$.

By optimizing the model parameters to a certain data set, the model will assign higher relative probabilities to HT curves that appear more frequently in the data set. This bias has a direct effect on the performance of the HT estimation and optimal stimulus selection methods based on the model. The more representative the data set used for learning is for the population on which the methods are applied, the better the performance will be.

3.3. Hearing Threshold Estimation Through Bayesian Inference

Given the (optimized) model and a collection of trials, the HT can be estimated by performing Bayesian inference. In Bayesian inference, the distributions of all random variables in the model are updated to reflect the information in the data, in this case a collection of trials. The result is a posterior probability distribution over the subject's HT curve, including uncertainty bands. Let $\mathcal{D} = [(x_1, h_1, y_1), \ldots, (x_N, h_N, y_N)]$ denote a data set containing N trials of the same subject. Applying Bayes' rule to the model from Equation (4) yields the following posterior distribution for HT t:

$$p(t \mid \mathcal{D}) \propto p(t) \cdot p(\mathcal{D} \mid t) \tag{5a}$$

$$= \sum_{c=1}^{C} \alpha_c(\mathcal{I}) \cdot \mathcal{GP}(t \mid m_c, k_c) \cdot p(\mathcal{D} \mid t) \tag{5b}$$

$$= \sum_{c=1}^{C} \alpha_c(\mathcal{I}) \cdot \frac{1}{A_c} \cdot p_c(t \mid \mathcal{D}), \tag{5c}$$

where $p_c(t \mid \mathcal{D})$ denotes the posterior distribution of t under mixture component c, and A_c are scaling factors to satisfy the second equality. This means that the posterior distribution of t is a weighted mixture of the posterior distributions of the individual GPs. Thanks to this property, Bayesian inference can be implemented in two steps:

1. Perform Bayesian inference for all C mixture components separately, yielding $p_c(t \mid \mathcal{D})$ and A_c.
2. Combine the individual posterior GPs according to Equation (5c).

Under our model, exact Bayesian inference in step 1 is intractable. However, various techniques are available to achieve approximate Bayesian inference, including variational Bayesian inference, Laplace inference and expectation propagation (3). Since a single component from our mixture model resembles a standard GP probit classifier, various GP software libraries can perform the approximate Bayesian inference from step 1 out-of-the-box. A complete derivation of an approximate Bayesian inference algorithm for a single-component using the Laplace method is available in (6).

Using Equation (5c), the posterior distribution of t can be written as a weighed sum of the individual posterior GPs:

$$p(t \mid \mathcal{D}) \propto \sum_{c=1}^{C} \alpha_c(\mathcal{I}) \cdot \frac{1}{A_c} \cdot p_c(t \mid \mathcal{D}) \tag{6a}$$

$$= \sum_{c=1}^{C} \pi_c \cdot p_c(t \mid \mathcal{D}), \tag{6b}$$

where π_c is the (unnormalized) posterior mixing weight for component c. Once the individual (approximate) posteriors have been obtained, π_c can be calculated according to:

$$\pi_c = \frac{\alpha_c(\mathcal{I})}{A_c}, \tag{7a}$$

$$A_c = \left(\int p(\mathcal{D} \mid f) p_c(f) \mathrm{d}f \right)^{-1}. \tag{7b}$$

A_c is known as the *marginal likelihood* of the data under component c, and it is usually returned by the software that implements the approximate inference.

3.4. Optimal Stimulus Selection

A central aspect of our Bayesian PTA method is to leverage an optimized probabilistic model of the procedure to repeatedly present the stimulus that will yield the most informative response. Information theory provides a fundamental mathematical framework to enable this. Given a (posterior) probabilistic model, the expected information content of a response about a variable in the model—in our case the HT—can be expressed analytically, allowing the input leading to the response to be optimized with respect to the expected information gain. The approach of actively selecting inputs to trigger responses to learn from is known in the literature as "active learning." This section provides the derivation of an optimal stimulus selection procedure based on the "Bayesian Active Learning by Disagreement" (BALD) framework from (21).

Given a set of trials $\mathcal{D}$ and posterior distribution $p(t \mid \mathcal{D})$, the goal is to find the stimulus (x_*, h_*) that "maximizes the decrease in expected posterior entropy" of t (21):

$$(x_*, h_*) = \arg\max_{(x,h)} \left(\mathrm{H}[t \mid x, h, \mathcal{D}] - \mathbb{E}_{y \sim P(y \mid x, h, \mathcal{D})} \mathrm{H}[t \mid y, x, h, \mathcal{D}] \right). \tag{8a}$$

In this expression, $\mathrm{H}[A \mid B]$ represents the Shannon entropy[1] of A given B. In the remainder of this section we work out

[1]Shannon entropy is a measure of uncertainty about the value of a random variable: $\mathrm{H}[A] \triangleq \mathbb{E}[-\log p(A)]$. A conditional entropy corresponds to the entropy of a conditional probability distribution: $\mathrm{H}[A \mid B] \triangleq \mathbb{E}[-\log p(A \mid B)]$.

the objective function under our model. The end result of the derivation can be found in Equation (15).

Inspection of the expression to be maximized reveals that is equal to the mutual information[2] of the hearing threshold t and the binary response y to a stimulus (x, h) under the posterior distribution: $\mathrm{I}[t; y \mid x, h, \mathcal{D}]$. Because mutual information is symmetric, t and y can be exchanged to get an equivalent expression that is easier to evaluate:

$$(x_*, h_*) = \arg\max_{(x,h)} \left(\mathrm{H}[t \mid x, h, \mathcal{D}] - \mathbb{E}_{y\sim P(y|x,h,\mathcal{D})}\mathrm{H}[t \mid y, x, h]\right) \tag{9a}$$

$$= \arg\max_{(x,h)} \mathrm{I}[t; y \mid x, h, \mathcal{D}] \tag{9b}$$

$$= \arg\max_{(x,h)} \mathrm{I}[y; t \mid x, h, \mathcal{D}] \tag{9c}$$

$$= \arg\max_{(x,h)} \left(\mathrm{H}[y \mid x, h, \mathcal{D}] - \mathbb{E}_{t\sim p(t|\mathcal{D})}\mathrm{H}[y \mid t, x, h]\right). \tag{9d}$$

Since y is a binary random variable, the entropy terms in Equation (9d) reduce to binary entropy terms[3].

The first term in Equation (9d) is obtained by writing out the binary entropy of the posterior predictive distribution of y, given by:

$$P(y \mid x, h; \mathcal{D}) = \int P(y \mid t_x, h) p(t_x \mid \mathcal{D}) \mathrm{d}t \tag{10a}$$

$$= \int P(y \mid t_x, h) \sum_{c=1}^{C} \pi_c \cdot p_c(t_x \mid \mathcal{D}) \mathrm{d}t \tag{10b}$$

$$= \sum_{c=1}^{C} \pi_c \int P(y \mid t_x, h) p_c(t_x \mid \mathcal{D}) \mathrm{d}t, \tag{10c}$$

where t_x denotes function value $t(x)$. Since $p_c(t \mid \mathcal{D})$ is a posterior GP, $p_c(t_x \mid \mathcal{D})$ is approximated by a Gaussian. Let the Gaussian approximate posterior distribution of $t(x)$ under component c be defined as:

$$p_c(t_x \mid \mathcal{D}) \approx \mathcal{N}\left(t_x \mid \mu_c, \sigma_c^2\right), \tag{11}$$

where μ_c and σ_c are returned by the GP approximate inference engine. Substituting Equation (11) and (4c) in Equation (10) and evaluating the integral yields[4]:

$$P(y \mid x, h; \mathcal{D}) \approx \sum_{c=1}^{C} \pi_c \int \Phi\left(\frac{y \cdot (h - t_x)}{\sigma_p}\right) \cdot \mathcal{N}(t_x \mid \mu_c, \sigma_c^2) \mathrm{d}t_x \tag{12a}$$

$$= \sum_{c=1}^{C} \pi_c \Phi\left(\frac{y \cdot (h - \mu_c)}{\sqrt{\sigma_p^2 + \sigma_c^2}}\right). \tag{12b}$$

With this, the first term in the objective function from Equation (9d) resolves to:

$$\mathrm{H}[y \mid x, h, \mathcal{D}] \approx \mathrm{h}\left(\sum_{c=1}^{C} \pi_c \Phi\left(\frac{h - \mu_c}{\sqrt{\sigma_p^2 + \sigma_c^2}}\right)\right). \tag{13}$$

The second term in Equation (9d) is intractable but can be approximated very well by replacing the binary entropy function by a squared exponential function as proposed in (21):

$$\begin{aligned} \mathbb{E}_{t\sim p(t|\mathcal{D})}\mathrm{h}[y \mid t, x, h] &\approx \int \mathrm{h}\left(\Phi\left(\frac{h - t_x}{\sigma_p}\right)\right) \sum_{c=1}^{C} \pi_c \mathcal{N}(t_x \mid \mu_c, \sigma_c^2) \mathrm{d}t_x \\ &= \sum_{c=1}^{C} \pi_c \int \mathrm{h}\left(\Phi\left(\frac{h - t_x}{\sigma_p}\right)\right) \mathcal{N}(t_x \mid \mu_c, \sigma_c^2) \mathrm{d}t_x \\ &\approx \sum_{c=1}^{C} \pi_c \int \exp\left(-\frac{(h - t_x)^2}{\sigma_p^2 \pi \ln 2}\right) \mathcal{N}(t_x \mid \mu_c, \sigma_c^2) \mathrm{d}t_x \\ &= \sum_{c=1}^{C} \frac{\pi_c K}{\sqrt{\sigma_c^2 + K^2}} \exp\left(\frac{-(h - \mu_c)^2}{2(\sigma_x^2 + K^2)}\right), \end{aligned} \tag{14}$$

with $K = \sigma_p \sqrt{\frac{\pi \ln 2}{2}}$. With this, the expression for the most informative stimulus from Equation (9d) evaluates to:

$$\begin{aligned} (x_*, h_*) = \arg\max_{(x,h)} \mathrm{h}\left(\sum_{c=1}^{C} \pi_c \Phi\left(\frac{h - \mu_c}{\sqrt{\sigma_p^2 + \sigma_c^2}}\right)\right) \\ - \sum_{c=1}^{C} \frac{\pi_c K}{\sqrt{\sigma_c^2 + K^2}} \exp\left(\frac{-(h - \mu_c)^2}{2(\sigma_x^2 + K^2)}\right). \end{aligned} \tag{15}$$

The most straightforward way to find (x_*, h_*) is to perform an (adaptive) grid search in the desired feasible set, or to

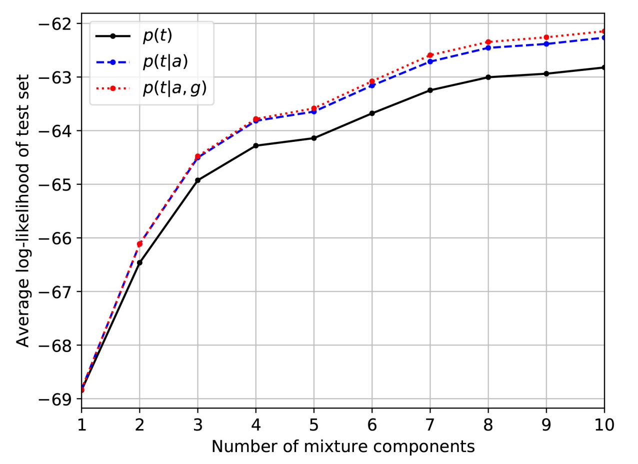


FIGURE 1 | Predictive accuracy of the optimized models as measured by the average log-likelihood of audiograms in the test set. The dashed and dotted lines denote the effect of conditioning the model on age (*a*) and/or gender (*g*).

[2]Mutual information is a symmetrical measure of dependence between two random variables, which can be expressed in terms of (conditional) entropies: $\mathrm{I}[A; B] = \mathrm{I}[B; A] = \mathrm{H}[A] - \mathrm{H}[A \mid B]$.

[3]Binary entropy function $\mathrm{h}(\cdot)$ relates the odds parameter of a binary random variable to its Shannon entropy: $x \sim \mathrm{Bernoulli}(\alpha) \Rightarrow \mathrm{H}[x] = \mathrm{h}(x) \triangleq -\alpha \log(\alpha) - (1 - \alpha) \log(1 - \alpha)$.

[4]A derivation of the solution to the integral can be found in section 3.9 of (3).

use another global optimization method. Since the objective in Equation (15) is relatively cheap to evaluate in terms of computations, this is no problem in practice. However, it is also possible to perform the optimization in a cheaper way by adding an additional approximation. Instead of using the Gaussian mixture posterior directly to obtain the objective function, one could first approximate the mixture posterior by a single Gaussian posterior through moment matching. Plugging this single Gaussian approximate posterior into the objective function eliminates the sums in Equation (15). In that case, x_* turns out to correspond to the transformed frequency on which the approximate posterior has the largest variance, which can be found using a one-dimensional grid search. Once x_* is known, h_* is equal to the posterior mean of the HT at x_* under the single Gaussian approximation (6). It is also possible to only use of the single Gaussian approximation to find x_*, and then solve for h_* under the full model using a simple line search.

4. RESULTS

4.1. Model Learning

To evaluate the predictive performance of the probabilistic model, we learned multiple models with the numbers of mixture components ranging from 1 to 10. All models are learned from the same training set. Training and test sets are obtained by randomly splitting the data set described in section 2.1 subject-wise, such that the training set contains the data of 80% of the subjects (70,590 subjects, 141,180 audiograms) while the test set covers the remaining 20% of subjects.

The quality of a probabilistic model is determined by the extent to which it can predict data in the test set. The better the model has captured the properties and statistics of the data set, the higher the (average) probabilities it assigns to records in the test set. The predictive accuracy of a model is typically measured by the average posterior log-likelihood of records in the test set.

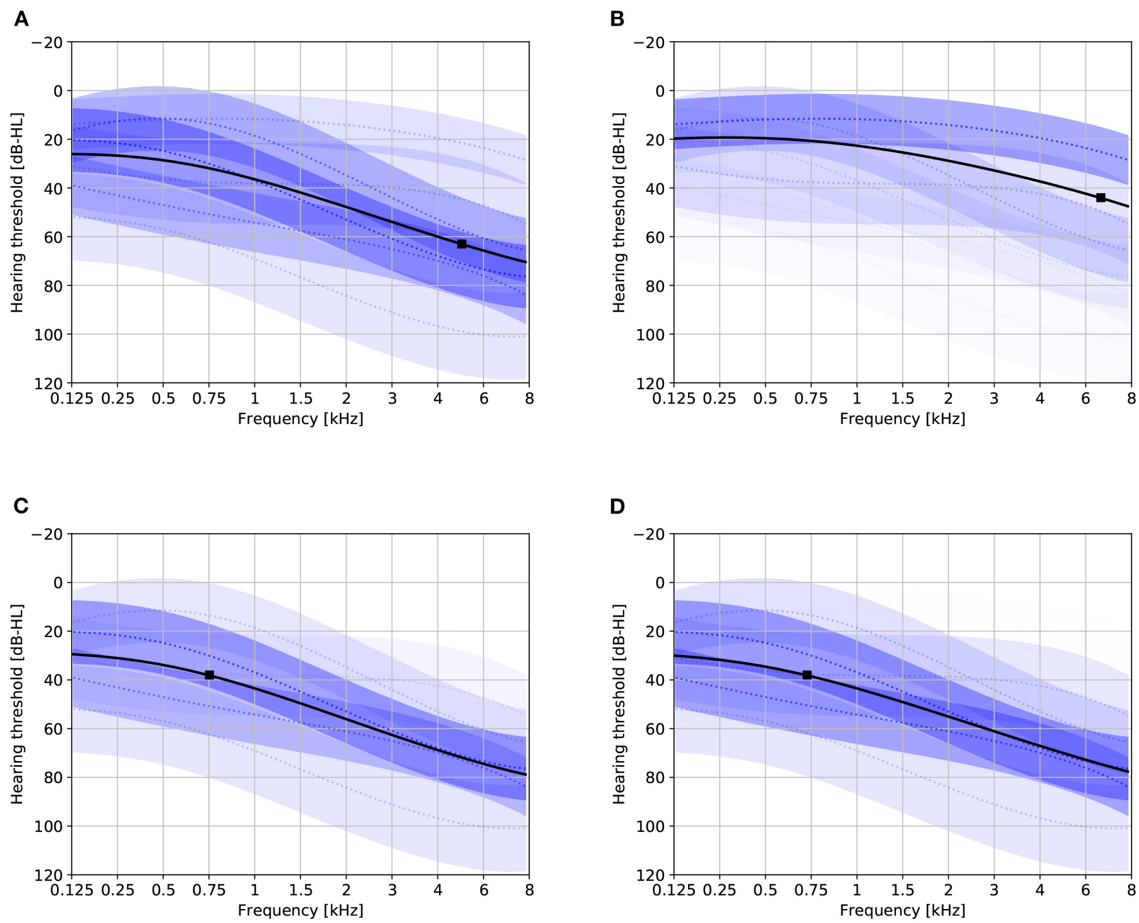


FIGURE 2 | Audiogram plots visualizing the learned model containing six mixture components with various conditionings. Solid black lines denote the initial HT estimate. Dotted blue lines and shaded areas depict the individual GP mixture components ±1 standard deviation (the transparency is proportional to the mixing weight). Solid black boxes indicate the optimal first stimulus. **(A)** Not conditioned on age or gender. **(B)** Age 40, gender unspecified. **(C)** Age 80, gender unspecified. **(D)** Age 80, female.

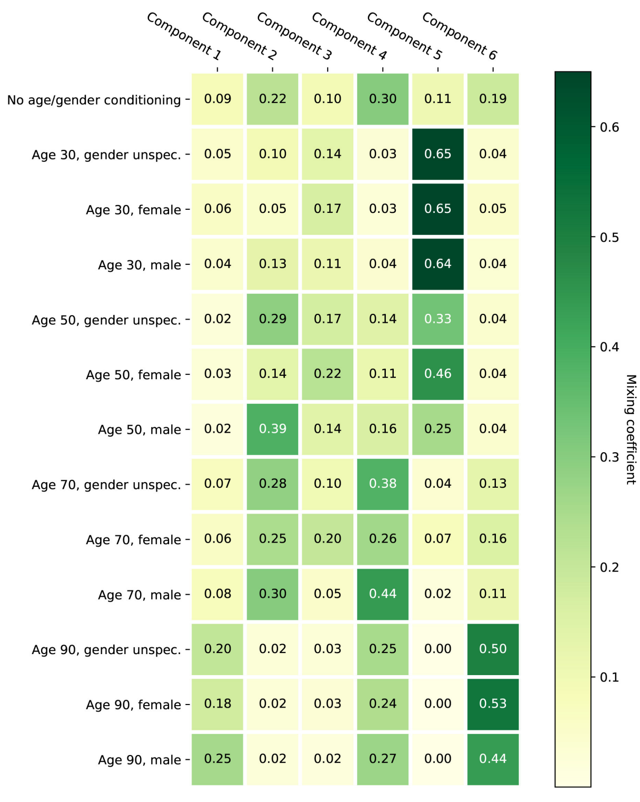


FIGURE 3 | A-priori mixing weights of the learned six-component model under various conditionings.

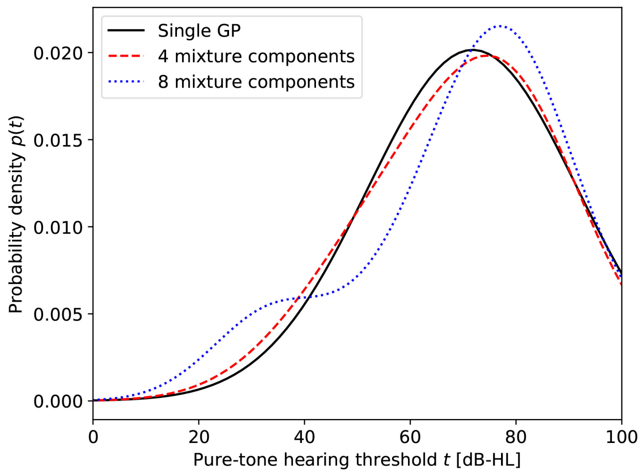


FIGURE 4 | Probability distributions of the HT at 8 kHz for different numbers of mixture components.

The posterior log-likelihood of a hearing threshold t specified at a set of transformed frequencies $\mathcal{X}$ is defined as:

$$\log p(t \mid \mathcal{D}) = \sum_{x \in \mathcal{X}} \log p(t_x \mid \mathcal{D}), \tag{16}$$

where $p(t_x \mid \mathcal{D})$ is the value of the posterior distribution of the hearing threshold evaluated at transformed frequency x.

Figure 1 shows the average log-likelihood of records in the test set as a function of the number of mixture components. Separate lines are used to show the effect of conditioning the models on age and/or gender. As is to be expected, increasing the number of mixture components monotonically increases the predictive accuracy, although the incremental benefit of extra components diminishes after about seven components. Conditioning on age or gender has a positive effect on performance if the number of mixture components is sufficiently large. Conditioning on age has a stronger impact than conditioning on gender.

Figure 2 provides a visualization of the learned model with six mixture components under various conditionings. Without conditioning on age and gender, the distribution of the HT should be representative for the complete population represented by the training set. The effect of conditioning on age is clearly visible: for age 40 the mixture components corresponding to mild hearing loss get assigned a higher weight compared to age 80. The effect of conditioning on gender is smaller. The plots also show that the first proposed stimulus can be different based on age and gender. **Figure 3** shows the a-priori component mixing weights of the six-component model under various conditionings, providing a visual overview of the relative importance of the different components per age and gender group.

To inspect the effect of using a mixture of GPs instead of a single GP, we evaluated the posterior probability distributions of the hearing threshold at fixed frequencies under the various models. **Figure 4** shows these probability distributions at 8 kHz for a selected number of models. The figure clearly illustrates the increased ability of the model to capture the non-Gaussian distribution of the hearing threshold as the number of mixture components is increased.

4.2. Bayesian PTA Simulations

To test the usefulness of our method, we performed PTA simulations on a random subset of 200 audiograms from the test set. Identical simulations are performed using both the learned single-component model and the learned 8-component model, which seems to provide a good trade-off between predictive accuracy and computational complexity judging from **Figure 1**. For each of the 200 randomly selected audiograms, a PTA simulation is performed in the following way.

1. Interpolate the HT in the audiogram (which is only defined on a subset of the standard audiometric frequencies) linearly, yielding a piecewise linear function of frequency.
2. Optionally condition the model on the age and gender corresponding to the audiogram.
3. Determine the optimal next stimulus according to Equation (9d).
4. Simulate the response based on the (interpolated) HT and the response model from Equation (2). Update the model with the response by performing Bayesian inference.
5. Go to step 3 and repeat until 25 responses have been collected.

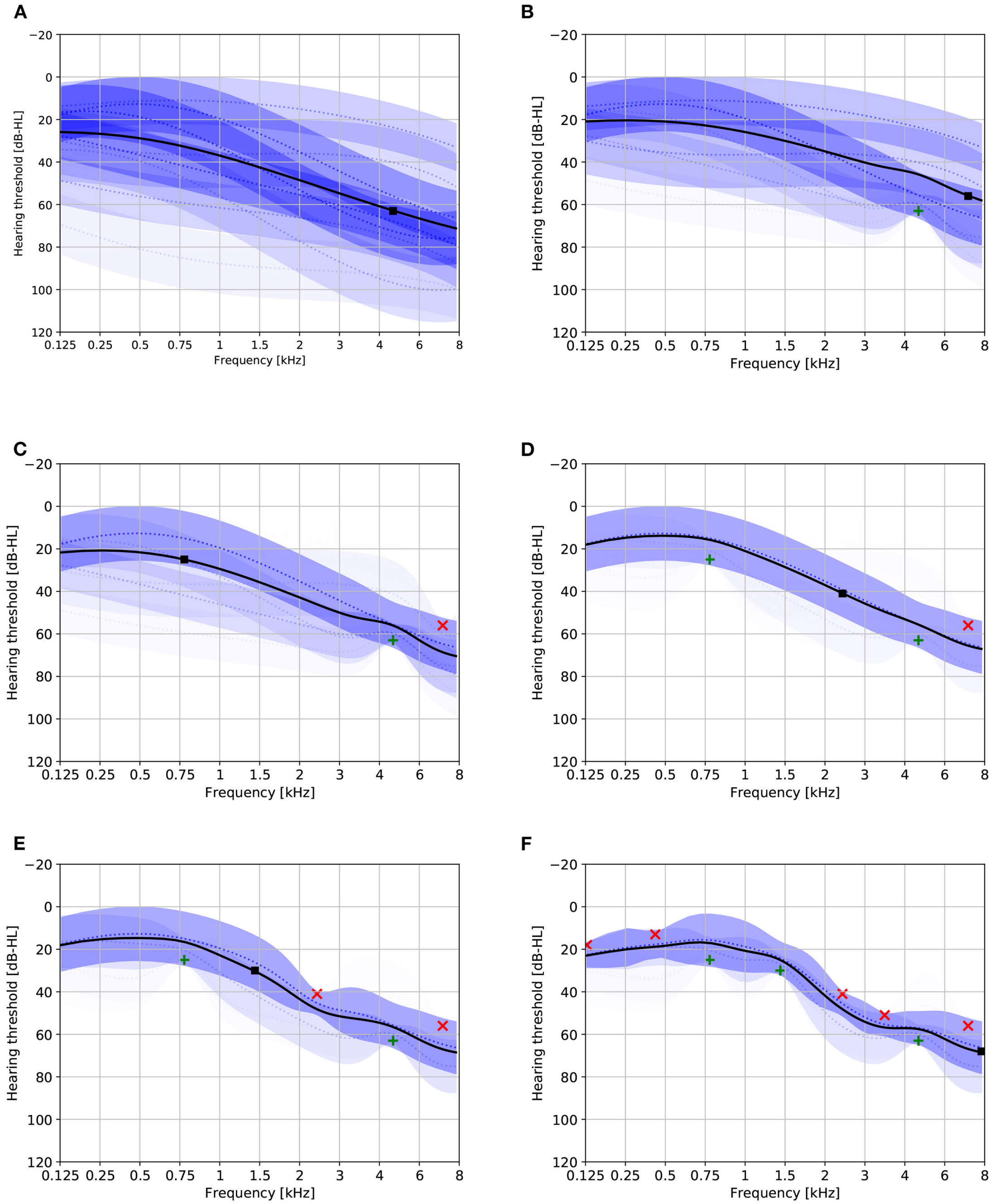


FIGURE 5 | Audiogram plots visualizing the progression of a single Bayesian PTA simulation under the eight-component model. Solid black lines denote the HT estimate. Dotted blue lines and shaded areas depict the individual GP mixture components ±1 standard deviation (the transparency is proportional to the mixing weight). Green and red crosses, respectively, depict audible and non-audible responses. Solid black boxes indicate the optimal next stimulus. **(A)** Audiogram based on zero responses. **(B)** Audiogram based on one response. **(C)** Audiogram based on two responses. **(D)** Audiogram based on three responses. **(E)** Audiogram based on four responses. **(F)** Audiogram based on eight responses.

Figure 5 provides a visualization of the progression of a single Bayesian PTA simulation under the 8-component model. A couple of observations can be made from this figure:

- The posterior mixing weights tend to converge toward either 0 or 1 as more responses are incorporated. Intuitively, this can be interpreted as the model first detecting which mixture component best matches the responses, and then refining the estimate based on the most dominant mixture component. It is this ability that enables the mixture model to attain a faster convergence rate than a single GP model.

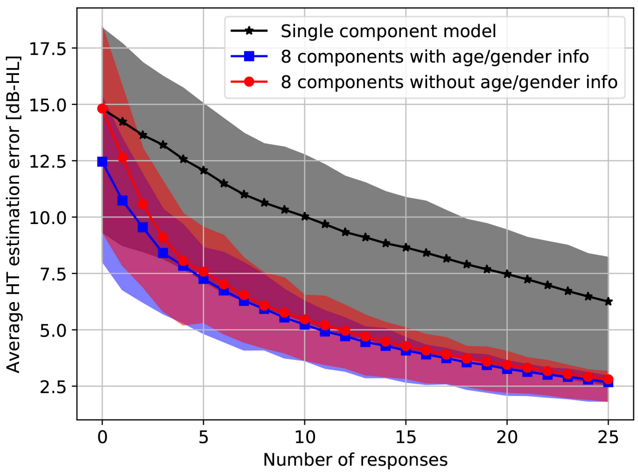


FIGURE 6 | Average absolute HT estimation error on the standard audiometric frequencies (125, 250, 500, 750, 1,000, 1,500, 2,000, 3,000, 4,000, 6,000, 8,000 Hz) as a function of the number of responses under the single-component and 8-component models. The results are averaged over 200 simulations on a random subset of the test set. Shaded areas span from the first quartile to the third quartile. Note that the single-component model cannot be conditioned on age or gender.

- The overall uncertainty about the HT as measured by the variance of the HT estimate tends to decrease as more responses are incorporated.
- The optimal stimuli proposed by the method seem to approximate a randomized grid search in the frequency dimension.

To quantify the performance of our methods in the simulations, we calculate the average absolute error of the HT estimates after every simulated response. The average absolute estimation error is obtained by averaging the absolute differences between the assumed HT and the posterior HT estimate at the following frequencies: 125, 250, 500, 750, 1,000, 1,500, 2,000, 3,000, 4,000, 6,000, 8,000 Hz. **Figure 6** shows the evolution of the estimation error as responses are added. Both the single-component model and the 8-component model result in a monotonic decrease in estimation error, although the decrease is significantly faster under the 8-component model. Using age and gender information consistently decreases the estimation error, with the effect being largest before any responses have been processed. As more responses are processed, the relative benefit of age and gender information diminishes.

5. DISCUSSION

We have demonstrated how a practical and data-efficient PTA method can be obtained by taking a probabilistic modeling approach. Any PTA method involves at least two parts: a method to select stimuli and a prescription for estimating the HT based on responses to said stimuli. Instead of defining these parts in a direct way, we have shown that if one starts with a probabilistic model of the response-generating process, both parts can be derived in a natural way based on formally defined objectives. In the case of stimulus selection, the concept of expected information gain can be used to derive a method that sequentially selects optimal stimuli (in terms of information retrieval rate about the HT) under a given model. The task of estimating the HT based on responses reduces to one of Bayesian posterior inference under the probabilistic modeling approach. Since both parts arise naturally given the underlying model, the model specification indirectly specifies the complete PTA method. As a result, improvements in the quality of the underlying model directly translate into an improved PTA method, either in terms of estimate convergence rate or robustness. Combining a probabilistic modeling approach with information gain maximization in the context of audiometry already has a long history (9, 10, 12). More recently, such approaches have been extended based on GP models (5–7, 13). Various studies have been conducted to experimentally validate these methods (14, 15).

The focus of this work was to increase the efficiency and accuracy of GP-based PTA methods by improving the quality of the underlying model. Toward this end, we proposed a more complex model, i.e., a finite mixture of GPs with mixing weights that depend on additional information about the subject. Moreover, we leverage data to optimize the parameters of the more complex model for a specific target population. The combination of a more complex model and a data-driven optimization leads to a model of higher quality, while preserving the ability to derive optimal stimulus selection and HT estimation based on Bayesian inference. Our simulations indicate that the improved model indeed yields a PTA method with a significantly faster convergence rate than that of an optimized single-GP model. The ability of the proposed model to be conditioned on age and gender also increases performance. The increased performance comes at the price of increased computational complexity. The computational complexity of the Bayesian inference algorithm is linear in terms of the number of mixture components. The same holds for the complexity of the optimal trial selection algorithm if the described approximation is applied. As a result, the computational complexity of our method with a mixture of K components requires roughly K times the amount of computations required under a single-GP model.

We identify multiple possible directions to improve upon the described methods. Firstly, making the user response model (i.e., the part of the model that specifies how a response to a stimulus is generated given the HT) dependent on frequency should increase the quality of the model, given the likeliness that such a relation indeed exists. Ideally, the parameters of a more complex user response model should be learned from data as well, which would require a data set containing raw audiometric test data. Secondly, the method could be made more robust to corrupted responses. In practice, it is to be expected that responses are sometimes inverted by accident, for example due to external disturbances, mistakes, or hardware malfunction. If corrupted responses occur, an optimal active learning method will have a hard time recovering unless the underlying model explicitly incorporates a data corruption aspect. Extending the model with a data corruption part will increase the robustness at the expense of slower convergence, since the assumed signal-to-noise ratio of the responses will decrease. Another option would be to try to detect corrupted responses post hoc, and then excluding them from the data. A third possible improvement is to exploit the correlation between the HTs at both ears of the same subject. Preliminary analysis of our data set indicates that there is a statistically significant correlation to be exploited. One could extend the model such that the result of the PTA procedure on the one ear could be used to improve the predictive accuracy about the HT at the other, leading to a speedup of the PTA procedure at the second ear.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The data set is proprietary. Requests to access these datasets should be directed to Bert de Vries, bert.de.vries@tue.nl.

AUTHOR CONTRIBUTIONS

MC and BV jointly developed the described methods. MC wrote the software implementation of the methods, implemented and executed the experiments, and wrote the manuscript. BV reviewed and revised the manuscript. Both authors contributed to the article and approved the submitted version.

FUNDING

This work was partly funded by GN Advanced Science.

ACKNOWLEDGMENTS

We thank GN Advanced Science for providing the anonymized data set of annotated audiometric records that was used to fit the presented models and to perform simulations.

REFERENCES

1. Yost WA. *Fundamentals of Hearing: An Introduction*. 3rd Edn. San Diego, CA: Academic Press (1994).
2. Carhart R, Jerger JF. Preferred method for clinical determination of pure-tone thresholds. *J Speech Hear Disord*. (1959) 24:330–45. doi: 10.1044/jshd.2404.330
3. Rasmussen CE, Williams CKI. *Gaussian Processes for Machine Learning*. MIT Press (2006). Available online at: http://www.gaussianprocess.org/gpml/chapters/
4. Cohn DA, Ghahramani Z, Jordan MI. Active learning with statistical models. *J Artif Intell Res*. (1996) 4:129–45. doi: 10.1613/jair.295
5. Gardner J, Malkomes G, Garnett R, Weinberger KQ, Barbour D, Cunningham JP. Bayesian active model selection with an application to automated audiometry. In: Cortes C, Lawrence ND, Lee DD, Sugiyama M, Garnett R, editors. *Advances in Neural Information Processing Systems 28*. Curran Associates, Inc. (2015). p. 2386–94. Available online at: http://papers.nips.cc/paper/5871-bayesian-active-model-selection-with-an-application-to-automated-audiometry.pdf
6. Cox M, de Vries B. A Bayesian binary classification approach to pure tone audiometry. *arXiv:151108670 [stat*]. (2015).
7. Song XD, Wallace BM, Gardner JR, Ledbetter NM, Weinberger KQ, Barbour DL. Fast, continuous audiogram estimation using machine learning. *Ear Hear*. (2015) 36:e326. doi: 10.1097/AUD.0000000000000186
8. Bisgaard N, Vlaming MSMG, Dahlquist M. Standard audiograms for the IEC 60118-15 measurement procedure. *Trends Amplif*. (2010) 14:113–20. doi: 10.1177/1084713810379609
9. Özdamar Ö, Eilers RE, Miskiel E, Widen J. Classification of audiograms by sequential testing using a dynamic Bayesian procedure. *J Acoust Soc Am*. (1990) 88:2171–9. doi: 10.1121/1.400114
10. King-Smith PE, Grigsby SS, Vingrys AJ, Benes SC, Supowit A. Efficient and unbiased modifications of the QUEST threshold method: theory, simulations, experimental evaluation and practical implementation. *Vis Res*. (1994) 34:885–912. doi: 10.1016/0042-6989(94)90039-6
11. Kontsevich LL, Tyler CW. Bayesian adaptive estimation of psychometric slope and threshold. *Vis Res*. (1999) 39:2729–37. doi: 10.1016/S0042-6989(98)00285-5
12. Remus JJ, Collins LM. Comparison of adaptive psychometric procedures motivated by the theory of optimal experiments: simulated and experimental results. *J Acoust Soc Am*. (2008) 123:315–26. doi: 10.1121/1.2816567
13. Gardner JR, Song X, Weinberger KQ, Barbour DL, Cunningham JP. Psychophysical detection testing with bayesian active learning. In: *UAI*. (2015). p. 286–95. Available online at: http://auai.org/uai2015/proceedings/papers/213.pdf
14. Schlittenlacher J, Turner RE, Moore BCJ. Audiogram estimation using Bayesian active learning. *J Acoust Soc Am*. (2018) 144:421–30. doi: 10.1121/1.5047436
15. Barbour DL, Howard RT, Song XD, Metzger N, Sukesan KA, DiLorenzo JC, et al. Online machine learning audiometry. *Ear Hear*. (2019) 40:918–26. doi: 10.1097/AUD.0000000000000669
16. Bezanson J, Karpinski S, Shah VB, Edelman A. Julia: a fast dynamic language for technical computing. *arXiv:12095145 [cs*]. (2012).
17. Traunmüller H. Analytical expressions for the tonotopic sensory scale. *J Acoust Soc Am*. (1990) 88:97–100. doi: 10.1121/1.399849
18. Wang S, Sekey A, Gersho A. Auditory distortion measure for speech coding. In: [*Proceedings] ICASSP 91: 1991 International Conference on Acoustics, Speech, and Signal Processing*. Toronto, ON: IEEE (1991). p. 493–496. Available online at: http://ieeexplore.ieee.org/document/150384/
19. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the *EM* Algorithm. *J R Stat Soc Ser B*. (1977) 39:1–22. doi: 10.1111/j.2517-6161.1977.tb01600.x
20. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. *TensorFlow: A System for Large-Scale Machine Learning*. (2016). p. 265–283. Available online at: https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi
21. Houlsby N, Huszár F, Ghahramani Z, Lengyel M. Bayesian active learning for classification and preference learning. *arXiv:11125745 [cs, stat*] (2011).

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.

frontiers in Digital Health

ORIGINAL RESEARCH
published: 18 August 2021
doi: 10.3389/fdgth.2021.723533

Using Machine Learning and the National Health and Nutrition Examination Survey to Classify Individuals With Hearing Loss

Gregory M. Ellis[1]* *and Pamela E. Souza*[1,2]

[1] *Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL, United States,* [2] *Knowles Hearing Center, Evanston, IL, United States*

OPEN ACCESS

Edited by:
Changxin Zhang,
East China Normal University, China

Reviewed by:
Raul Sanchez-Lopez,
Technical University of Denmark, Denmark
Shekh Md Mahmudul Islam,
University of Dhaka, Bangladesh

***Correspondence:**
Gregory M. Ellis
gregory.ellis@northwestern.edu

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: *10 June 2021*
Accepted: *27 July 2021*
Published: *18 August 2021*

Citation:
Ellis GM and Souza PE (2021) Using Machine Learning and the National Health and Nutrition Examination Survey to Classify Individuals With Hearing Loss. Front. Digit. Health 3:723533. doi: 10.3389/fdgth.2021.723533

Even before the COVID-19 pandemic, there was mounting interest in remote testing solutions for audiology. The ultimate goal of such work was to improve access to hearing healthcare for individuals that might be unable or reluctant to seek audiological help in a clinic. In 2015, Diane Van Tasell patented a method for measuring an audiogram when the precise signal level was unknown (patent US 8,968,209 B2). In this method, the slope between pure-tone thresholds measured at 2 and 4 kHz is calculated and combined with questionnaire information in order to reconstruct the most likely audiograms from a database of options. An approach like the Van Tasell method is desirable because it is quick and feasible to do in a patient's home where exact stimulus levels are unknown. The goal of the present study was to use machine learning to assess the effectiveness of such audiogram-estimation methods. The National Health and Nutrition Examination Survey (NHANES), a database of audiologic and demographic information, was used to train and test several machine learning algorithms. Overall, 9,256 cases were analyzed. Audiometric data were classified using the Wisconsin Age-Related Hearing Impairment Classification Scale (WARHICS), a method that places hearing loss into one of eight categories. Of the algorithms tested, a random forest machine learning algorithm provided the best fit with only a few variables: the slope between 2 and 4 kHz; gender; age; military experience; and self-reported hearing ability. Using this method, 54.79% of the individuals were correctly classified, 34.40% were predicted to have a milder loss than measured, and 10.82% were predicted to have a more severe loss than measured. Although accuracy was low, it is unlikely audibility would be severely affected if classifications were used to apply gains. Based on audibility calculations, underamplification still provided sufficient gain to achieve ∼95% correct

(Speech Intelligibility Index ≥ 0.45) for sentence materials for 88% of individuals. Fewer than 1% of individuals were overamplified by 10 dB for any audiometric frequency. Given these results, this method presents a promising direction toward remote assessment; however, further refinement is needed before use in clinical fittings.

Keywords: audiology, remote audiology, machine learning, CDC, NHANES, centers for disease control and prevention, national health and nutrition examination survey

INTRODUCTION

Several factors have been pushing audiologists toward telehealth, the most obvious of which is the COVID-19 pandemic. The pandemic closed the physical doors of audiology clinics around the world, requiring healthcare professionals to come up with alternatives to traditional in-person clinical approaches. Regardless of the pandemic, a shift to telehealth is necessary to reach underserved communities and individuals far away from audiology clinics.

One way to provide more convenient, accessible care for patients is to have them complete hearing tests in their own home. Testing hearing in the home is not a new concept. Computer-based or cellular phone-based hearing screenings (i.e., evaluating whether the participant can hear a preset level, and referring for further testing if they cannot) have been used successfully [e.g., (1–3)]. However, it is still more difficult to estimate hearing thresholds outside of an audiology testing center. Some at-home tests rely on a fairly traditional approach to audiometric testing, examining thresholds at octave frequencies between 250 and 8,000 Hz by providing a calibrated tablet and headphones. One such test, the Home Hearing Test, has been shown to produce reliable results in the home (4, 5). For a more thorough review of automated and in-home audiometric testing, please see Pragt et al. (6).

It is difficult to devise at-home hearing testing when the patient uses their own home computer or cell phone with earphones because that equipment will produce unknown presentation levels [for recent review of such approaches, see (7)]. A method for determining a patient's audiogram with limited audiological information was patented by Diane Van Tasell in 2015 (patent US 8,968,209 B2). In this method, pure-tone thresholds are measured at 2 kHz and 4 kHz. Rather than attempting to measure precise hearing thresholds at those frequencies, the slope between 2 and 4 kHz is calculated and combined with questionnaire information. Together, these data are used to reconstruct the most likely audiogram for that listener from a database of options. The method was intended to overcome the limitations of presenting accurate signal levels when using uncalibrated equipment. An approach like the Van Tasell method is desirable because it is relatively quick (only two thresholds in each ear are measured) and feasible to do in a patient's home on uncalibrated equipment where the exact levels of presented stimuli are unknown.

A similar in-home test would also be useful for experimental procedures. A large, diverse pool of subjects can be recruited and tested quickly by using remote testing. If the population of interest for a study is people with hearing impairment, it may be important to apply gain to the stimuli being tested. In this case, an estimate of the participant's hearing loss is necessary. Because a precise threshold cannot be guaranteed to be measured in the home for the reasons listed above, a remote testing solution that does not rely on precise threshold measurements is desirable.

Put plainly, the problem that needs to be solved is this: how can a person's audiometric thresholds be accurately predicted with limited information? Machine learning excels when using a set of features (variables) to categorize an unseen case. In order to do this, a machine learning algorithm is trained on a set of sample data, then it is asked to categorize a set of test data. By way of example, suppose a machine learning algorithm were trained to categorize objects as either an animal, a plant, or a mineral based on the object's features (e.g., shape, color, and size). If the algorithm was asked to categorize a strawberry, it would use the features it was trained on to make its best guess. Then the algorithm would—hopefully—correctly categorize the strawberry as a plant. The accuracy of any given machine learning algorithm is dependent on the particular cases it receives when it is being trained and how generally the algorithm is able to apply what it "learned" during the training phase. A large, diverse dataset tends to provide strong fits for a machine learning approach.

Fortunately, a large, diverse dataset of audiologic information exists in the public domain: the National Health and Nutrition Examination Survey (NHANES). NHANES is a complex survey that is collected biennially in the United States. Each survey cycle examines roughly 10,000 individuals from the United States civilian non-institutionalized population. Participants in the survey are given questionnaires, some are interviewed, and some receive medical examinations including audiometric tests. The NHANES database provides a rich source of pure tone audiometric and demographic data from individuals in the United States.

Audiometric data were categorized in order to facilitate most machine learning approaches (8). There are two major ways to categorize hearing losses that the authors are aware of today: the Wisconsin Age-Related Hearing Impairment Classification Scale (WARHICS) (9) and the IEC 60118-15 standard audiograms (10). Because the IEC standard audiograms are based on data from Stockholm (10) and the WARHICS classes were based on data collected in the United States (9), the WARHICS classes were used in the present study.

The goal of the present study was to determine how accurately a machine learning algorithm can predict a person's audiometric configuration given limited information about that person's demographics, hearing loss, and self-reported difficulty hearing. An additional goal was to apply this approach in a hypothetical speech test remotely administered, and to quantify the degree to which mismatches between the observed and predicted audiometric configurations would affect speech intelligibility. Three machine learning algorithms were trained using the following features: age, gender, previous military experience, the slope between 2,000 and 4,000 Hz pure tone thresholds, and self-reported amount of hearing difficulty.

MATERIALS AND METHODS

Procedure

All data preprocessing and analysis was done in R (11) using the lattice (12), caret (13), Metrics (14), and tidyverse (15) packages.

Data were downloaded from the National Health and Nutrition Examination Survey (NHANES) database (https://www.cdc.gov/nchs/nhanes/index.htm). NHANES is a complex survey that studies the United States civilian non-institutionalized population. As a part of this survey, participants in most survey cycles receive audiometric evaluations. The large sample size and diverse population make NHANES an excellent dataset for examining audiometric patterns within the population surveyed. Because pure-tone thresholds were necessary for the present analysis, sample sets that did not include audiometric measurements were excluded. The sample sets that included audiometric data are those from 1999–2012 and 2015–2016. This span of years resulted in 71,963 cases.

Because of the complex survey design, special care needs to be taken when merging several datasets. These datasets were merged following the procedures outlined on the NHANES website in order to preserve sample weights. Sample weights are an important part of a complex survey, as they account for factors that make the selected sample more representative of the targeted population. Sample weights in the NHANES database take into account three major components. First, the sample weights account for the probability that a particular individual was selected to participate in the survey. Second, adjustments are made for non-response rates. Third, adjustments are made to account for oversampling of particular genders, age groups, and ethnic backgrounds.

It is also important to choose the appropriate set of weights. According to the NHANES site, a researcher must choose the weight that includes the smallest possible subpopulation that includes all of the variables of interest. The cases with audiometric data are the smallest subpopulation in the present study and the audiometric data were collected in the mobile exam center (MEC). Therefore, the MEC weights were used for the present study. Eight NHANES cycles were combined for this dataset. Based on the guidelines laid out in the NHANES tutorials, a combined weight was created by multiplying the weights from 1999–2002 by 0.25 and the weights for all other years by 0.125. These new weights were saved and used in analysis.

The survey questions asked of participants also changed over the years. The question of "General condition of hearing" (Which statement best describes your hearing (without a hearing aid)?) had six possible answers from 1999–2004 (AUQ130), eight possible answers from 2005–2010 (AUQ131), and was given a new designation starting in 2011 (AUQ054). The data needed to be adjusted for these changes. From 1999–2004, participants could answer: "Good," "Little trouble," "Lot of trouble," "Deaf," "Don't know," or by refusing to answer the question. Starting in 2005, participants were given two new answers to the question: "Excellent," and "Moderate hearing trouble." When the data were merged, the class of answer from 1999–2004 was unchanged, though it should be noted that some number of participants that responded "Good" from 1999–2004 may have chosen "Excellent" if it were an option for them. A similar argument applies to the "Moderate hearing trouble" response added in 2005. For ease of reading, all three versions of this question (AUQ130, AUQ131, and AUQ054) will be referred to as "the question regarding hearing condition."

Next, data were cleaned to ensure all cases had the following data: audiological thresholds for both ears at audiometric frequencies between 0.5 and 8 kHz, military experience, age, gender, and a response to the question regarding hearing condition. Of those 72,509 cases, 62,087 cases (85.6%) did not have audiological data because they did not participate in the MEC portion of their NHANES cycle. In addition to those missing audiological data, 1,164 cases (1.6%) were missing military status data. Two cases were missing answers to the question regarding hearing condition. All of these cases were dropped from the analysis resulting in 9,256 individuals with complete data for the variables listed above (12.8% of the original sample). Wisconsin Age-Related Hearing Impairment Classification Scale (WARHICS) classes were calculated for each ear for each person and saved as a separate variable for the two ears (WARHICS left and WARHICS right). The WARHICS subcategories were not included in this analysis because the subclasses were subsumed by the major classes in a previous study (9), and because the main classes were sufficient for the goals of the present study. See **Figure 1** for a visualization of the WARHICS classes as they would appear on an audiogram.

Demographics

Of the 9,256 valid cases from 1999–2012 and 2015–2016, 4,156 cases (44.9%) were male and 5,100 cases (55.1%) were female. Eight hundred seventy-seven cases had military experience (9.5%), 8,377 cases had no military experience (90.5%), one refused to answer the question and one responded "I don't know." Age ranged from 17 to 85 years. The distribution of ages is plotted in **Figure 2**. See **Table 1** for a breakdown of WARHICS class for left and right ears. **Table 2** shows the distribution of answers to the question "Which statement best describes your hearing (without a hearing aid)?"

ANALYSIS

Three machine learning algorithms were trained on the dataset to predict WARHICS class: random forest (RF), support vector machines with a radial kernel (SVM Radial), and k-nearest

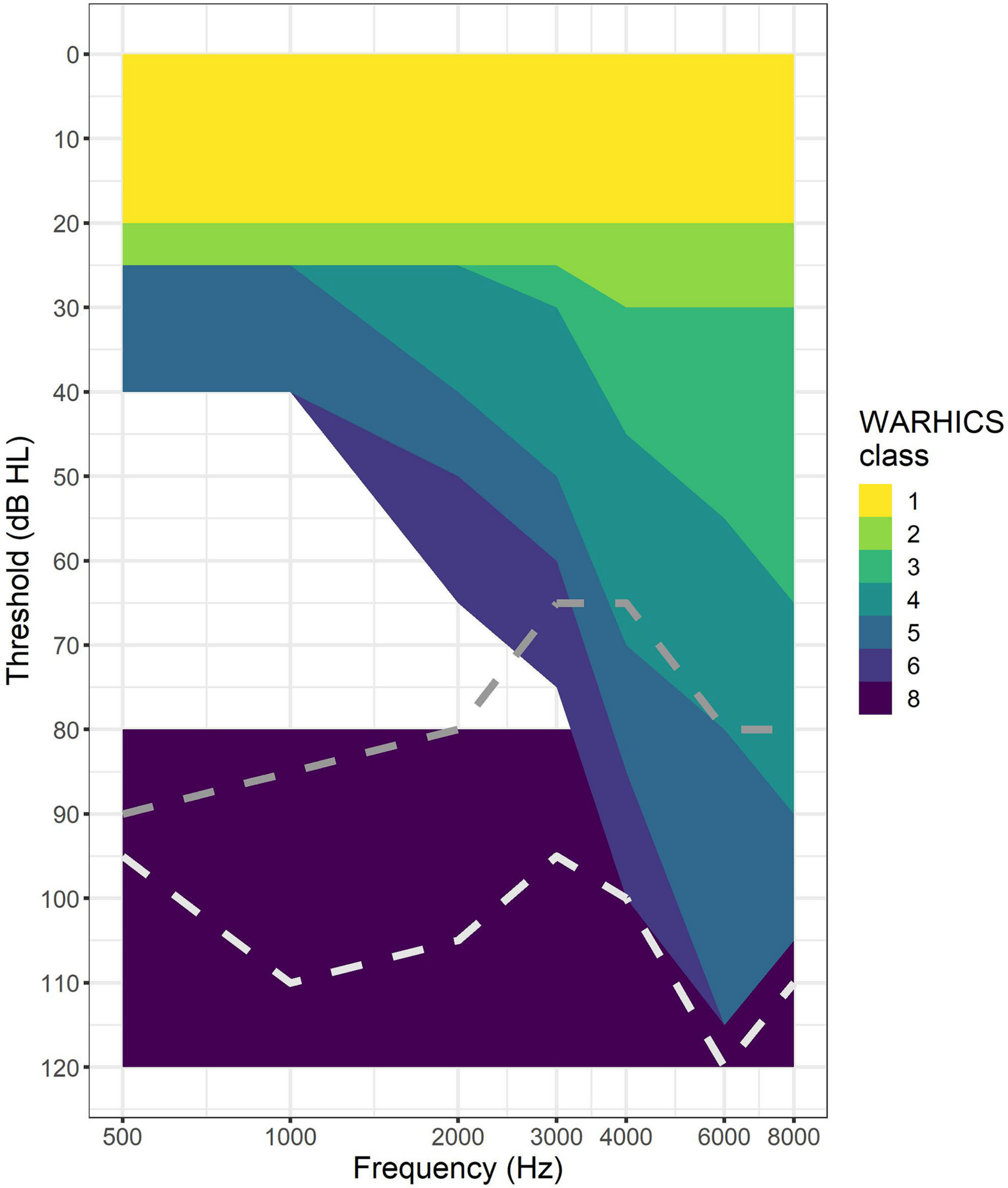


FIGURE 1 | WARHICS classes plotted on an audiogram. Shaded regions represent regions in which an individual's audio must fall to be classified as that WARHICS class. Gray dotted line is an example from WARHICS class 7 (not 1–6 and at least one threshold <= 80 dB). White dotted line is an example from WARCHIS class 8 (all thresholds >= 80 dB).

neighbors (KNN). Accuracy was used to assess the efficacy of the algorithms.

The 9,256 data points for the left ears were split into two sub-datasets: one for training (80% of the data: 7,407 cases) and one for validation (20% of the data: 1,849 cases). Only three cases were classified as WARHICS class 8, so one of these cases was forced into the validation dataset. The other two cases classified as WARHICS class 8 were used in the training dataset. The right ear data ($N = 9{,}256$) were used for a second round of validation and testing.

RESULTS

The three machine learning algorithms were assessed based on the time they took to run using a 2.8 GHz 11th Generation Intel® Core™ i7 processor with no parallel processing, the overall accuracy, and learning curves. See **Table 3** for run time, accuracy, and final parameters fit. Learning curves for the three algorithms are plotted in **Figure 3**. WARHICS class 8 was excluded from the learning curves because that class was rare.

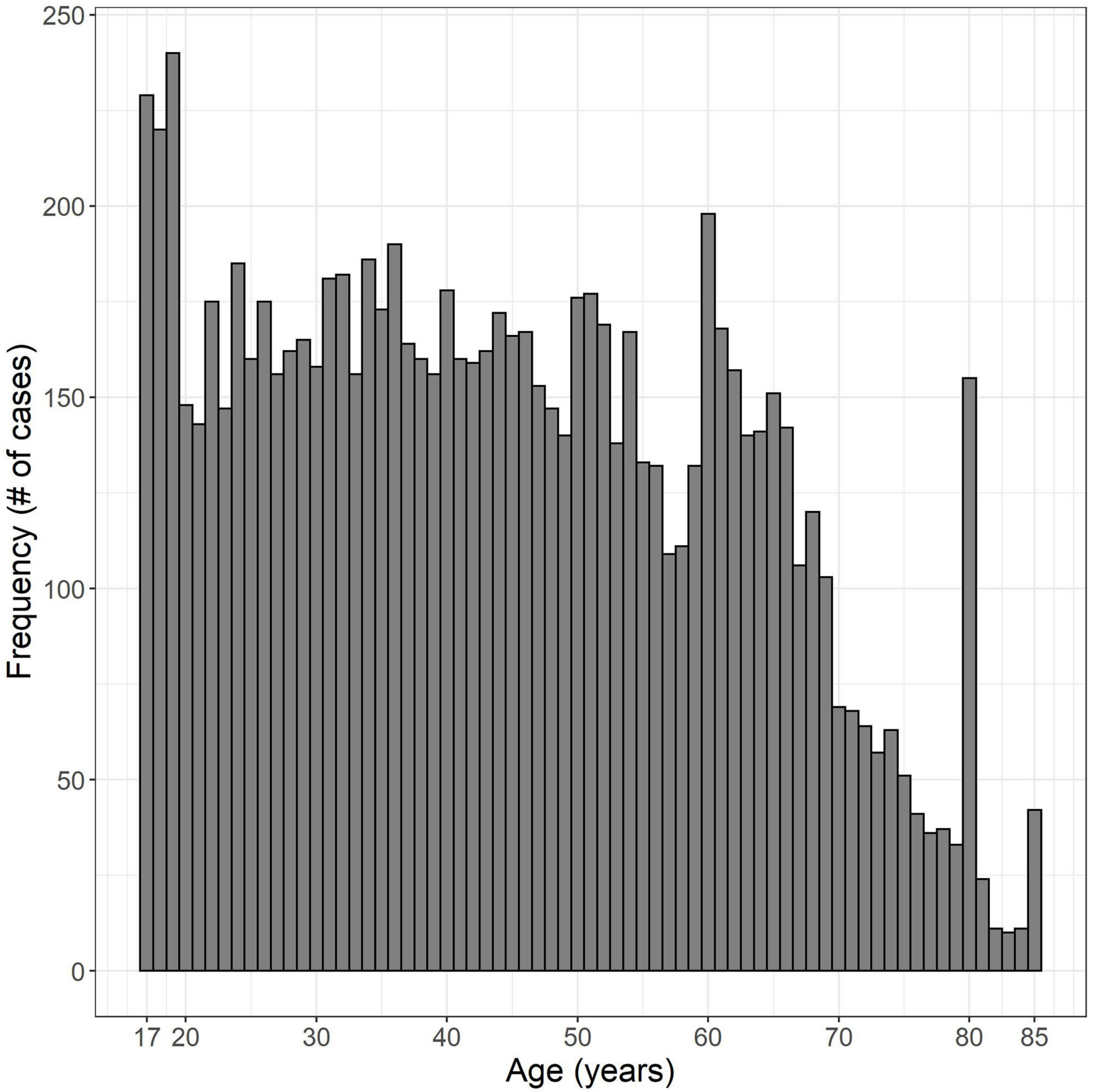


FIGURE 2 | Histogram of ages examined. Data were obtained from the National Health and Nutrition Examination survey.

TABLE 1 | Distribution of WARHICS classes in left and right ears.

WARHICS class	Frequency (L)	Percentage (L)	Frequency (R)	Percentage (R)
1	3,761	40.62%	3,955	42.72%
2	1,754	18.95%	1,711	18.48%
3	1,505	16.26%	1,499	16.19%
4	931	10.06%	822	8.88%
5	760	8.21%	781	8.44%
6	210	2.27%	177	1.91%
7	334	3.61%	313	3.38%
8	3	0.03%	0	0.00%

TABLE 2 | Distribution of responses to the question "Which statement best describes your hearing (without a hearing aid)?"

Response	Frequency	Percentage
Excellent	5,027	54.30%
Good	3,103	33.52%
A little trouble	769	8.31%
Moderate hearing trouble	251	2.71%
A lot of trouble	100	1.08%
Deaf	5	0.05%
Refused	0	0.00%
Don't Know	1	0.01%

TABLE 3 | Run time, accuracy, and the final fit parameters for a random forest (RF), support vector machines with a radial kernel (SVM Radial), and k-nearest neighbors.

Algorithm	Run Time	Accuracy	Parameters
RF	224.91 s	0.54	mtry = 2
SVM Radial	460.99 s	0.51	sigma = 0.23, C = 1
KNN	3.69 s	0.45	k = 9

Run time is included here for completeness, but a machine learning algorithm could be implemented for the purposes discussed in this paper using a remote server to bypass the run time.

The learning curves and accuracy indicate that the random forest algorithm is the best algorithm among the three. Although run time is another typical metric for measuring machine learning algorithms, it is not an important factor here. Run time information is only relevant for assessing the algorithms if they needed to be run each time they had to categorize a new case. In the applications discussed in the present study, this would not be the case because the chosen algorithm could be implemented on a remote server and called when needed. The run time is reported here for completeness. The learning curves show a normal pattern of results and a good fit for both the RF and KNN algorithms. The large jump in performance around 3,500 trials and the wide gap in performance at the end of the training indicate that the SVM Radial model is not a good fit for these data. Based on run time, accuracy, and learning curves, RF was used to predict WARHICS class.

RF prediction efficacy was assessed using confusion matrices. The left ear validation dataset saw significantly higher accuracy than the no information rate (Acc = 0.5462, NIR = 0.4067, $p < 0.001$). Cohen's kappa was calculated as 0.35 which signifies a fair agreement between the reference and prediction (16). The confusion matrix from which these values were calculated is shown in **Table 4** along with within-class precision and recall calculated following the guidelines laid out in Sokolova and Lapalme (17). Overall, the model performs best at classifying individuals with no clinical hearing loss (WARHICS class 1). The algorithm performs less well at identifying individuals that fall into WARHICS class 2 and WARHICS class 5.

Right ear data were used to test the RF algorithm. The same model trained on 80% of the left ear data was used to predict the classification for all 9,258 cases of right ear data. Again, accuracy was significantly greater than the no information rate (Acc = 0.5583, NIR = 0.4273, $p < 0.001$). Cohen's kappa showed fair agreement between the machine learning algorithm and the reference classifications (Cohen's kappa = 0.3573). The confusion matrix from which these values were calculated is shown in **Table 5** along with within-class precision and recall. The algorithm shows a similar pattern of results for the right ear data as it did for the left ear validation dataset, though the algorithm seems to have more success classifying listeners in WARHICS class 5 for the right ears than it did for the left.

DISCUSSION

There are several ways to assess the real-world efficacy of a machine learning algorithm. At the end of the day, we want to know how accurate the algorithm is; however, "accuracy" can be conceived of in different ways. We will explore two rules for assessing accuracy: a strict rule, and a practical rule.

The strict rule states that *any* mismatch between the predicted WARHICS class and the reference WARHICS class is a miss. For example, if the listener has a reference WARHICS class of 5 and they are categorized as WARHICS 5, this is a hit. If they are categorized as WARHICS 4, this is a miss. The practical rule is based on the difference in expected speech audibility due to a mismatch between the predicted and reference WARHICS classes. For details on calculation of this this rule, please see the **Supplementary Materials**.

By the strict rule, the machine learning algorithm correctly categorizes a loss roughly 55% of the time—about 1 in 2 individuals. This is certainly below the desired success rate, but this is due to the intentional lack of information provided to the machine learning algorithm. If all pure tone frequencies were included, the machine learning algorithm would have been significantly more accurate; however, this was not the goal of the present study. The intent was to see how accurately the machine learning algorithm could predict audiometric configurations given *limited* information that one might expect to have when using uncalibrated equipment in a person's home, similar to the approach suggested by Van Tasell. A

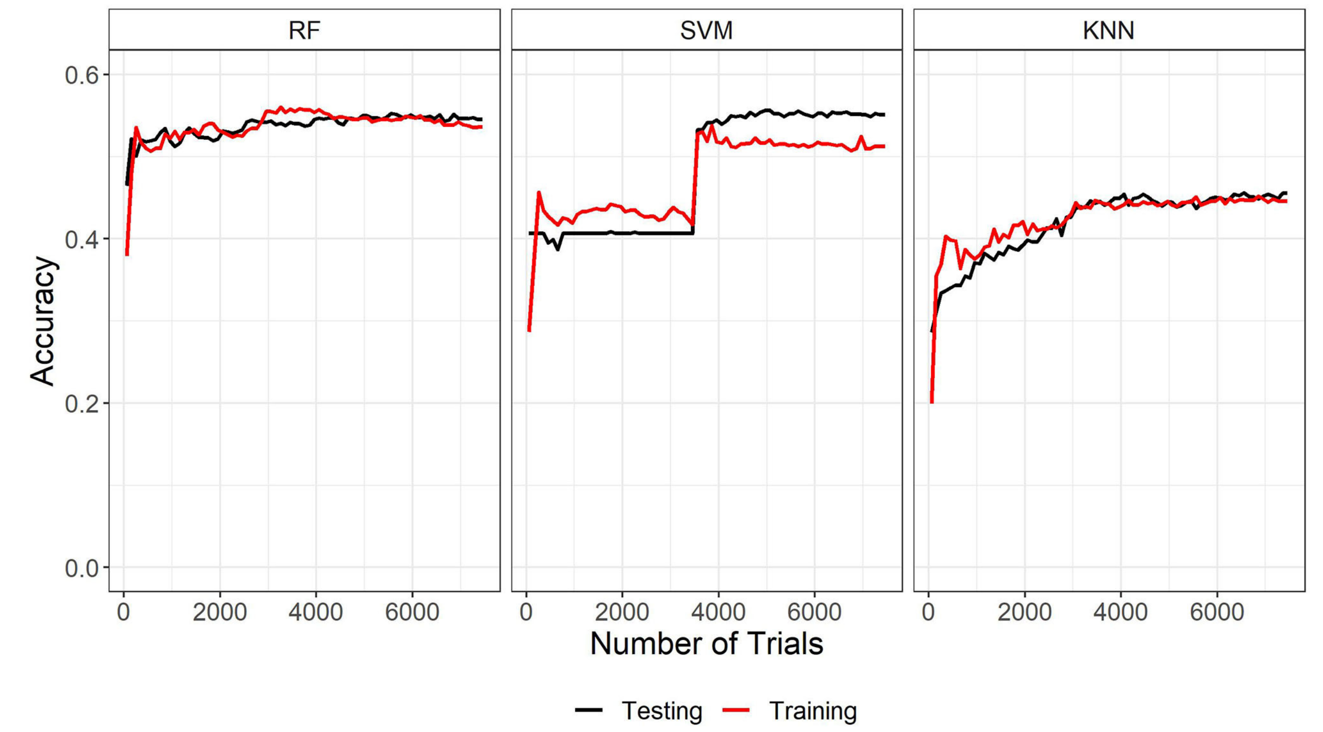


FIGURE 3 | Learning curves of the three machine learning algorithms trained on the left ear data. RF, Random Forest; SVM, Support vector machines with radial kernel; KNN, k-nearest neighbors.

TABLE 4 | Confusion matrix results of left ear machine learning predictions.

		Reference								
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Precision**
Prediction	1	**705**	244	92	13	31	2	5	0	0.65
	2	11	**13**	8	2	1	0	1	0	0.36
	3	36	92	**168**	72	55	2	14	0	0.38
	4	0	0	26	**89**	40	19	10	0	0.48
	5	0	0	5	8	**11**	10	12	0	0.24
	6	0	0	0	1	5	**5**	2	0	0.38
	7	0	1	1	1	9	4	**22**	1	0.56
	8	0	0	0	0	0	0	0	**0**	0
Recall		0.9	0	0.6	0.5	0.07	0.12	0.3	0	

Columns represent the reference WARHICS classes from the NHANES dataset. Rows represent the predicted WARHICS classes from the RF algorithm. Numbers along the diagonal (in bold) count the number of correct classifications. Numbers above the diagonal count the number of classifications where the predicted WARHICS class is less severe than the reference WARHICS class (underprediction of loss). Numbers below the diagonal count the number of classifications where the predicted WARHICS class is more severe than the reference WARHICS classification (overprediction of loss). Overall, the machine learning algorithm correctly predicted 54.79% of losses, underpredicted 34.40% of losses, and overpredicted the remaining 10.82% of losses. Within-class precision and recall for these data are presented in the margins. The precision and recall values for WARHICS class 8 are anomolous because there was only 1 case in the validation dataset and it was misclassified. They are included here for completeness.

different machine learning approach achieved around 90% accuracy across different audiometric configurations using judgments provided by three licensed audiologists about the configuration, severity, and symmetry of participant's losses (18). However, such an approach requires more resources and is subject to variability according to the experts being consulted. An advantage of the method tested here, despite its lower accuracy by the strict rule, is its ability to be fully automated and implemented in remotely-conducted auditory experiments where expert judgment cannot be easily applied.

Given these results, the practical rule may be the appropriate way to describe the results of the present experiment if the machine learning solution presented here were used to predict thresholds for a speech intelligibility experiment. By the practical rule, the machine learning algorithm succeeds 88.3% of the time. This success rate is much better than the strict rule partly due in part to a laxer criterion for

TABLE 5 | Confusion matrix results of right ear machine learning predictions.

		Reference								
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Precision**
Prediction	1	**3,712**	1,185	482	73	140	0	34	0	0.66
	2	46	**56**	59	13	6	0	3	0	0.31
	3	195	455	**803**	300	291	8	75	0	0.38
	4	1	10	139	**374**	197	92	39	0	0.44
	5	0	2	5	45	**97**	37	41	0	0.43
	6	0	0	0	5	14	**25**	16	0	0.42
	7	1	3	9	12	36	15	**105**	0	0.58
	8	0	0	0	0	0	0	0	**0**	0
Recall		0.94	0.03	0.5	0.5	0.12	0.14	0.34	0	

*This table is laid out the same way as **Table 4**. Overall, the machine learning algorithm correctly predicted 55.88% of losses, underpredicted 33.39% of losses, and overpredicted the remaining 10.73% of losses. Within-class precision and recall for these data are presented in the margins.*

counting a success. However, the practical rule does as its name implies: uses a practical threshold for success based on the audibility that would be achieved for presented speech stimuli. Eighty eight point three percent of the cases would still be predicted to score 95% correct on sentence materials, even when underamplified. If a machine learning solution were used in this context, a researcher may be able to identify whether a listener received the correct gain or not. A researcher might be able to identify which of the remaining 11.7% were misclassified by looking at volume control (presumably, listeners that were overamplified would turn the volume down to a comfortable level), or by devising a threshold test at the outset of the experiment to identify those that were underamplified. Such methods are speculative here and would need to be refined further in the future.

User-operated tests could be applied inside and outside of the clinic. In the clinic, it could be used to improve efficiency. An audiologist that needs to only measure two or three air conduction thresholds in conjunction with a short questionnaire would save a substantial amount of time, improving the efficiency of clinic operations. The saved time could then be used for other diagnostic tests or counseling. This is consistent with calls to action for audiologists to focus on more sophisticated measures, expert interpretation, and patient counseling, vs. spending a majority of their appointment time manually adjusting the levels produced by a pure-tone audiometer (19, 20). With regard to in-home testing, measurement of audiometric thresholds is becoming a reality with devices like the AMTAS Home Hearing Test (4, 5). Such in-home devices are expensive and must be physically provided to the patient if it is important that the test be calibrated to provide accurate results. However, if a patient were provided with an online link via their home computer, a first fit could be estimated with only two pure tone thresholds, a short questionnaire, and without the need for precisely calibrated presentation levels. If a method was able to accurately predict an individual's WARHICS class, a hearing aid might then be provided with the initial frequency-gain response set according to the predicted audiogram and with a margin of adjustment considered acceptable for the user. The margin of adjustment would likely cover the range of the WARHICS class assigned to the patient. Such a range would acknowledge the fact that the machine learning solution presented here does not predict a specific audiogram, but rather a range of possible audiograms. Setting a range of adjustment values could be a potential solution to this problem. Support for this method comes from a recent paper suggesting that hearing aids set by the user using a smartphone app can provide outcomes that are as good as—or better than—those provided by the traditional audiologic best practices (21). Using machine learning to restrict the adjustment range could speed up the process of self-fitting for the patient. A combination of user-adjusted response and response constraints based on predicted audiogram would guard against situations where the user chooses a response that is inadequate or inappropriate for their hearing loss.

As a caution, if in-home testing becomes a broadly accepted option in the future, careful steps will need to be taken to make sure that patients have a pathway for follow-up likely including a full audiogram, medical care, and that treatable audiologic disorders are not missed. One questionnaire, the Consumer Ear Disease Risk Assessment (CEDRA), effectively screens for serious audiologic disorders (22). CEDRA or a similar questionnaire could be used as a supplement during at-home hearing screening. In view of data that perceived hearing disability is not strongly related to pure-tone thresholds [e.g., (23)], additional information may be needed to guide provision of amplification once a hearing loss has been identified. Nonetheless, recent developments in auditory science (accelerated by effects of COVID-19 on elective medical care) suggest that remote, at-home or other user-centered assessment techniques will play a role in future treatment options. That said, the present machine learning algorithm is not ready for deployment on a massive scale in clinical settings. The solution presented here would need to be fine-tuned, validated, and likely included in a battery of other tests.

For research studies, a researcher using the approach described here might be able to administer in-home speech tests to individuals with hearing loss without needing detailed knowledge of the participant's computer, headphones, or sound card. Although environmental factors (e.g., road noise, background voices, construction, pets, etc.) cannot be controlled for using this method, the experimenter can coarsely estimate the class of a participant's loss and apply the appropriate gain. In view of the known difficulties in accurately predicting loudness perception from pure-tone thresholds (24), it would be prudent of that experimenter to include a restricted volume adjustment for the participant (perhaps one that maintains an acceptable SII, as described above) in the case of loudness discomfort. Such an approach would benefit the field of hearing research by greatly expanding the sample size and sample demographics without incurring much extra cost.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://wwwn.cdc.gov/nchs/nhanes/default.aspx National Health and Nutrition Examination Survey (NHANES) years 1999-2012 and 2015-2016.

AUTHOR CONTRIBUTIONS

GE: responsible for data processing, analysis, figures, tables, and Discussion. PS: responsible for Introduction, Discussion, framing, and editing. Both authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by NIDCD R01 DC006014. The data used for this paper were acquired from the National Health and Nutrition Examination Survey provided by the Centers for Disease Control and Prevention (https://wwwn.cdc.gov/nchs/nhanes/default.aspx).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fdgth.2021.723533/full#supplementary-material

REFERENCES

1. Brown L, Mahomed-Asmail F, De Sousa KC, Swanepoel DW. Performance and reliability of a smartphone digits-in-noise test in the sound field. *Am J Audiol.* (2019) 28:736–41. doi: 10.1044/2019_AJA-HEAL18-18-0161
2. Jansen S, Luts H, Wagener KC, Frachet B, Wouters J. The French digit triplet test: a hearing screening tool for speech intelligibility in noise. *Int J Audiol.* (2010) 49:378–87. doi: 10.3109/14992020903431272
3. Smits C, Kapteyn TS, Houtgast T. Development and validation of an automatic speech-in-noise screening test by telephone. *Int J Audiol.* (2004) 43:15–28. doi: 10.1080/14992020400050004
4. Margolis RH, Killion MC, Bratt GW, Saly GL. Validation of the home hearing test. *J Am Acad Audiol.* (2016) 27:416–20. doi: 10.3766/jaaa.15102
5. Mosley CL, Langley LM, Davis A, McMahon CM, Tremblay KL. Reliability of the home hearing test: implications for public health. *J Am Acad Audiol.*(2019) 30:208–16. doi: 10.3766/jaaa.17092
6. Pragt L, Wasmann JA, Eikelboom RD. Automated and machine learning approaches in diagnostic hearing assessment: a scoping review. *PsyArXiv.* (2021) 1–20. doi: 10.31234/osf.io/4phfx
7. Irace AL, Sharma RK, Reed NS, Golub JS. Smartphone-based applications to detect hearing loss: a review of current technology. *J Am Geriatr Soc.* (2021) 69:307–16. doi: 10.1111/jgs.16985
8. Chao W-L. *Machine Learning Tutorial. Digital Image and Signal Processing*. (2011). https://www.semanticscholar.org/paper/Machine-Learning-Tutorial-Chao/e74d94c407b599947f9e6262540b402c568674f6#citing-papers
9. Cruickshanks KJ, Nondahl DM, Fischer ME, Schubert CR, Tweed TS. A novel method for classifying hearing impairment in epidemiological studies of aging: the wisconsin age-related hearing impairment classification scale. *Am J Audiol.* (2020) 29:59–67. doi: 10.1044/2019_AJA-19-00021
10. Bisgaard N, Vlaming MS, Dahlquist M. Standard audiograms for the IEC 60118-15 measurement procedure. *Trends Amplif.* (2010) 14:113–20. doi: 10.1177/1084713810379609
11. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing (2021). Available online at: http://www.R-project.org (accessed July 19, 2021).
12. Sarkar D. *Lattice: Multivariate Data Visualization With R*. New York, NY: Springer. Available online at: http://lmdvr.r-forge.r-project.org (accessed July 19, 2021).
13. Kuhn M. Caret package. *J Stat Softw.* (2008) 28:1–26. doi: 10.18637/jss.v028.i05
14. Hammer B. *Metrics: Evaluation Metrics for Machine Learning*. (2018). Available online at: https://github.com/mfrasco/Metrics (accessed July 19, 2021).
15. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the tidyverse. *J Open Source Softw.* (2019) 4:1686. doi: 10.21105/joss.01686
16. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* (1977) 33:159–74. doi: 10.2307/2529310
17. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. *Inf Process Manage.* (2009) 45:427–37. doi: 10.1016/j.ipm.2009.03.002
18. Charih F, Bromwich M, Mark AE, Lefrancois R, Green JR. Data-driven audiogram classification for mobile audiometry. *Sci Rep.* (2020) 10:3962. doi: 10.1038/s41598-020-60898-3
19. Hall JWI. *Rethinking Your Diagnostic Audiology Batter: Using Value-Added Tests*. AudiologyOnline. (2017). Available online at: www.audiologyonline.com (accessed June 7, 2021).
20. Margolis R. *Automated Audiometry: Then and Now. Audiology Practices*. (2021). Available online at: https://www.audiologypractices.org/automated-audiometry-then-and-now (accessed June 7, 2021).
21. Sabin AT, Van Tasell DJ, Rabinowitz B, Dhar S. Validation of a self-fitting method for over-the-counter hearing aids. *Trends Hear.* (2020) 24:2331216519900589. doi: 10.1177/2331216519900589

22. Klyn NAM, Kleindienst Robler S, Bogle J, Alfakir R, Nielsen DW, Griffith JW, et al. CEDRA: a tool to help consumers assess risk for ear disease. *Ear Hear.* (2019) 40:1261–6. doi: 10.1097/AUD.0000000000000731
23. Ferguson MA, Kitterick PT, Chong LY, Edmondson-Jones M, Barker F, Hoare DJ. Hearing aids for mild to moderate hearing loss in adults. *Cochrane Database Syst Rev.* (2017) 2017:CD012023. doi: 10.1002/14651858.CD012023.pub2
24. Hawkins DB. The effect of signal type on the loudness discomfort level. *Ear Hear.* (1980) 1:38–41. doi: 10.1097/00003446-198001000-00006

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.

frontiers in Digital Health

ORIGINAL RESEARCH
published: 20 August 2021
doi: 10.3389/fdgth.2021.722186

Check for updates

Real-World Hearing Aid Usage Patterns and Smartphone Connectivity

Jeppe Høy Christensen[1]*, *Gabrielle H. Saunders*[2], *Lena Havtorn*[1] *and Niels H. Pontoppidan*[1]

[1] Eriksholm Research Centre, Oticon A/S, Snekkersten, Denmark, [2] Manchester Centre for Audiology and Deafness, School of Health Sciences, University of Manchester, Manchester, United Kingdom

Data for monitoring individual hearing aid usage has historically been limited to retrospective questionnaires or data logged intrinsically in the hearing aid cumulatively over time (e. g., days or more). This limits the investigation of longitudinal interactions between hearing aid use and environmental or behavioral factors. Recently it has become possible to analyze remotely logged hearing aid data from in-market and smartphone compatible hearing aids. This can provide access to novel insights about individual hearing aid usage patterns and their association to environmental factors. Here, we use remotely logged longitudinal data from 64 hearing aid users to establish basic norms regarding smartphone connectivity (i.e., comparing remotely logged data with cumulative true hearing aid on-time) and to assess whether such data can provide representative information about ecological usage patterns. The remotely logged data consists of minute-by-minute timestamped logs of cumulative hearing aid on-time and characteristics of the momentary acoustic environment. Using K-means clustering, we demonstrate that hourly hearing aid usage patterns (i.e., usage as minutes/hour) across participants are separated by four clusters that account for almost 50% of the day-to-day variation. The clusters indicate that hearing aids are worn either sparsely throughout the day; early morning to afternoon; from noon to late evening; or across the day from morning to late evening. Using linear mixed-effects regression modeling, we document significant associations between daily signal-to-noise, sound intensity, and sound diversity with hearing aid usage. Participants encounter louder, noisier, and more diverse sound environments the longer the hearing aids are worn. Finally, we find that remote logging via smartphones underestimates the daily hearing aid usage with a pooled median of 1.25 h, suggesting an overall connectivity of 85%. The 1.25 h difference is constant across days varying in total hearing aid on-time, and across participants varying in average daily hearing aid-on-time, and it does not depend on the identified patterns of daily hearing aid usage. In sum, remote data logging with hearing aids has high representativeness and face-validity, and can offer ecologically true information about individual usage patterns and the interaction between usage and everyday contexts.

Keywords: hearing aids, smartphone connectivity, usage patterns, acoustic environment, K-means clustering

Edited by:
Fan-Gang Zeng, University of California, Irvine, United States

Reviewed by:
Shuo Wang, Capital Medical University, China
Stefan Launer, Sonova, Switzerland

***Correspondence:**
Jeppe Høy Christensen
jych@eriksholm.com

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: *08 June 2021*
Accepted: *27 July 2021*
Published: *20 August 2021*

Citation:
Christensen JH, Saunders GH, Havtorn L and Pontoppidan NH (2021) Real-World Hearing Aid Usage Patterns and Smartphone Connectivity. Front. Digit. Health 3:722186. doi: 10.3389/fdgth.2021.722186

INTRODUCTION

The real world benefit obtained from hearing aids varies considerably across individuals (1). This is thought to be due to individual differences in cognition and working memory (2), variability in hearing aid programming (3, 4), and most relevant to this paper, differences in contextual exposure and needs (5, 6), which can be influenced by both environmental factors (7), individual preferences (8, 9), and listening intentions (6). For example, an individual who needs to hear speech in complex listening environments, such as during a large meeting in a noisy office, will obtain less real world hearing aid benefit than an individual whose listening needs are lower demand such as one-on-one conversations in a quiet room (10).

In the past it has been difficult to assess real world hearing aid use and outcomes because assessments have had to rely on retrospective reports from users combined with limited information collected via intrinsic hearing aid data logging or self-reports. The perceived total use-time is often over-estimated (11–13) or inaccurate depending on hearing aid experience (14, 15). With intrinsic data logging, information about use time, program usage, time spent streaming data, sound pressure level input, listening environment classification, directional microphone settings, and signal-to-noise ratio is stored within the hearing aid (16). However, because space for data storage within the hearing aid is limited, intrinsically-logged data are saved cumulatively over time (17, 18). This not only limits the temporal resolution of the data but also means it is not possible to link patterns of hearing aid usage to specific sound environments and listening conditions.

Ecological momentary assessment (EMA) is another approach that has been used to examine real world hearing aid use and benefit. With EMA, participants describe real world experiences in real time in their own natural environments (19). Recently EMA has been used in several hearing-related studies [see review by Holube et al. (20)]. Most recently it has been used to examine how momentary contextual factors influence subjective ratings of hearing aid outcome (5, 21), differences between listening behaviors of people seeking hearing aids and those who already use them, with a view to predicting who will and will not become a successful hearing aid user (22), and to compare *in-situ* vs. retrospective reports of hearing aid outcome (21, 23). In many hearing-related studies, the participant's EMA reports have been linked in time to an analysis of the sound environment collected by the hearing aids. In some studies, an EMA survey is triggered when a predetermined acoustic environment is encountered (5, 24), while in others, a sound analysis is conducted when a randomly prompted or voluntarily initiated EMA assessment occurs (21, 25). In both instances, there is a direct link between an acoustic analysis of the sound environment and responses to an EMA survey. However, EMA is somewhat intrusive, requiring participants to be willing to answer surveys on multiple occasions during the day. Further, recent work has shown that in certain situations EMA surveys go uncompleted more frequently than in other situations. Specifically, Schinkel-Bielefeld et al. (26) found that participants oftentimes skipped EMA surveys in situations when it was considered inappropriate to respond such as during a conversation or a church service, while Wu et al. (27) showed that surveys were most often not completed in noisier situations containing speech, in which directional microphones and noise reduction algorithms are typically enabled (27). As noted by Wu et al., this will lead to biases in interpretation of subjective EMA reports.

An approach that is unobtrusive and minimizes user burden is to use ongoing remote data logging in which data collected by the hearing aids are automatically and continuously transferred from the aids to a smartphone via a Bluetooth connection. Because smartphones usually have ample space for data storage, fine-grained temporal data can be collected for numerous acoustic parameters. Remote data logging has the potential to provide new insights into ways in which hearing aids are being used in real life (11, 28), and provides new opportunities for research and for development. For example, it has been used to establish evidence regarding daily hearing aid usage for public health decision-making (29, 30), to augment EMA with hearing devices with acoustic information (21, 25) and to support development of advanced hearing aid technology (31–34). Further, by combining remotely logged hearing aid data with that collected from fitness trackers it has been shown that heart rate is significantly and continuously moderated by dimensions of the ambient acoustic environment (32). Finally, remote data logging has also been introduced in a commercially available hearing aid for users to track their own hearing aid usage and sound exposure [HearingFitness™, (35)].

Remote logging is in its infancy and as such there are many unknowns about its reliability and validity. For example, for stable flow of data, most remote logging requires a constant Bluetooth connection between the smartphone and the hearing aids. However, Bluetooth connections can be unstable and not all hearing aid users always keep their smartphone close by, which can then lead to data loss. A thorough investigation of the validity and representativeness of remote data logging is therefore needed to validate its use in audiological research and clinical work.

In this study, we investigate the representativeness of remote data logging to understand whether it provides a quantitative account of hearing aid usage and its association with everyday contextual factors, so that in the future, individual deviations from group-level patterns can be identified and used to support patients and hearing care professionals. We compare information obtained through remote data logging with that obtained through intrinsic data logging to assess the extent to which remotely-logged data reflect daily hearing aid on-time and sound exposure. Our analysis leverages an observational and longitudinal dataset from in-market hearing aid users. The dataset has been reported upon and validated elsewhere (32) and similar data are publicly available online (31).

METHODS

Participants and Ethics

Participants were users of Oticon hearing aids (Oticon A/S, Smørum, Denmark) who had signed up to use the HearingFitness™ feature via the Oticon ON™ remote control app and used it at least once with their hearing aids between

May the 15th and September the 30th 2019. The Oticon ON™ remote control app can be used by users of Oticon Opn™ and newer hearing aids and it provides an interface to keep track of battery status, changing listening program, adjusting volume etc. When signing up to use the HearingFitness™ feature via the app, participants gave informed consent for data to be collected, stored, and used for research purposes on aggregated levels. No other action is required by the participants when using the feature. Note that no personal identifiers nor qualitative characteristics (e.g., age, hearing loss) are being collected. However, since the participants represent a random sample of typical hearing aid users, we speculate that 6 in 10 are male, are aged around 74 years based on hearing aid user surveys (36).

No ethical approval is necessary for this study according to Danish National Scientific Ethical Committee (https://www.nvk.dk/forsker/naar-du-anmelder/hvilke-projekter-skal-jeg-anmelde).

Data

We extracted a convenience sample of remote logging from 64 hearing aid users. The data, stored in the HearingFitness™ database, consist of minute-based timestamped remote data logs of ambient sound pressure levels (SPLs) and signal-to-noise levels (SNRs) estimated from within the hearing aid processing (35). In addition, for each log, the intrinsically accumulated hearing aid on-time (in seconds) since last clinical visit is stored. More details about the dataset can be found in Christensen et al. (32). In all following analyses, only data logged between 6:00 and 00:00 are used. This is done to minimize confounding effects from night-time logging occurring while the hearing aids were not actively worn (e.g., by forgetting to turn off hearing aids at night). In addition, only participants with at least 2 days of remote data logging were included to enable a comparison of remote logging and the intrinsically accumulated hearing aid on-time. After filtering, the data represents bi-lateral hearing aid usage from 62 users across a combined total of 1,099 days. When separated by hearing aid side, the data consist of 2,054 days of usage.

Pre-processing

Daily hearing aid usage can be estimated from two sources: the data log timestamps obtained through remote datalogging (henceforward referred to as T_{remote}), and the intrinsically accumulated hearing aid on-time (henceforward referred to as $T_{intrinsic}$). From the remote logs, each timestamp represents 60 s of hearing aid on-time and smartphone connectivity. Thus, T_{remote} is estimated by counting the number of timestamps within a selected time-window that are longer than 1 min. For example, counting 40 unique timestamps in 1 h equates to 40 min of hearing-aid use in that hour. For data to be saved in a remote log, smartphone connectivity to the hearing aid is required. On the other hand, the intrinsically accumulated hearing aid on-time represents absolute hearing aid on-time regardless of smartphone connectivity. More specifically, from a series of N consecutive data logs with accumulated on-time times $t_1, \ldots, t_N$ spanning time-window T, the total on-time is estimated as $T_{intrinsic} = \sum_{n=2}^{N} t_n - t_{n-1}$, which compensates for potential gaps from inter-log times longer than 1 min (e.g., due to momentarily lost Bluetooth connectivity). A comparison of hearing aid usage from the two estimators (T_{remote} and $T_{intrinsic}$) within a time-window provides insight into the amount of data lost due to a lack of smartphone connection. Connectivity is thereby defined as the proportion of time a hearing aid is connected to a smartphone for the duration of the inspected on-time.

Statistical Analysis

Clustering of Usage Patterns

We used K-means unsupervised clustering to identify archetype usage patterns, where a usage pattern is defined by the minutes of hearing aid usage per hour during a day from 6:00 to 00:00. Thus, the input data consisted of vectors $X_d = \{x_i; i \in N\}$ with variables x_i defined by usage in minutes, x, per hour, i, and days $d = \{1, .., D\}$ pooled among all participants in the sample. The K-means algorithm then assign each X_d to a cluster centroid so that the Euclidian distances between all X_d's and centroids are minimized (i.e., the total sum of squares distance) while iteratively selecting cluster centroids that minimize the intra-cluster variation (i.e., the within-cluster sum of squares distance). The optimal number of clusters was selected based on the elbow method (37), which selects the number of clusters that lead to only a minor change in the total sum of squares with the addition of more cluster centroids. Finally, the clustering was evaluated with the Silhouette Coefficient (38) assessing how densely clustered each X_d is around the centroids. The Silhouette Coefficient (SC) is calculated for each vector X_d as:

$$SC = \frac{b - a}{\max(a, b)},$$

where a is the mean Euclidian distance between a vector and all other vectors in the same cluster, and b is the mean Euclidian distance between a vector and all other vectors in the next nearest cluster. In general, the SC is bounded between -1 for incorrect clustering and $+1$ for highly dense clustering. Coefficients around zero indicate overlapping clusters (that is, the distance from a vector to the two clusters is equal) and the coefficient is positive and higher when clusters are dense and well-separated.

The K-means optimization was implemented in R using "cluster" package [version 2.1.0, (39)].

Associating Hearing Aid Usage With Ambient Sound Characteristics

The associations between daily hearing aid usage and parameters of the acoustic environment (SPL and SNR) were tested using a linear-mixed model (LMM). LMMs are ideal for regressing longitudinal and hierarchical multi-level data allowing for random offsets and slopes from grouping variables (40). The model included total daily hearing aid usage as the dependent variable in hours. The independent predictors consisted of the daily (logarithmic) mean SPL–i.e., the equivalent continuous mean SPL (SPL L_{eq}), daily median SNR, and the daily standard deviation of the SPL (SPLSD). These predictors represent the intensity, the quality, and the loudness diversity of the daily sound exposure. For simplicity, we did not include the daily standard deviation of the SNR because SNR is a relative measure (i.e., a difference between noise and signal) and, thus, daily variations in SNR cannot be easily interpreted.

The random effect's structure accounted for the day of the week, hearing aid laterality, and individual offsets in daily usage (i.e., random intercepts) and individual sensitivity toward acoustic characteristics (i.e., random slopes). Accounting for laterality effectively ensure that coefficients from the LMMs are estimated based on the average ambient sound sensed between the left and right hearing aid, while accounting for individual sensitivity toward acoustic characteristics with random slopes ensure that individual differences in e.g., loudness growths functions, do not affect the results.

Besides inspecting coefficient magnitude and confidence intervals, significance of predictors was assessed by likelihood ratio testing against an intercept-only model. Prior to modeling, extreme outliers in the acoustic characteristics (i.e., values <1% quantile and larger than the 99% quantile) were removed to ensure normality of the residuals. This removed 147 observations (hearing aid usage days) in total.

To assess the degree of multicollinearity within the model, the generalized variance inflation factor (GVIF) was computed using the 'car' package in R [version 3.0.8, (41)]. The GVIF is a generalization of the variance inflation factor (VIF) that can be applied to categorical explanatory variables (42). Values of GVIF <4 are usually considered to be acceptable (43).

Effects size estimates were computed by separating the explained variance by fixed effects alone or by the full model using the pseudo-R-squared for Generalized Mixed-Effect models implemented in the "MuMIn" package [version 1.43, (44)] in R. LMMs were fitted in R using the "lmerTest" package [version 3.1, (45)].

RESULTS

Identifying Hearing Aid Usage Patterns

The fine-grained temporal structure of the data collected by remote logging opens the possibility of examining hourly usage patterns to investigate how hearing aid use varies with e.g., time of day. Here, a K-means clustering algorithm is applied to the pooled hourly usage patterns among all participants and days. Note that we only include usage data from the hearing aid that had the highest daily total. The "elbow" approach, i.e., visual inspection of between-cluster variance vs. cluster number, suggested the existence of four clusters (**Figure 1A**). These clusters explain 48.5% of the total variance observed between the daily usage patterns in the data. As evident from **Figure 1B**, the four clusters can be well-separated when inspecting the two most contributing dimensions of a principal component's analysis of the usage patterns. Moreover, the overall mean Silhouette Coefficient (**Figure 1C**) was 0.37 ($SD = 0.24$) and separated by cluster 1 to 4 it was 0.21 ($SD = 0.18$), 0.18 ($SD = 0.18$), 0.58 ($SD = 0.18$), and 0.39 ($SD = 0.16$). Thus, the Silhouette Coefficients across pooled usage days were predominantly positive indicating successful clustering. **Figure 1D** shows the average time-course of usage patterns belonging to each cluster, which were associated with 19.1, 21.5, 33.5, and 25.9%, respectively, of the 1,099 days pooled among participants. Cluster 1 indicates a pattern of use predominantly in the morning morning/noon, cluster 2 with predominant use in the afternoon/evening, cluster 3 with low use throughout the day in brief epochs of time, and cluster 4 with constant usage throughout the day.

Modeling Daily Hearing Aid Usage

Together with usage data, each remote log contains estimates of the ambient acoustic environment. Such data enables investigations of how environmental factors influence hearing aid usage—that is, an examination of true ecological hearing aid use.

The daily SPL L_{eq}, daily median SNR, and the daily SPLSD were used as predictors to daily total hearing aid usage with usage estimated from remote logging. **Figure 2** shows histograms of each acoustic parameter. The SPL L_{eq}s and SPLSDs are approximately normally distributed with a grand median of 68.1 dB ($SD = 6.9$ dB) and 10.56 dB ($SD = 2.6$ dB), respectively. The daily median SNRs are right skewed with a grand median of 3.6 dB ($SD = 1.4$ dB), a minimum of 1.6 dB and maximum of 8.6 dB. In fact, only 6.4% of the raw logged SNRs were zero or negative.

The full LMM model ($n = 1{,}907$ days) explained 35% of the variance in day-to-day total hearing aid usage and of those 15%-point was related to the acoustic predictors alone. In addition, the acoustic predictors significantly improve the model's account of daily total hearing aid usage compared to an intercept-only model [Likelihood ratio test: $\chi^2_{(12)} = 229.38$, $p < 0.001$] and there are significant main effects associated with each. The fitted coefficients for SPL L_{eq} and SPLSD are positive [$\beta = 0.89$, 95% CI = (+0.52 to +1.27), $p < 0.001$; $\beta = 1.30$, 95% CI = (+1.00 to +1.60), $p < 0.001$], while SNR had a negative association with daily total usage [$\beta = -0.98$, 95% CI = (−1.43 to −0.53), $p < 0.001$]. Day of the week (implemented as a random effect) was significant [Likelihood ratio test: $\chi^2_{(1)} = 18.85$, $p < 0.001$] and the largest difference in the intercept of total daily hearing aid usage was Monday and Tuesday (+0.65 and +0.45 h) and Friday (−1.21 h). The VIFs were all <1.44. These results suggest that daily total usage is higher on days with more intense and more diverse sound environments, while days with easier listening environments (more positive SNRs) exhibited lower total usage. The lack of covariance between the acoustic predictors indicated by the low VIF suggests that these effects were independent of each other.

Representativeness of Remote Data Logging

While the analysis presented in the preceding sections highlights the potential of using remote logging to understand hearing aid user's ecology, the face-validity and representativeness of such data is still unknown since its generation requires a constant stable Bluetooth connection between hearing aid(s) and a smartphone. Here, we assess the extent to which the minute-based remote data logs correspond to true hearing aid on-time by directly comparing the daily total and the participant-average hearing aid usage measured by the intrinsic accumulation ($T_{intrinsic}$) with the data log timestamps (T_{remote}). Note that $T_{intrinsic}$ is not subjected to data loss, thus, it reflects true hearing aid on-time from the first to the last remote log each day. In

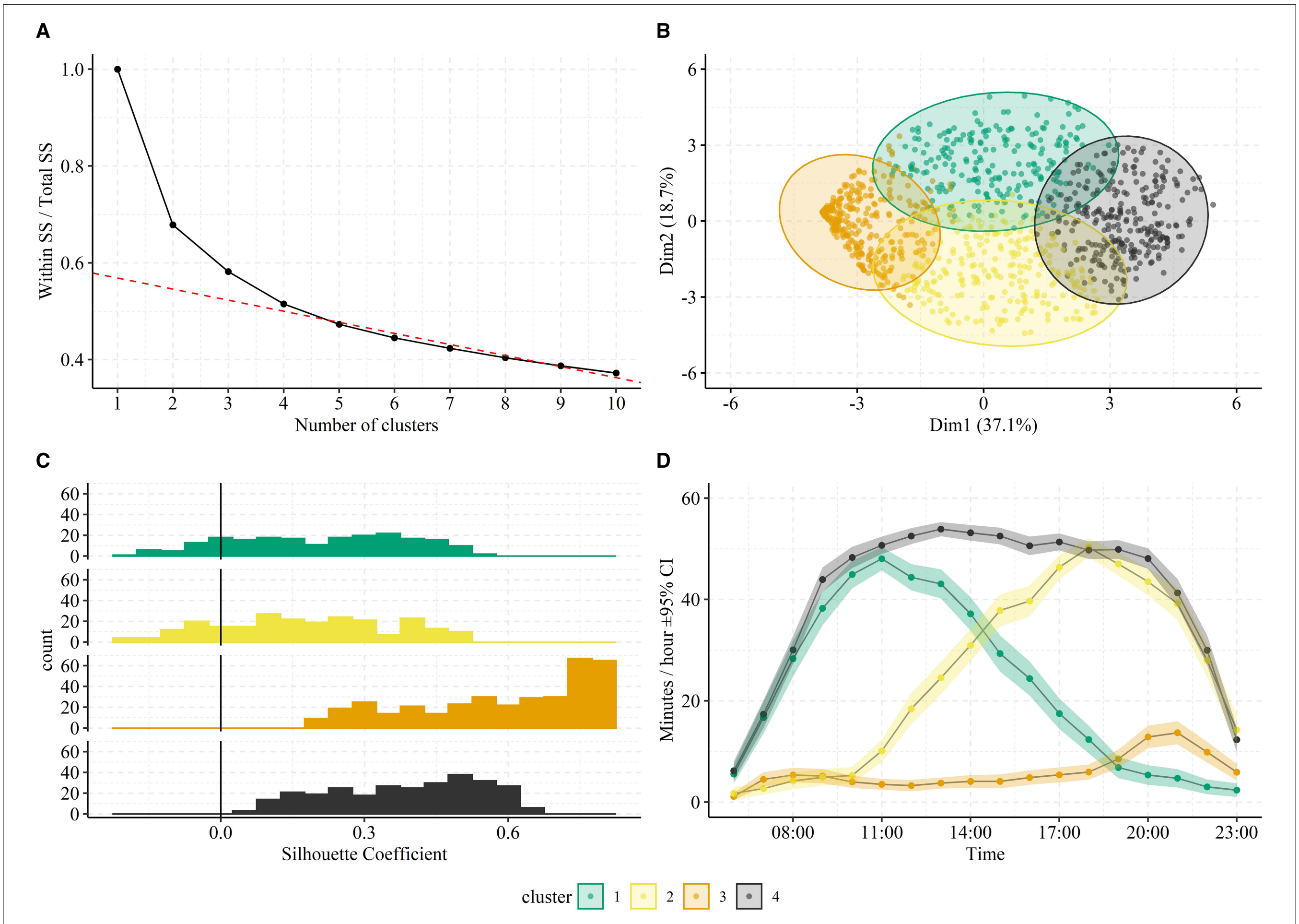


FIGURE 1 | Clustering of pooled hearing aid usage patterns. In **(A)**, the total within-cluster sum of squares (SS) normalized by the total SS is plotted against number of clusters. The "elbow" approach suggests selecting four clusters since adding additional clusters result in minor improvements. This is illustrated by the red dotted line, which represents the best linear fit to >4 clusters. In **(B)**, the two most contributing dimensions of a principal component analysis of the usage patterns are plotted against each other with separation of the four identified clusters (colors). In **(C)**, the average time-course of usage patterns belonging to each cluster is shown. The shaded area in represents the 95% confidence intervals across usage patterns belonging to each cluster. The clusters one to four were associated with 19.1, 21.5, 33.5, and 25.9% of the participants days, respectively. In **(D)**, distributions of Silhouette Coefficients are shown for each cluster. The vertical line at 0.0 represents the threshold for poor clustering (<0).

Figure 3 we plot hearing aid usage accumulated across 24 h (6:00 to 6:00 next day) on a random day for two participants to illustrate differing patterns of connectivity. Participant S43 exhibits a stable connectivity, where the periods without remote data logs correspond closely to times when the hearing aids were turned off. In contrast, in participant S64 the hearing aid usage computed via remote data logs diverge from that obtained through the intrinsically accumulated data from 16:00 onwards. It should thus be clear that the total daily hearing aid usage will differ depending on the estimation used.

Pooled across all participants, individual days of data, and hearing aid laterality (totaling 2,054 observations) the median difference ($T_{remote} - T_{intrinsic}$) between daily usage estimates is 1.25 h (SD = 2.53 h). **Figure 4** shows histograms of daily usage from both estimators (**Figures 4A,B**, respectively), separated by hearing aid side together with a 2D histogram comparing the daily usage (across hearing aid side) from the two estimators directly (**Figure 4C**).

When averaged across each participants' data and then across all participants (see **Figure 5A**) the grand median difference in usage estimates (50% quantile in **Figure 5C**) is 1.59 h (SD = 1.26 h). The fact that the pooled median difference in usage estimates is ∼0.3 h lower than the grand median difference in usage estimates suggests that some participants consistently exhibit poorer connectivity than others. This might also explain the rather large difference in the two histograms of daily usage (**Figures 4A,B**)—that is, it might be driven by few participants experiencing many days with poor connectivity. However, from the histogram in **Figure 5A**, data loss from poor connectivity "left-shifts" the distribution of average daily hearing aid usage to lower values but doesn't otherwise change its shape. In addition, the scatterplot on **Figure 5B** shows that most participants

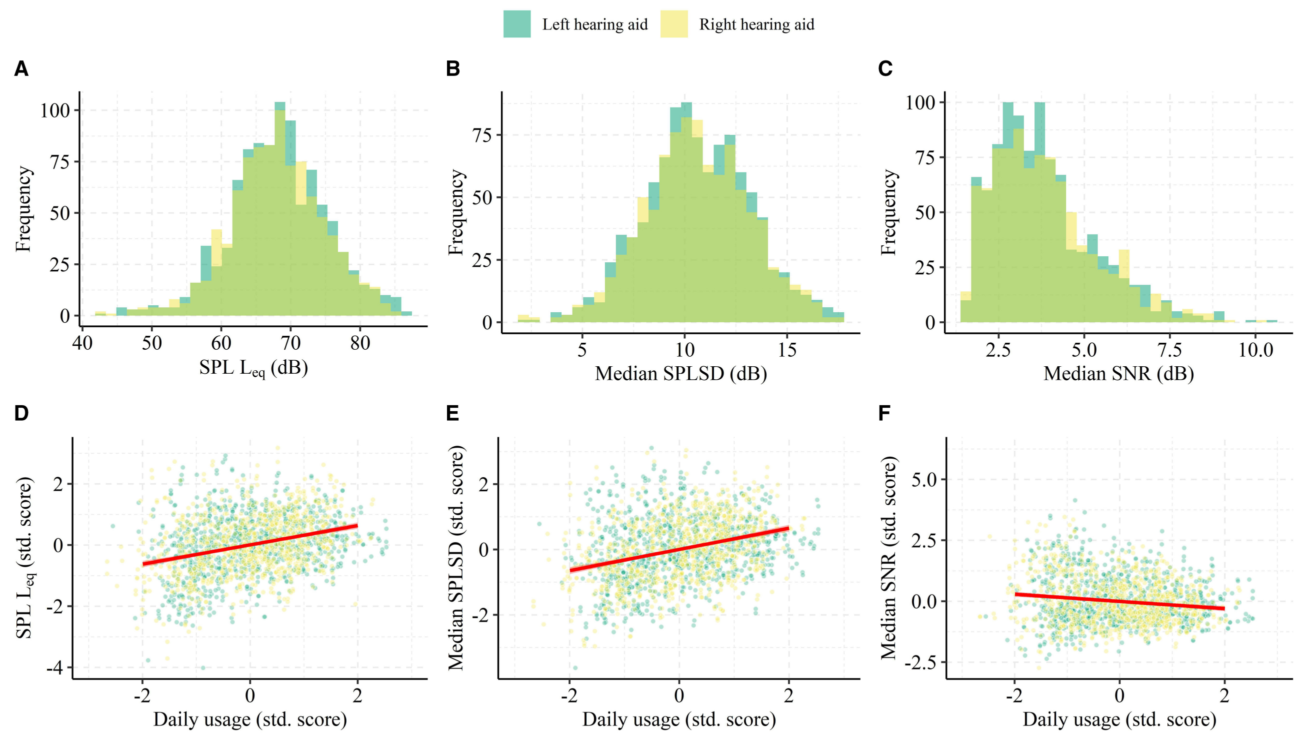


FIGURE 2 | Data for LMM modeling of the association between daily hearing aid usage and the ambient acoustic environment. In **(A–C)**, histograms of daily SPL L_{eq}, the standard deviation of the SPL (SPLSD), and the daily median SNR pooled from all participants. In **(D–F),** scatterplots of the acoustic predictors and the daily usage, both in participant-specific standardized score (i.e., Z-score). For visualization of the trends, the red line represents the best linear fit to data within the interval −2:2 in Z-score.

experience data loss (all points are above the dashed line), that the relative data loss does not depend on average daily usage (points fall parallel to the diagonal dashed line), and that no clear outliers are present. These results suggest that the loss of data in the remote logs due to connectivity is in fact not specific to certain participants but rather it occurs generally. When inspecting the participant-specific cumulative distribution functions (**Figure 5C**) most participants follow a similar curve, but participants with few days of data logging (identified by traces with large steps on **Figure 5C**) exhibit either extremely poor or good connectivity.

Finally, when inspecting the distributions of estimated daily usage in **Figure 4C**, it seems that days with higher hearing aid on-time exhibit a larger difference in usage estimates. That is, loss of remote logs from poor connectivity might be a function of daily hearing aid on-time rather than a constant. Intuitively, this makes sense in that the more time the hearing aids are on, the greater the likelihood that there will be connectivity issues during that time. In **Figure 4C**, the trend can be seen by the deviation of the mode (i.e., darker yellow squares) of the daily usage from remote logging above the diagonal. This is further corroborated by comparing the histograms in **Figures 4A,B**. The largest visible difference is occurring between 10 and 16 h of daily usage. To investigate this, the daily difference in usage estimates were computed relative to the proportion of daily hearing aid on-time (ΔT_{rel}) by: $\Delta T_{rel} = \frac{(T_{intr} - T_{remote})}{T_{intr}}$, where T_{intr} is the daily usage from intrinsic logging (i.e., the true hearing aid on-time) and T_{remote} is the daily usage estimated from remote logs. Next, we computed the average ΔT_{rel} stratified by participant and hearing aid on-time (from intrinsic accumulation) in discrete bins. If connectivity is a fixed ratio of hearing aid on-time, we would expect a constant ΔT_{rel} across all bins. On the other hand, if connectivity is a constant, we would expect a declining ΔT_{rel} with increasing hearing aid on-time.

Figure 6 shows boxplots of ΔT_{rel} for each bin. Across all bins, the median ΔT_{rel} is 0.15, which equals an overall connectivity [calculated as $100\bullet(1 - \Delta T_{rel})$] of 85% of the time a hearing aid is on. However, there is a significant change in ΔT_{rel} with each step of on-time (LMM regression adjusted by participant, [$F_{(1,1,776.2)} = 86.88$, $p < 0.001$], suggesting that connectivity first decrease slightly and is lowest between 3 and 7 h of daily usage and then continuously increase. In addition, visual inspection suggests that when on-times are longer than 5 h the inter-individual variability decrease, indicating more stable connectivity patterns across participants. Thus, connectivity is a constant when on-time is >5 h and therefore affect days with higher daily usage less than days with lower daily usage. We also assessed if the connectivity depended on the daily usage patterns identified by clustering (**Figure 1**). For cluster 1 to 4 the connectivity was 87.01% ($SD = 23.9\%$), 85.7% ($SD = 20.8\%$), 85.1% ($SD = 21.9\%$), and 83.8% ($SD = 23.0\%$), respectively.

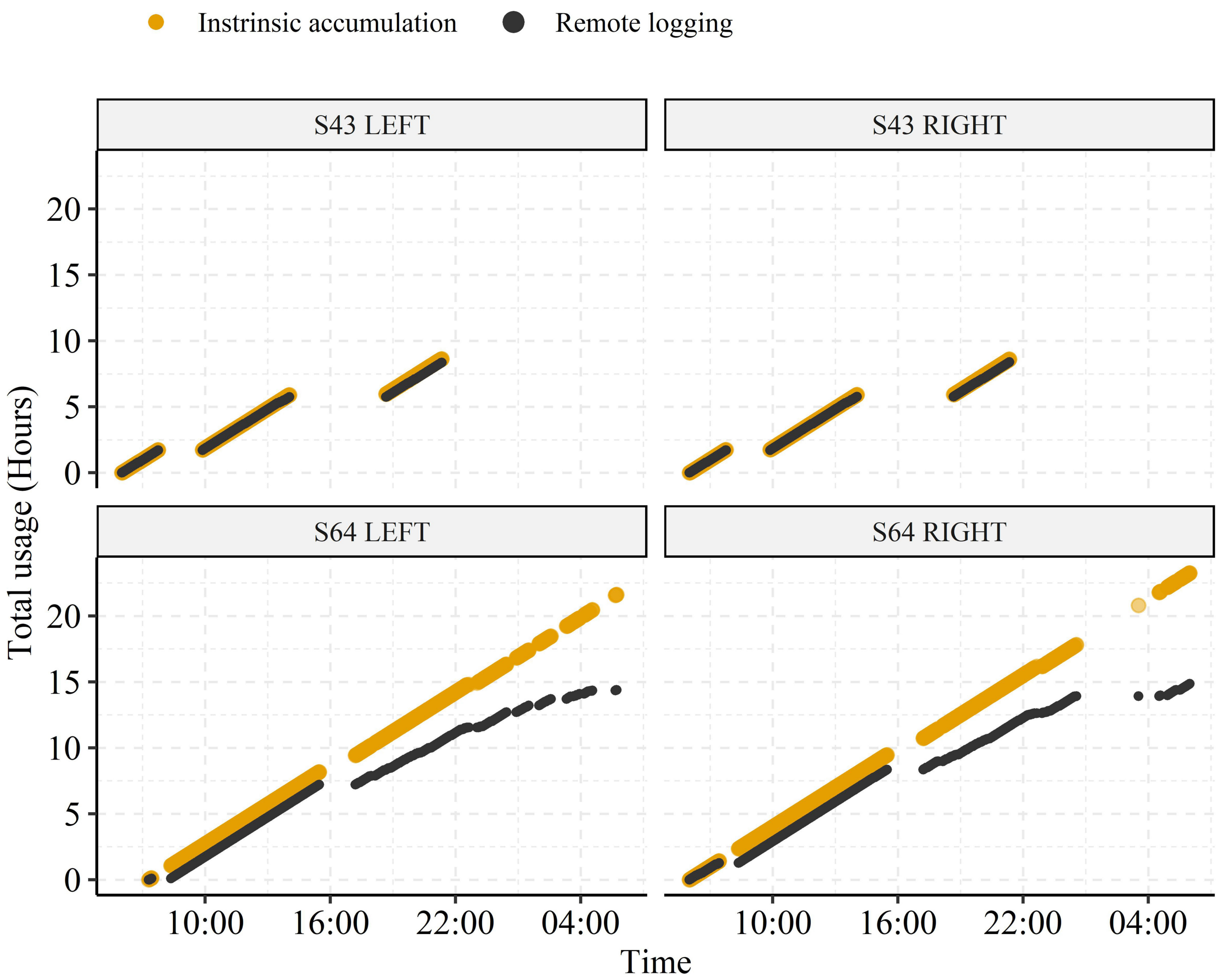


FIGURE 3 | Representative examples of hearing aid usage accumulated throughout the day from two participants (S43 and S64) across 24 h (06:00 to 06:00 next day). The usage is estimated from either intrinsic accumulation ($T_{intrinsic}$) or from the minute-based remote datalogging (T_{remote}).

DISCUSSION

In this study, we analyzed a longitudinal real-world dataset to demonstrate how remote data logging with hearing aids can produce behavioral and ecological insights into everyday and hourly hearing aid usage. At the same time, we evaluated the validity and representativeness of such data logging by estimating connectivity (i.e., the proportion of time a hearing aid is Bluetooth connected to a smartphone for the duration of the total hearing aid on-time). The data consists of minute-based remote logs collected via Bluetooth transfer of data from hearing aids to smartphones of hearing aid on-time and measures of the ambient acoustic environment sensed by hearing aid microphones.

The K-means algorithm applied to the pooled usage data estimated from remote data logs (**Figure 1**) identified four distinct clusters of hourly hearing aid usage patterns. Importantly, the Silhouette Coefficients (**Figure 1C**) demonstrate that the archetypical usage patterns represent an acceptable clustering of the data (i.e., values predominantly above zero). However, there are some days that do not cluster well (especially in cluster 1 and 2) suggesting that, besides the four archetypical patterns, few days exhibit usage patterns that do not categorically fall into one of the four clusters. Interestingly though, the principal component analysis of all patterns (**Figure 1B**) show only a small degree of overlap among clusters, which suggests that the four patterns occur independently. That is, participants predominantly use their hearing aids according to only one of these patterns on any given day.

Our examination of the relationship between characteristics of the ambient acoustic environment and the daily total hearing aid usage showed significant associations. We saw that days with higher usage were associated with higher ambient sound levels (SPL L_{eq}), greater sound diversity (SPLSD), and more difficult listening conditions (lower SNR). These effects can perhaps be attributed to the fact that the longer a hearing aid is worn, the higher is the probability of experiencing varied acoustic

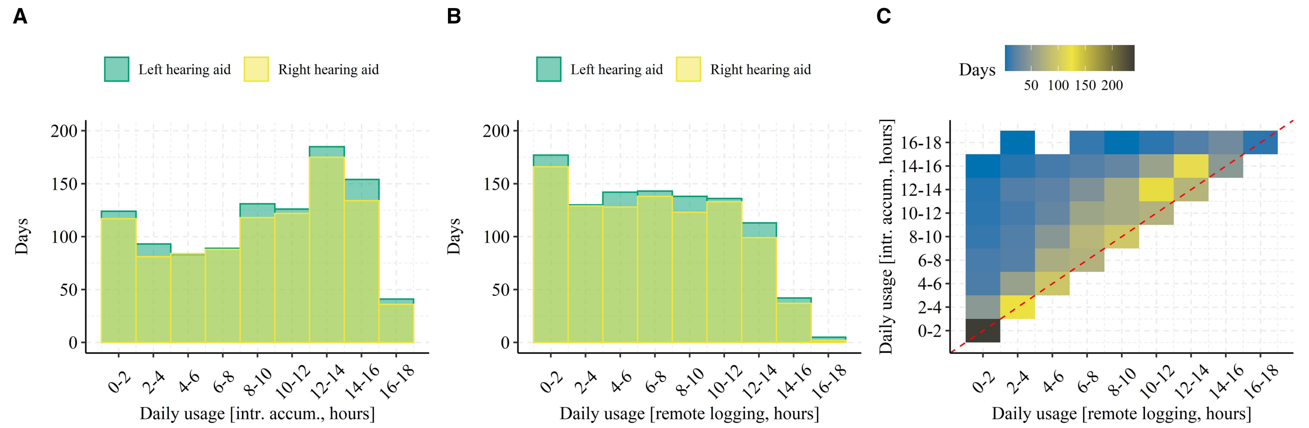


FIGURE 4 | Pooled daily hearing aid usage among all participant days. In **(A)**, a histogram of daily usage estimated from intrinsic accumulation and pooled among all participants, days, and separated by hearing aid side is shown. In **(B)**, the same is shown but for daily usage estimate from remote logs. In **(C)**, the estimated daily usage is shown as a 2D density histogram with the intrinsic accumulation on the y-axis and the connectivity-dependent remote logging on the x-axis. Data falling on the diagonal (red dashed line) represent days with no or <2 h of connectivity issues (i.e., comparable estimation of daily usage from intrinsic accumulation and remote logging).

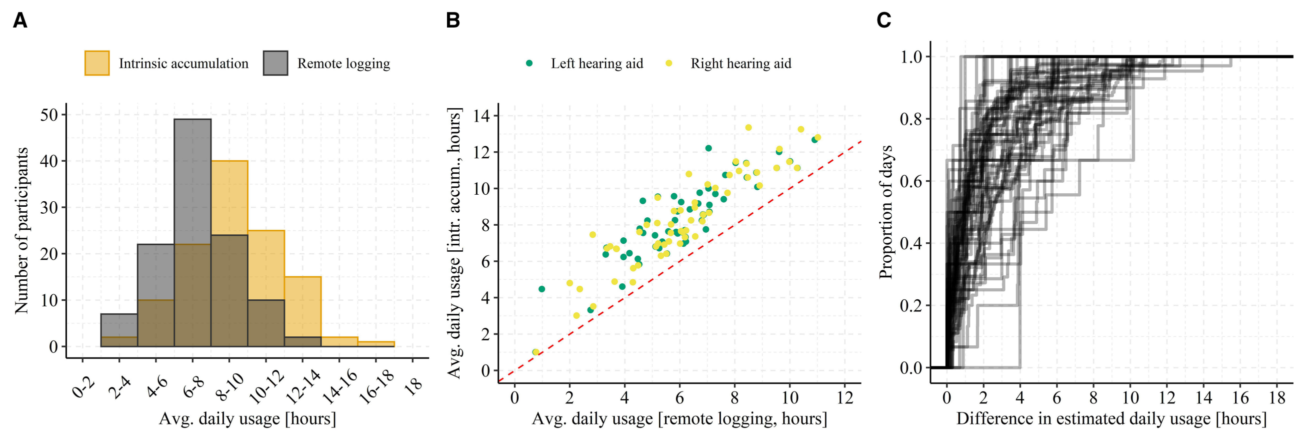


FIGURE 5 | Average daily hearing aid usage. In **(A)**, a histogram of average daily usage (across left and right hearing aid) binned by 2 h intervals and separated by estimation method (i.e., intrinsic accumulation or remote logging). In **(B)**, a scatterplot of average daily usage for both left and right hearing aid from each participant with usage estimated from intrinsic accumulation on the y-axis and from remote logging on the x-axis is shown. The red dashed line represents a y = x relationship. In **(C)**, each trace represents one participant's cumulative distribution function for the daily difference between usage estimates. Note that traces with big jumps ("steps") on the y-axis are indicative of participants with only few days of data logging.

environments that include high sound pressure levels and poor signal to noise ratios. However, this might also reflect the fact that individuals choose to wear their hearing aids when their communication needs are greatest, which tends to be in situations in which sound levels are higher and the acoustic environment is more complex, such as at home in the kitchen, in a restaurant, or in meetings (46). Indeed, while past research has had to rely on participant report, data show that active communication often takes place in "noisy" situations. For instance, the participants of Walden et al. (47) reported that 63% of active listening situations involved the presence of noise, Keidser (48) participants reported that 26% of their time was spent talking in quiet with 24% talking when noise was present. Sound recordings show similar. Wagener et al. (49) asked participants to make short recordings of "different situations from your daily life." They found that about half of the recordings involved conversation, with 11.5% taking place "without background noise," 18% "with background noise-2 persons," and 10% "with background noise with more than two persons." Using EMA and simultaneous sound recordings Wu et al. (50) and Timmer et al. (24) found most EMA reports to be provided for listening situations with low SPLs and high SNRs. However, as noted in the introduction, data collected via EMA are biased toward quieter listening situations because participants often choose not to provide EMA responses when in social situations (26, 27), thus these data neither support nor refute our findings. This also highlights the value of remote data logging to understand the interactions between hearing aid use and the acoustic environment used here. In fact, we

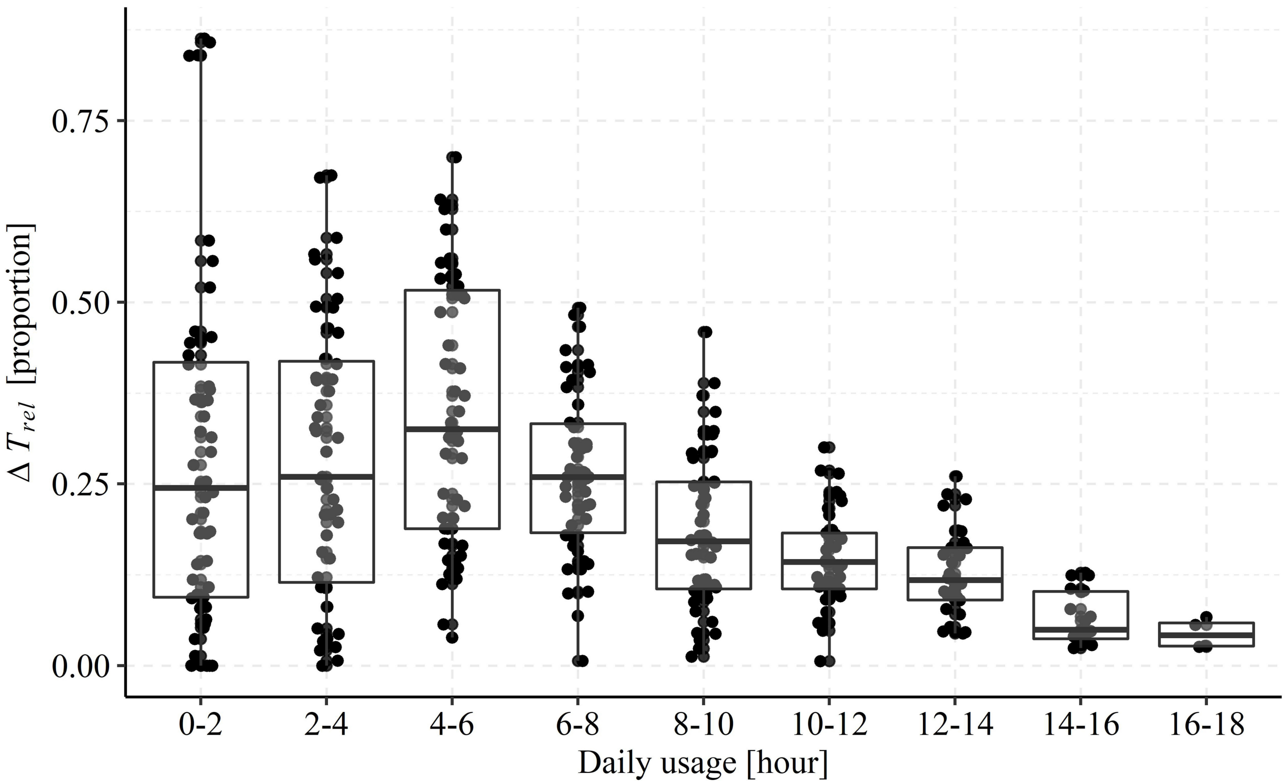


FIGURE 6 | Boxplots and single observations of mean relative difference in usage estimates stratified by the true daily on-time. Each point represents the mean relative difference (ΔT_{rel}) for one participant across all days registered with a total on-time falling into the bins on the x-axis. Each box encapsulates the interquartile range (25th percentile to 75th percentile) and the horizontal line represents the median. The whiskers extend to the largest value within 1.5 times the interquartile range above the 75th percentile (top) and smallest value within 1.5 times the interquartile range below the 25th percentile (bottom).

argue that given the shortcomings of self-reports and the lack of fine temporal information from intrinsically logged data remote data logging with smartphones using Bluetooth-enabled hearing aids provide a valid way to accurately map daily usage patterns from populations of hearing aid users on a minute-based unobtrusive basis. This is further corroborated by the fact that the representativeness of remote logging is high in the currently examined sample of hearing aid users. We found a median difference of 1.25 h between the true daily hearing aid on-time and the daily usage estimated from the remote data logs when pooled across all 2,054 observations (participants, days, hearing aid side), which corresponds to an overall connectivity of 85%. The absolute difference of 1.25 h seemed to be constant regardless of total hearing aid on-time >5 h (**Figure 6**), type of usage pattern (**Figure 1**), and participant-specific average daily hearing aid use (**Figure 5B**). We assume this difference is due to periods with a lack of smartphone connectivity, which can occur when phone reception is poor, Bluetooth is disabled, or when the hearing aids are out of range of the smartphone. The latter is likely due to the participant not carrying their smartphone with them when using hearing aids. In sum, remotely logged data are a more accurate reflection of hearing aid usage for individuals who wear their hearing aids for longer each day than it is for those who wear their hearing aids for less time. Thus, for detailed and accurate investigations into daily usage patterns, data from days with high connectivity loss should be discarded by limiting analysis of data to individuals who wear their hearing aids for a considerable period of time. Finally, the participant-specific average daily hearing aid usage and the difference between estimators (see **Figure 5**) suggests that those participants that only contributed with few days of hearing aid data exhibited large inter-individual variability in their connectivity (**Figure 5C**), but that the average usage across days was not predictive of connectivity (**Figure 5B**).

Clinical Relevance

Use of remote data logging potentially has benefits on an individual patient level as well as the population level discussed above. It has the potential to provide deeper insights into an individual's listening lifestyle and how and when they use their hearing aids than has been previously possible. As such, it could then be used by the audiologist to provide counseling at a more fine-grained level than intrinsic data logging permits (16). Perhaps more interestingly and relevant to this paper, at a technological level, the hearing aid could provide automated messages when atypical hearing aid use is detected and/or automatically change hearing aid setting when specific combinations of acoustic parameters and time of day/day of the week are encountered. To "calibrate" these changes to meet patient needs, there could be a process in which there is a period in which patients provide subjective input in order that the hearing aid can "learn" how best to adapt setting to the acoustic environment.

LIMITATIONS

In the current study, we accessed intrinsic logging of hearing aid on-time. However, hearing care professionals can usually also access an intrinsic log of accumulated sound exposure through the clinical computer. It would therefore have been interesting to compare the distributions of the sound environments for the remotely-logged and intrinsically-logged data. However, these data are inaccessible.

Moreover, while we expect the participants in the study to be representative of in-market hearing aid users, we acknowledge that they had all actively signed up for an advanced data tracking feature via the Oticon ON™ remote control smartphone app. Thus, the participants might be more tech-savvy than the average hearing aid user, which might have biased their hearing aid usage patterns.

Lastly, the presented results rely on data from only 62 hearing aid users, which might not be an adequately large sample to generalize insights from. However, the data collected from each participant is rich (minute-based logging), which means that derived insights accurately reflect individual behavior.

CONCLUSION

Remote data logging using smartphone-enabled hearing aids can provide rich data regarding hearing aid usage and the ambient acoustic environment in which they are used. The data have high face-validity and smartphone connectivity is generally high (>85%). Days with poor connectivity can be identified and filtered out using statistical methods prior to assessing hearing aid usage patterns and their association to environmental factors.

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 non-disclosure agreement.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JC drafted the manuscript and performed that data analysis. GS wrote significant contributions to the manuscript. LH and NP critically reviewed the manuscript. JC, LH, and NP conceptualized the study. All authors contributed to the article and approved the submitted version.

FUNDING

GS was supported by the NIHR Manchester Biomedical Research Center.

REFERENCES

1. Humes L, Ahlstrom J, Bratt G, Peek B. Studies of hearing-aid outcome measures in older adults: a comparison of technologies and an examination of individual differences. *Semin Hear.* (2009) 30:112–28. doi: 10.1055/s-0029-1215439
2. Akeroyd MA. Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *Int J Audiol.* (2008) 47:S53–71. doi: 10.1080/14992020802301142
3. Humes LE. Factors underlying the speech-recognition performance of elderly hearing-aid wearers. *J Acoust Soc Am.* (2002) 112:1112–1132. doi: 10.1121/1.1499132
4. Woods WS, Kalluri S, Pentony S, Nooraei N. Predicting the effect of hearing loss and audibility on amplified speech reception in a multi-talker listening scenario. *J Acoust Soc Am.* (2013) 133:4268–78. doi: 10.1121/1.4803859
5. Jenstad LM, Singh G, Boretzki M, DeLongis A, Fichtl E, Ho R, et al. Ecological momentary assessment: a field evaluation of subjective ratings of speech in noise. *Ear Hear.* (2021) 42:520–30. doi: 10.1097/AUD.0000000000001071
6. Williger B, Lang FR. Hearing aid use in everyday life: managing contextual variability. *Gerontology.* (2015) 61:158–65. doi: 10.1159/000366062
7. Larson VD. Efficacy of 3 Commonly used hearing aid circuitsa crossover trial. *JAMA.* (2000) 284:1806–13. doi: 10.1001/jama.284.14.1806
8. Marozeau J, Florentine M. Loudness growth in individual listeners with hearing losses: a review. *J Acoust Soc Am.* (2007) 122:EL81–7. doi: 10.1121/1.2761924
9. Oetting D, Hohmann V, Appell JE, Kollmeier B, Ewert SD. Restoring perceived loudness for listeners with hearing loss. *Ear Hear.* (2018) 39:664–78. doi: 10.1097/AUD.0000000000000521
10. Wong LLN, Hickson L, McPherson B. Hearing aid satisfaction: what does research from the past 20 years Say? *Trends Amplif.* (2003) 7:117–61. doi: 10.1177/108471380300700402
11. Laplante-Lévesque A, Nielsen C, Jensen LD, Naylor G. Patterns of hearing aid usage predict hearing aid use amount (data logged and self-reported) and overreport. *J Am Acad Audiol.* (2014) 25:187–98. doi: 10.3766/jaaa.25.2.7
12. Solheim J, Hickson L. Hearing aid use in the elderly as measured by datalogging and self-report. *Int J Audiol.* (2017) 56:472–9. doi: 10.1080/14992027.2017.1303201
13. Walker EA, Spratford M, Moeller MP, Oleson J, Ou H, Roush P, et al. Predictors of hearing aid use time in children with mild-to-severe hearing loss. *Lang Speech Hear Serv Sch.* (2013) 44:73–88. doi: 10.1044/0161-1461(2012/12-0005)
14. Houmøller SS, Wolff A, Narne V, Loquet G, Hougaard DD, Hammershøi D, et al. Hearing aid satisfaction and differences between self-reported and data logged hearing aid usage time for experienced and first-time hearing aid users. In: *International Symposium on Auditory and Audiological Research (2019): Auditory Learning in Biological and Artificial Systems*. (2019). Available

online at: https://portal.findresearcher.sdu.dk/da/publications/hearing-aid-satisfaction-and-differences-between-self-reported-an
15. Taubman LB, Palmer CV, Durrant JD, Pratt S. Accuracy of hearing aid use time as reported by experienced hearing aid wearers. *Ear Hear.* (1999) 20:299–305. doi: 10.1097/00003446-199908000-00003
16. Saunders GH, Bott A, Tietz LHB. Hearing care providers' perspectives on the utility of datalogging information. *Am J Audiol.* (2020) 29:610–22. doi: 10.1044/2020_AJA-19-00089
17. Banerjee S. Hearing aids in the real world: typical automatic behavior of expansion, directionality, noise management. *J Am Acad Audiol.* (2011) 22:34–48. doi: 10.3766/jaaa.22.1.5
18. Camacho SK, Luis F. *DataLogging—Hearing Aid Behavior in the Real World Sejal Kuvadia Luis F. Camacho*. (2017). Available online at: https://www.audiologyonline.com/articles/datalogging-hearing-aid-behavior-in-20258
19. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol.* (2008) 4:1–32. doi: 10.1146/annurev.clinpsy.3.022806.091415
20. Holube I, von Gablenz P, Bitzer J. Ecological momentary assessment in hearing research: current state, challenges, future directions. *Ear Hearing.* (2020) 41:79S—90S. doi: 10.1097/AUD.0000000000000934
21. Andersson KE, Andersen LS, Christensen JH, Neher T. Assessing real-life benefit from hearing-aid noise management: SSQ12 questionnaire versus ecological momentary assessment with acoustic data-logging. *Am J Audiol.* (2021) 30:93–104. doi: 10.1044/2020_AJA-20-00042
22. von Gablenz P, Kowalk U, Bitzer J, Meis M, Holube I. Individual hearing aid benefit in real life evaluated using ecological momentary assessment. *Trends Hearing.* (2021) 25:233121652199028. doi: 10.1177/2331216521990288
23. Wu, Y.-H., Stangl E, Chipara O, Gudjonsdottir A, Oleson J, et al. Comparison of *in-situ* and retrospective self-reports on assessing hearing aid outcomes. *J Am Acad Audiol.* (2020) 31:746–62. doi: 10.1055/s-0040-1719133
24. Timmer BHB, Hickson L, Launer S. Ecological momentary assessment: feasibility, construct validity, future applications. *Am J Audiol.* (2017) 26:436–42. doi: 10.1044/2017_AJA-16-0126
25. Bosman AJ, Christensen JH, Rosenbom T, Patou F, Janssen A, Hol MK. Investigating real-world benefits of high-frequency gain in bone-anchored users with ecological momentary assessment and real-time data logging. *J Clin Med.* (2021).
26. Schinkel-Bielefeld N, Kunz P, Zutz A, Buder B. Evaluation of hearing aids in everyday life using ecological momentary assessment: what situations are we missing? *Am J Audiol.* (2020) 29:591–609. doi: 10.1044/2020_AJA-19-00075
27. Wu, Y.-H., Xu J, Stangl E, Pentony S, Vyas D, et al. Why ecological momentary assessment surveys go incomplete: when it happens and how it impacts data. *J Am Acad Audiol.* (2021) 32:16–26. doi: 10.1055/s-0040-1719135
28. de Graaff F, Huysmans E, Ket JC, Merkus P, Goverts ST, Leemans CR, et al. Is there evidence for the added value and correct use of manual and automatically switching multimemory hearing devices? A scoping review. *Int J Audiol.* (2018) 57:176–83. doi: 10.1080/14992027.2017.1385864
29. Gutenberg J, Pontoppidan NH, Laplante-Lévesque A. (2018). EHealth and big data usage in audiology EVOTION: a european research project generating big data evidence [Poster]. In: *Medical Informatics Europe Conference 2018* (Gothenburg).
30. Saunders GH, Christensen JH, Gutenberg J, Pontoppidan NH, Smith A, Spanoudakis G, et al. Application of big data to support evidence-based public health policy decision-making for hearing. *Ear Hear.* (2020) 41:1057–63. doi: 10.1097/AUD.0000000000000850
31. Christensen JH, Rossing R, Pontoppidan NH, Anisetti M, Bamiou, D.-E., et al. Fully synthetic longitudinal real-world data from hearing aid wearers for public health policy modeling. *Front Neurosci.* (2019) 13:850. doi: 10.3389/fnins.2019.00850
32. Christensen JH, Saunders GH, Porsbo M, Pontoppidan NH. The everyday acoustic environment and its association with human heart rate: evidence from real-world data logging with hearing aids and wearables. *R Soc Open Sci.* (2021) 8:201345. doi: 10.1098/rsos.201345
33. Johansen B, Flet-Berliac YPR, Korzepa MJ, Sandholm P, Pontoppidan NH, Petersen MK, et al. Hearables in hearing care: discovering usage patterns through iot devices. In: *International Conference on Universal Access in Human-Computer Interaction.* Vancouver, BC (2017). p. 39–49. doi: 10.1007/978-3-319-58700-4_4
34. Pontoppidan NH, Li X, Bramsløw L, Johansen B, Nielsen C, Hafez A, et al. Data-driven hearing care with time stamped data-logging. In: *Proceedings of the International Symposium on Auditory and Audiological Research.* (2018). Available online at: https://proceedings.isaar.eu/index.php/isaarproc/article/view/2017-15/332
35. Laplante-Lévesque A, Pontoppidan NH, Mazevski A, Schum D, Behrens T, Porsbo M. *Data-Driven Hearing Care With HearingFitness^{TM}: Oticon Shares Openly Its Innovative Vision and Roadmap (White Paper No. 27896).* Smoerum: Oticon A/S (2017).
36. Kochkin S. Marketrak VIII: the efficacy of hearing aids in achieving compensation equity in the workplace. *Hear J.* (2010) 63:19–28. doi: 10.1097/01.HJ.0000389923.80044.e6
37. Kodinariya TM, Makwana PR. Review on determining number of cluster in K-means clustering. *Int J.* (2013) 1:90–5.
38. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math.* (1987) 20:53–65. doi: 10.1016/0377-0427(87)90125-7
39. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. *cluster: Cluster Analysis Basics and Extensions*. R Package Version 1.14.4. (2013).
40. Harrison XA, Donaldson L, Correa-Cano ME, Evans J, Fisher DN, Goodwin CED, et al. A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ.* (2018) 6:e4794. doi: 10.7717/peerj.4794
41. Fox J. *car: Companion to Applied Regression (R Package Version 1.2-14).* (2009). Available online at: http://Cran. r-Project. Org/Web/Packages/Car/
42. Fox J, Monette G. Generalized collinearity diagnostics. *J Am Stat Assoc.* (1992) 87:178–83. doi: 10.1080/01621459.1992.10475190
43. O'brien RM. A caution regarding rules of thumb for variance inflation factors. *Qual Quant.* (2007) 41:673–90. doi: 10.1007/s11135-006-9018-6
44. Barton K, Barton MK. *Package 'Mumin'. Version* (2015).
45. Pinheiro J, Bates D, DebRoy S, Sarkar D, Team RC. *nlme: Linear and Nonlinear Mixed Effects Models. R Package Version* (2013).
46. Smeds K, Wolters F, Rung M. Estimation of signal-to-noise ratios in realistic sound scenarios. *J Am Acad Audiol.* (2015) 26:183–6. doi: 10.3766/jaaa.26.2.7
47. Walden BE, Surr RK, Cord MT, Dyrlund O. Predicting hearing aid microphone preference in everyday listening. *J Am Acad Audiol.* (2004) 15:365–96.
48. Keidser G. Many factors are involved in optimizing environmentally adaptive hearing aids. *Hear J.* (2009) 62:26–8.
49. Wagener KC, Hansen M, Ludvigsen C. Recording and classification of the acoustic environment of hearing aid users. *J Am Acad Audiol.* (2008) 19:348–70.
50. Wu YH, Stangl E, Chipara O, Hasan SS, Welhaven A, Oleson J. Characteristics of real-world signal-to-noise ratios and speech listening situations of older adults with mild-to-moderate hearing loss. *Ear Hear.* (2018) 39:293.

Author Disclaimer: The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Conflict of Interest: JC, LH, and NP are employed by Oticon A/S.

The remaining author declares 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.

frontiers in Neuroscience

REVIEW
published: 31 August 2021
doi: 10.3389/fnins.2021.723877

Can Haptic Stimulation Enhance Music Perception in Hearing-Impaired Listeners?

Mark D. Fletcher[1,2]*

[1] University of Southampton Auditory Implant Service, Faculty of Engineering and Physical Sciences, University of Southampton, Southampton, United Kingdom, [2] Institute of Sound and Vibration Research, Faculty of Engineering and Physical Sciences, University of Southampton, Southampton, United Kingdom

Cochlear implants (CIs) have been remarkably successful at restoring hearing in severely-to-profoundly hearing-impaired individuals. However, users often struggle to deconstruct complex auditory scenes with multiple simultaneous sounds, which can result in reduced music enjoyment and impaired speech understanding in background noise. Hearing aid users often have similar issues, though these are typically less acute. Several recent studies have shown that haptic stimulation can enhance CI listening by giving access to sound features that are poorly transmitted through the electrical CI signal. This "electro-haptic stimulation" improves melody recognition and pitch discrimination, as well as speech-in-noise performance and sound localization. The success of this approach suggests it could also enhance auditory perception in hearing-aid users and other hearing-impaired listeners. This review focuses on the use of haptic stimulation to enhance music perception in hearing-impaired listeners. Music is prevalent throughout everyday life, being critical to media such as film and video games, and often being central to events such as weddings and funerals. It represents the biggest challenge for signal processing, as it is typically an extremely complex acoustic signal, containing multiple simultaneous harmonic and inharmonic sounds. Signal-processing approaches developed for enhancing music perception could therefore have significant utility for other key issues faced by hearing-impaired listeners, such as understanding speech in noisy environments. This review first discusses the limits of music perception in hearing-impaired listeners and the limits of the tactile system. It then discusses the evidence around integration of audio and haptic stimulation in the brain. Next, the features, suitability, and success of current haptic devices for enhancing music perception are reviewed, as well as the signal-processing approaches that could be deployed in future haptic devices. Finally, the cutting-edge technologies that could be exploited for enhancing music perception with haptics are discussed. These include the latest micro motor and driver technology, low-power wireless technology, machine learning, big data, and cloud computing. New approaches for enhancing music perception in hearing-impaired listeners could substantially improve quality of life. Furthermore, effective haptic techniques for providing complex sound information could offer a non-invasive, affordable means for enhancing listening more broadly in hearing-impaired individuals.

Keywords: neuroprosthetic, cochlear implant, hearing aid, tactile aid, electro-haptic stimulation, pitch, multi-sensory, sensory substitution

OPEN ACCESS

Edited by:
Qinglin Meng,
South China University of Technology, China

Reviewed by:
Yonghee Oh,
University of Florida, United States
James W. Dias,
Medical University of South Carolina, United States

Correspondence:
Mark D. Fletcher
M.D.Fletcher@soton.ac.uk

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Neuroscience

Received: *11 June 2021*
Accepted: *11 August 2021*
Published: *31 August 2021*

Citation:
Fletcher MD (2021) Can Haptic Stimulation Enhance Music Perception in Hearing-Impaired Listeners? Front. Neurosci. 15:723877. doi: 10.3389/fnins.2021.723877

INTRODUCTION

Cochlear implants (CIs) recover hearing for severely-to-profoundly hearing-impaired individuals by electrically stimulating the cochlea. They deploy an array of up to 22 microelectrodes, replacing the approximately 3,500 hair cells that transfer sound to the brain in normal-hearing listeners. Despite the fact that only limited sound information can be provided through this small number electrodes, CIs have been remarkably successful at recovering access to speech in quiet listening conditions (Zeng et al., 2008). However, CI users typically have impaired speech recognition in background noise (Fletcher et al., 2019, 2020b), as well as substantially reduced sound-localization accuracy (Dorman et al., 2016; Fletcher et al., 2020a) and music enjoyment (McDermott, 2004; Drennan et al., 2015). Hearing-aid (HA) users and other hearing-impaired listeners have similar performance limitations, though typically to a lesser extent (Looi et al., 2008; Dorman et al., 2016; Miller et al., 2016).

Several studies have recently shown that haptic stimulation can enhance CI listening by allowing access to sound features that are poorly transferred through electrical CI stimulation (see Fletcher, 2020; Fletcher and Verschuur, 2021). This “electro-haptic stimulation” can substantially improve speech-in-noise performance (Huang et al., 2017; Fletcher et al., 2018, 2019, 2020b), sound localization (Fletcher and Zgheib, 2020; Fletcher et al., 2020a), and melody recognition (Huang et al., 2019; Luo and Hayes, 2019), as well as discrimination of basic sound features such as pitch (Fletcher et al., 2020c). The impressive performance found in studies of haptic sound-localization and haptic enhancement of pitch discrimination suggests that it could also assist HA users (Fletcher and Zgheib, 2020; Fletcher et al., 2020a,c). There is also evidence that haptic stimulation can improve timbre discrimination (Russo et al., 2012) and music appreciation (Nanayakkara et al., 2009) in HA users. Music represents the biggest challenge for signal processing as it is often an extremely complex acoustic signal that contains several simultaneous harmonic and inharmonic sounds. Progress in enhancing music perception could therefore have strong implications for enhancing listening in the complex auditory environments in which hearing-impaired listeners often struggle to understand speech, such as busy offices, classrooms, or restaurants.

This review will focus on the use of haptic stimulation to enhance music perception in hearing-impaired listeners. Most people in the deaf community report being involved in music activities (Darrow, 1993) and music is central to many significant events, such as weddings and funerals, as well as to media, such as film. It is an important part of interactions with children (Hallam, 2010), can strongly influence the mood of films and the audience’s connection to the characters (Hoeckner et al., 2011), and can even bias shopping habits (North et al., 1999). As will be discussed, music perception is highly limited in many hearing-impaired listeners. This review first assesses the limits of music perception in hearing-impaired listeners, the suitability of the tactile system for transferring musical signals, and the evidence that audio and haptic inputs are integrated in the brain. It then discusses the existing haptic systems for enhancing music perception, the evidence of their utility, and the signal-processing approaches that could be deployed on future devices. Finally, it reviews the cutting-edge technologies that could be utilized for haptic enhancement of music perception.

IS HAPTIC STIMULATION SUITABLE FOR ENHANCING MUSIC PERCEPTION?

Music Perception in Hearing-Impaired Listeners

When considering whether a haptic system might enhance music perception in hearing-impaired listeners, it is important to first establish the limits of music listening when hearing is impaired. It has been reported that, after a CI is implanted, only around 15% of adults enjoy listening to music (Philips et al., 2012) and around 70% are disappointed by how music sounds (Mirza et al., 2003). On a 10-point visual analog scale, CI users rated their musical enjoyment at 8.7 on average prior to hearing loss and at just 2.6 after implantation (Mirza et al., 2003). Low music appreciation has also been found for HA users, with those that have the most severe hearing loss reporting the lowest music appreciation (Looi et al., 2019). Some hearing-impaired listeners describe music as sounding “dissonant,” “out-of-tune,” “fuzzy,” and “tinny” (Uys et al., 2012; Jiam et al., 2017).

Numerous studies have explored which of the auditory features within musical pieces can be effectively extracted by hearing-assistive device users. CI users typically perform well at basic rhythm (Cooper et al., 2008; Kim et al., 2010), tempo (Kong et al., 2004), and meter (Cooper et al., 2008) perception tasks (although there is evidence that they perform less well for more complex rhythms (Gfeller et al., 2000; Petersen et al., 2012; Jiam and Limb, 2019). In contrast, CI users perform poorly for spectral and spectro-temporal features, such as pitch (Galvin et al., 2007; Cooper et al., 2008), harmony (Brockmeier et al., 2011), melody (Galvin et al., 2007; Zeng et al., 2014), and timbre (Gfeller et al., 2002c; Drennan and Rubinstein, 2008; Nimmons et al., 2008). CI users also have poorer spectral and temporal modulation detection thresholds than normal-hearing listeners (Choi et al., 2018).

HA users have similar spectral and temporal modulation thresholds to normal-hearing listeners (Choi et al., 2018; Looi et al., 2019) and, like CI users, tend not to have deficits with basic rhythm perception (Looi et al., 2019). HA users have been found to have subnormal pitch, melody, and timbre perception (Choi et al., 2018; Looi et al., 2019). However, HA users tend to perform much better than CI users on music perception tasks, such as instrument identification, melody recognition, and pitch discrimination (Gfeller and Lansing, 1991, 1992; Gfeller et al., 1998, 2002a,c; Fujita and Ito, 1999; Leal et al., 2003). It should, however, be noted that there is substantial variance between individual CI and HA users.

Vision plays an important role in music perception for hearing-impaired listeners. Viewing the performer and reading lyrics can increase their musical enjoyment (Gfeller et al., 2000; Looi and She, 2010) and raves targeted at the deaf community frequently include musical visualization. Furthermore, the size of sung musical intervals can be determined when only viewing the singer's face (without audio), with larger intervals associated with more head movement, eyebrow raising, and mouth opening (Thompson and Russo, 2007; Abel et al., 2016). Viewing a singer's face with accompanying audio can also bias the perception of pitch interval size (Thompson et al., 2010), with the mouth apparently increasing in significance as audio signal-to-noise ratios become more challenging (Russo et al., 2011). For musical instruments, visual influences have been observed on timbre perception (Saldana and Rosenblum, 1993), as well as on loudness (Rosenblum and Fowler, 1991) and duration (Schutz and Lipscomb, 2007; Schutz and Kubovy, 2009) perception for rhythms.

Several other factors are known to have important influences on music perception for hearing-impaired listeners. For example, the age at which hearing impairment occurred, the amount of residual hearing retained for CI users, and the efficiency of sequential cognitive processing are predictive of pitch and timbre perception (Gfeller et al., 2000, 2008, 2010; O'Connell et al., 2017). Age is also important, with younger CI users listening to music more often and tending to have better timbre perception (Gfeller et al., 2008, 2010; Drennan et al., 2015). More listening hours and musical training have both been linked to higher acuity and music appraisal scores (Gfeller et al., 2002b, 2008, 2010, 2011; Fu and Galvin, 2007; Galvin et al., 2009; Chen et al., 2010; Looi and She, 2010; Driscoll, 2012). However, no strong relationship has been found between perceptual accuracy and music appraisal or enjoyment (Gfeller et al., 2008; Drennan et al., 2015).

Limits of Haptic Sensitivity Compared to Hearing-Impaired Listening

To establish how haptic stimulation might effectively augment listening, this section compares the sensitivity of the tactile system to the impaired auditory system. First, sensitivity to frequency, intensity, and temporal features will be considered (for a detailed review in the context of speech perception, see Fletcher and Verschuur, 2021).

While frequency discrimination for CI and other hearing-impaired listeners is poorer than for normal-hearing listeners (Moore, 1996; Turgeon et al., 2015), it is better than for haptic stimulation (Goff, 1967; Rothenberg et al., 1977). Because of this poor frequency resolution, several systems for transmitting sound information through haptic stimulation have mapped sound frequency information to location on the skin using an array of haptic stimulators, each triggered by a different pitch or frequency band (Guelke and Huyssen, 1959; Brooks and Frost, 1983; Fletcher et al., 2020c). Using this approach, high-resolution pitch information has been transferred through haptic stimulation (Fletcher et al., 2020c). This could be important for enhancing music perception in hearing-impaired listeners.

The dynamic range of the tactile system at the arm, wrist, and hand is similar to that available to HA users and is around four times larger than that available through electrical CI stimulation (Verrillo et al., 1969; Moore et al., 1985; Zeng and Galvin, 1999; Zeng et al., 2002; Fletcher et al., 2021a,b). CI users are able to discriminate approximately 20 different intensity steps across their dynamic range (Kreft et al., 2004; Galvin and Fu, 2009). For HA users and for haptic stimulation at the arm, wrist, or hand, approximately 40 different steps can be discriminated (Hall and Fernandes, 1983; Gescheider et al., 1996; Fletcher et al., 2021a,b). Interestingly, there is evidence that congenitally deaf people have higher tactile sensitivity than those with normal hearing (Levanen and Hamdorf, 2001), which may mean that the available dynamic range is larger than has been estimated previously in studies using participants with no known hearing impairment. The tactile system therefore seems well suited to deliver sound intensity information to CI users and could provide additional intensity information for at least a subset of HA users.

As highlighted above, CI users typically perform well when extracting temporal sound features. Temporal gap detection thresholds for hearing-impaired listeners and CI users are typically only slightly worse than those for normal-hearing listeners (Moore and Glasberg, 1988; Garadat and Pfingst, 2011). Gap detection thresholds for the tactile system are worse than for most hearing-impaired listeners (Gescheider, 1966, 1967) and tactile signals are more susceptible to masking from temporally remote maskers (Elliot, 1962; Gescheider et al., 1989; Shannon, 1990). Haptic stimulation may therefore not be suitable for providing complex temporal information.

The tactile system has been shown to be highly sensitive to amplitude modulation (Weisenberger, 1986). For a carrier tone at 250 Hz – the frequency at which tactile sensitivity is highest (Verrillo et al., 1969) and a common characteristic frequency for compact motors – amplitude modulation sensitivity was found to be high across the range of frequency modulations most important for speech and music (Drullman et al., 1994; Ding et al., 2017). Sensitivity was reduced when the carrier tone frequency was reduced to 100 Hz (around the lowest characteristic frequency for a compact motor). At modulation frequencies most important to music and speech, amplitude modulation sensitivity for a 250-Hz carrier is below that for an auditory tone carrier at 250 Hz (Zwicker, 1952), but similar to auditory sensitivity for a narrowband noise centred at 200 Hz (Viemeister, 1979), in normal-hearing listeners. This suggests that amplitude modulation is a highly viable route through which sound information can be transferred through haptic stimulation, particularly for CI users, who have reduced sensitivity to amplitude modulation (Choi et al., 2018).

Besides transferring sound information through stimulation at a single site or at adjacent sites, recent studies have shown that sound location information can be transferred through across-limb stimulation (Fletcher and Zgheib, 2020; Fletcher et al., 2020a, 2021a,b). CI and HA users have reduced sound localization accuracy compared to normal hearing listeners (Dorman et al., 2016); using this approach, large improvements in sound localization accuracy for CI users were shown, with accuracy reaching levels that could be beneficial to HA users. In this approach, the sound received by devices behind each ear was converted to haptic stimulation on each wrist (Fletcher and Zgheib, 2020; Fletcher et al., 2020a). This meant that time and intensity differences between the ears, which are critical sound localization cues, were available through time and intensity differences across the wrists. Recently, the tactile system has been shown to be highly sensitive to intensity differences across the arms and wrists, but insensitive to time differences (Fletcher et al., 2021a,b). Strikingly, sensitivity to tactile intensity differences across the limbs matched the sensitivity of the auditory system to intensity differences across the ears. Given that instruments in most musical pieces are mapped to a left-right spatial location using only amplitude panning, this high sensitivity to across-limb tactile intensity differences might be exploited to improve localization and segregation of musical instruments.

Multisensory Integration of Auditory and Haptic Signals

Effective integration of haptic and auditory inputs in the brain is likely to be crucial to haptic augmentation of musical listening. Encouragingly, projections from tactile brain regions have been observed at all stages along the auditory pathway (Aitkin et al., 1981; Foxe et al., 2000; Shore et al., 2000, 2003; Caetano and Jousmaki, 2006; Allman et al., 2009; Meredith and Allman, 2015). Furthermore, physiological studies have shown that the responses of large numbers of auditory cortical neurons can be modulated by input from tactile pathways (Lakatos et al., 2007; Meredith and Allman, 2015) and neuroimaging studies have shown that haptic stimulation can activate auditory cortex (Schurmann et al., 2006); interestingly, stronger activation has been found for deaf participants than for normal-hearing subjects (Levanen and Hamdorf, 2001; Auer et al., 2007). One study in normal-hearing subjects tracked the time course of cortical activation for haptic stimulation on the fingertip (Caetano and Jousmaki, 2006). Initial responses peaked in primary tactile cortical brain regions around 60 ms after the stimulus onset. This was followed by transient responses to the haptic signal in auditory cortex between 100 and 200 ms after onset, before a sustained response was seen between 200 and 700 ms after onset. This could indicate that tactile responses feed forward from tactile brain regions to influence auditory brain regions.

Behavioral studies also offer a range of evidence that haptic and auditory input is integrated. For example, haptic stimulation has been shown to improve sound detection (Schurmann et al., 2004), modulate perceived loudness (Gillmeister and Eimer, 2007; Merchel et al., 2009), and influence syllable perception (Gick and Derrick, 2009). Other studies have shown that tactile feedback from a musical instrument can influence a performer's perception of sound quality (Fontana et al., 2017). Audio and haptic stimulation have also been effectively combined to improve speech-in-noise performance (Drullman and Bronkhorst, 2004; Huang et al., 2017; Fletcher et al., 2018, 2019, 2020b) and sound localization (Fletcher et al., 2020a).

When considering whether haptic and audio input will be integrated to improve music perception, individual characteristics such as age at which hearing loss occurred, length of time spent with hearing loss, and length of time spent with a hearing-assistive device may be critical. It has been observed that those who receive a CI after a few years of deafness integrate audio and visual information less effectively than those who are implanted shortly after deafness (Bergeson et al., 2005; Schorr et al., 2005; Tremblay et al., 2010). It is possible that a similar limitation will be seen for audio-haptic integration. Some studies have also shown evidence that audio-haptic integration is reduced in congenitally deaf CI recipients compared to late-deafness CI recipients (Landry et al., 2013; Nava et al., 2014). Future work should establish whether benefit of haptic stimulation to music perception is dependent on these factors.

Age may also be important. Haptic stimulation has been shown to improve performance when combined with auditory stimulation in both young (Drullman and Bronkhorst, 2004; Fletcher et al., 2018; Ciesla et al., 2019) and older (Huang et al., 2017; Fletcher et al., 2019, 2020a,b) adults, although these groups have not been directly compared. Several studies have shown evidence that multisensory integration increases in older adults (Laurienti et al., 2006; Rouger et al., 2007; Diederich et al., 2008; Strelnikov et al., 2009, 2015; de Dieuleveult et al., 2017) and there is also evidence that young brains are particularly open to integrating multisensory stimuli (Lewkowicz and Ghazanfar, 2006). It is therefore possible that older adults and children will benefit most from haptic enhancement of music perception.

Auditory deprivation has been associated with increased sensitivity to visual (Finney et al., 2001, 2003) and tactile (Auer et al., 2007) stimuli in auditory brain regions. During early development, substantial neural pruning occurs based on the sensory input received. If auditory input is limited or extinguished by congenital or early-onset deafness, this process can be disrupted and non-auditory inputs can take over auditory brain areas (Quartz and Sejnowski, 1997; Sharma et al., 2007; Glennon et al., 2020). If auditory pathways later receive new sensory stimulation (e.g., because a CI has been fitted), this is thought to compete for neural resources in auditory brain regions with the other sensory inputs that have become established (Sharma et al., 2007; Glennon et al., 2020). This may explain why early implantation is associated with better speech performance (Robbins et al., 2004; Svirsky et al., 2004; Kral, 2009; Tajudeen et al., 2010) and why more visual takeover of auditory brain regions is associated with poorer speech outcomes (Lee et al., 2001; Sandmann et al., 2012; Zhou et al., 2018). The influence of auditory-derived haptic stimulation on this process is unknown, but it may be that such an input would allow auditory brain areas to tune to critical auditory features, such as the amplitude envelope, in the absence of auditory input. Such a process might allow auditory input to compete for neural resources more effectively once input has been restored and might facilitate more effective audio-haptic integration. Future work should explore these possibilities.

Visual input is thought to provide missing speech and sound location information when the audio signal is degraded, to calibrate auditory neural responses, and to guide auditory perceptual learning (Rouger et al., 2007; Bernstein et al., 2013; Strelnikov et al., 2013; Isaiah et al., 2014). As discussed, audio-derived haptic stimulation has been shown to provide missing speech and sound location information when audio is degraded (e.g., Fletcher et al., 2019, 2020a) and to improve lip-reading ability in the absence of auditory stimulation (e.g., De Filippo, 1984; Brooks et al., 1986b; Hanin et al., 1988; Cowan et al., 1991; Reed et al., 1992). However, it has not yet been established whether haptic stimulation can calibrate auditory neural responses or guide auditory perceptual learning. There are relatively few studies of tactile influences on auditory cortex, but one has shown tactile stimulation can enhance responses to auditory input by modulating the rhythm of ambient neural responses in auditory cortex (Lakatos et al., 2007). This might reflect a critical mechanism for haptic enhancement of music perception.

Training is important both for integration of audio and haptic information and for extraction of information from haptic stimulation. Studies with haptic devices for providing speech information when no auditory information is available have shown continued benefits of training throughout long-term training regimes (Sparks et al., 1979; Brooks et al., 1985). Other studies have also shown the importance of training for maximizing haptic sound-localization accuracy (Fletcher and Zgheib, 2020; Fletcher et al., 2020a) and for improving speech-in-noise performance in CI users (Fletcher et al., 2018, 2019, 2020b).

The delay in arrival time of the haptic signal relative to the audio signal is also likely to be important for maximizing integration. A study using broadband signals showed that audio and haptic signals were judged to be simultaneous if the haptic signal onset was delayed from the audio by up to around 25 ms (Altinsoy, 2003). Another study with musical instruments found that the delay at which audio and haptic signal were no longer judged to be simultaneous varied across musical instruments, with attack time seemingly an important factor (Kim et al., 2006). It should be noted that there is significant evidence of rapid temporal recalibration, whereby stimulation from two modalities (including audio and tactile) that are consistently delayed by tens of milliseconds rapidly become perceived as synchronized, provided that they are highly correlated (Navarra et al., 2007; Keetels and Vroomen, 2008; van der Burg et al., 2013). There is evidence that integration occurs even for substantially delayed audio and haptic stimulation. Haptic stimulation has been shown to influence vowel perception, with no statistically significant reduction in this effect when the haptic signal onset was delayed from the audio onset by up to 100 ms (Gick et al., 2010). If haptic signal delays of several tens of milliseconds do not reduce the benefits of haptic stimulation, sophisticated real-time signal-processing strategies could be deployed to enhance music perception.

CURRENT SYSTEMS FOR IMPROVING MUSIC PERCEPTION USING HAPTIC STIMULATION

A range of systems have been developed to enhance music perception using haptic stimulation. At the largest scale, these include systems used for delivering whole-body vibration, such as those used at Deaf Raves, where music containing a lot of low-frequency energy is played at a high intensity. There is evidence that whole-body low-frequency vibration, which is also common during live pop or organ concerts, can play a significant role in the quality of the concert experience (Merchel and Altinsoy, 2014). There is also evidence that vibrating floors can improve the synchronization of dancing to music for hearing-impaired listeners (Shibasaki et al., 2016; Tranchant et al., 2017).

In addition to these large-scale systems, several smaller systems built into chairs have been developed. These typically use a multi-band filtering approach similar to that used in devices to improve access to speech cues in hearing-impaired people (e.g., Brooks et al., 1986a; Fletcher et al., 2019; reviewed in Fletcher, 2020; Fletcher and Verschuur, 2021). In this approach, the audio signal is separated into multiple frequency bands, with each band represented by a haptic stimulator at a different location on the skin. One example is the Emoti-Chair, which has eight haptic stimulators at different body locations (Karam et al., 2009, 2010). Users of the Emoti-Chair were shown to be able to discriminate between a cello, piano, and trombone (matched for fundamental frequency, duration, and intensity), and to be able to discriminate bright from dull timbres (varying only by spectral centroid) (Russo et al., 2012).

Another chair system developed by Jack et al. (2015) also splits the sound into frequency bands that are mapped to different haptic stimulators (see **Figure 1A**). In addition to haptic stimulation transferring information about energy within each frequency band, the bandwidth of haptic stimulation at each stimulator is modulated to deliver timbre information (spectral flatness). While subjective reports when using this system were favorable, formal behavioral testing was not performed. They did note, however, that highly rhythmic music tended to be received more positively than music that relied heavily on harmonic motion.

A final example is the haptic chair built by Nanayakkara et al. (2009), which delivered unprocessed music through contact loudspeakers targeting the feet, back, arms, and hands. In their study with 43 young hearing-impaired listeners (with their hearing aids switched off), participants rated their musical experience considerably higher with vibration through the chair than without. However, there were several limitations to the study, including the absence of control for novelty or placebo effects and the possible influence of audio from the contact loudspeakers.

Other medium-scale wearable systems have also been developed, typically deployed using suits or vests. One system uses a suit with 13 haptic stimulators placed around the body and maps different musical instruments to different stimulators (Gunther et al., 2003). A major limitation of this approach is that it requires access to each instrument within a musical piece, which is not typically possible. No formal testing of this haptic suit was performed, although informal feedback from individuals using it as part of an art exhibition was reported to be favorable.

Another wearable system, the LIVEJACKET, which uses a vest with 22 haptic stimulators attached to the arms and torso has also been developed (Hashizume et al., 2018). Like the haptic suit, the LIVEJACKET presents different musical instruments through different haptic stimulators. Survey results suggested the LIVEJACKET enhanced the musical experience for normal-hearing participants. However, critical experimental controls were not in place and, like for the haptic suit, access to each instrument within the musical piece is required.

Finally, there are a range of more compact wearable systems. One such system is the Pump-and-Vibe (Haynes et al., 2021), which is worn on the arm (**Figure 1B**). The Pump-and-Vibe has eight vibration motors mounted on the forearm arm and an air pump on the upper arm to modulate pressure ("squeeze"). Squeeze is thought to more effectively elicit emotional responses than vibration (Tsetserukou, 2010) and has been deployed in a number of previous devices for various applications (e.g., Chinello et al., 2014; Gupta et al., 2017; Moriyama et al., 2018; Stephens-Fripp et al., 2018; Pezent et al., 2019). The Pump-and-Vibe system aimed to increase the emotional impact of music. The rhythm of the bass was mapped to changes in the amount of squeeze. The squeeze system used was unable to track fast rhythms, so these were mapped to three vibrotactile motors at the top of the forearm. Melody information was mapped to the remaining five motors, with pitch mapped to the location of stimulation along the arm. For vibration, intensity changes were mapped to co-varying haptic frequency and amplitude changes. Sound information was extracted from music using a process involving an online audio-to-MIDI converter. It is not clear how effective this conversion will be for different music types. A qualitative assessment of the Pump-and-Vibe evaluated the mood evoked by a musical piece for audio alone, haptic alone, and haptic and audio together in young participants with no specified hearing impairment (Haynes et al., 2021). Results suggested that the system could evoke moods and influence the mood evoked by audio.

Other examples of more compact systems are the Mood Glove and the mosaicOne series of devices. The Mood Glove (**Figure 1C**) has eight motors, with five mounted on the back of the hand and three on the palm (Mazzoni and Bryan-Kinns, 2016). Stimulation frequency and intensity are adjusted to portray different moods in musical pieces. A study of the device reported that low-frequency pulses could induce a feeling of calmness and higher-frequency pulses a feeling of excitement (Mazzoni and Bryan-Kinns, 2016). However, the Mood Glove requires the intended mood created by each section of the musical piece to be extracted and provided to the device, which was achieved in the study through manual labeling. This requirement substantially limits the potential for real-world use.

The mosaicOne_B, has two sets of six haptic stimulators arranged along the top and underside of the forearm (Fletcher et al., 2020c). It maps the fundamental frequency of sound (an acoustic correlate of pitch) to location on the skin. Using this device, participants were able to discriminate fundamental frequency differences of just 1.4%. This is markedly better than can be achieved by most CI users (Kang et al., 2009; Drennan et al., 2015) and would allow discrimination of the smallest fundamental frequency changes found in most western melodies. The mosaicOne_B incorporates a novel noise-reduction strategy that was found to be highly effective, with discrimination performance retained even with high levels of background noise. However, it is important to note that the background noise used was inharmonic, while many musical pieces contain multiple simultaneous harmonic sounds. Further work is required to establish the resilience of the mosaicOne_B against harmonic background noise. Furthermore, development is required to allow the device to extract multiple pitches simultaneously, for tracking of multiple simultaneous harmonic instruments. Musical experience was not formally tested using this device, but users reported enhanced musical enjoyment (when listening and feeling pop music) in informal testing by the author of this review with several normal-hearing listeners. Another version of the device, the mosaicOne_C (**Figure 1D**), has also been developed, which uses a similar approach to that described above, but with shakers spaced around the wrist (Fletcher, 2020; Fletcher and Verschuur, 2021). This device has not yet been subjected to behavioral testing.

Two further studies reported behavioral results for wearable devices. One wrist-worn device extracted the fundamental frequency, like the mosaicOne_B, but mapped it to changes in the frequency and amplitude of the haptic signal (which varied together), rather than spatial location (Luo and Hayes, 2019). Critically, unlike for the mosaicOne_B, this meant that intensity information could not be delivered. Another device delivered the low-frequency portion of the audio signal through haptic stimulation on the fingertip (Huang et al., 2019). Encouragingly, both systems were shown to improve melody recognition. However, the effectiveness of these devices in the presence of background noise has not been tested, and the effect on music appreciation also remains to be established.

In addition to devices developed to augment music perception, several devices have been developed to aid those with sensory impairments by substituting one sense with another. An early example of a sensory substitution device is the Teletactor, developed in the 1920s, which transferred sound to deaf listeners through tactile stimulation on the hand (Gault, 1924, 1926). The principle has since been applied across a number of senses, with systems developed to substitute vision with tactile (Bach-Y-Rita et al., 1969), vestibular with tactile (Bach-Y-Rita et al., 2005), and vision with audio (Meijer, 1992). While these devices have shown promising results, few have found widespread use. Several factors have likely led to this. For example, many systems are highly restrictive, such as the BrainPort (Bach-Y-Rita et al., 2003, 2005) that stimulates the tongue, leaving users unable to speak or eat whilst using the device. Limitations in technology have also often heavily limited discreetness, comfort, and effectiveness. For example, the tactile aids for hearing that were developed in the 1980s and 1990s (before being superseded by CIs (see Fletcher and Verschuur, 2021)) were often large, had short battery lives, and could only perform crude signal processing. However, many of these technological limitations have since been overcome (Fletcher, 2020).

Some of the key design considerations when developing a modern haptic device for enhancing listening are discussed by Fletcher (2020). However, when developing a device for those with hearing-impairment, close engagement with the intended users (such as the deaf community) will be critical for ensuring maximum uptake. Fletcher (2020) advocates a wrist-worn device because they are easy to self-fit, offer a relatively large design space, and because wrist-worn devices, such as smartwatches and exercise trackers, are commonplace and therefore aesthetically

FIGURE 1 | Examples of haptic devices for enhancing music perception. Panel **(A)** Haptic chair developed at Queen Mary University of London (United Kingdom) by Jack and colleagues. Image reproduced with permission of Robert Jack and Andrew McPherson. Panel **(B)** The Pump-and-Vibe, developed at University of Bristol (United Kingdom) by Haynes and colleagues. Adapted from an image reproduced with permission of Alice Haynes. Panel **(C)** The Mood Glove, developed at Queen Mary University of London (United Kingdom) by Antonella Mazzoni. Image reproduced with her permission. Panel **(D)** The mosaicOne_C, developed at the University of Southampton (United Kingdom) by Samuel Perry and Mark Fletcher as part of the Electro-Haptics Research Project. Image reproduced with their permission.

acceptable. Indeed, technology for enhancing music perception using haptics could in future be embedded into smartwatches and exercise trackers.

HAPTIC SIGNAL-PROCESSING APPROACHES

Music is commonly accessed through streaming services. This opens the possibility of using signal-processing approaches that cannot be applied in real-time or that are non-causal (require the ability to look ahead). It also opens the possibility of using pre-trained machine-learning algorithms that are selected between based on metadata sent through the streaming service. These algorithms could be trained using the numerous high-quality musical corpora available, which can be supplemented using advanced automated music generation algorithms (Herremans and Chuan, 2020). So-called "near real-time" algorithms, which have processing delays of no more than a few seconds, may be of particular interest as such a delay before playback might be tolerable if clear enhancement of music experience could be demonstrated. Nevertheless, since a substantial portion of music is not streamed (e.g., at a concert or as background music in a shop), real-time signal-processing approaches are still preferred. Current evidence suggests that large delays of haptic stimulation from audio stimulation might be tolerable, which would allow sophisticated real-time signal-processing approaches to be deployed (see section "Multisensory Integration of Auditory and Haptic Signals"). Both real-time and offline approaches should therefore be considered.

It is important to first establish the goal when converting audio to haptics for music enhancement. One approach is to remove elements that reduce clarity when audio is transferred at a low-resolution (e.g., through a CI). One example of this is spectral complexity reduction, in which the frequency spectrum is sparsened and simplified, using methods such as principal component analysis (Nagathil et al., 2017; Gauer et al., 2019). Spectrally reduced musical pieces have been shown to be preferred for CI listening (Nagathil et al., 2017) and a similar approach might be trialed for haptic enhancement of music perception. An alternative approach is to enhance perception of certain instruments within a multi-instrument piece. It has been observed that CI and HA users find musical pieces with multiple instruments less pleasant than pieces with a single instrument (Looi et al., 2007) and that CI users prefer pop music with the vocal level substantially increased (Buyens et al., 2014). It may therefore be desirable to separate instruments and use haptic stimulation to enhance one or a small subset.

Source Separation

Some basic methods for separating sound sources have already been used for converting audio to haptic stimulation. One haptic signal-processing approach uses an expander, which amplifies loud sounds, to extract speech from background noise when the signal-to-noise ratio (SNR) is positive (i.e., the speech is louder than the noise; Fletcher et al., 2018, 2019). This simple real-time approach improves speech-in-noise performance for CI users at positive SNRs but is not expected to be suitable for enhancing music, where the SNRs for individual instruments are typically less favorable. Another approach used pitch extraction methods to separate harmonic and inharmonic sounds (Fletcher et al., 2020c). Pitch extraction is often susceptible to background noise (Jouvet and Laprie, 2017), but the proposed approach was shown to be robust to inharmonic noise (Fletcher et al., 2020c).

However, this and other pitch extraction approaches for enhancing music perception using haptics (e.g., Luo and Hayes, 2019), are not designed to accommodate musical pieces with multiple simultaneous harmonic sounds. More advanced multi-pitch extraction methods will likely be required if they are to be effective across a range of musical pieces.

A range of noise-reduction techniques are deployed in hearing-assistive devices to extract speech from background noise, and these might also have utility for haptic signal-processing strategies. One commonly used group of techniques focus on the temporal domain. These exploit the fact that the amplitude envelope of speech tends to have a lower modulation frequency and depth than environmental noise (Ding et al., 2017; Lakshmi et al., 2021). These techniques classify speech signals as having a modulation rate less than around 10–30 Hz and a modulation depth greater than around 15 dB (e.g., Schum, 2003). Another commonly used group of techniques focus on the spectral domain. These estimate the spectrum of the background noise and subtract this from the speech-in-noise signal. To determine when only background noise is present, these spectral subtraction techniques typically employ a voice detector (Boll, 1979; Ephraim and Malah, 1984). Another approach, that is less commonly used in modern hearing-assistive devices, focuses on harmonic structure. Unlike many noise signals, speech contains harmonics with strong co-modulation. Synchrony detection algorithms classify the signal as speech if it has highly synchronous energy fluctuations across frequency bands (Schum, 2003). The latest noise-reduction strategies in hearing-assistive devices often deploy multiple noise-reduction approaches, as well as using environmental classification methods and adaptive filtering (Ricketts and Hornsby, 2005; Peeters et al., 2009). These techniques might be adapted to focus on the typical characteristics of musical instruments (e.g., Ding et al., 2017), although it should be noted that these approaches were developed to extract a single sound source and that musical instruments often share temporal and spectral characteristics. Furthermore, a recent meta-analysis found no significant improvement in speech intelligibility with digital noise-reduction algorithms in HA users, although subjective outcomes, such as sound quality, did show moderate improvement (Lakshmi et al., 2021).

Many HAs have dedicated signal-processing settings for music listening. While manufacturers often do not reveal exactly how these differ from those for improving speech-in-noise performance, they often appear to reduce or remove the noise-reduction applied and use slower-acting compression (Moore, 2016). In a survey of HA users, no clear difference in music experience was found between those with a dedicated music setting on their HA and those without (Madsen and Moore, 2014).

More advanced methods for separating sound sources in musical pieces have also been developed. One approach attempts to separate harmonic and percussive sounds (Buyens et al., 2014, 2015). While this approach may have utility for haptic signal-processing, its potential is significantly limited by the fact that it cannot separate common key instruments, such as vocals and bass, from each other. Another method using non-negative matrix factorization has shown potential for separating and enhancing vocals, although notable distortions and artifacts were observed (Pons et al., 2016). More advanced machine-learning-based source separation methods have also been tested and were found to outperform non-negative matrix factorization (Gajecki and Nogueira, 2018). Deep convolutional auto encoders, which combine denoising auto encoding and convolutional neural networks, performed extremely well, but only when the audio processed was similar to that used to train the algorithm. Multilayer perceptrons and deep recurrent neural networks, on the other hand, performed well across a range of data. The authors concluded that multilayer perceptrons were most suitable because they were faster to compute, although none of the techniques tested were implemented in real-time. A recent study developed a real-time multilayer perceptron method, which was shown to be effective in isolating vocals and to be robust to background noise and reverb that would be encountered with live audio (Tahmasebi et al., 2020). Advanced source separation approaches like these could be critical to maximizing the effectiveness of haptic devices for enhancing music perception.

Feature Extraction

In addition to deciding the source or sources to be separated, it will be important to determine which sound features should be provided through haptic stimulation. Features shown to enhance speech perception when presented through haptic stimulation, such as amplitude envelope (e.g., Brooks and Frost, 1983; Fletcher et al., 2019) and fundamental frequency (e.g., Huang et al., 2017), should be explored. The utility of other features, like those used by the Moving Picture Expert Group for audio content, should also be investigated as they could provide additional information, such as timbre (as in, for example, Jack et al., 2015). These include: spectral features, such as centroid, spread, and flatness; harmonic features, such as centroid, spread, variation, and deviation; and temporal features, such as centroid and log attack time (see Zhang and Ras, 2007).

The optimal features to extract are likely to differ across instruments and musical styles. For example, vocals in rap music might require rhythmic information through features such as amplitude envelope, whereas vocals in show tunes may benefit more from pitch-based features, such as fundamental frequency. For a non-harmonic instrument like a snare drum, pitch-based features cannot be extracted and features like spectral spread or spectral centroid might be most appropriate.

Sound classification algorithms will be important to any approach that selects features based on instrument type or musical style. A range of methods for music genre classification have shown promise, including ensemble classifiers and methods that implement sound source segregation approaches, such as non-negative matrix factorization (Silla et al., 2007; Pérez-García et al., 2010; Rosner and Kostek, 2018). Several instrument classification approaches have also shown promise, including advanced methods using deep convolutional neural networks (Benetos et al., 2006; Gomez et al., 2018; Solanki and Pandey, 2019; Racharla et al., 2020). Establishing the most effective classification approaches and auditory features to provide through haptic stimulation will be a critical part of future research in this area.

Haptic Mapping

Having separated the instruments and extracted sound features, the next consideration will be how to map these to haptic stimulation. Haptic music-enhancement approaches should take advantage of the tactile system's large dynamic range (Verrillo et al., 1969; Fletcher et al., 2021a,b) and high sensitivity to intensity differences, both at a single site and across sites (Gescheider et al., 1996; Fletcher et al., 2021a,b). As discussed (see section "Limits of Haptic Sensitivity Compared to Hearing-Impaired Listening"), this might include spatially mapping instruments using amplitude panning across sites, such as the wrists (Fletcher and Zgheib, 2020; Fletcher et al., 2020a,b), that mimics amplitude panning of instruments within a musical piece. Stimulus features (such as fundamental frequency) might also be mapped to changes in spatial location on the skin to maximize information transfer (e.g., Brooks and Frost, 1983; Karam et al., 2010; Fletcher et al., 2020c).

IMPORTANT CUTTING-EDGE TECHNOLOGIES

Modern haptic devices can take advantage of critical recent advances in technology (see Fletcher, 2020 for a detailed review). These include: haptic motor and driver technology to deliver high-fidelity stimulation with low power usage; battery technology, to increase the potential power usage and reduce the necessity for frequent charging; manufacturing techniques, such as 3D printing, to facilitate the development of comfortable, aesthetically acceptable, and easy to self-fit devices; wireless technologies, to allow audio streaming from remote microphones and other devices and to link processing across multiple stimulation points on the body; and microprocessors to allow advanced signal-processing. Future devices might also take advantage of flexible microprocessor technology, which is currently being developed (Biggs et al., 2021). This could allow additional signal-processing capacity to be built into device components that need to be flexible, such as straps.

Several other recent and ongoing technological developments could be exploited to maximize haptic enhancement of music perception. One example is big data systems that have the capacity to collect data from devices as they are being used in the real world. This technology is currently being exploited in the EVOTION platform (funded by the European Union) and the HearingFitness program (developed by Oticon Medical), which use big data collected from devices in the real world to inform policy-making (Gutenberg et al., 2018; Dritsakis et al., 2020; Saunders et al., 2020). In future, the technology might also be used to optimize haptic signal-processing. **Figure 2** shows an example remote data processing pipeline. In this pipeline, audio is streamed to the haptic device from a hearing-assistive device to ensure maximum correlation between the audio and haptic signals (see Fletcher, 2020). Audio statistics, such as spectral flatness and short-term energy, are then extracted by the haptic device and transferred to a smartphone. The smartphone also has an app to collect user feedback, for example ratings of sound quality and music enjoyment, and to link clinical data such as hearing-assistive device type and hearing-loss profile. Audio statistics and user data are stored on the smartphone and uploaded to a remote server or The Cloud when a WIFI connection is established (to reduce power consumption and mobile data usage). The data is processed remotely to update models and derive optimized signal-processing parameters. These models could be optimized for each individual or be used as part of a big data approach for optimizing signal-processing globally, for subgroups of users, or for different music types. Once updated signal-processing parameters are determined, these are transferred to the haptic device via the smartphone.

To implement a remote data processing pipeline of this sort, exploitation of cutting-edge technology and further research are required. It should be noted that, in practice, simpler systems that collect user feedback to optimize new iterations of algorithms might be developed before a full pipeline like that proposed is implemented. One key technology for the proposed pipeline is wireless data streaming. This can be achieved using the latest Bluetooth Low Energy technology, which allows multiple simultaneous data streams, has low power usage, and is already integrated into many of the latest hearing-assistive devices. Another critical element is the development of a smartphone app for collecting user feedback, which must have a high level of data security and privacy. User feedback is likely to be important as music perception varies substantially across hearing-impaired listeners due to factors such as previous musical experience (Galvin et al., 2009; Gfeller et al., 2015). The app developed for the proposed system can build on existing apps that are already deployed in the growing field of telemedicine to collect real-world user feedback for optimization of hearing-assistive devices, such as ReSound Assist (Convery et al., 2020). Finally, future research will be required to determine the optimal audio statistics to be extracted and sent for remote processing, as well as the most effective approaches for processing this data and deriving optimal signal-processing parameters. The recent expansion in remote data collection and analysis capacity through systems such as Cloud computing will be critical in allowing big data to be processed with sophisticated models.

In addition to user- and stimulus-based optimization of signal processing, steps should be taken to ensure that haptic stimulation is perceived as uniformly as possible across users. One simple way to do this is to adjust the stimulation intensity based on each user's detection thresholds (as is done for hearing-assistive devices). It may also be important to adapt the intensity based on the fitting of the device on the body. The fitting (e.g., how tightly the device is strapped on) can substantially alter the amount of pressure applied to the haptic motor and the coupling with the skin. Techniques have recently been developed to estimate the pressing force on the motor and dynamically calibrate it (Dementyev et al., 2020). Such techniques should be explored for future haptic devices for enhancing music perception.

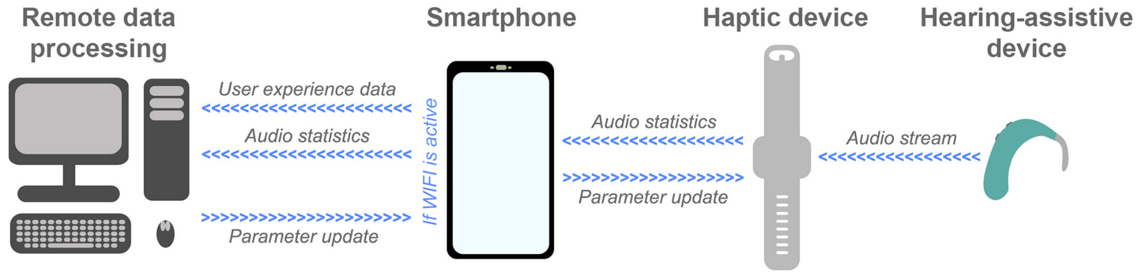


FIGURE 2 | Schematic representation of an example future remote data-processing pipeline for haptic signal-processing optimization (described in the text). Audio is streamed from a hearing-assistive device to a haptic device that extracts audio statistics and sends them to a smartphone. A smartphone app also collects user feedback about their experience using the haptic device. Audio statistics and user experience data are then sent for remote data processing, where optimized signal-processing parameters are determined. Finally, these updated parameters are sent to the smartphone and uploaded to the haptic device.

DISCUSSION

Music perception is often significantly impaired in those with hearing loss. Critical factors are the loss of ability to discriminate sounds of different frequencies and a reduction in dynamic range. Recently, it has been shown that haptic devices can be highly effective at providing intensity (Fletcher and Zgheib, 2020; Fletcher et al., 2020a, 2021a,b) and frequency information (Fletcher et al., 2020c), and can support perception of complex signals such as speech (Huang et al., 2017; Fletcher et al., 2018, 2019, 2020b). However, despite the large number of haptic systems that have been developed for enhancing music perception, there is a lack of robust data on whether haptic devices can effectively improve music perception for hearing-impaired listeners. Whilst haptic stimulation has vast potential to enhance music perception, a significant research program is required to provide a clear evidence base.

Several critical technologies have been developed in recent years, which can be exploited in future haptic devices. These allow faithful haptic signal reproduction, advanced signal processing, wireless communication between hardware components (such as smartphones, microphones, and haptic devices), long battery lives, and rapid prototyping and manufacturing. These technologies give scope for vast improvements to current haptic devices for enhancing hearing. In addition, several key emerging technologies and methods have been identified, which further expand the potential for haptic enhancement of music perception. These include cloud computing and cutting-edge machine-learning approaches. Exploitation of these new technologies could considerably increase haptic enhancement of listening and allow a dramatic expansion in access to music and other media for hearing-impaired listeners.

Another consideration raised in this review is the interaction between haptic, audio, and visual stimulation. It was highlighted that significant sound information from music is accessible through vision, particularly pitch interval size and direction. Future work should establish whether critical sound information, such as pitch, provided through haptic, audio, and visual modalities can be effectively combined to enhance discrimination. It will also be critical to explore how providing sound information through non-auditory senses can alter auditory perception. This could determine whether future research on haptic enhancement aims to restore conventional music perception or whether it instead seeks to offer an alternative way to experience music.

In addition to enhancing music listening, there is significant potential for haptics to be used for enhancing musical performance in hearing-impaired individuals. Of particular interest might be enhancement of vocal performance. CI users often have considerable difficulties when singing, particularly in producing the correct pitch (Xu et al., 2009; Mao et al., 2013). There have been some promising results when providing pitch information to hearing-impaired listeners through haptic stimulation to improve singing (Sakajiri et al., 2010, 2013; Shin et al., 2020; Hopkins et al., 2021). Future work should establish the effectiveness of the alternative pitch-based haptic stimulation approach suggested by Fletcher et al. (2020c), which was shown to provide high-resolution pitch information. These pitch-based approaches might also be highly effective for speech rehabilitation. Congenitally deaf individuals often struggle to acquire and maintain normal speech (Smith, 1975; Gold, 1980), and those who suffer hearing loss later in life often also experience a reduction in vocal control, often including greater pitch variability (Lane and Webster, 1991).

This review has discussed the enormous potential of haptic stimulation to enhance music listening. It is estimated that around 1.6 billion people across the world have hearing loss, with this number expected to increase rapidly (Haile et al., 2021). Alongside this growth in the number of people who need support with hearing impairment is a rapid growth in technologies that could improve and expand this support. The use of haptic stimulation to enhance listening for those with hearing impairment offers an opportunity to exploit many of these recently developed technologies. The time therefore seems right for a major expansion of research into haptic enhancement of listening.

If effective and accessible systems are developed, as well as directly enhancing music enjoyment, they could substantially improve access to and enjoyment of media (such as films and documentaries), video games, and social events, such as weddings. Furthermore, given that music is an extremely challenging signal because of its complexity, progress in this area could have substantial benefits for enhancing communication and spatial awareness in complex everyday acoustic environments. Thanks to inexpensive core technologies, haptic devices could become widely accessible, including in low- and middle-income countries, and bring substantial improvements in quality of life for those with hearing impairment.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

MF's salary was funded by the William Demant Foundation. The article processing fee was funded by the Signal Processing, Audio and Hearing Group at the University of Southampton - Institute of Sound and Vibration Research.

ACKNOWLEDGMENTS

My deepest thanks to Anders de Broojinan and Samir W. Poiré for illuminating discussions and comments on the text.

REFERENCES

Abel, M. K., Li, H. C., Russo, F. A., Schlaug, G., and Loui, P. (2016). Audiovisual interval size estimation is associated with early musical training. *PLoS One* 11:e0163589. doi: 10.1371/journal.pone.0163589

Aitkin, L. M., Kenyon, C. E., and Philpott, P. (1981). The representation of the auditory and somatosensory systems in the external nucleus of the cat inferior colliculus. *J. Comp. Neurol.* 196, 25–40. doi: 10.1002/cne.901960104

Allman, B. L., Keniston, L. P., and Meredith, M. A. (2009). Adult deafness induces somatosensory conversion of ferret auditory cortex. *Proc. Natl. Acad. Sci. U S A.* 106, 5925–5930. doi: 10.1073/pnas.0809483106

Altinsoy, M. E. (2003). "Perceptual aspects of auditory-tactile asynchrony," in *Proceedings of the 10th International Congress on Sound and Vibration*, (Stockholm: Institut für Kommunikationsakustik).

Auer, E. T. Jr., Bernstein, L. E., Sungkarat, W., and Singh, M. (2007). Vibrotactile activation of the auditory cortices in deaf versus hearing adults. *Neuroreport* 18, 645–648. doi: 10.1097/WNR.0b013e3280d943b9

Bach-Y-Rita, P., Collins, C. C., Saunders, F. A., White, B., and Scadden, L. (1969). Vision substitution by tactile image projection. *Nature* 221, 963–964. doi: 10.1038/221963a0

Bach-Y-Rita, P., Danilov, Y., Tyler, M., and Grimm, R. J. (2005). Late human brain plasticity: vestibular substitution with a tongue BrainPort human-machine interface. *Intellectica* 40, 115–122. doi: 10.3406/intel.2005.1362

Bach-Y-Rita, P., Tyler, M. E., and Kaczmarek, K. A. (2003). Seeing with the brain. *Int. J. Hum-Comput. Int.* 15, 285–295. doi: 10.1207/S15327590ijhc1502_6

Benetos, E., Kotti, M., and Kotropoulos, C. (2006). "Musical instrument classification using non-negative matrix factorization algorithms and subset feature selection," in *Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing*, (Toulouse: Institute of Electrical and Electronics Engineers Inc), 221–224. doi: 10.1109/ICASSP.2006.1661252

Bergeson, T. R., Pisoni, D. B., and Davis, R. A. (2005). Development of audiovisual comprehension skills in prelingually deaf children with cochlear implants. *Ear Hear.* 26, 149–164. doi: 10.1097/00003446-200504000-200504004

Bernstein, L. E., Auer, E. T. Jr., Eberhardt, S. P., and Jiang, J. (2013). Auditory perceptual learning for speech perception can be enhanced by audiovisual training. *Front. Neurosci.* 7:34. doi: 10.3389/fnins.2013.00034

Biggs, J., Myers, J., Kufel, J., Ozer, E., Craske, S., Sou, A., et al. (2021). A natively flexible 32-bit Arm microprocessor. *Nature* 595, 532–536. doi: 10.1038/s41586-021-03625-w

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

Brockmeier, S. J., Fitzgerald, D., Searle, O., Fitzgerald, H., Grasmeder, M., Hilbig, S., et al. (2011). The MuSIC perception test: a novel battery for testing music perception of cochlear implant users. *Cochlear Implants Int.* 12, 10–20. doi: 10.1179/146701010X12677899497236

Brooks, P. L., and Frost, B. J. (1983). Evaluation of a tactile vocoder for word recognition. *J. Acoust. Soc. Am.* 74, 34–39. doi: 10.1121/1.389685

Brooks, P. L., Frost, B. J., Mason, J. L., and Chung, K. (1985). Acquisition of a 250-word vocabulary through a tactile vocoder. *J. Acoust. Soc. Am.* 77, 1576–1579. doi: 10.1121/1.392000

Brooks, P. L., Frost, B. J., Mason, J. L., and Gibson, D. M. (1986a). Continuing evaluation of the Queen's University tactile vocoder. I: identification of open set words. *J. Rehabil. Res. Dev.* 23, 119–128.

Brooks, P. L., Frost, B. J., Mason, J. L., and Gibson, D. M. (1986b). Continuing evaluation of the Queen's University tactile vocoder II: identification of open set sentences and tracking narrative. *J. Rehabil. Res. Dev.* 23, 129–138.

Buyens, W., Van Dijk, B., Moonen, M., and Wouters, J. (2014). Music mixing preferences of cochlear implant recipients: a pilot study. *Int. J. Audiol.* 53, 294–301. doi: 10.3109/14992027.2013.873955

Buyens, W., Van Dijk, B., Wouters, J., and Moonen, M. (2015). A stereo music preprocessing scheme for cochlear implant users. *IEEE Trans. Biomed. Eng.* 62, 2434–2442. doi: 10.1109/TBME.2015.2428999

Caetano, G., and Jousmaki, V. (2006). Evidence of vibrotactile input to human auditory cortex. *Neuroimage* 29, 15–28. doi: 10.1016/j.neuroimage.2005.07.023

Chen, J. K., Chuang, A. Y., Mcmahon, C., Hsieh, J. C., Tung, T. H., and Li, L. P. (2010). Music training improves pitch perception in prelingually deafened children with cochlear implants. *Pediatrics* 125, e793–e800. doi: 10.1542/peds.2008-3620

Chinello, F., Aurilio, M., Pacchierotti, C., and Prattichizzo, D. (2014). *The HapBand: A Cutaneous Device for Remote Tactile Interaction*. Berlin: Springer, 284–291.

Choi, J. E., Won, J. H., Kim, C. H., Cho, Y. S., Hong, S. H., and Moon, I. J. (2018). Relationship between spectrotemporal modulation detection and music perception in normal-hearing, hearing-impaired, and cochlear implant listeners. *Sci. Rep.* 8:800. doi: 10.1038/s41598-017-17350-w

Ciesla, K., Wolak, T., Lorens, A., Heimler, B., Skarzynski, H., and Amedi, A. (2019). Immediate improvement of speech-in-noise perception through multisensory stimulation via an auditory to tactile sensory substitution. *Restor. Neurol Neurosci.* 37, 155–166. doi: 10.3233/RNN-190898

Convery, E., Keidser, G., Mclelland, M., and Groth, J. (2020). A smartphone app to facilitate remote patient-provider communication in hearing health care: usability and effect on hearing aid outcomes. *Telemed. J. Health* 26, 798–804. doi: 10.1089/tmj.2019.0109

Cooper, W. B., Tobey, E., and Loizou, P. C. (2008). Music perception by cochlear implant and normal hearing listeners as measured by the montreal battery for evaluation of Amusia. *Ear Hear.* 29, 618–626. doi: 10.1097/AUD.0b013e318174e787

Cowan, R. S., Blamey, P. J., Sarant, J. Z., Galvin, K. L., Alcantara, J. I., Whitford, L. A., et al. (1991). Role of a multichannel electrotactile speech processor in a cochlear implant program for profoundly hearing-impaired adults. *Ear Hear.* 12, 39–46. doi: 10.1097/00003446-199102000-199102005

Darrow, A. A. (1993). The role of music in deaf culture: implications for music educators. *J. Res. Music Educ.* 41, 93–110. doi: 10.2307/3345402

de Dieuleveult, A. L., Siemonsma, P. C., Van Erp, J. B., and Brouwer, A. M. (2017). Effects of aging in multisensory integration: a systematic review. *Front. Aging Neurosci.* 9:80. doi: 10.3389/fnagi.2017.00080

De Filippo, C. L. (1984). Laboratory projects in tactile aids to lipreading. *Ear Hear.* 5, 211–227. doi: 10.1097/00003446-198407000-198407006

Dementyev, A., Olwal, A., and Lyon, R. F. (2020). "Haptics with input: back-EMF in linear resonant actuators to enable touch, pressure and environmental awareness," in *Proceedings of the UIST '20: The 33rd Annual ACM Symposium on User Interface Software and Technology*, (New York, NY: Association for Computing Machinery).

Diederich, A., Colonius, H., and Schomburg, A. (2008). Assessing age-related multisensory enhancement with the time-window-of-integration model. *Neuropsychologia* 46, 2556–2562. doi: 10.1016/j.neuropsychologia.2008.03.026

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., and Poeppel, D. (2017). Temporal modulations in speech and music. *Neurosci. Biobehav. Rev.* 81, 181–187. doi: 10.1016/j.neubiorev.2017.02.011

Dorman, M. F., Loiselle, L. H., Cook, S. J., Yost, W. A., and Gifford, R. H. (2016). Sound source localization by normal-hearing listeners, hearing-impaired listeners and cochlear implant listeners. *Audiol. Neurootol.* 21, 127–131. doi: 10.1159/000444740

Drennan, W. R., Oleson, J. J., Gfeller, K., Crosson, J., Driscoll, V. D., Won, J. H., et al. (2015). Clinical evaluation of music perception, appraisal and experience in cochlear implant users. *Int. J. Audiol.* 54, 114–123. doi: 10.3109/14992027.2014.948219

Drennan, W. R., and Rubinstein, J. T. (2008). Music perception in cochlear implant users and its relationship with psychophysical capabilities. *J. Rehabil. Res. Dev.* 45, 779–789. doi: 10.1682/jrrd.2007.08.0118

Driscoll, V. D. (2012). The effects of training on recognition of musical instruments by adults with cochlear implants. *Semin. Hear.* 33, 410–418. doi: 10.1055/s-0032-1329230

Dritsakis, G., Trenkova, L., Sliwinska-Kowalska, M., Brdaric, D., Pontoppidan, N. H., Katrakazas, P., et al. (2020). Public health policy-making for hearing loss: stakeholders' evaluation of a novel eHealth tool. *Health Res. Policy Syst.* 18:125. doi: 10.1186/s12961-020-00637-632

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

Drullman, R., Festen, J. M., and Plomp, R. (1994). Effect of temporal envelope smearing on speech reception. *J. Acoust. Soc. Am.* 95, 1053–1064. doi: 10.1121/1.408467

Elliot, L. L. (1962). Backward and forward masking of probe tones of different frequencies. *J. Acoust. Soc. Am.* 34, 1116–1117. doi: 10.1121/1.1918254

Ephraim, Y., and Malah, D. (1984). Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. *IEEE Trans. Acoustics Speech Signal Proc.* 32, 1109–1121. doi: 10.1109/Tassp.1984.1164453

Finney, E. M., Clementz, B. A., Hickok, G., and Dobkins, K. R. (2003). Visual stimuli activate auditory cortex in deaf subjects: evidence from MEG. *Neuroreport* 14, 1425–1427. doi: 10.1097/00001756-200308060-200308064

Finney, E. M., Fine, I., and Dobkins, K. R. (2001). Visual stimuli activate auditory cortex in the deaf. *Nat. Neurosci.* 4, 1171–1173. doi: 10.1038/nn763

Fletcher, M. D. (2020). Using haptic stimulation to enhance auditory perception in hearing-impaired listeners. *Expert Rev. Med. Dev.* 18, 63–74. doi: 10.1080/17434440.2021.1863782

Fletcher, M. D., Cunningham, R. O., and Mills, S. R. (2020a). Electro-haptic enhancement of spatial hearing in cochlear implant users. *Sci. Rep.* 10:1621. doi: 10.1038/s41598-020-58503-58508

Fletcher, M. D., Song, H., and Perry, S. W. (2020b). Electro-haptic stimulation enhances speech recognition in spatially separated noise for cochlear implant users. *Sci. Rep.* 10:12723.

Fletcher, M. D., Thini, N., and Perry, S. W. (2020c). Enhanced pitch discrimination for cochlear implant users with a new haptic neuroprosthetic. *Sci. Rep.* 10:10354. doi: 10.1038/s41598-020-67140-67140

Fletcher, M. D., Hadeedi, A., Goehring, T., and Mills, S. R. (2019). Electro-haptic enhancement of speech-in-noise performance in cochlear implant users. *Sci. Rep.* 9:11428. doi: 10.1038/s41598-019-47718-z

Fletcher, M. D., Mills, S. R., and Goehring, T. (2018). Vibro-tactile enhancement of speech intelligibility in multi-talker noise for simulated cochlear implant listening. *Trends Hear.* 22:2331216518797838. doi: 10.1177/2331216518797838

Fletcher, M. D., and Verschuur, C. A. (2021). Electro-haptic stimulation: a new approach for improving cochlear-implant listening. *Front. Neurosci.* 15:581414. doi: 10.3389/fnins.2021.581414

Fletcher, M. D., and Zgheib, J. (2020). Haptic sound-localisation for use in cochlear implant and hearing-aid users. *Sci. Rep.* 10:14171. doi: 10.1038/s41598-020-70379-70372

Fletcher, M. D., Zgheib, J., and Perry, S. W. (2021a). Sensitivity to haptic sound-localisation cues. *Sci. Rep.* 11:312. doi: 10.1038/s41598-020-79150-z

Fletcher, M. D., Zgheib, J., and Perry, S. W. (2021b). Sensitivity to haptic sound-localization cues at different body locations. *Sensors (Basel)* 21:3770. doi: 10.3390/s21113770

Fontana, F., Papetti, S., Jarvelainen, H., and Avanzini, F. (2017). Detection of keyboard vibrations and effects on perceived piano quality. *J. Acoust. Soc. Am.* 142:2953. doi: 10.1121/1.5009659

Foxe, J. J., Morocz, I. A., Murray, M. M., Higgins, B. A., Javitt, D. C., and Schroeder, C. E. (2000). Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Brain Res. Cogn. Brain Res.* 10, 77–83. doi: 10.1016/s0926-6410(00)00024-20

Fu, Q. J., and Galvin, J. J. III (2007). Perceptual learning and auditory training in cochlear implant recipients. *Trends Amplif.* 11, 193–205. doi: 10.1177/1084713807301379

Fujita, S., and Ito, J. (1999). Ability of nucleus cochlear implantees to recognize music. *Ann. Otol. Rhinol. Laryngol.* 108, 634–640. doi: 10.1177/000348949910800702

Gajecki, T., and Nogueira, W. (2018). Deep learning models to remix music for cochlear implant users. *J. Acoust. Soc. Am.* 143:3602. doi: 10.1121/1.5042056

Galvin, J. J. III, and Fu, Q. J. (2009). Influence of stimulation rate and loudness growth on modulation detection and intensity discrimination in cochlear implant users. *Hear. Res.* 250, 46–54. doi: 10.1016/j.heares.2009.01.009

Galvin, J. J. III, Fu, Q. J., and Nogaki, G. (2007). Melodic contour identification by cochlear implant listeners. *Ear Hear.* 28, 302–319. doi: 10.1097/01.aud.0000261689.35445.20

Galvin, J. J. III, Fu, Q. J., and Shannon, R. V. (2009). Melodic contour identification and music perception by cochlear implant users. *Ann. N. Y. Acad. Sci.* 1169, 518–533. doi: 10.1111/j.1749-6632.2009.04551.x

Garadat, S. N., and Pfingst, B. E. (2011). Relationship between gap detection thresholds and loudness in cochlear-implant users. *Hear. Res.* 275, 130–138. doi: 10.1016/j.heares.2010.12.011

Gauer, J., Nagathil, A., Martin, R., Thomas, J. P., and Volter, C. (2019). Interactive evaluation of a music preprocessing scheme for cochlear implants based on spectral complexity reduction. *Front. Neurosci.* 13:1206. doi: 10.3389/fnins.2019.01206

Gault, R. H. (1924). Progress in experiments on tactile interpretation of oral speech. *J. Ab. Soc. Psychol.* 19, 155–159.

Gault, R. H. (1926). Touch as a substitute for hearing in the interpretation and control of speech. *Arch. Otolaryngol.* 3, 121–135.

Gescheider, G. A. (1966). Resolving of successive clicks by the ears and skin. *J. Exp. Psychol.* 71, 378–381. doi: 10.1037/h0022950

Gescheider, G. A. (1967). Auditory and cutaneous temporal resolution of successive brief stimuli. *J. Exp. Psychol.* 75, 570–572. doi: 10.1037/h0025113

Gescheider, G. A., Bolanowski, S. J. Jr., and Verrillo, R. T. (1989). Vibrotactile masking: effects of stimulus onset asynchrony and stimulus frequency. *J. Acoust. Soc. Am.* 85, 2059–2064. doi: 10.1121/1.397858

Gescheider, G. A., Zwislocki, J. J., and Rasmussen, A. (1996). Effects of stimulus duration on the amplitude difference limen for vibrotaction. *J. Acoust. Soc. Am.* 100, 2312–2319. doi: 10.1121/1.417940

Gfeller, K., Christ, A., Knutson, J. F., Witt, S., Murray, K. T., and Tyler, R. S. (2000). Musical backgrounds, listening habits, and aesthetic enjoyment of adult cochlear implant recipients. *J. Am. Acad. Audiol.* 11, 390–406.

Gfeller, K., Driscoll, V., Kenworthy, M., and Van Voorst, T. (2011). Music therapy for preschool cochlear implant recipients. *Music Ther. Perspect.* 29, 39–49. doi: 10.1093/mtp/29.1.39

Gfeller, K., Guthe, E., Driscoll, V., and Brown, C. J. (2015). A preliminary report of music-based training for adult cochlear implant users: rationales and development. *Cochlear Implants Int.* 16(Suppl. 3), S22–S31. doi: 10.1179/1467010015Z.000000000269

Gfeller, K., Jiang, D., Oleson, J. J., Driscoll, V., and Knutson, J. F. (2010). Temporal stability of music perception and appraisal scores of adult cochlear implant recipients. *J. Am. Acad. Audiol.* 21, 28–34. doi: 10.3766/jaaa.21.1.4

Gfeller, K., Knutson, J. F., Woodworth, G., Witt, S., and Debus, B. (1998). Timbral recognition and appraisal by adult cochlear implant users and normal-hearing adults. *J. Am. Acad. Audiol.* 9, 1–19.

Gfeller, K., and Lansing, C. (1992). Musical perception of cochlear implant users as measured by the primary measures of music audiation - an item analysis. *J. Music Ther.* 29, 18–39. doi: 10.1093/jmt/29.1.18

Gfeller, K., and Lansing, C. R. (1991). Melodic, rhythmic, and timbral perception of adult cochlear implant users. *J. Speech Hear. Res.* 34, 916–920. doi: 10.1044/jshr.3404.916

Gfeller, K., Oleson, J., Knutson, J. F., Breheny, P., Driscoll, V., and Olszewski, C. (2008). Multivariate predictors of music perception and appraisal by adult cochlear implant users. *J. Am. Acad. Audiol.* 19, 120–134. doi: 10.3766/jaaa.19.2.3

Gfeller, K., Turner, C., Mehr, M., Woodworth, G., Fearn, R., Knutson, J. F., et al. (2002a). Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults. *Cochlear Implants Int.* 3, 29–53. doi: 10.1179/cim.2002.3.1.29

Gfeller, K., Witt, S., Adamek, M., Mehr, M., Rogers, J., Stordahl, J., et al. (2002b). Effects of training on timbre recognition and appraisal by postlingually deafened cochlear implant recipients. *J. Am. Acad. Audiol.* 13, 132–145.

Gfeller, K., Witt, S., Woodworth, G., Mehr, M. A., and Knutson, J. (2002c). Effects of frequency, instrumental family, and cochlear implant type on timbre recognition and appraisal. *Ann. Otol. Rhinol. Laryngol.* 111, 349–356. doi: 10.1177/000348940211100412

Gick, B., and Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature* 462, 502–504. doi: 10.1038/nature08572

Gick, B., Ikegami, Y., and Derrick, D. (2010). The temporal window of audio-tactile integration in speech perception. *J. Acoust. Soc. Am.* 128, 342–346. doi: 10.1121/1.3505759

Gillmeister, H., and Eimer, M. (2007). Tactile enhancement of auditory detection and perceived loudness. *Brain Res.* 1160, 58–68. doi: 10.1016/j.brainres.2007.03.041

Glennon, E., Svirsky, M. A., and Froemke, R. C. (2020). Auditory cortical plasticity in cochlear implant users. *Curr. Opin. Neurobiol.* 60, 108–114. doi: 10.1016/j.conb.2019.11.003

Goff, G. D. (1967). Differential discrimination of frequency of cutaneous mechanical vibration. *J. Exp. Psychol.* 74, 294–299. doi: 10.1037/h0024561

Gold, T. (1980). Speech production in hearing-impaired children. *J. Commun. Disord.* 13, 397–418. doi: 10.1016/0021-9924(80)90042-90048

Gomez, J. S., Abeßer, J., and Cano, E. (2018). "Jazz solo instrument classification with convolutional neural networks, source separation, and transfer," in *Proceedings of the 19th ISMIR Conference*, (Paris: ISMIR).

Guelke, R. W., and Huyssen, R. M. J. (1959). Development of apparatus for the analysis of sound by the sense of touch. *J. Acoust. Soc. Am.* 31, 799–809. doi: 10.1121/1.1907789

Gunther, E., Davenport, G., and O'modhrain, M. S. (2003). Cutaneous grooves: composing for the sense of touch. *J. New Music Res.* 32, 369–381. doi: 10.1076/jnmr.32.4.369.18856

Gupta, A., Irudayaraj, A. A. R., and Balakrishnan, R. (2017). "HapticClench: investigating squeeze sensations using memory alloys," in *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, (Québec City, QC: Association for Computing Machinery).

Gutenberg, J., Katrakazas, P., Trenkova, L., Murdin, L., Brdaric, D., Koloutsou, N., et al. (2018). Big data for sound policies: toward evidence-informed hearing health policies. *Am. J. Audiol.* 27, 493–502. doi: 10.1044/2018_AJA-IMIA3-18-13

Haile, L. M., Kamenov, K., Briant, P. S., Orji, A. U., Steinmetz, J. D., Abdoli, A., et al. (2021). Hearing loss prevalence and years lived with disability, 1990-2019: findings from the global burden of disease study 2019. *Lancet* 397, 996–1009. doi: 10.1016/S0140-6736(21)00516-X

Hall, J. W., and Fernandes, M. A. (1983). Monaural and binaural intensity discrimination in normal and cochlear-impaired listeners. *Audiology* 22, 364–371. doi: 10.3109/00206098309072796

Hallam, S. (2010). The power of music: its impact on the intellectual, social and personal development of children and young people. *Int. J. Music Ed.* 28, 269–289. doi: 10.1177/0255761410370658

Hanin, L., Boothroyd, A., and Hnath-Chisolm, T. (1988). Tactile presentation of voice fundamental frequency as an aid to the speechreading of sentences. *Ear Hear.* 9, 335–341. doi: 10.1097/00003446-198812000-198812010

Hashizume, S., Sakamoto, S., Suzuki, K., and Ochiai, Y. (2018). "LIVEJACKET: wearable music experience device with multiple speakers," in *Distributed, Ambient and Pervasive Interactions: Understanding Humans, Dapi 2018*, eds N. Streitz and S. Konomi (Berlin: Springer), 359–371. doi: 10.1007/978-3-319-91125-0_30

Haynes, A., Lawry, J., Kent, C., and Rossiter, J. (2021). FeelMusic: enriching our emotive experience of music through audio-tactile mappings. *Multimodal Technol. Interact.* 5, 1–21. doi: 10.3390/mti5060029

Herremans, D., and Chuan, C. H. (2020). The emergence of deep learning: new opportunities for music and audio technologies. *Neural Comput. Appl.* 32, 913–914. doi: 10.1007/s00521-019-04166-4160

Hoeckner, B., Wyatt, E. W., Decety, J., and Nusbaum, H. (2011). Film music influences how viewers relate to movie characters. *Psych. Aesthet. Creat. Arts* 5, 146–153. doi: 10.1037/a0021544

Hopkins, C., Fulford, R., Seiffert, G., and Ginsborg, J. (2021). Perception and learning of relative pitch by musicians using the vibrotactile mode. *Musicae Scientiae* 10298649211015278. doi: 10.1177/10298649211015278 [Epub ahead of print].

Huang, J., Lu, T., Sheffield, B., and Zeng, F. (2019). Electro-tactile stimulation enhances cochlear-implant melody recognition: effects of rhythm and musical training. *Ear Hear.* 41, 160–113. doi: 10.1097/AUD.0000000000000749

Huang, J., Sheffield, B., Lin, P., and Zeng, F. G. (2017). Electro-tactile stimulation enhances cochlear implant speech recognition in noise. *Sci. Rep.* 7:2196. doi: 10.1038/s41598-017-02429-2421

Isaiah, A., Vongpaisal, T., King, A. J., and Hartley, D. E. (2014). Multisensory training improves auditory spatial processing following bilateral cochlear implantation. *J. Neurosci.* 34, 11119–11130. doi: 10.1523/JNEUROSCI.4767-13.2014

Jack, R., Mcpherson, A., and Stockman, T. (2015). "Designing tactile musical devices with and for deaf users: a case study," in *Proceedings of the International Conference on the Multimodal Experience of Music*, (Sheffield: University of Sheffield), 1–7.

Jiam, N. T., Caldwell, M. T., and Limb, C. J. (2017). What does music sound like for a cochlear implant user? *Otol. Neurotol.* 38, 240–247. doi: 10.1097/MAO.0000000000001448

Jiam, N. T., and Limb, C. J. (2019). Rhythm processing in cochlear implant-mediated music perception. *Ann. N. Y. Acad. Sci.* 1453, 22–28. doi: 10.1111/nyas.14130

Jouvet, D., and Laprie, Y. (2017). "Performance analysis of several pitch detection algorithms on simulated and real noisy speech data," in *Proceedings of the 2017 25th European Signal Processing Conference*, (Greece), 1614–1618.

Kang, R., Nimmons, G. L., Drennan, W., Longnion, J., Ruffin, C., Nie, K., et al. (2009). Development and validation of the University of Washington clinical assessment of music perception test. *Ear Hear.* 30, 411–418. doi: 10.1097/AUD.0b013e3181a61bc0

Karam, M., Branje, C., Nespoli, G., Thompson, N., Russo, F. A., and Fels, D. I. (2010). "The emoti-chair: an interactive tactile music exhibit," in *Proceedings of the CHI'10 Extended Abstracts on Human Factors in Computing Systems*, (Atlanta, GA), 3069–3074. doi: 10.1145/1753846.1753919

Karam, M., Russo, F. A., and Fels, D. I. (2009). Designing the model human cochlea: an ambient crossmodal audio-tactile display. *IEEE Trans. Haptics* 2, 160–169. doi: 10.1109/TOH.2009.32

Keetels, M., and Vroomen, J. (2008). Temporal recalibration to tactile-visual asynchronous stimuli. *Neurosci. Lett.* 430, 130–134. doi: 10.1016/j.neulet.2007.10.044

Kim, I., Yang, E., Donnelly, P. J., and Limb, C. J. (2010). Preservation of rhythmic clocking in cochlear implant users: a study of isochronous versus anisochronous beat detection. *Trends Amplif.* 14, 164–169. doi: 10.1177/1084713810387937

Kim, S., Martens, W. L., and Walker, K. (2006). "Perception of simultaneity and detection of asynchrony between audio and structural vibration in multimodal music reproduction," in *Proceedings of the 120th Convention of the Audio Engineering Society*, (Paris), 6825.

Kong, Y. Y., Cruz, R., Jones, J. A., and Zeng, F. G. (2004). Music perception with temporal cues in acoustic and electric hearing. *Ear Hear.* 25, 173–185. doi: 10.1097/01.aud.0000120365.97792.2f

Kral, A. (2009). [Early hearing experience and sensitive developmental periods]. *HNO* 57, 9–16. doi: 10.1007/s00106-008-1877-1879

Kreft, H. A., Donaldson, G. S., and Nelson, D. A. (2004). Effects of pulse rate and electrode array design on intensity discrimination in cochlear implant users. *J. Acoust. Soc. Am.* 116, 2258–2268. doi: 10.1121/1.1786871

Lakatos, P., Chen, C. M., O'connell, M. N., Mills, A., and Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* 53, 279–292. doi: 10.1016/j.neuron.2006.12.011

Lakshmi, M. S. K., Rout, A., and O'donoghue, C. R. (2021). A systematic review and meta-analysis of digital noise reduction hearing aids in adults. *Disabil. Rehabil. Assist. Technol.* 16, 120–129. doi: 10.1080/17483107.2019.1642394

Landry, S. P., Guillemot, J. P., and Champoux, F. (2013). Temporary deafness can impair multisensory integration: a study of cochlear-implant users. *Psychol. Sci.* 24, 1260–1268. doi: 10.1177/0956797612471142

Lane, H., and Webster, J. W. (1991). Speech deterioration in postlingually deafened adults. *J. Acoust. Soc. Am.* 89, 859–866. doi: 10.1121/1.1894647

Laurienti, P. J., Burdette, J. H., Maldjian, J. A., and Wallace, M. T. (2006). Enhanced multisensory integration in older adults. *Neurobiol. Aging* 27, 1155–1163. doi: 10.1016/j.neurobiolaging.2005.05.024

Leal, M. C., Shin, Y. J., Laborde, M. L., Calmels, M. N., Verges, S., Lugardon, S., et al. (2003). Music perception in adult cochlear implant recipients. *Acta Otolaryngol.* 123, 826–835. doi: 10.1080/00016480310000386

Lee, D. S., Lee, J. S., Oh, S. H., Kim, S. K., Kim, J. W., Chung, J. K., et al. (2001). Cross-modal plasticity and cochlear implants. *Nature* 409, 149–150. doi: 10.1038/35051653

Levanen, S., and Hamdorf, D. (2001). Feeling vibrations: enhanced tactile sensitivity in congenitally deaf humans. *Neurosci. Lett.* 301, 75–77. doi: 10.1016/s0304-3940(01)01597-x

Lewkowicz, D. J., and Ghazanfar, A. A. (2006). The decline of cross-species intersensory perception in human infants. *Proc. Natl. Acad. Sci. U S A.* 103, 6771–6774. doi: 10.1073/pnas.0602027103

Looi, V., Mcdermott, H., Mckay, C., and Hickson, L. (2007). Comparisons of quality ratings for music by cochlear implant and hearing aid users. *Ear Hear.* 28, 59S–61S. doi: 10.1097/AUD.0b013e31803150cb

Looi, V., Mcdermott, H., Mckay, C., and Hickson, L. (2008). Music perception of cochlear implant users compared with that of hearing aid users. *Ear Hear.* 29, 421–434. doi: 10.1097/AUD.0b013e31816a0d0b

Looi, V., Rutledge, K., and Prvan, T. (2019). Music appreciation of adult hearing aid users and the impact of different levels of hearing loss. *Ear Hear.* 40, 529–544. doi: 10.1097/AUD.0000000000000632

Looi, V., and She, J. (2010). Music perception of cochlear implant users: a questionnaire, and its implications for a music training program. *Int. J. Audiol.* 49, 116–128. doi: 10.3109/14992020903405987

Luo, X., and Hayes, L. (2019). Vibrotactile stimulation based on the fundamental frequency can improve melodic contour identification of normal-hearing listeners with a 4-channel cochlear implant simulation. *Front. Neurosci.* 13:1145. doi: 10.3389/fnins.2019.01145

Madsen, S. M. K., and Moore, B. C. J. (2014). Music and hearing aids. *Trends Hear.* 18:2331216514558271. doi: 10.1177/2331216514558271

Mao, Y., Zhang, M., Nutter, H., Zhang, Y., Zhou, Q., Liu, Q., et al. (2013). Acoustic properties of vocal singing in prelingually-deafened children with cochlear implants or hearing aids. *Int. J. Pediatr. Otorhinolaryngol.* 77, 1833–1840. doi: 10.1016/j.ijporl.2013.08.022

Mazzoni, A., and Bryan-Kinns, N. (2016). Mood glove: a haptic wearable prototype system to enhance mood music in film. *Entertainment Comp.* 17, 9–17. doi: 10.1016/j.entcom.2016.06.002

McDermott, H. J. (2004). Music perception with cochlear implants: a review. *Trends Amplif.* 8, 49–82. doi: 10.1177/108471380400800203

Meijer, P. B. (1992). An experimental system for auditory image representations. *IEEE Trans. Biomed. Eng.* 39, 112–121. doi: 10.1109/10.121642

Merchel, S., and Altinsoy, M. E. (2014). The influence of vibrations on musical experience. *J. Aud. Eng. Soc.* 62, 220–234. doi: 10.17743/jaes.2014.0016

Merchel, S., Leppin, A., and Altinsoy, M. E. (2009). "Hearing with your body: the influence of whole-body vibrations on loudness perception," in *Proceedings of the 16th International Congress on Sound and Vibration*, (Kraków).

Meredith, M. A., and Allman, B. L. (2015). Single-unit analysis of somatosensory processing in the core auditory cortex of hearing ferrets. *Euro. J. Neurosci.* 41, 686–698. doi: 10.1111/ejn.12828

Miller, C. W., Bates, E., and Brennan, M. (2016). The effects of frequency lowering on speech perception in noise with adult hearing-aid users. *Int. J. Audiol.* 55, 305–312. doi: 10.3109/14992027.2015.1137364

Mirza, S., Douglas, S. A., Lindsey, P., Hildreth, T., and Hawthorne, M. (2003). Appreciation of music in adult patients with cochlear implants: a patient questionnaire. *Cochlear Implants Int.* 4, 85–95. doi: 10.1179/cim.2003.4.2.85

Moore, B. C. (1996). Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids. *Ear Hear.* 17, 133–161. doi: 10.1097/00003446-199604000-199604007

Moore, B. C., and Glasberg, B. R. (1988). Gap detection with sinusoids and noise in normal, impaired, and electrically stimulated ears. *J. Acoust. Soc. Am.* 83, 1093–1101. doi: 10.1121/1.396054

Moore, B. C., Glasberg, B. R., Hess, R. F., and Birchall, J. P. (1985). Effects of flanking noise bands on the rate of growth of loudness of tones in normal and recruiting ears. *J. Acoust. Soc. Am.* 77, 1505–1513. doi: 10.1121/1.392045

Moore, B. C. J. (2016). Effects of sound-induced hearing loss and hearing aids on the perception of music. *J. Aud. Eng. Soc.* 64, 112–123. doi: 10.17743/jaes.2015.0081

Moriyama, T. K., Nishi, A., Sakuragi, R., Nakamura, T., and Kajimoto, H. (2018). "Development of a wearable haptic device that presents haptics sensation of the finger pad to the forearm," in *Proceedings of the 2018 IEEE Haptics Symposium HAPTICS*, (Piscataway, NJ: IEEE) 180–185. doi: 10.1109/HAPTICS.2018.8357173

Nagathil, A., Weihs, C., Neumann, K., and Martin, R. (2017). Spectral complexity reduction of music signals based on frequency-domain reduced-rank approximations: an evaluation with cochlear implant listeners. *J. Acoust. Soc. Am.* 142:1219. doi: 10.1121/1.5000484

Nanayakkara, S., Taylor, E., Wyse, L., and Ong, S. H. (2009). "An enhanced musical experience for the deaf: design and evaluation of a music display and a haptic chair," in *Proceedings of the 27th Annual Chi Conference on Human Factors in Computing Systems*, (New York, NY: Association for Computing Machinery), 337–346. doi: 10.1145/1518701.1518756

Nava, E., Bottari, D., Villwock, A., Fengler, I., Buchner, A., Lenarz, T., et al. (2014). Audio-tactile integration in congenitally and late deaf cochlear implant users. *PLoS One* 9:e99606. doi: 10.1371/journal.pone.0099606

Navarra, J., Soto-Faraco, S., and Spence, C. (2007). Adaptation to audiotactile asynchrony. *Neurosci. Lett.* 413, 72–76. doi: 10.1016/j.neulet.2006.11.027

Nimmons, G. L., Kang, R. S., Drennan, W. R., Longnion, J., Ruffin, C., Worman, T., et al. (2008). Clinical assessment of music perception in cochlear implant listeners. *Otol. Neurotol.* 29, 149–155. doi: 10.1097/mao.0b013e31812f7244

North, A. C., Hargreaves, D. J., and Mckendrick, J. (1999). The influence of in-store music on wine selections. *J. App. Psych.* 84, 271–276.

O'Connell, B. P., Dedmon, M. M., and Haynes, D. S. (2017). Hearing preservation cochlear implantation: a review of audiologic benefits, surgical success rates, and variables that impact success. *Curr. Otorhinolaryngol. Rep.* 5, 286–294. doi: 10.1007/s40136-017-0176-y

Peeters, H., Kuk, F., Lau, C. C., and Keenan, D. (2009). Subjective and objective evaluation of noise management algorithms. *J. Am. Acad. Aud.* 20, 89–98. doi: 10.3766/jaaa.20.2.2

Pérez-García, T., Pérez-Sancho, C., and Iñesta, J. M. (2010). "Harmonic and instrumental information fusion for musical genre classification," in *Proceedings of the 3rd international workshop on Machine learning and music*, (New York, NY), 49–52. doi: 10.1145/1878003.1878020

Petersen, B., Mortensen, M. V., Hansen, M., and Vuust, P. (2012). Singing in the key of life: a study on effects of musical ear training after cochlear implantation. *Psychomusicol. Music Mind Brain* 22, 134–151.

Pezent, E., Israr, A., Samad, M., Robinson, S., Agarwal, P., Benko, H., et al. (2019). "Tasbi: multisensory squeeze and vibrotactile wrist haptics for augmented and virtual reality," in *Proceedings of the 2019 IEEE World Haptics Conference (WHC)*, (Tokyo: IEEE), 1–6. doi: 10.1109/WHC.2019.8816098

Philips, B., Vinck, B., De Vel, E., Maes, L., D'haenens, W., Keppler, H., et al. (2012). Characteristics and determinants of music appreciation in adult CI users. *Eur. Arch. Otorhinolaryngol.* 269, 813–821. doi: 10.1007/s00405-011-1718-1714

Pons, J., Janer, J., Rode, T., and Nogueira, W. (2016). Remixing music using source separation algorithms to improve the musical experience of cochlear implant users. *J. Acoust. Soc. Am.* 140:4338. doi: 10.1121/1.4971424

Quartz, S. R., and Sejnowski, T. J. (1997). The neural basis of cognitive development: a constructivist manifesto. *Behav. Brain Sci.* 20, 537–556. doi: 10.1017/s0140525x97001581

Racharla, K., Kumar, V., Jayant, C. B., Khairkar, A., and Harish, P. (2020). "Predominant musical instrument classification based on spectral features," in *Proceedings of the 7th International Conference on Signal Processing and Integrated Networks (SPIN)*, (Noida), 617–622. doi: 10.1109/SPIN48934.2020.9071125

Reed, C. M., Delhorne, L. A., and Durlach, N. A. (1992). "Results obtained with Tactaid II and Tactaid VII," in *Proceedings of the The 2nd International Conference on Tactile Aids, Hearing Aids, and Cochlear Implants*, eds A. Risberg, S. Felicetti, G. Plant, and K. E. Spens (Stockholm: Royal Institute of Technology), 149–155.

Ricketts, T. A., and Hornsby, B. W. (2005). Sound quality measures for speech in noise through a commercial hearing aid implementing digital noise reduction. *J. Am. Acad. Audiol.* 16, 270–277. doi: 10.3766/jaaa.16.5.2

Robbins, A. M., Koch, D. B., Osberger, M. J., Zimmerman-Phillips, S., and Kishon-Rabin, L. (2004). Effect of age at cochlear implantation on auditory skill development in infants and toddlers. *Arch. Otolaryngol. Head Neck Surg.* 130, 570–574. doi: 10.1001/archotol.130.5.570

Rosenblum, L. D., and Fowler, C. A. (1991). Audiovisual investigation of the loudness-effort effect for speech and nonspeech events. *J. Exp. Psychol. Hum. Percept. Perform.* 17, 976–985. doi: 10.1037//0096-1523.17.4.976

Rosner, A., and Kostek, B. (2018). Automatic music genre classification based on musical instrument track separation. *J. Intell. Inf. Syst.* 50, 363–384. doi: 10.1007/s10844-017-0464-465

Rothenberg, M., Verrillo, R. T., Zahorian, S. A., Brachman, M. L., et al. (1977). Vibrotactile frequency for encoding a speech parameter. *J. Acoust. Soc. Am.* 62, 1003–1012. doi: 10.1121/1.381610

Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., and Barone, P. (2007). Evidence that cochlear-implanted deaf patients are better multisensory integrators. *Proc. Natl. Acad. Sci. U S A.* 104, 7295–7300. doi: 10.1073/pnas.0609419104

Russo, F. A., Ammirante, P., and Fels, D. I. (2012). Vibrotactile discrimination of musical timbre. *J. Exp. Psychol. Hum. Percept. Perform.* 38, 822–826. doi: 10.1037/a0029046

Russo, F. A., Sandstrom, G. M., and Maksimowski, M. (2011). Mouth versus eyes: gaze fixation during perception of sung interval size. *Psychomusicoloy Music Mind Brain* 21, 98–107. doi: 10.1037/h0094007

Sakajiri, M., Nakamura, K., Fukushima, S., Ifukube, T., and Miyoshi, S. (2010). "Voice pitch control using tactile feedback or the deafblind or the hearing impaired persons to assist their singing," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, (Istanbul: IEEE), doi: 10.1109/ICSMC.2010.5642329

Sakajiri, M., Nakamura, K., Fukushima, S., Miyoshi, S., and Ifukube, T. (2013). "Accuracy of voice pitch control in singing using tactile voice pitch feedback display," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, (Manchester: IEEE), 4201–4206. doi: 10.1109/Smc.2013.716

Saldana, H. M., and Rosenblum, L. D. (1993). Visual influences on auditory pluck and bow judgments. *Percept. Psychophys.* 54, 406–416. doi: 10.3758/bf03205276

Sandmann, P., Dillier, N., Eichele, T., Meyer, M., Kegel, A., Pascual-Marqui, R. D., et al. (2012). Visual activation of auditory cortex reflects maladaptive plasticity in cochlear implant users. *Brain* 135, 555–568. doi: 10.1093/brain/awr329

Saunders, G. H., Christensen, J. H., Gutenberg, J., Pontoppidan, N. H., Smith, A., Spanoudakis, G., et al. (2020). Application of big data to support evidence-based public health policy decision-making for hearing. *Ear Hear.* 41, 1057–1063. doi: 10.1097/AUD.0000000000000850

Schorr, E. A., Fox, N. A., Van Wassenhove, V., and Knudsen, E. I. (2005). Auditory-visual fusion in speech perception in children with cochlear implants. *Proc. Natl. Acad. Sci. U S A.* 102, 18748–18750. doi: 10.1073/pnas.0508862102

Schum, D. J. (2003). Noise-reduction circuitry in hearing aids (2) goals and current strategies. *Hear. J.* 56, 32–41. doi: 10.1097/01.HJ.0000294049.75608.1c

Schurmann, M., Caetano, G., Hlushchuk, Y., Jousmaki, V., and Hari, R. (2006). Touch activates human auditory cortex. *Neuroimage* 30, 1325–1331. doi: 10.1016/j.neuroimage.2005.11.020

Schurmann, M., Caetano, G., Jousmaki, V., and Hari, R. (2004). Hands help hearing: facilitatory audiotactile interaction at low sound-intensity levels. *J. Acoust. Soc. Am.* 115, 830–832. doi: 10.1121/1.1639909

Schutz, M., and Kubovy, M. (2009). Deconstructing a musical illusion: point-light representations capture salient properties of impact motions. *Canadian Acoust* 37, 23–28.

Schutz, M., and Lipscomb, S. (2007). Hearing gestures, seeing music: vision influences perceived tone duration. *Perception* 36, 888–897. doi: 10.1068/p5635

Shannon, R. V. (1990). Forward masking in patients with cochlear implants. *J. Acoust. Soc. Am.* 88, 741–744. doi: 10.1121/1.399777

Sharma, A., Gilley, P. M., Dorman, M. F., and Baldwin, R. (2007). Deprivation-induced cortical reorganization in children with cochlear implants. *Int. J. Audiol.* 46, 494–499. doi: 10.1080/14992020701524836

Shibasaki, M., Kamiyama, Y., and Minamizawa, K. (2016). "Designing a haptic feedback system for hearing-impaired to experience tap dance," in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, (Tokyo), 97–99.

Shin, S., Oh, C., and Shin, H. (2020). "Tactile tone system: a wearable device to assist accuracy of vocal pitch in cochlear implant users," in *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, (Greece), doi: 10.1145/3373625.3418008

Shore, S. E., El Kashlan, H., and Lu, J. (2003). Effects of trigeminal ganglion stimulation on unit activity of ventral cochlear nucleus neurons. *Neuroscience* 119, 1085–1101. doi: 10.1016/s0306-4522(03)00207-200

Shore, S. E., Vass, Z., Wys, N. L., and Altschuler, R. A. (2000). Trigeminal ganglion innervates the auditory brainstem. *J. Comp. Neurol.* 419, 271–285.

Silla, C. N., Kaestner, C. A., and Koerich, A. L. (2007). "Automatic music genre classification using ensemble of classifiers," in *Proceedings of the IEEE International conference on systems, man and cybernetics*, (Montreal), 1687–1692. doi: 10.1007/BF03192561

Smith, C. R. (1975). Residual hearing and speech production in deaf children. *J. Speech Hear. Res.* 18, 795–811. doi: 10.1044/jshr.1804.795

Solanki, A., and Pandey, S. (2019). Music instrument recognition using deep convolutional neural networks. *Int. J. Inf. Tecnol.* 8, 1076–1079. doi: 10.1007/s41870-019-00285-y

Sparks, D. W., Ardell, L. A., Bourgeois, M., Wiedmer, B., and Kuhl, P. K. (1979). Investigating the MESA (multipoint electrotactile speech aid): the transmission of connected discourse. *J. Acoust. Soc. Am.* 65, 810–815. doi: 10.1121/1.382502

Stephens-Fripp, B., Mutlu, R., and Alici, G. (2018). "Applying mechanical pressure and skin stretch simultaneously for sensory feedback in prosthetic hands," in *Proceedings of the 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*, (Enschede).

Strelnikov, K., Rouger, J., Barone, P., and Deguine, O. (2009). Role of speechreading in audiovisual interactions during the recovery of speech comprehension in deaf adults with cochlear implants. *Scand. J. Psychol.* 50, 437–444. doi: 10.1111/j.1467-9450.2009.00741.x

Strelnikov, K., Rouger, J., Demonet, J. F., Lagleyre, S., Fraysse, B., Deguine, O., et al. (2013). Visual activity predicts auditory recovery from deafness after adult cochlear implantation. *Brain* 136, 3682–3695. doi: 10.1093/brain/awt274

Strelnikov, K., Rouger, J., Lagleyre, S., Fraysse, B., Demonet, J. F., Deguine, O., et al. (2015). Increased audiovisual integration in cochlear-implanted deaf patients: independent components analysis of longitudinal positron emission tomography data. *Eur. J. Neurosci.* 41, 677–685. doi: 10.1111/ejn.12827

Svirsky, M. A., Teoh, S. W., and Neuburger, H. (2004). Development of language and speech perception in congenitally, profoundly deaf children as a function of age at cochlear implantation. *Audiol. Neurootol.* 9, 224–233. doi: 10.1159/000078392

Tahmasebi, S., Gajecki, T., and Nogueira, W. (2020). Design and evaluation of a real-time audio source separation algorithm to remix music for cochlear implant users. *Front. Neurosci.* 14:434. doi: 10.3389/fnins.2020.00434

Tajudeen, B. A., Waltzman, S. B., Jethanamest, D., and Svirsky, M. A. (2010). Speech perception in congenitally deaf children receiving cochlear implants in the first year of life. *Otol. Neurotol.* 31, 1254–1260. doi: 10.1097/MAO.0b013e3181f2f475

Thompson, W. F., and Russo, F. A. (2007). Facing the music. *Psychol. Sci.* 18, 756–757. doi: 10.1111/j.1467-9280.2007.01973.x

Thompson, W. F., Russo, F. A., and Livingstone, S. R. (2010). Facial expressions of singers influence perceived pitch relations. *Psychon. Bull. Rev.* 17, 317–322. doi: 10.3758/PBR.17.3.317

Tranchant, P., Shiell, M. M., Giordano, M., Nadeau, A., Peretz, I., and Zatorre, R. J. (2017). Feeling the beat: bouncing synchronization to vibrotactile music in hearing and early deaf people. *Front. Neurosci.* 11:507. doi: 10.3389/fnins.2017.00507

Tremblay, C., Champoux, F., Lepore, F., and Theoret, H. (2010). Audiovisual fusion and cochlear implant proficiency. *Restor. Neurol. Neurosci.* 28, 283–291. doi: 10.3233/RNN-2010-2498

Tsetserukou, D. (2010). "HaptiHug: a novel haptic display for communication of hug over a distance," in *Haptics: Generating and Perceiving Tangible Sensations. EuroHaptics 2010. Lecture Notes in Computer Science*, eds J. B. F. Van Erp, W. M. Bergmann Tiest, and F. C. T. Van Der Helm (Berlin: Springer).

Turgeon, C., Champoux, F., Lepore, F., and Ellemberg, D. (2015). Deficits in auditory frequency discrimination and speech recognition in cochlear implant users. *Cochlear Implants Int.* 16, 88–94. doi: 10.1179/1754762814Y.0000000091

Uys, M., Pottas, L., Vinck, B., and Van Dijk, C. (2012). The influence of non-linear frequency compression on the perception of music by adults with a moderate to sever hearing loss: subjective impressions. *South Afr. J. Commun. Disord* 59, 53–67.

van der Burg, E., Alais, D., and Cass, J. (2013). Rapid recalibration to audiovisual asynchrony. *J. Neurosci.* 33, 14633–14637. doi: 10.1523/JNEUROSCI.1182-13.2013

Verrillo, R. T., Fraioli, A. J., and Smith, R. L. (1969). Sensation magnitude of vibrotactile stimuli. *Percept. Psychophys.* 6, 366–372. doi: 10.3758/BF03212793

Viemeister, N. F. (1979). Temporal modulation transfer functions based upon modulation thresholds. *J. Acoust. Soc. Am.* 66, 1364–1380. doi: 10.1121/1.383531

Weisenberger, J. M. (1986). Sensitivity to amplitude-modulated vibrotactile signals. *J. Acoust. Soc. Am.* 80, 1707–1715. doi: 10.1121/1.394283

Xu, L., Zhou, N., Chen, X., Li, Y., Schultz, H. M., Zhao, X., et al. (2009). Vocal singing by prelingually-deafened children with cochlear implants. *Hear. Res.* 255, 129–134. doi: 10.1016/j.heares.2009.06.011

Zeng, F. G., and Galvin, J. J. III (1999). Amplitude mapping and phoneme recognition in cochlear implant listeners. *Ear Hear.* 20, 60–74. doi: 10.1097/00003446-199902000-199902006

Zeng, F. G., Grant, G., Niparko, J., Galvin, J., Shannon, R., Opie, J., et al. (2002). Speech dynamic range and its effect on cochlear implant performance. *J. Acoust. Soc. Am.* 111, 377–386. doi: 10.1121/1.1423926

Zeng, F. G., Rebscher, S., Harrison, W., Sun, X., and Feng, H. (2008). Cochlear implants: system design, integration, and evaluation. *IEEE Rev. Biomed. Eng.* 1, 115–142. doi: 10.1109/RBME.2008.2008250

Zeng, F. G., Tang, Q., and Lu, T. (2014). Abnormal pitch perception produced by cochlear implant stimulation. *PLoS One* 9:e88662. doi: 10.1371/journal.pone.0088662

Zhang, X., and Ras, Z. W. (2007). "Analysis of sound features for music timbre recognition," in *Proceedings of the International Conference on Multimedia and Ubiquitous Engineering (MUE'07)*, (Seoul: IEEE).

Zhou, X., Seghouane, A. K., Shah, A., Innes-Brown, H., Cross, W., Litovsky, R., et al. (2018). Cortical speech processing in postlingually deaf adult cochlear implant users, as revealed by functional near-infrared spectroscopy. *Trends Hear.* 22:2331216518786850. doi: 10.1177/2331216518786850

Zwicker, E. (1952). Die grenzen der hörbarkeit der amplitudenmodulation und der Frequenz-modulation eines tones. *Acustica* 2, 125–133.

Conflict of Interest: The author declares 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.

ORIGINAL RESEARCH
published: 01 September 2021
doi: 10.3389/fdgth.2021.724539

Measuring Distortion-Product Otoacoustic Emission With a Single Loudspeaker in the Ear: Stimulus Design and Signal Processing Techniques

*Wei-Chen Hsiao, Yung-Ching Chen and Yi-Wen Liu**

Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan

The distortion-product otoacoustic emission (DPOAE) is a backward propagating wave generated inside the cochlea during the wave amplification process. The DPOAE signal can be detected rapidly under relatively noisy conditions. In recent years, the earphone industry demonstrated interest in adopting DPOAE as an add-on feature to make their product "intelligent" of inner-ear status. However, a technical challenge remains to be tackled—the loudspeaker in an earphone generates its own cubic distortion at the same frequency as DPOAE. Unfortunately, the intensity of loudspeaker distortion is typically comparable to that of the DPOAE, if not higher. In this research, we propose two strategies, namely *compensation* and *cancellation*, to enable DPOAE measurement with a single loudspeaker. The compensation strategy exploits the part of the growth function of the loudspeaker distortion which is almost linear, and thus suppresses the distortion it generates while retaining a larger portion of DPOAE in the residual signal. The cancellation strategy utilizes a one-dimensional Volterra filter to remove the cubic distortion from the loudspeaker. Testing on normal-hearing ears shows that the compensation strategy improved the DPOAE-to-interference ratio by approximately 7 dB, resulting in a cross-correlation of 0.62 between the residual DPOAE level and the true DPOAE level. Meanwhile, the cancellation strategy directly recovered both the magnitude and the phase of DPOAE, reducing the magnitude estimation error from 15.5 dB to 3.9 dB in the mean-square sense. These pilot results suggest that the cancellation strategy may be suitable for further testing with more subjects.

Keywords: hearing, otoacoustic emissions, intermodulation distortion, nonlinear signal processing, Volterra filtering

Edited by:
Qinglin Meng,
South China University of Technology, China

Reviewed by:
Ivan Miguel Pires,
Universidade da Beira Interior, Portugal
Shixiong Chen,
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences (CAS), China

****Correspondence:***
Yi-Wen Liu
ywliu@ee.nthu.edu.tw

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: *13 June 2021*
Accepted: *09 August 2021*
Published: *01 September 2021*

Citation:
Hsiao WC, Chen YC and Liu YW (2021) Measuring Distortion-Product Otoacoustic Emission With a Single Loudspeaker in the Ear: Stimulus Design and Signal Processing Techniques. Front. Digit. Health 3:724539. doi: 10.3389/fdgth.2021.724539

1. INTRODUCTION

Otoacoustic emissions (OAEs) are sounds generated in the cochlea that propagate backward to emit from the ear (1). OAEs can be classified into two types (2)—spontaneous OAEs (SOAEs) and evoked OAEs. SOAEs occur in the absense of external stimulus, and evoked OAEs can be regarded as acoustic responses to external stimulus. Within the family of evoked OAEs, the distortion-product OAE (DPOAE) is widely used as an objective tool for detecting hearing impairment associated with outer hair cell (OHC) dysfunctions (3, 4). To measure DPOAE, a pair of primary tones at frequencies $f_1 < f_2$ are delivered to an earphone inserted to the ear canal. With appropriately chosen intensities and frequencies of the primary tones [e.g., $f_2/f_1 = 1.22$, (5)], the most prominent distortion product would occur at $2f_1 - f_2$ and it can be recorded from a microphone in the ear canal. Because the primary tones' excitation patterns mainly overlap near the f_2 characteristic place in the cochlea (6), the sound-pressure level (SPL) of DPOAE at $f_{\mathrm{DP}} = 2f_1 - f_2$ represents the cochlea's ability to process signals normally at frequency f_2. Thus, DPOAE serves as a robust and non-invasive tool for assessing cochlear functions in a frequency-specific manner (4). It has been applied clinically for hearing screening (7, 8), and diagnosis of acute hearing loss (9) and other kinds of hearing impairment (10, 11).

Typically, a clinical DPOAE probe consists of two loudspeakers and one microphone; for each ear, the primary tones at f_1 and f_2 are separately delivered to the two speakers to avoid generating intermodulation distortion (IMD) electrically (12). As a rare exception, a single-speaker configuration was adopted for measuring vibration caused by DPOAE on insect tympanal organs (13); however, it was emphasized that one should avoid over-driving the speaker and thus producing IMD artifacts (14). In the field of cochlear neurophysiology, nonetheless, a combination of 5–7 tones with carefully arranged frequencies could be delivered simultaneously to a single speaker to elicit auditory-nerve responses (15); in their study, loudspeaker IMD was not a concern because the neural response by nature contains strong quadratic-distortion components which actually facilitate efficient estimation of the cochlear tuning curve at the auditory-nerve level.

Recently, the two-loudspeaker hardware design has been adopted by a commercial headphone that promotes at-home DPOAE measurement as a means of providing personalized frequency response adjustment (16). The two-speaker design seems necessary because, even with a high-quality headphone or earphone, the total harmonic distortion (THD) can reach 3% when driven to its full dynamic range (17). This THD level is acceptable for listening to music; however, when delivering two pure tones simultaneously, we found that the distortion generated by such speakers would significantly interfere with the DPOAE from the ear since the cubic distortion of the speaker also occurs at f_{DP}.

Nevertheless, human DPOAE and loudspeaker IMD have different generation mechanisms even though they may occur at the same frequency. For example, the DPOAE signal is comprised of a direct component plus a reflective component (18, 19); the direct component travels back from the f_2 characteristic place in the cochlea, while the reflective component travels further to the f_{DP} place and changes direction due to *coherent reflection* (20). The two components thus have different latency in the range of 5–20 ms, which allows them to be separated via envelope-tracking techniques (21). Also, they may superpose constructively or destructively depending on their relative phase. In comparison, the loudspeaker IMD is perhaps elicited nearly instantaneously, so we expect that its latency and rate of growth with respect to the primary tone levels L_1 and L_2 might differ from that of DPOAE. In this research, we seek to exploit these differences and develop methods for estimation of DPOAE levels using a *single speaker*, despite of interference from loudspeaker IMD.

In particular, we propose stimulus design and signal processing strategies that handle the interference issues due to loudspeaker IMD. The first strategy is called *compensation* and it involves finding a combination of L_1 and L_2 such that the IMD level grows *almost linearly* with respect to simultaneous increment in (L_1, L_2). The second strategy is called *cancellation* and it utilizes 3rd-order one-dimensional Volterra filter (22) to subtract the loudspeaker IMD from the signal. The organization of the remaining part of this paper follows the standard order of Methods, Results, Discussion, and Conclusion.

2. METHODS

In this section, we first review the mathematics of IMD generated by two tones. Typical spectrums of DPOAE and loudspeaker IMD will be shown so we can examine the similarities and differences. Then, the compensation strategy and the cancellation strategy will be described. This section ends with brief descriptions of the recording equipment and the human subjects who participated in the testing.

2.1. IMD and Mathematical Notations

Assuming that an acoustic or electrical stimulus, called the input signal, contains two frequency components f_1 and f_2, so the signal can be expressed as follows,

$$x(t) = A_1 \cos(2\pi f_1 t + \phi_1) + A_2 \cos(2\pi f_2 t + \phi_2), \tag{1}$$

where A_1, A_2 and ϕ_1, ϕ_2 denote the amplitude and phase for two components, respectively. Assume that the stimulus is delivered to a nonlinear system G so that the response $y(t)$ can be denoted as $y(t) = G\big(x(t)\big)$. When G is instantaneous, it can be expanded by Taylor's series near the origin; that is,

$$G(\eta) = G(0) + \sum_{k=1}^{\infty} \frac{G^{(k)}(0)}{k!}\eta^k. \tag{2}$$

By setting $\eta = x(t)$ and through simple trigonometry, one can show that

$$y(t) = \mathrm{Re}\Big[\sum_m \sum_n B_{m,n} e^{j2\pi(mf_1+nf_2)t}\Big], \tag{3}$$

where $B_{m,n}$ are complex-valued coefficients, and m and n sum over all integers such that $mf_1 + nf_2 > 0$. The components $B_{m,0}$ or $B_{0,n}$ are referred to as harmonics; the additional components $B_{m,n}$ when m and n are both nonzero are called the intermodulation products and they can be classified by their order $|m| + |n|$. In the context of DPOAE measurement with a single loudspeaker, $y(t)$ thus contains not only the primary frequencies f_1 and f_2 but also higher-order components at $mf_1 + nf_2$. Empirically, the 3rd-order intermodulation products $B_{2,-1}$ and $B_{-1,2}$ are most prominent from a loudspeaker (see **Figure 1A**), and a few

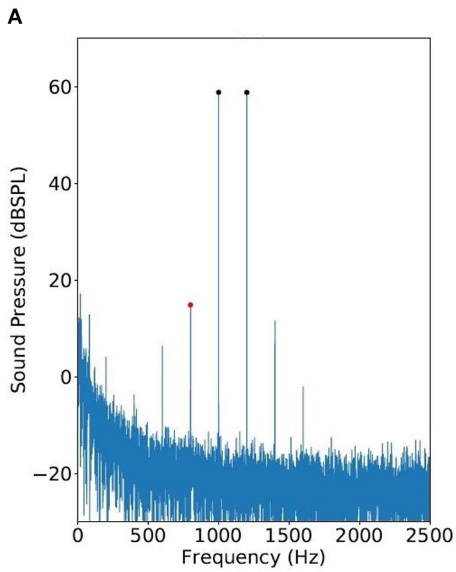


FIGURE 1 | Spectrums of modulation distortion produced by a single speaker **(A)** vs. true DPOAE obtained with a two-speaker probe **(B)**. The primary frequencies are $f_1 = 1,000$ Hz and $f_2 = 1,200$ Hz.

other components such as $B_{3,-2}$, $B_{-2,3}$ can also be identified above the noise floor. In comparison, from the DPOAE spectrum (measured by the conventional two-speaker approach), the $B_{-1,2}$ component corresponding to frequency $2f_2 - f_1$ usually cannot be detected (see **Figure 1B**). The reason is because, even though the intermodulation product at $2f_2 - f_1$ is generated due to OHC nonlinearity near the f_2 place, it is prohibited from backward propagation along the basilar membrane (23).

By inspecting **Figure 1**, note that the loudspeaker components occurring at $2f_1 - f_2$ has the same frequency as DPOAE. This 3rd-order component from the loudspeaker is referred to as IMD3 hereafter, and we shall investigate how to estimate DPOAE regardless of the presence of IMD3.

2.2. The Compensation Strategy

The DPOAE level, denoted as L_{DP}, depends systematically on parameters (L_1, L_2, f_1, f_2). The relation $L_{\mathrm{DP}} = L_{\mathrm{DP}}(L_1, L_2, f_1, f_2)$ was comprehensively measured from a cohort of 20 normal-hearing human subjects (24) with a purpose to recommend the optimal choice of $L_1 = L_1^{\mathrm{opt}}$ that maximizes L_{DP} given L_2. When L_1 increases beyond L_1^{opt}, L_{DP} starts to decrease due to two-tone suppression (25). The same phenomenon has also been reproduced *in silico* by simulation of cochlear mechanics (23). In this section, we report on how differently the IMD3 level depends on the parameters, and hence devise a way to suppress IMD3 by considering two sets of primary-tone level (L_1, L_2) jointly.

2.2.1. Growth Function of IMD3

In contrast to cochlear mechanics, the loudspeaker nonlinearity does not demonstrate two-tone suppression; for instance, **Figure 2** shows IMD3 level as a function of (L_1, L_2) with $f_1 = 1,000$ Hz and several different ratios f_2/f_1. The results were obtained by delivering the primary tones to one of the two loudspeakers of a DPOAE probe (see section 2.4 for details) and measuring the response inside a syringe of approximately 2.0 cc. For any fixed L_2, as L_1 increases, we do not find a clear L_1^{opt} beyond which L_{IMD3} starts to decrease, and this is quite unlike what was observed in human subjects with the most commonly used frequency ratio $f_2/f_1 \approx 1.2$ (24, **Figure 1**).

The contour plot of IMD3 also differs from that of human DPOAE in the rate of growth with respect to L_1 and L_2. In particular, when L_1 and L_2 increases proportionately as they move toward the top-right corner of the plot along the straight lines $L_1 = L_2 + 10$ or $L_1 = L_2$, the rate of growth of L_{IMD3} with respect to L_2 seems to be close to 1.0 dB/dB for a wide range of L_2 and across different f_2/f_1 ratio (see the "growth functions" in **Figure 3**). The slope of these growth functions are shown in **Figure 4**, and the path $L_1 = L_2$ happens to have the slope that is closest to 1.0 dB/dB across different primary-frequency ratios. In comparison, the average human DPOAE growth rate when $L_1 = L_2$ falls in the range of $0.3 - 0.5$ if $f_2/f_1 \approx 1.2$ (24, **Figure 1**). By exploiting this difference between the growth function of loudspeaker IMD3 and human DPOAE, we present a method that suppresses the IMD3 level while partially retaining the DPOAE.

2.2.2. Signal Acquisition Protocol

To leverage the part of IMD3 growth function that is almost linear (i.e., 1 dB/dB slope), we can devise the following performance metric,

$$J = \frac{\left|\beta G_{\mathrm{DP}}(P_1, P_2) - G_{\mathrm{DP}}(\beta P_1, \beta P_2)\right|}{\left|\beta G_{\mathrm{IMD3}}(P_1, P_2) - G_{\mathrm{IMD3}}(\beta P_1, \beta P_2)\right|}, \quad (4)$$

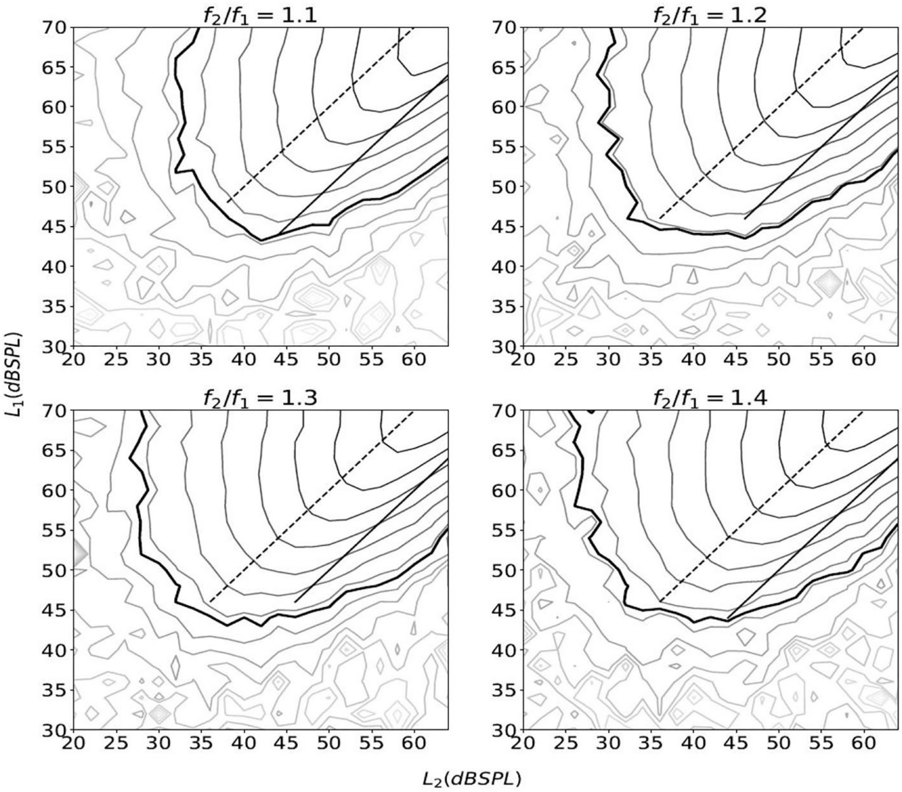


FIGURE 2 | The equal-level contour plots of IMD3 as a function of (L_1, L_2). The thick line marks L_{IMD3} =0 dB SPL, and other lines are in 4-dB steps. The IMD3 level were obtained by varying L_1 from 30 to 70 dB SPL and L_2 from 20 to 64 dB SPL in 2 dB steps. For each combination of (L_1, L_2), the stimulus lasted for 1.0 s with a 2.5-ms raised cosine ramp for the rising and falling edges. The stimulation was repeated five times and the average IMD3 level is shown. The solid line and the dashed line represent $L_1 = L_2$ and $L_1 = L_2 + 10$ (dB), respectively.

where (P_1, P_2) denotes primary-tone sound pressure in Pa, $\beta > 1$ denotes a scaling factor, and G_{IMD3} and G_{DP} denote the sound pressure of IMD3 and DPOAE in Pa, respectively. Conceptually, the goal of the compensation strategy is to choose (P_1, P_2) and β such that J is maximized. However, since the numerator (referred to as *DPOAE residual*) would vary among individuals, we seek to minimize the denominator in J, referred to as the *IMD3 residual*. Based on the results shown in **Figure 4**, we selected $L_1 = L_2$ (i.e., $P_1 = P_2$) for the remaining parts of this paper. As implied by **Figure 3**, choosing L_1 anywhere between 45 dB to 65 dB SPL should work well in reducing the IMD3 residual since $G_{\mathrm{IMD3}}(\beta P_1, \beta P_2) \approx \beta G_{\mathrm{IMD3}}(P_1, P_2)$. In contrast, we expect a larger proportion of DPOAE would remain in the DPOAE residual because the rate of growth against (P_1, P_2) is sub-linear (i.e., $<$ 1 dB/dB).

Based on the above-mentioned concept, we propose the following signal acquisition protocol.

- Step 1: Calibrate the stimulus levels A_1 and A_2 in Equation (1) such that $P_1 = P_2$ in the ear canal.
- Step 2: Transform the recorded signal to the frequency domain, and calculate the magnitude at $2f_1 - f_2$, which is the vector sum of IMD3 and DPOAE. Denote the result as $Y(P_1, P_2)$.
- Step 3: Repeat Step 2 with increased primary-tone levels βP_1 and βP_2. Denote the result as $Y(\beta P_1, \beta P_2)$.
- Step 4: Calculate the magnitude of residual at $2f_1 - f_2$, defined as $|\beta Y(P_1, P_2) - Y(\beta P_1, \beta P_2)|$.

To summarize, the goal of the signal acquisition protocol is to keep a large portion of DPOAE in the residual while maximally suppressing IMD3 at the same time.

2.3. The Cancellation Strategy

To describe the cancellation strategy, since digital adaptive filtering techniques are involved, we change the time variable from t to the integer index n (not to be confused with the index n in Equation 3). We follow a standard digital signal processing notation in defining $y[n] = y(nT)$ (26), where $y(t)$ is a continuous-time signal, T denotes the sampling period, and $y[n]$ denotes the result after sampling in time.

The idea behind this strategy is to cancel IMD3 instantaneously. To achieve this goal, we utilize two techniques, namely a phase-controlled exponential swept-sine chirp and the one-dimensional Volterra filters (ODVFs), to adjust the input signal before sending it to the speaker. The workflow is shown in **Figure 5** where $H(\omega)$ denotes a frequency

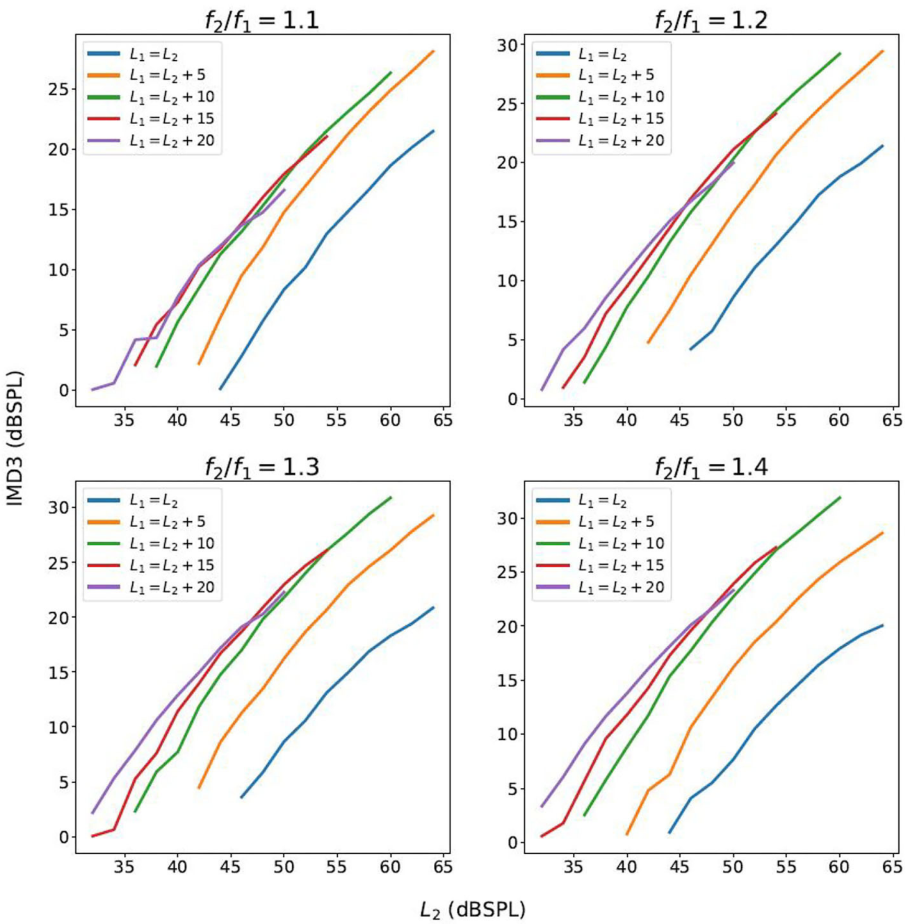


FIGURE 3 | The growth function of the IMD3 level with respect to L_2, as L_1 and L_2 increase proportionately.

response measured by the phase-controlled exponential swept-sine chirp (to be described in section 2.3.1), and the normalization factor ensures that the input to ODVF is limited to the range $[-1, 1]$. The input signal which contains two pure tones for DPOAE measurement is first transformed from time domain to frequency domain through DFT, then transformed back to time domain after multiplying with the frequency response of the single speaker $H(\omega)$. The filter coefficients of ODVF are meant to be obtained *offline* by adaptive LMS algorithm (to be described in section 2.3.2); when applied online, the ODVF filter coefficients are fixed. Subsequently, the multiplication by $H^{-1}(\omega)$ is to compensate the gain and phase change due to multiplication by $H(\omega)$. This workflow produces an adjusted input signal to be delivered to the single speaker for the purpose of measuring DPOAE.

The rationale of this design is to artificially generate intermodulation distortions by the ODVF filter but with an inverted polarity, so they can cancel the real intermodulation distortion generated by the loudspeaker. We shall see in section 2.3.2 that the ODVF is trained off-line to emulate the situation of using two separate loudspeakers.

2.3.1. Linear System Estimation by Phase Controlled Exponential Swept-Sine Chirp

A phase-controlled exponential swept-sine chirp (27) is used to obtain the linear coupling response from the loudspeaker to the microphone. This chirp exhibits an instantaneous frequency that increases exponentially with time as follows,

$$s[n] = A \sin\left(\frac{\pi L}{2^Q \ln(2^Q)}\left(2^{\frac{Q}{N}}\right)^n\right), \tag{5}$$

where A is the amplitude of the sine wave, Q is an integer number of octaves, L is the ideal chirp length, and N is the real chirp length which is L rounded to the nearest integer. We also need an inverse chirp to convolve with, which can be expressed as

$$s^{-1}[n] = \frac{Q \ln 2}{A^2(1 - 2^{-Q})} \cdot s[N-n]\left(2^{\frac{Q}{N}}\right)^{-n}. \tag{6}$$

The result of convolving $s[n]$ and $s^{-1}[n]$ is approximately a Dirac delta impulse, and the gain is 0 dB for all frequencies.

We set the chirp's frequency to glide from 47 to 24,000 Hz and the chirp length was 10.5 s. Then, we delivered the chirp to

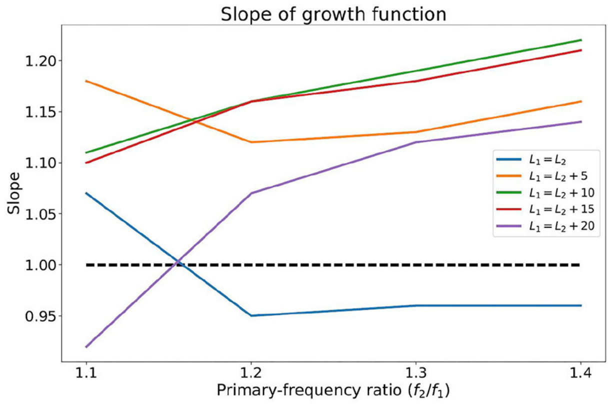


FIGURE 4 | The slope of the growth functions of the IMD3 shown in **Figure 3** with various primary-frequency ratios and primary-level differences.

drive the single speaker and recorded the sound simultaneously. Subsequently, the recorded signal was convolved with the inverse chirp (Equation 6) to obtain the impulse response $h[n]$ that characterizes the linear coupling from the speaker to the microphone. The Fourier transform of $h[n]$ is denoted as $H(\omega)$ in **Figure 5**.

2.3.2. IMD3 Cancellation by One-Dimensional Volterra Filtering

Since the intermodulation is due to loudspeaker nonlinearity, the behavior of its inverse system can ideally be characterized by Volterra series expansion (28). However, the full-scale Volterra series expansion requires estimate of multi-variate kernel functions which may be computationally impractical to implement and its estimation might be slow in convergence. In this research, we adopted a simplified version called one-dimensional Volterra filters (ODVF) (22) — assume that the inverse system can be modeled as follows,

$$y[n] = \sum_{i=0}^{M_1-1} h_1[i]x[n-i] + \sum_{i=0}^{M_2-1} h_2[i](x[n-i])^2 + \cdots + \sum_{i=0}^{M_p-1} h_p[i](x[n-i])^p + \cdots \quad (7)$$

where $x[n]$ and $y[n]$ denote the input and output of the inverse system, respectively, $h_p[i]$ denotes the pth-order kernel of ODVF, and M_p is the length of the pth-order kernel. Since this research focuses on canceling a cubic distortion, a partial ODVF retaining only the 1st-order and 3rd-order kernels was used; that is,

$$y[n] \approx \sum_{i=0}^{M_1-1} h_1[i]x[n-i] + \sum_{i=0}^{M_3-1} h_3[i](x[n-i])^3. \quad (8)$$

The filter coefficients could be obtained by the adaptive least mean square (LMS) method (29). The LMS method involves updating the filter coefficients constantly as the time index n proceeds. First, denote the filter coefficients in the following vector form,

$$\mathbf{h}_{1,n} = (h_1[0], h_1[1], \ldots, h_1[M_1-1])^T \in \mathbb{R}^{M_1},$$

and similarly,

$$\mathbf{h}_{3,n} = (h_3[0], h_3[1], \ldots, h_3[M_3-1])^T \in \mathbb{R}^{M_3}.$$

Note that the subscript n indicates that both vectors are updated as n increases. Then, an approximation error signal can be defined as follows,

$$e[n] = d[n] - y[n] \approx d[n] - \mathbf{h}_{1,n}^T\mathbf{x}[n] - \mathbf{h}_{3,n}^T\mathbf{x}_3[n], \quad (9)$$

where $d[n]$ is a desired signal to be defined shortly, and vectors $\mathbf{x}[n]$ and $\mathbf{x}_3[n]$ are defined as follows,

$$\mathbf{x}[n] = (x[n], x[n-1], \ldots, x[n-M_1+1])^T \in \mathbb{R}^{M_1}$$

and

$$\mathbf{x}_3[n] = \left((x[n])^3, (x[n-1])^3, \ldots, (x[n-M_3+1])^3\right)^T \in \mathbb{R}^{M_3}.$$

Finally, the update equations are given as follows,

$$\mathbf{h}_{1,n+1} = \mathbf{h}_{1,n} + \alpha e[n]\mathbf{x}[n], \quad (10)$$

and similarly,

$$\mathbf{h}_{3,n+1} = \mathbf{h}_{3,n} + \alpha e[n]\mathbf{x}_3[n], \quad (11)$$

where $\alpha = 6 \times 10^{-3}$ denotes a stepsize that was chosen empirically.

Figure 6 shows how the ODVF coefficients were obtained in this research. The stimulus $s[n]$ contained two primary tones, and we set $d[n]$, the desired signal, to be the signal recorded by sending the two tones to separate speakers of a reliable reference probe, so $d[n]$ was free of IMD. Meanwhile, the input $x[n]$ to the ODVF was the signal recorded by using one single speaker while $y[n]$ denotes the output of the ODVF. Here, we emphasize that the adaptive procedure was performed offline just for one time, and it was not necessary to repeat the procedure when measuring DPOAE from individual ears. In practice, we first recorded $x[n]$ and $d[n]$ separately in the same 2-cc syringe with the same stimulus $s[n]$. Then, the filter coefficients $\mathbf{h}_{1,0}$ were initialized at 1 and $\mathbf{h}_{3,0}$ were initialized at 0, and we computed the updates iteratively according to Equations (9–11). After the filter coefficients converged, we could expect that the variance of $e[n]$ would be minimized and $y[n]$ should approximate $d[n]$, which is a cubic distortion-free signal, in a stochastic sense.

In practice, we found that the 1st-order kernel $h_1[n]$ tended to converge to a band-pass filter around the primary-tone frequencies which simply rejected all the intermodulation components linearly. This caused the "training" of $\mathbf{h}_{3,n}$ to fail. Therefore, we set the 1st-order kernel length M_1 to 1 to ensure that $h_3[n]$ learns to cancel the cubic distortion.

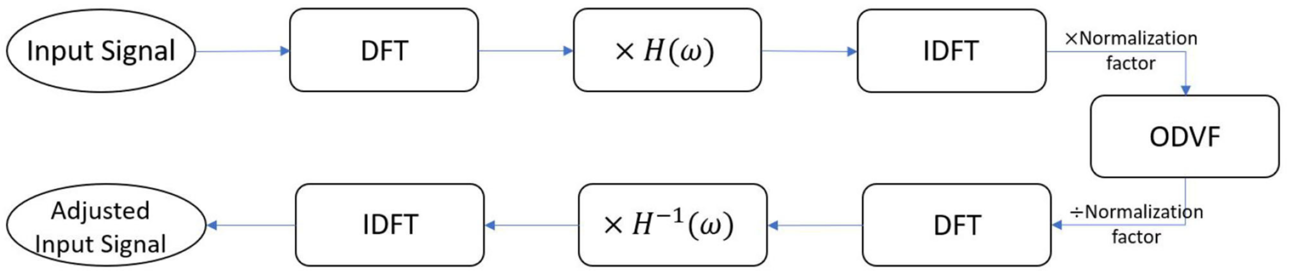


FIGURE 5 | Block diagram of the IMD3 cancellation method. DFT, discrete Fourier transform; IDFT, inverse DFT; ODVF, one-dimensional Volterra filter. $H(\omega)$ represents the linear coupling from the speaker to the microphone, which was measured offline by a sweep-sine chirp method described in section 2.3.1.

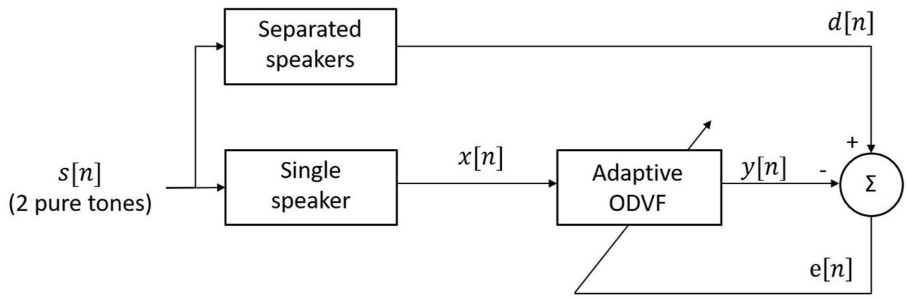


FIGURE 6 | Block diagram for computing the ODVF coefficients.

2.4. Equipment

All recordings were collected using a Python script that controls the ER-10C DPOAE probe-microphone (Etymotic Research) system via a 24-bit soundcard (Fireface UFX II, RME). The sampling frequency was set to 48 kHz. All recordings were done in a sound-proof room with the noise floor of approximately 19-21 dB SPL (30).

2.5. Human Subjects

Twelve subjects between age 22 and 32 participated in the research, including 8 males and 4 females. The compensation strategy was tested on the data from 5 of the subjects, while the cancellation strategy was tested on the data from 7 subjects. All the subjects did not have ear infection or report any hearing problems at the time of experiment. The recruitment of human subjects was approved by the IRB of National Tsing Hua University (No. 10912HE101).

3. RESULTS

Here we report the efficacy of applying the compensation and cancellation strategies.

3.1. The Compensation Strategy

We first tested the signal acquisition protocol in a 2-cc syringe. The parameters being tested were $\beta = 1.5, 2.0$ or 3.0, and $P_1 = P_2 = 6.3, 11.3$, and 20.0 mPa, which correspond to 50, 55, and 60 dB SPL, respectively. The primary frequencies were $f_1 = 1,000$ and $f_2 = 1,200$ Hz. We quantify the performance of IMD3 suppression by the following index,

TABLE 1 | The suppression index K for different combination of parameters.

	$L_1 = L_2$ (dB SPL)		
	50	**55**	**60**
$\beta = 1.5$	21.7	21.9	22.2
$\beta = 2.0$	16.2	15.9	16.5
$\beta = 3.0$	12.0	13.1	12.4

$$K = 20\log_{10}\left|\frac{G_{\mathrm{IMD3}}(P_1, P_2)}{\beta G_{\mathrm{IMD3}}(P_1, P_2) - G_{\mathrm{IMD3}}(\beta P_1, \beta P_2)}\right|. \quad (12)$$

Here, K is just the IMD3 to IMD3 residual ratio in dB scale.

The values K for different combinations of β and L_1 are listed in **Table 1**. Note that K is quite insensitive to change in L_1, while $\beta = 1.5$ gives the highest K. Therefore, $\beta = 1.5$ was chosen for testing in the ear. Also, $L_1 = L_2 = 60$ dB SPL was selected in order to maximize DPOAE and its residual.

DPOAEs residuals were recorded from five subjects for their left and right ears using the compensation strategy. Two protocols were considered. The first one is called the "stereo" protocol which uses separate speakers to obtain the ground truth of DPOAE and the residual thereof after applying the compensation strategy. The second is called "mono" protocol and it uses a single speaker mentioned in section 2.2.2 to obtain

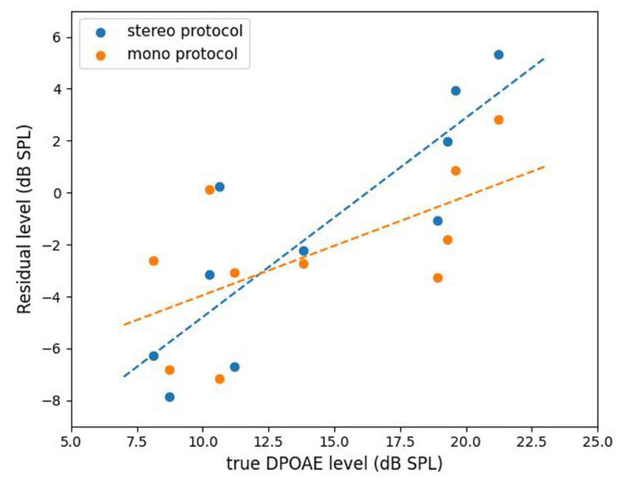


FIGURE 7 | The residual level vs. the true DPOAE level with stereo and mono protocols, respectively. The dashed lines are least-square fits to the data.

the DPOAE residual subject to the IMD3 interference. The residual obtained with the stereo protocol should be regarded as a performance upper-bound for the compensation strategy since the IMD3 component is negligible when both speakers are used. We then study the correlation between the residual level obtained by both protocols and the true DPOAE level to evaluate the effectiveness of the compensation strategy.

The results of mono vs. stereo protocols are plotted in **Figure 7**. The x-axis is the true DPOAE level obtained with two speakers, and the y-axis represents the DPOAE residual levels obtained with the two protocols. The dashed lines show the results of linear regression. For the stereo protocol, the regression line is $y = 0.77x - 12.46$, and for the mono protocol, the regression line is $y = 0.38x - 7.76$. By using the stereo protocol, the correlation between DPOAE residual and true DPOAE level, both in dB SPL, equals to 0.86 with a high significance ($p < 0.001$). However, the DPOAE residual levels were 15 dB lower than the real DPOAE levels in average. This indicates that the DPOAE was suppressed by about 80% after applying the compensation strategy. Note that, nevertheless, the IMD3 component was suppressed by 22 dB under the same settings (see **Table 1**). Thus, we can say that the compensation strategy improved the DPOAE to IMD3 ratio by 7 dB in average. By using the mono protocol, the correlation between the residual and the true DPOAE level is lowered to 0.62. Nevertheless, with a $p < 0.01$, the correlation is deemed significant for this particular set of data, in the sense that the null hypothesis (no correlation) is rejected.

3.2. The Cancellation Strategy

We obtained different sets of ODVF coefficients at $f_1 = 1{,}000$, 1,200, 1,500, and 2,000 Hz, respectively, while $f_2/f_1 = 1.2$ was fixed. The amplitude of both tones was set to 20 mPa (60 dBSPL), and the recordings ran for 10.5 s with 2.5 ms raised cosine ramp for the rising and falling edges of the stimulation. Empirically, the recording time was sufficiently long to ensure convergence of the filter coefficients. **Figure 8** shows the resulting 3rd-order filter coefficients $h_3[n]$ for $f_1 = 1{,}000$ and 2,000 Hz, respectively. These $h_3[n]$ coefficients were subsequently used for validating the proposed cancellation strategy depicted in **Figure 5**.

Following the workflow described in **Figure 5**, we first tested the adjusted input signal in the 2-cc syringe. The results are shown in **Figure 9**. **Figure 9A** is the signal recorded by using a single speaker to play two primary tones. Therefore, the spectrum contains intermodulation distortion. **Figure 9B** is the signal recorded by using separated speakers and it therefore does not contain intermodulation distortion, and **Figure 9C** is the signal by using the adjusted input signal produced by the workflow of **Figure 5**. The black dots represent the magnitude of primary tones f_1 and f_2, the red dot represents the magnitude of the IMD3. All the three recordings ran for 2.5 s with 2.5 ms raised cosine ramp for the rising as well as falling edges of the stimulation. The result shows that IMD3 is largely reduced to submerge below the noise floor. Note that the $2f_2 - f_1$ component at 1400 Hz is also suppressed, though not as perfectly as at $2f_1 - f_2$. The fifth-order distortions at 600 and 1600 Hz remain unchanged.

Then we applied the same recording procedure in human ears. **Figures 10**, **11** show some typical results. **Figure 10A** contains DPOAE interfered with the original IMD3, **Figure 10B** shows the ideal signal recorded by using separate speakers for the primary tones, and the results are regarded as the true DPOAE signal to compare against. **Figure 10C** is the signal recorded by using the adjusted input signal described in **Figure 5**; the signal contains DPOAE interfered with the remaining IMD3.

We also extended the experiments to $f_1 = 1,200, 1,500$, and $2,000$ Hz while $f_2/f_1 = 1.2$. The results without (**Figures 12A,C**) and with (**Figures 12B,D**) the IMD3 cancellation strategy are shown in **Figure 12**. In **Figures 12A,B**, the horizontal axis is the true DPOAE level, and the vertical axis is the magnitude at f_{DP}. In **Figures 12C,D**, the horizontal axis is the true DPOAE phase, and the vertical axis is the phase measured at f_{DP}. The linear regression lines are also shown for visualization.

Without applying IMD3 cancellation strategy, the regression lines of the magnitude and the phase are $y = 0.12x + 19.87$ and $y = 1.03x - 0.25$, respectively, and the root mean square errors (RMSE) are 15.52 dB and 1.50 rad, respectively. After applying IMD3 cancellation strategy, the regression lines of the magnitude and the phase become $y = 0.76x + 2.64$ and $y = 1.02x - 0.06$, respectively, and the RMSE are reduced to 3.88 dB and 0.76 rad, respectively.

After applying the IMD3 cancellation strategy, we calculated the correlation between the estimated and the true DPOAE sound pressure levels, and the correlation between the estimated and the true DPOAE phase, respectively. Subsequently, we ran the Wald test with t-distribution to evaluate the significance of correlation. The correlation between the estimated and true DPOAE levels equals to 0.81 with a high significance ($p < 0.001$); also, the correlation between the estimated and the true DPOAE phase equals to 0.93 with a high significance ($p < 0.001$).

In contrast, without applying the IMD3 cancellation strategy, the correlation between the measured DPOAE and the true DPOAE sound pressure levels equals to 0.25 with a low

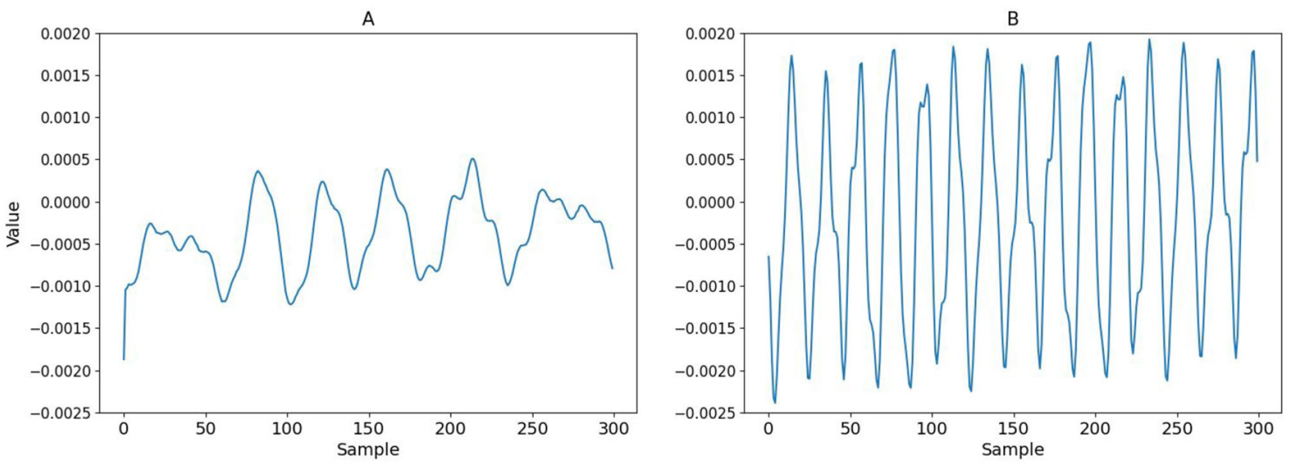


FIGURE 8 | The 3rd-order filter coefficients $h_3[n]$ under **(A)** $f_1 = 1,000$ Hz and **(B)** $f_1 = 2,000$ Hz. The length of the filter was set to 300, which corresponds to 6.25 ms at 48 kHz sampling rate.

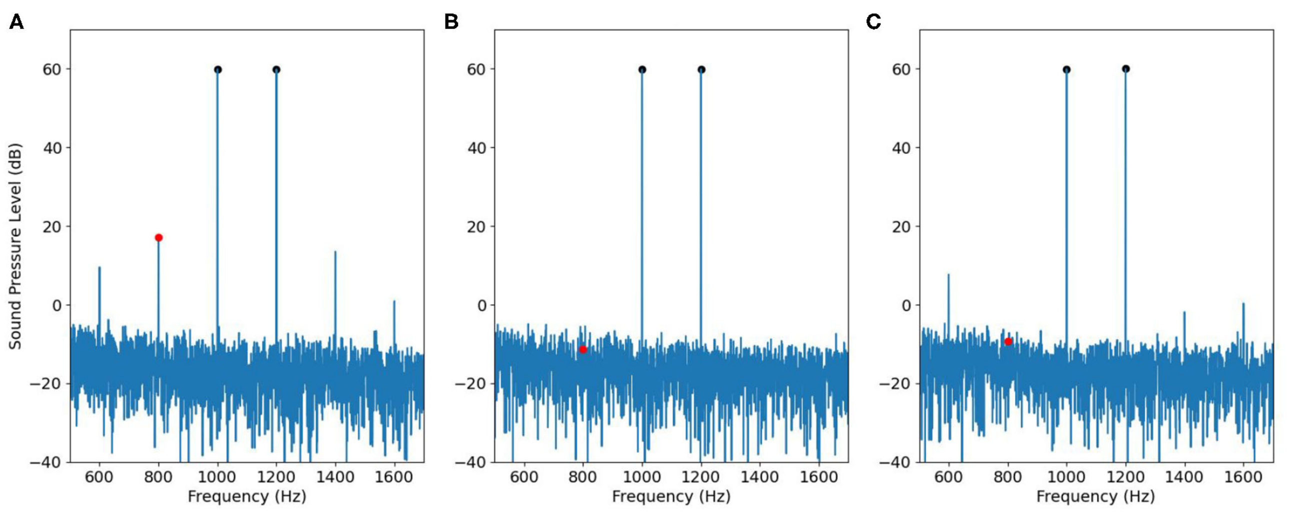


FIGURE 9 | Effectiveness of ODVF in canceling IMD3 in a syringe, visualized in the frequency domain. **(A)** The signal recorded by using a single speaker. **(B)** The signal recorded by using two speakers. **(C)** The signal recorded by using the adjusted input signal in **Figure 5**.

significance ($p = 0.057$). Although the phase estimation error is higher without applying the cancellation strategy, the cross-correlation between the measured and the true DPOAE phase was still high (0.80) and significant ($p < 0.001$).

The number of data points turns out to be 55. We recruited 7 subjects, test both ears at 4 frequencies, resulting in 56 DPOAE magnitudes and phases. However, the DPOAE level of one of the ears was below the noise floor when $f_1 = 1,000$ Hz. So that single point of data was abandoned.

4. DISCUSSION

As true wireless, noise-cancellation earphones are gaining popularity in recent years, the ear canal also becomes an over-booked space for various body sensors to enter and make the earphone intelligent (31, 32). Since active noise-cancellation earphones are indeed equipped an internal microphone[1], there is no reason why the microphone cannot measure DPOAE. The main factor that hinders such application might be the interference due to loudspeaker IMD. As much as we are aware of, consumer earphones are not typically designed to have two speakers in one ear[2], so sending the primary tones to separate speakers would not be a choice. In this research, we demonstrated

[1]For example *AirPods Pro and AirPods Max Active Noise Cancellation and Transparency mode*, https://support.apple.com/en-us/HT210643.

[2]For example “Qualcomm QCC3056 is an ultra-low power, single-chip solution, optimized for use in truly wireless earbuds and hearables.” The chip features mono audio playback. https://www.qualcomm.com/products/qcc3056.

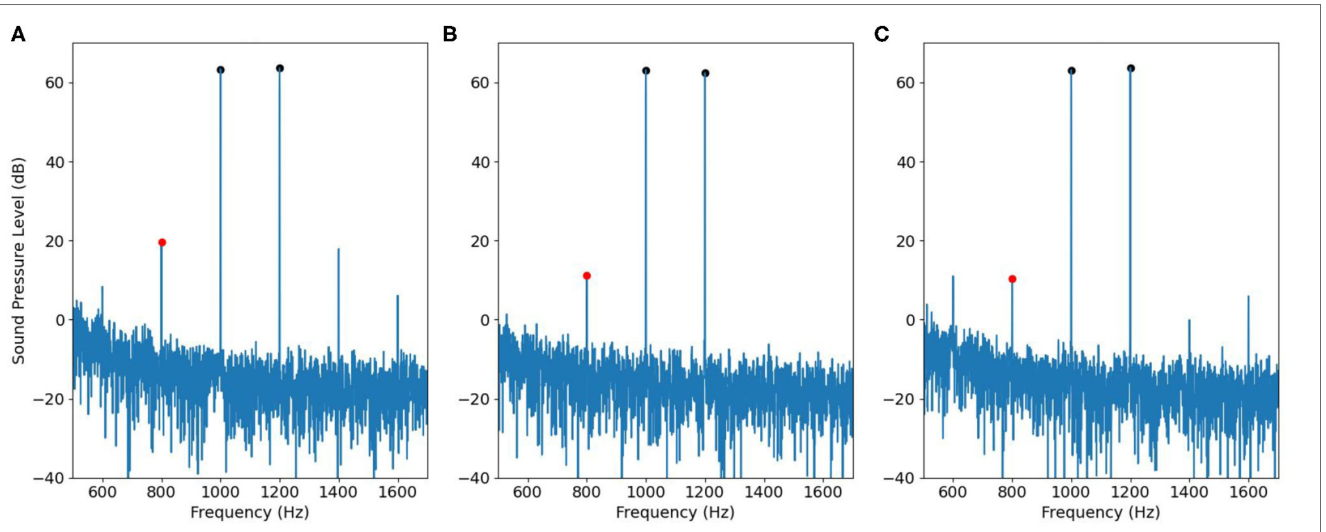


FIGURE 10 | An example of DPOAE estimation by the cancellation strategy when f_1 is 1,000 Hz, viewed in the frequency domain. **(A)** The signal recorded by using a single speaker. **(B)** The desired signal recorded by using two speakers. **(C)** The signal recorded by using adjusted input signal in **Figure 5**.

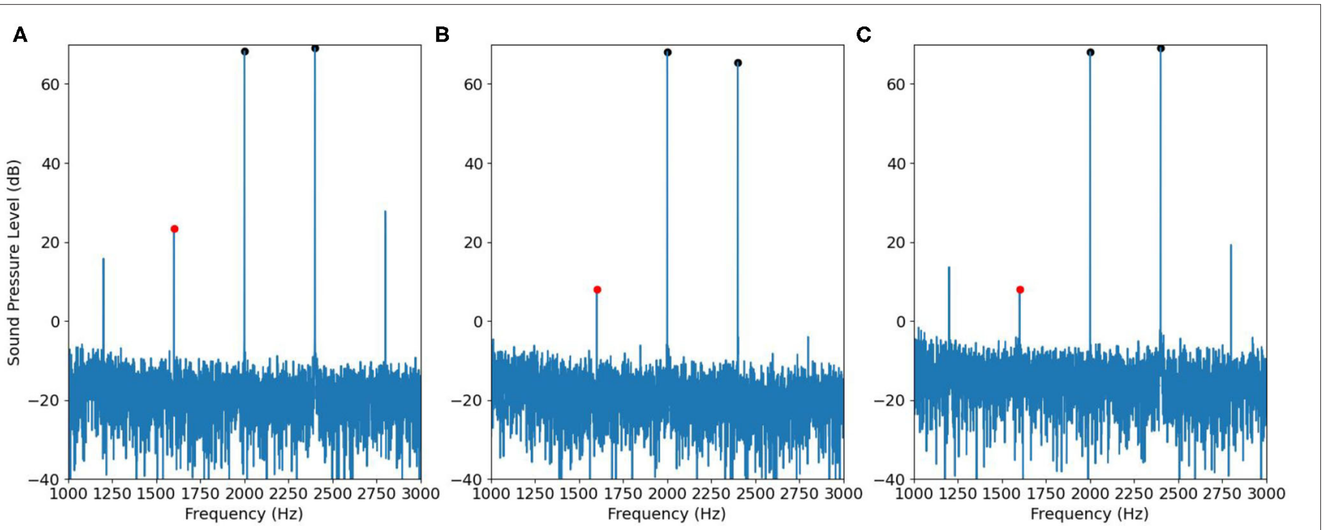


FIGURE 11 | Similar to **Figure 10**, but with f_1 = 2,000 Hz.

that "wrongly" using one single loudspeaker to play the f_1 and f_2 tones may still work as long as we can cancel the IMD3 it generates. Hence, we hope that this research can serve as a feasibility study for the earphone industry to promote DPOAE as a service to consumers of active noise cancellation earphones. We envision that making DPOAE available at home could also enrich any remote hearing care program in the future.

Among the two proposed strategies, cancellation outperforms compensation in producing a more accurate prediction of the true DPOAE level. On one hand, the cancellation strategy achieves a higher cross-correlation to the true DPOAE level (0.81) than the compensation strategy (0.62); one the other hand, it also provides a direct estimate of the DPOAE magnitude and phase, instead of just a residual. It remains to be seen in the future whether similar results would be obtained with a larger sample size.

The usage of Volterra filters also brings up many research questions. For instance, **Figure 8** shows that the 3rd-order function $h_3[n]$ are different for different choices of (f_1, f_2). Presently, we are uncertain whether (a) this is a limitation due to omitting all the off-diagonal elements of the Volterra filter to make it one-dimensional, or (b) is it possible to apply certain transformation akin to pre-whitening so the adaptive system eventually "learns" a universal ODVF for canceling all the cubic distortions given any input signal. Apparently, there is still much room to explore on this topic.

Meanwhile, as much as the compensation strategy is concerned, it is surprising that we found a large region on the

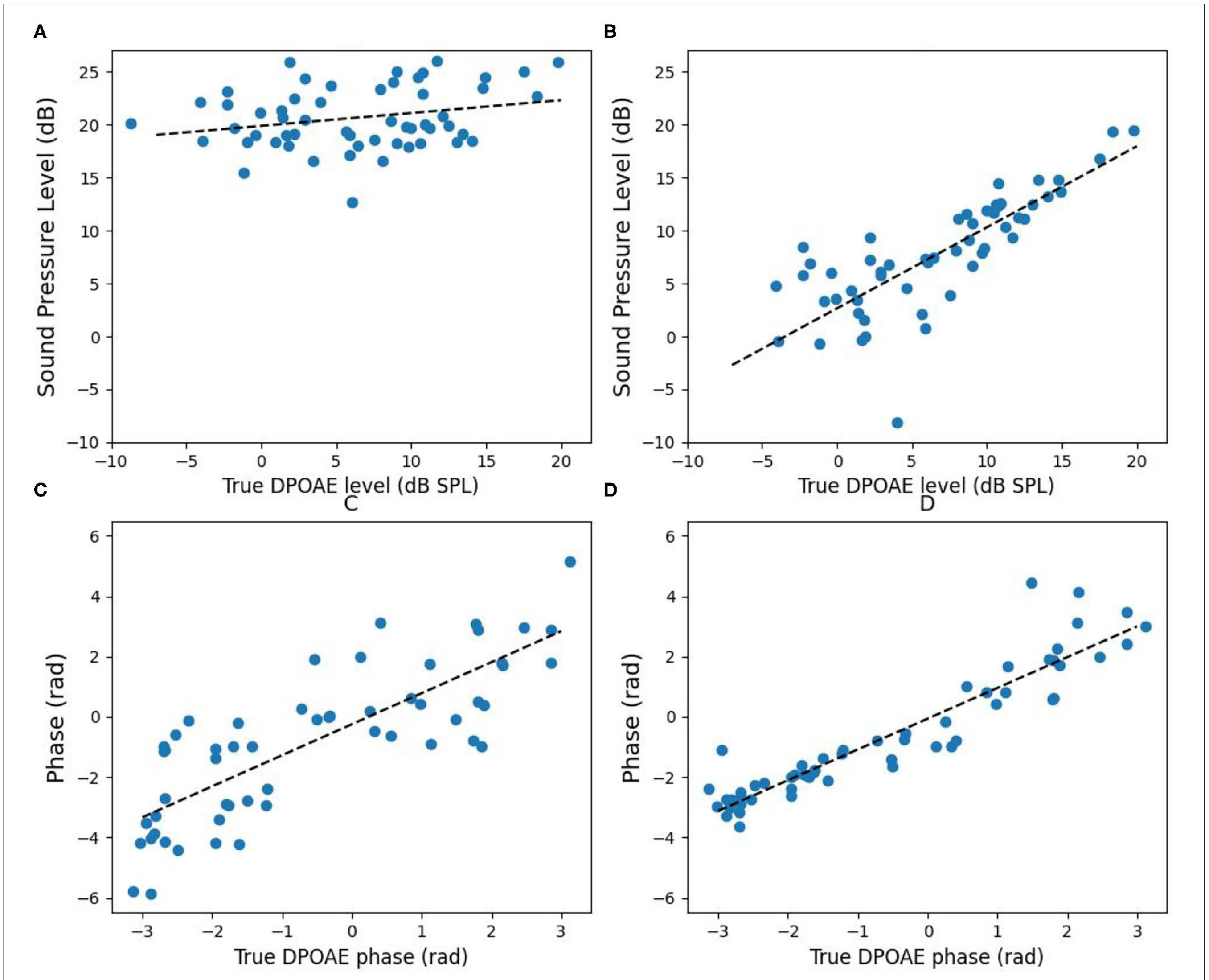


FIGURE 12 | Comparison of DPOAE data without and with the IMD3 cancellation strategy. **(A)** The sound pressure level without applying the IMD3 cancellation strategy; **(B)** the sound pressure level after applying the IMD3 cancellation strategy; **(C)** the phase without, and **(D)** the phase with the IMD3 cancellation strategy. In each plot, the dashed line shows the result of linear regression.

(L_1, L_2) plane where the IMD3 level grows *quasi-linearly* when L_1 and L_2 increase proportionately. In practice, it may be interesting to see if a similar property can be observed in other DPOAE probes or consumer earphones. We speculate that the quasi-linear growth is an epiphenomenon because cubic distortion is by nature a third-order component. Under the light of Taylor expansion in Equation (2), we can expect IMD3 to demonstrate a 3 dB/dB growth when the stimulus is at low intensity—and so does DPOAE (23). As the intensity of η increases, higher order terms $G^{(k)}(0)\eta^k/k!$ begin affecting the growth function. In particular, all the odd-order terms should jointly reduce the slope of growth of $B_{2,-1}$ and account for its saturation (it is straightforward to show that the even-order terms do not contribute to $B_{2,-1}$). So the fact that we observe nearly 1 dB/dB growth at the working range of (L_1, L_2) may just be a coincidence.

Some other techniques might be worth consideration for estimating the DPOAE level under IMD3 interference. As mentioned in section 1, DPOAE itself has two components — the direct one and the reflection. Based on the difference in latency, Vetešník et al. (21) designed short pips to elicit DPOAE and applied envelope tracking techniques to *separate* the two components. If the loudspeaker distortion is generated within a shorter time before DPOAE emerges, one might be able to identify an early peak by tracking the instantaneous amplitude at the $2f_1 - f_2$ frequency. Thus, the loudspeaker IMD3 and DPOAE might be separated in the time domain. This may require careful re-thinking of the stimulus design and is left for future exploration.

5. CONCLUSIONS

We proposed two strategies to estimate the DPOAE level subject to interference from the loudspeaker IMD3. The compensation strategy was designed to suppress IMD3 based on its quasi-linear growth with respect to primary-tone levels, in contrast to DPOAE's sub-linear growth. Results show that, although the DPOAE level was also suppressed by 80%, the residual level correlates to the true DPOAE level. The cancellation strategy adjusted the input signal nonlinearly to emulate distortion-free stimulation. It thus recovered both the DPOAE magnitude and phase directly. Overall, this research suggests that it might be feasible to use a single-loudspeaker probe to measure DPOAE. Testing with a larger sample of human subjects as well as various types of earphones shall ensue to evaluate whether commercial noise-cancellation earphones could be utilized to allow sufficiently accurate DPOAE measurement at home.

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 Research Ethic Committee, National Tsing Hua University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YCC discovered the almost linear growth in IMD3, and came up with the compensation strategy. WCH tested the compensation strategy, and conceived the cancellation strategy using ODVF and then implemented and tested the cancellation strategy. YWL supervised the research and is responsible for the manuscript preparation process. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the Ministry of Science and Technology of Taiwan (grant no. 107-2221-E-007-093-MY2).

ACKNOWLEDGMENTS

We thank Drs. Yu-Ting Lin and Tang-Chuan Wang for regular exchange of research ideas and offering critiques.

REFERENCES

1. Kemp DT. Stimulated acoustic emissions from within the human auditory system. *J Acoust Soc Am*. (1978) 64:1386–91. doi: 10.1121/1.382104
2. Shera CA, Guinan JJ Jr. Evoked otoacoustic emissions arise by two fundamentally different mechanisms: a taxonomy for mammalian OAEs. *J Acoust Soc Am*. (1999). 105:782–98. doi: 10.1121/1.426948
3. Zurek PM, Clark WW, Kim DO. The behavior of acoustic distortion products in the ear canals of chinchillas with normal or damaged ears. *J Acoust Soc Am*. (1982) 72:774–80. doi: 10.1121/1.388258
4. Lonsbury-Martin B, Martin G. The clinical utility of distortion-product otoacoustic emissions. *Ear Hear*. (1990) 11:144–54. doi: 10.1097/00003446-199004000-00009
5. Harris F, Lonsbury-Martin B, Stagner B, Coats A, Martin G. Acoustic distortion products in humans: systematic changes in amplitude as a function of f2/f1 ratio. *J Acoust Soc Am*. (1989) 85:220–9. doi: 10.1121/1.397728
6. Kanis LJ, de Boer E. Two-tone suppression in a locally active nonlinear model of the cochlea. *J Acoust Soc Am*. (1994) 96:2156–65. doi: 10.1121/1.410157
7. Barker SE, Lesperance MM, Kileny PR. Outcome of newborn hearing screening by ABR compared with four different DPOAE pass criteria. *Amer J Audiol*. (2000) 9:142–8. doi: 10.1044/1059-0889(2000/017)
8. Sanford CA, Keefe DH, Liu YW, Fitzpatrick D, McCreery RW, Lewis DE, et al. Sound-conduction effects on DPOAE screening outcomes in newborn infants: test performance of wideband acoustic transfer functions and 1-kHz tympanometry. *Ear Hear*. (2009) 30:635–52. doi: 10.1097/AUD.0b013e3181b61cdc
9. Brown DJ, Gibson WP. On the differential diagnosis of Ménière's disease using low-frequency acoustic biasing of the 2f1-f2 DPOAE. *Hear Res*. (2011) 282:119–27. doi: 10.1016/j.heares.2011.09.002
10. Shiomi Y, Tsuji J, Naito Y, Fujiki N, Yamamoto N. Characteristics of DPOAE audiogram in tinnitus patients. *Hear Res*. (1997) 108:83–8. doi: 10.1016/S0378-5955(97)00043-9
11. Job A, Nottet JB. DPOAEs in young normal-hearing subjects with histories of otitis media: evidence of sub-clinical impairments. *Hear Res*. (2002) 167:28–32. doi: 10.1016/S0378-5955(02)00330-1
12. Whitehead ML, Stagner BB, Lonsbury-Martin BL, Martin GK. Measurement of otoacoustic emissions for hearing assessment. *IEEE Eng Med Biol Mag*. (1994) 13:210–26. doi: 10.1109/51.281681
13. Moir HM, Jackson JC, Windmill JFC. No evidence for DPOAEs in the mechanical motion of the locust tympanum. *J Exp Biol*. (2011) 214:3165–72. doi: 10.1242/jeb.056465
14. Moir HM, Jackson JC, Windmill JFC. Response to "Measurement of sensitive distortion-product otoacoustic emissions in insect tympanal organs." *J Exp Biol*. (2012) 215:567. doi: 10.1242/jeb.068098
15. van der Heijden M, Joris PX. Cochlear phase and amplitude retrieved from the auditory nerve at arbitrary frequencies. *J Neurosci*. (2003) 23:9194–8. doi: 10.1523/JNEUROSCI.23-27-09194.2003
16. Campbell LJ, Slater KD. *Personalization of Auditory Stimulus*. United States: 9497530B1 (2016).
17. der Aerschot MV, Swanepoel DW, Mahomed-Asmail F, Myburgh HC, Eikelboom RH. Affordable headphones for accessible screening audiometry: an evaluation of the sennheiser HD202 II supra-aural headphone. *Int J Audiol*. (2016) 55:616–22. doi: 10.1080/14992027.2016.1214756
18. Kalluri R, Shera CA. Distortion-product source unmixing: a test of the two-mechanism model for DPOAE generation. *J Acoust Soc Am*. (2001) 109:622–37. doi: 10.1121/1.1334597
19. Shaffer LA, Withnell RH, Dhar S, Lilly DJ, Goodman SS, Harmon KM. Sources and mechanisms of DPOAE generation: implications for the prediction of auditory sensitivity. *Ear Hear*. (2003) 24:367–79. doi: 10.1097/01.AUD.0000090439.16438.9F
20. Zweig G, Shera CA. The origin of periodicity in the spectrum of evoked otoacoustic emissions. *J Acoust Soc Am*. (1995) 98:2018–47. doi: 10.1121/1.413320

21. Vetešník A, Turcanu D, Dalhoff E, Gummer AW. Extraction of sources of distortion product otoacoustic emissions by onset-decomposition. *Hear Res*. (2009) 256:21–38. doi: 10.1016/j.heares.2009.06.002
22. Mu Y, Ji P, Ji W, Wu M, Yang J. Modeling and compensation for the distortion of parametric loudspeakers using a one-dimension volterra filter. *IEEE/ACM Trans Audio Speech Lang Proc*. (2014) 22:2169–81. doi: 10.1109/TASLP.2014.2363414
23. Liu YW, Neely ST. distortion-product emissions from a cochlear model with nonlinear mechanoelectrical transduction in outer hair cells. *J Acoust Soc Am*. (2010) 127:2420–32. doi: 10.1121/1.3337233
24. Johnson TA, Neely ST, Garner CA, Gorga MP. Influence of primary-level and primary-frequency ratios on human distortion product otoacoustic emissions. *J Acoust Soc Am*. (2006) 119:418–28. doi: 10.1121/1.2133714
25. Hall JL. Two-tone distortion products in a nonlinear model of the basilar membrane. *J Acoust Soc Am*. (1974) 56:1818–28. doi: 10.1121/1.1903519
26. Oppenheim AV, Schafer RW. *Discrete-time Signal Processing, 3rd Edn*. Upper Saddle River, NJ: Pearson (2010).
27. Vetter K, di Rosario S. ExpoChirpToolbox: a pure data implementation of ESS impulse response measurement. In: *Proceeding Pure Data Convention*. Weimar (2011). Available online at: https://www.uni-weimar.de/kunst-und-gestaltung/wiki/images/ExpoChirpToolbox-_a_Pd_implementation_of_ESS_impulse_response_measurement.pdf
28. Lesiak C, Krener A. The existence and uniqueness of Volterra series for nonlinear systems. *IEEE Trans Autom Control*. (1978) 23:1090–5. doi: 10.1109/TAC.1978.1101898
29. Widrow B, Stearns S. *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall (1985).
30. Liu TC, Liu YW, Wu HT. Denoising click-evoked otoacoustic emission signals by optimal shrinkage. *J Acoust Soc Am*. (2021) 149:2659–70. doi: 10.1121/10.0004264
31. Venema B, Blanik N, Blazek V, Gehring H, Opp A, Leonhardt S. Advances in reflective oxygen saturation monitoring with a novel in-ear sensor system: results of a human hypoxia study. *IEEE Trans Biomed Eng*. (2012) 59:2003–10. doi: 10.1109/TBME.2012.2196276
32. Goverdovsky V, Looney D, Kidmose P, Mandic DP. In-Ear EEG From viscoelastic generic earpieces: robust and unobtrusive 24/7 monitoring. *IEEE Sens J*. (2016) 16:271–7. doi: 10.1109/JSEN.2015.2471183

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.

frontiers in Digital Health

ORIGINAL RESEARCH
published: 03 September 2021
doi: 10.3389/fdgth.2021.725130

Clustering Users Based on Hearing Aid Use: An Exploratory Analysis of Real-World Data

Alessandro Pasta[1,2], Tiberiu-Ioan Szatmari[1,2], Jeppe Høy Christensen[3], Kasper Juul Jensen[2], Niels Henrik Pontoppidan[3], Kang Sun[1,3] and Jakob Eg Larsen[1]*

[1] Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kongens Lyngby, Denmark, [2] Demant A/S, Smørum, Denmark, [3] Eriksholm Research Centre, Oticon A/S, Snekkersten, Denmark

Edited by:
Jing Chen,
Peking University, China

Reviewed by:
Pasin Israsena,
National Electronics and Computer Technology Center, Thailand
Parisis Gallos,
National and Kapodistrian University of Athens, Greece

***Correspondence:**
Alessandro Pasta
alpas@dtu.dk

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: 14 June 2021
Accepted: 26 July 2021
Published: 03 September 2021

Citation:
Pasta A, Szatmari T-I, Christensen JH, Jensen KJ, Pontoppidan NH, Sun K and Larsen JE (2021) Clustering Users Based on Hearing Aid Use: An Exploratory Analysis of Real-World Data. Front. Digit. Health 3:725130. doi: 10.3389/fdgth.2021.725130

While the assessment of hearing aid use has traditionally relied on subjective self-reported measures, smartphone-connected hearing aids enable objective data logging from a large number of users. Objective data logging allows to overcome the inaccuracy of self-reported measures. Moreover, data logging enables assessing hearing aid use with a greater temporal resolution and longitudinally, making it possible to investigate hourly patterns of use and to account for the day-to-day variability. This study aims to explore patterns of hearing aid use throughout the day and assess whether clusters of users with similar use patterns can be identified. We did so by analyzing objective hearing aid use data logged from 15,905 real-world users over a 4-month period. Firstly, we investigated the daily amount of hearing aid use and its within-user and between-user variability. We found that users, on average, used the hearing aids for 10.01 h/day, exhibiting a substantial between-user (SD = 2.76 h) and within-user (SD = 3.88 h) variability. Secondly, we examined hearing aid use hourly patterns by clustering 453,612 logged days into typical days of hearing aid use. We identified three typical days of hearing aid use: full day (44% of days), afternoon (27%), and sporadic evening (26%) day of hearing aid use. Thirdly, we explored the usage patterns of the hearing aid users by clustering the users based on the proportion of time spent in each of the typical days of hearing aid use. We found three distinct user groups, each characterized by a predominant (i.e., experienced ∼60% of the time) typical day of hearing aid use. Notably, the largest user group (49%) of users predominantly had full days of hearing aid use. Finally, we validated the user clustering by training a supervised classification ensemble to predict the cluster to which each user belonged. The high accuracy achieved by the supervised classifier ensemble (∼86%) indicated valid user clustering and showed that such a classifier can be successfully used to group new hearing aid users in the future. This study provides a deeper insight into the adoption of hearing care treatments and paves the way for more personalized solutions.

Keywords: data logging, user clustering, ensemble classification, hearing aid use amount, hearing aid use patterns, hearing aids, personalization

INTRODUCTION

It is estimated that, globally, 430 million people have disabling hearing loss, i.e., a hearing loss greater than 35 decibels (dB) in the better hearing ear (1). By 2050 over 700 million people are expected to have disabling hearing loss (1). Untreated hearing loss has repercussions at an individual level. It is associated with poorer cognitive and psychological status, resulting in increased risk of depression, dementia, falls, and quality of life (2–4). Hearing loss negatively impacts education, employment, and household income (1, 5). Additionally, untreated hearing loss has a negative impact on society and the economy. Older adults with untreated hearing loss experience higher health care costs and utilization patterns compared with adults without hearing loss (4). The World Health Organization (1) estimates that untreated hearing loss poses an annual global cost of US$ 980 billion, including health sector costs, costs of educational support, loss of productivity, and societal costs.

The adoption of hearing aids (HAs) has been shown to have a positive impact on the quality of life of users (6, 7) and to mitigate the effect on their household income (5). The success of HA provision as a treatment for hearing loss depends on the fact that the patient is provided with a favorable change in their condition, but also on the patient compliance with the intervention program (8). Perez and Edmonds (8) conducted a systematic review to identify and evaluate how studies have measured and reported the use of HAs in older adults. A limited number of studies (5 out of 64) were found to assess HA use based on objective measures, such as data logging and battery consumption. Most of the studies assessed HA use through self-reported measures, such as standardized questionnaires, custom questionnaires, interviews, and diaries. However, self-reported measures have been shown to diverge from objective measures, leading to inaccurate and overreported HA use (9–12). In addition to avoiding such recall bias, objective data logging enables measuring HA use with a greater temporal resolution and longitudinally (13). The widespread adoption of smartphones among older adults (14) and the introduction of smartphone-connected hearing aids make it possible to objectively assess the HA use of a larger number of users than ever before (15).

When evaluating HA usage, the amount of HA use time is commonly regarded as an indicator of treatment success (16) and frequently investigated (9, 12, 17–19). Although the amount of HA use time generally correlates with HA satisfaction (20), this metric might not provide a complete picture. Indeed, frequent use does not necessarily equate with benefit (21). A previous study found that some HA users reported low HA use time and high HA satisfaction, while other users reported high HA use time and low HA satisfaction (22). Furthermore, HA use time provides information about how much the HA has been used during the day, but it is not informative of when and how the HA has been used. For instance, two users might exhibit the same amount of use time (e.g., 8 h), but use the HAs at different times of the day (e.g., from 8:00 to 16:00 and from 15:00 to 23:00) or in different ways (e.g., on-off usage or continuous usage). For these reasons, in addition to the amount of HA use time, other patterns of HA use should be analyzed (11). However, methods possessing low temporal resolution (e.g., self-reports or accumulated use time across a day or a week) do not account for the hourly and daily variability in HA use. Smartphone-connected HAs enable continuous data logging, thereby making it possible to assess the hourly HA use and more accurately identify recurrent use patterns.

Additionally, the HA industry is currently predominantly accommodating for the average user (23). However, the amount of HA use varies widely among HA users (9, 19, 24). Similarly, the pattern of HA use has been reported to vary among HA users. Laplante-Lévesque et al. (11) clustered 171 HA owners, showing that 57% had, on average, a continuous HA use during the day, while 43% had an on-off HA use. A qualitative study (16) reported that optimal HA use depends on the individual needs of the HA owners and does not necessarily correspond to wearing the HAs most of the time. Some HA users reported that they do not depend on their HAs and that they experience situations which they can successfully attend without HAs. Therefore, it is of interest to objectively measure and investigate the HA use of a large number of HA owners, in order to identify and quantify different types of users based on their HA use patterns. This potentially enables gaining deeper insight into the adoption of hearing care treatments and paves the way for more personalized solutions (25).

Finally, when comparing users based on their HA use, the average individual use is usually considered. This means that the within-user variability in HA use is often disregarded (11, 19, 24). However, HA users might exhibit different HA use patterns from one day to another and two HA users with the same average use might behave differently. For instance, two users might exhibit the same average amount of use time throughout the logged days (e.g., 8 h), but one might use the HA constantly (e.g., 8 h each day) while the other might exhibit more variation among the days (e.g., alternating days with 2 and 14 h of use). Therefore, when comparing users based on HA use, it is desirable to adopt a metric that goes beyond the average use per user and that considers the within-user variability.

In this study, we analyze the objective HA use data logged from 15,905 real-world users over a 4-month period. Firstly, we investigate the daily amount of HA use and its within-user and between-user variability. Secondly, we examine HA use hourly patterns by clustering the 453,612 logged days to identify typical days of HA use. Thirdly, we explore the usage patterns of the HA users and investigate whether we can cluster the users based on how they used the HAs during the logged days. When performing the user clustering, instead of representing each user by her average HA use pattern, we consider the proportion of time spent in each of the typical days of HA use. Finally, we validate the HA user clustering by training a supervised classifier to predict the cluster to which each user belongs.

MATERIALS AND METHODS

Participants and Apparatus

This study used data from a large-scale internal database, which logs the HA use of HA owners who have signed up for the HearingFitness™ feature (25) *via* the Oticon ON™ smartphone app. The participants were the users of Oticon Opn™ hearing aids who used the HearingFitness™ feature for at least 10 days in the period between June and September 2020.

Data and Data Pre-processing

When the HAs are connected to the smartphone, the HearingFitness™ feature logs timestamped data about the HA use every 10 min. Based on HA use time (i.e., inferred from time counters embedded in the HAs) and connection status, an estimate of hourly HA use (measured in min/h) is computed. For binaural HA users, if the HA use amount was different between the right and left ear, this study selected the larger value, as done by Laplante-Lévesque et al. (11) and Walker et al. (27). If temporary disconnections occur, replacements for the missing data are injected by analyzing the time counters embedded in the HAs. When the disconnected use is full-time use (e.g., 120 min of use during 2 h of disconnection), the HA use during disconnection is simply assigned to the hours of disconnection. When the disconnected use is on-off use (i.e., not full-time use), the minutes of use are evenly distributed among the hours of disconnection (e.g., 60 min of use during 2 h of disconnection result in 30 min/h use for 2 h). The raw data set comprised 1,160,520 days of HA use from 32,216 users. In order to preserve representative patterns of HA use throughout the day, days with on-off use during temporary disconnections longer than 2 h were removed. Additionally, 12,876 days with more than 60 min/h were removed. This is likely a consequence of use time estimation when disconnections occur. Moreover, since this study focuses on analyzing HA use, only days with at least 60 min of HA use were included. Furthermore, only data related to HA use between 6:00 and 23:59 were included. Finally, to ensure that users' behavior was inferred from a representative sample of days, only users with at least 10 days of HA use were included. The cleaned data set comprised 453,612 days of HA use from 15,905 users (28.5 days per user on average).

Data Analysis

Figure 1 provides an overview of the flow of the data analysis we performed, presenting the main steps undertaken. More details on each step are provided below.

Exploring the Amount of Hearing Aid Use

We explored the amount of HA use (measured in hours/day), by computing summary statistics of the 453,612 logged days (mean, SD) and of the amount of HA use for each user (mean, between-user SD, quartiles). Furthermore, we analyzed the within-user daily variability (SD) in HA use amount. Independent sample t-tests were performed to compare the within-user SD of medium users (i.e., users with average HA use amount between Q_1 and Q_3) with that of light and heavy users (i.e., users with average HA use amount, respectively, below Q_1 and above Q_3). Cohen's d was computed to assess the magnitude of the differences (45). A polynomial linear regression was fitted to model the relationship between average amount of HA use per user (x) and within-user SD (y).

Clustering Days of Hearing Aid Use

We examined patterns of HA use by clustering the 453,612 logged days into typical days of HA use. The input data consisted of a $453,612 \times 18$ matrix

$$\mathbf{A}_{\mathbf{r}\times\mathbf{c}} = \mathbf{A}_{453612\times 18} = \begin{bmatrix} a_{11} & \cdots & a_{1c} \\ \vdots & \ddots & \vdots \\ a_{r1} & \cdots & a_{rc} \end{bmatrix}$$
$$= \left(a_{ij}\right) \in [0, 60] \quad i = 1, \ldots, r; j = 1, \ldots, c$$

where each row i represents a day of HA use, each column j represents an hour of the day (from 6 to 23) and a_{ij} is the amount of HA use (from 0 to 60 min) in the day i and hour j. The k-means clustering technique was applied (28), since it is suitable for large data sets. K-means aims to partition the observations in k clusters by minimizing the within-cluster variance (i.e., square Euclidean distances). The k-means++ initialization algorithm (29) was applied, which seeks to spread out the k initial clusters to avoid poor approximation. The optimal value of k was determined using the elbow method (30), which aims to select a number of clusters so that adding another cluster does not substantially increase the explained variation. The resulting clusters were evaluated by conducting a Silhouette analysis (31), which aims to evaluate the between-clusters dispersion (i.e., separation) and the within-cluster dispersion (i.e., cohesion). A Silhouette Coefficient (ranging from -1 to $+1$) was calculated for each observation and constitutes a measure of how similar an observation is to its own cluster compared to the next nearest cluster. Furthermore, principal component analysis was performed to visualize the observations in a lower dimensional space. Subsequently, for each cluster, we identified and removed observations that were abnormally distant from the other observations (i.e., below $Q_1 - 1.5 \cdot IQR$ and above $Q_3 + 1.5 \cdot IQR$). This was done in order not to include days of HA use that exhibited atypical patterns and were not well-represented by the cluster centroids. The association between the type of day of HA use and the day of the week was tested by performing a χ^2 test of independence and computing Cramer's V. The clustering and related analyses were performed in Python, using the scikit-learn library (32).

Clustering Users

We explored the behavior of HA users by clustering the 15,905 users based on the proportion of time spent in each of the typical days of HA use. The input data consisted of a $15,905 \times c$ matrix

$$\mathbf{B}_{r\times c} = \mathbf{B}_{15905\times c} = \begin{bmatrix} b_{11} & \cdots & b_{1c} \\ \vdots & \ddots & \vdots \\ b_{r1} & \cdots & b_{rc} \end{bmatrix}$$
$$= \left(b_{ij}\right) \in [0, 1] \quad i = 1, \ldots, r; j = 1, \ldots, c$$

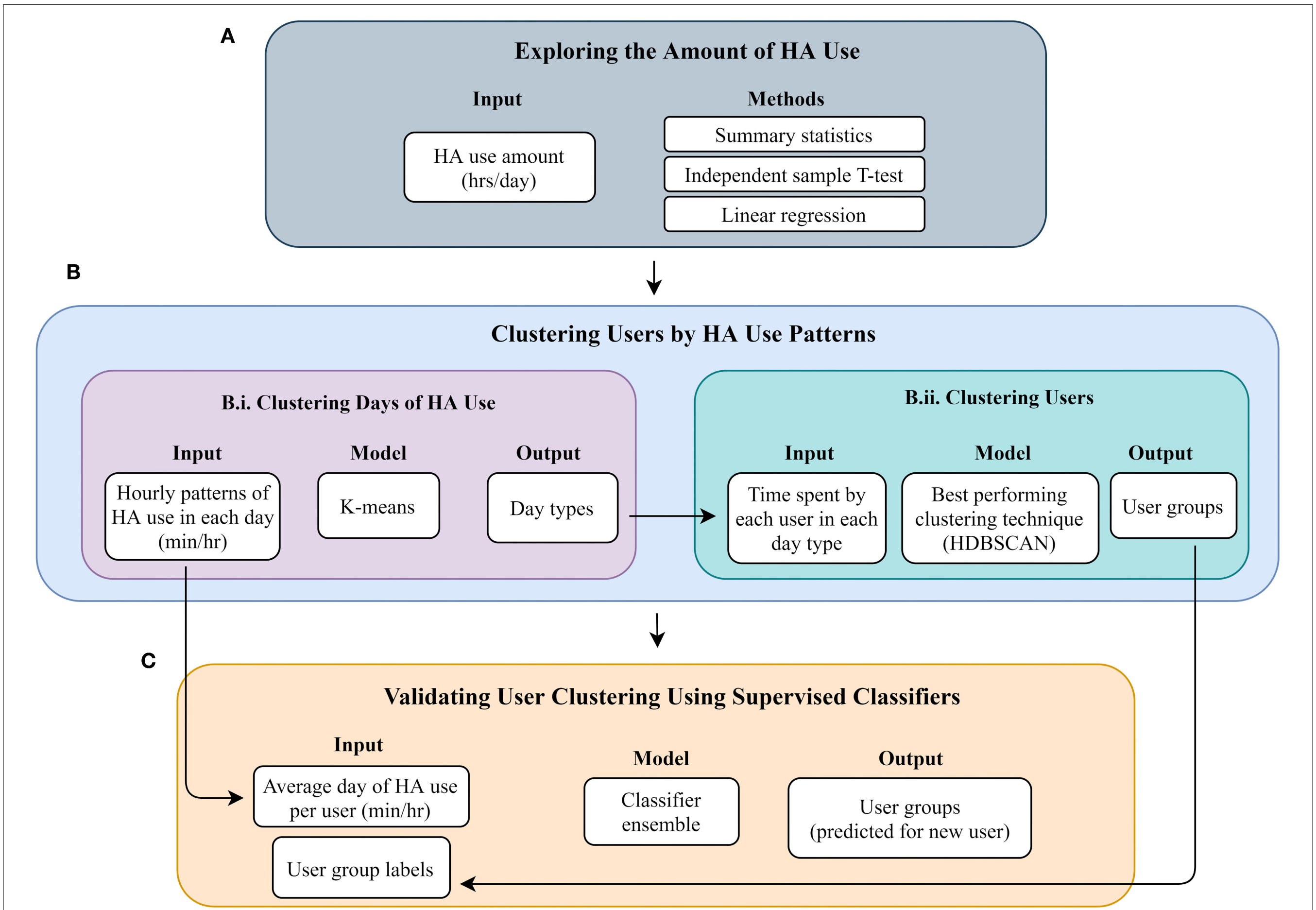


FIGURE 1 | Flow of data analysis. **(A)** Exploring the amount of HA use. **(B)** Clustering users by HA use patterns. **(C)** Validating user clustering using supervised classifiers.

where each row i represents a HA user, each column j represents one of the c typical days of HA use (referring to the clusters found *via* section Clustering Days of Hearing Aid Use) and b_{ij} is the proportion of days belonging to day type j for user i. Different clustering techniques were evaluated: k-means with k-means++ initialization algorithm, Hierarchical Agglomerative Clustering (HAC) with Ward's method, HAC with Pearson correlation and average linkage method, and Hierarchical Density-Based Spatial Clustering (HDBSCAN). HAC (33) initially treats each observation as a cluster and then builds nested clusters by successively merging pairs of the most similar clusters. HDBSCAN (34) groups observations that are in a dense region while marking the observations in sparse regions as noise. It expands on a different density-based technique, DBSCAN (35), by converting it into a hierarchical clustering technique, followed by extracting a flat clustering based on cluster stability. For k-means, the optimal value of clusters was determined using the elbow method (30). For HAC, the optimal value of clusters was determined using the dendrogram. HDBSCAN, instead, infers the optimal number of clusters based on the data. For each clustering technique, three internal validation metrics were computed: Silhouette score (31), Calińksi-Harabasz score (36), and Davies-Bouldin score (37). The Calińksi-Harabasz score is defined as a ratio of separation and cohesion. The Davies-Bouldin score measures the average similarity between each cluster and its most similar one, by comparing the distance between clusters with the size of the clusters themselves. Based on the three metrics, the best performing clustering technique was selected. The clustering was performed in Python, using the scikit-learn (32) and hdbscan (38) libraries.

Validating User Clustering Using Supervised Classifiers

We validated the HA user clustering by training an ensemble of supervised classifiers to predict the cluster label for individual users based on the average day of HA use for each user. The input data for classification consisted of a $15,905 \times 18$ matrix:

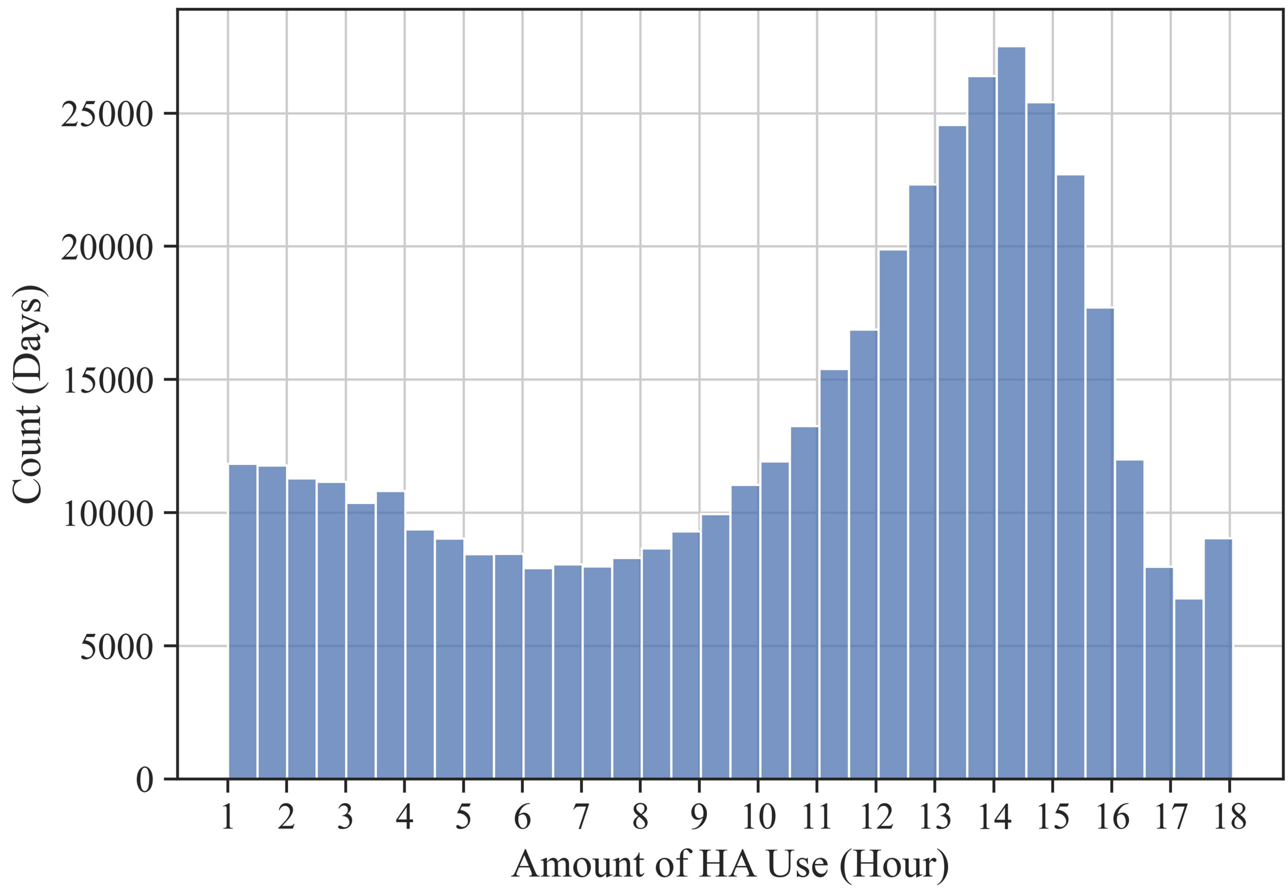


FIGURE 2 | Count of logged days by the amount of HA use time. Due to the data cleaning criteria (i.e., only days with at least 1 h of HA use were included; only data related to HA use in the 18 h between 6:00 and 23:59 were included), the amount of HA use (x-axis) ranges from 1 to 18 h.

$$\mathbf{D}_{\mathbf{r}\times\mathbf{c}} = \mathbf{D}_{15905\times 18} = \begin{bmatrix} d_{11} \cdots d_{1c} \\ \vdots \quad \ddots \quad \vdots \\ d_{r1} \cdots d_{rc} \end{bmatrix}$$
$$= \left(d_{ij}\right) \in [0, 60] \quad i = 1, \ldots, r; j = 1, \ldots, c$$

where each row i represents the average day of a HA user, each column j represents the hour of the day (from 6 to 23) and d_{ij} is the average amount of HA use (from 0 to 60 min) for user i in the hour j. This data was further split into separate training and testing data sets with an 80/20 split. To reduce bias (39), three classifiers were chosen from different families: multiclass logistic regression (regression), an XGBoost classifier (decision trees) (40) and a fully connected (FC) neural network classifier (41). The following individual parameters were chosen:

- Multiclass logistic regression: L2 penalty and "newton-cg" solver.
- XGBoost: estimators = 100, max depth = 5, gamma = 0, alpha = 0.1.
- FC neural network: four-layer network (128-64-32-4), ReLU activation, cross-entropy loss, Adam optimizer; trained for 25 epochs.

In order to reduce bias (39), a classification ensemble was defined, which assigns each user to a group by majority voting between the three classifiers. In cases where no majority could be defined, the group was decided by the best performing individual classifier. Two metrics were used to gauge each model's performance: accuracy, and Area Under the Receiver Operating Characteristic (ROC-AUC). Accuracy is obtained by calculating the ratio of correct test predictions to the total amount of samples in the testing set. ROC-AUC helps visualize the relationship between sensitivity (i.e., True Positive Rate) and specificity (i.e., False Positive Rate) for a binary classification problem. The ROC-AUC value ranges from 0 to 1 and represents the ability of a classifier to distinguish between classes at various thresholds. If the current classification task operates with more than two classes (i.e., multiclass classification), the individual classes are first binarized. The score of the individual classes is calculated, then a micro-average is computed by aggregating the contributions of all classes to compute the average metric. Finally, a macro-average is calculated by computing the metric independently for each class and then calculating the average. The training and evaluation of the supervised classification ensemble was performed in Python, using scikit-learn (32),

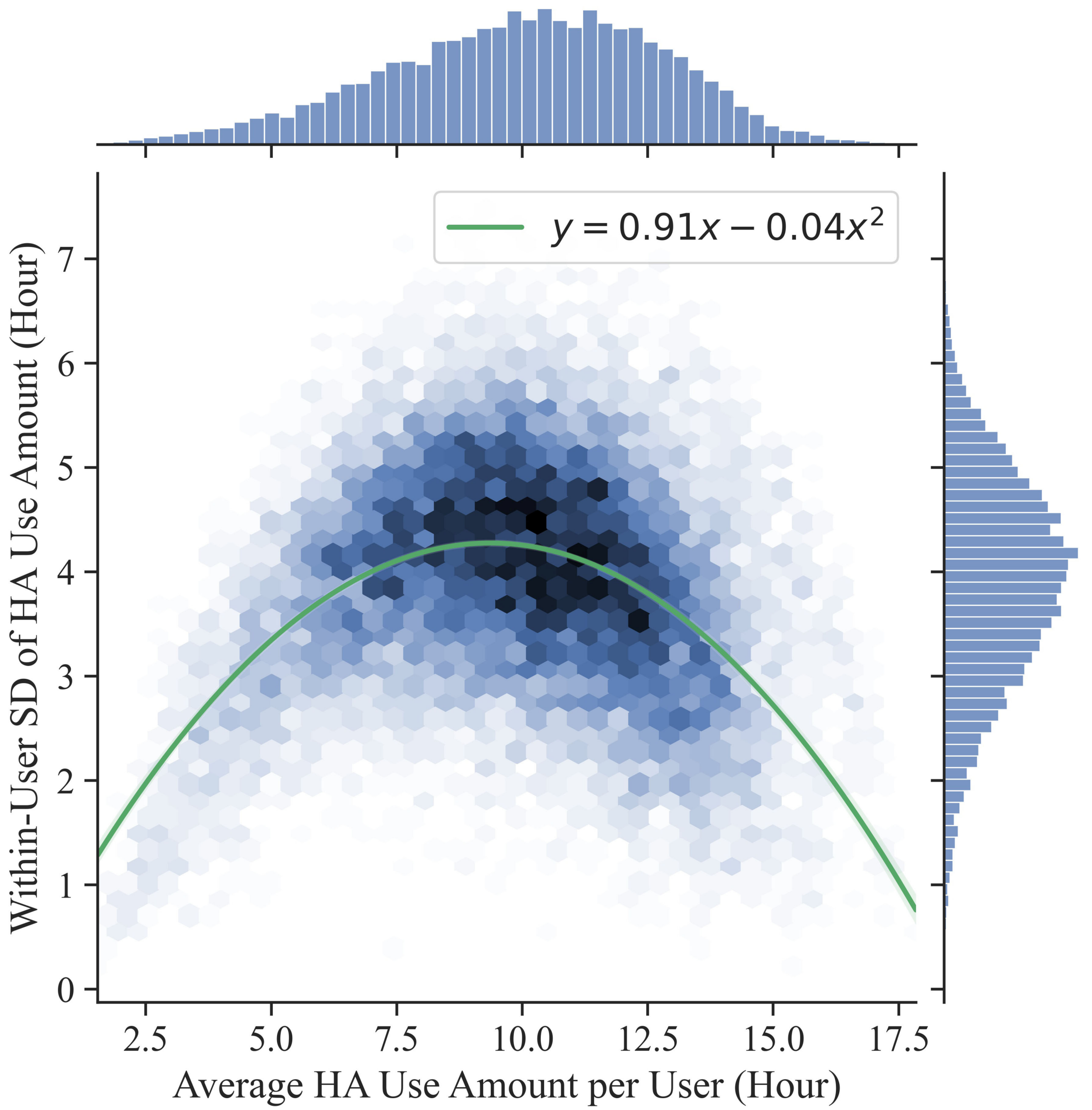


FIGURE 3 | Distribution of users by their average amount of HA use and their variability among the logged days (i.e., within-user SD). A second order linear regression model (line ± 99% confidence interval) was fitted to the data to model the relationship between average HA use (x) and within-user SD (y).

XGBoost (40), PyTorch (42), Yellowbrick (43), and scikit-plot libraries (44).

RESULTS

Exploring the Amount of Hearing Aid Use

The clean data set comprised 453,612 days of HA use from 15,905 users. The amount of HA use, defined as hours of HA use per day, was assessed to describe usage. **Figure 2** shows the frequency distribution of HA use amount during the pooled logged days. The data represents the HA use in days of connected use. On average, a day of HA use amounted to 10.55 h. However, the days were not normally distributed around the mean. The amount of HA use widely varied throughout the logged days (SD = 4.71 h), with a mode around 14 h of use and a smaller peak around 1 h of use.

On average, 28 days (SD = 18 days) were logged for each user. We investigated the extent to which users used the HAs differently among each other (i.e., between-user variability) and the extent to which the same user used the HAs consistently throughout the logged days (i.e., within-user variability). For each user, the average amount of HA use and the within-user standard deviation (SD) among the days of HA use were computed. **Figure 3** shows the distribution of the 15,905 HA users. We firstly investigated the between-user variability in the amount of HA use (x-axis in **Figure 3**). Users had an average amount of HA use of 10.01 h, with a SD of 2.76 h (Coefficient of variation = 0.276). The middle 50% of users (*medium users,*

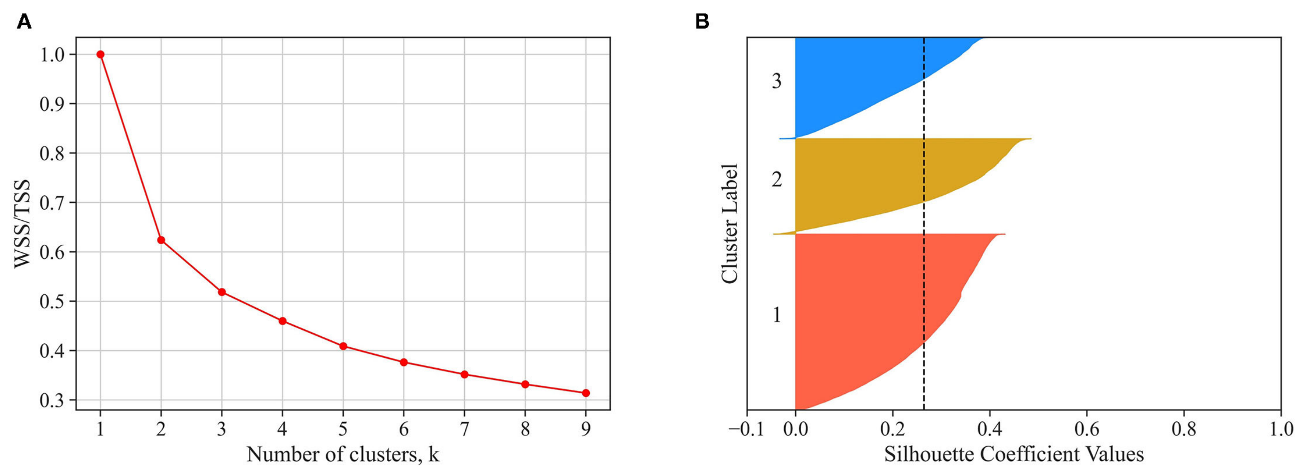


FIGURE 4 | In **(A)**, the ratio of within sum of squares (WSS) to the total sum of squares (TSS) is displayed as a function of the number of clusters. The elbow plot suggests selecting three clusters since adding an additional cluster does not substantially increase the explained variation. In **(B)**, the Silhouette coefficient value of each observation is displayed for the three-cluster solution (i.e., for each cluster, the observations are ordered by their Silhouette coefficient value and displayed in ascending order as horizontal stacked lines). The average silhouette score is reported (dashed line). The three clusters have predominantly positive scores, suggesting valid clustering.

between the first and third quartiles) ranged from 8.18 to 12.04 h (group mean = 10.16) of average HA use. The fact that the remaining 50% of the users exhibited an average HA use either below 8.18 (*light users*; group mean = 6.32) or above 12.04 (*heavy users*; group mean = 13.37) h indicates a substantial between-user variability.

Additionally, we investigated the within-user variability in the amount of HA use (y-axis in **Figure 3**). The average within-user SD was 3.88 h, indicating that the same user tended to use the hearing aids for varying durations throughout the logged days. A significantly larger within-user SD was observed for the medium users compared to both the light users (Two-sample t-test: $t = 23.06$, $p < 0.001$; Effect size: $d = 0.44$) and the heavy users (Two-sample t-test: $t = 41.85$, $p < 0.001$; Effect size: $d = 0.81$). This proves that both light users and heavy users were more consistent than medium users throughout the logged days (i.e., lower within-user SD) and constitutes an indication of users consistently displaying diverse behaviors in terms of HA use. The relationship between average HA use (x) and within-user SD (y) was modeled by fitting a second order linear regression model to the data. The line of best fit ($R^2 = 0.2$) was described by the equation $y = 0.91x - 0.04x^2$. The maximum of the curve is around 10 h of HA use, indicating that the within-user SD increases with the amount of HA use for users using the HAs up to 10 h and it decreases for users using the HAs more than 10 h.

Clustering Days of Hearing Aid Use

The substantial within-user variability in HA use suggests that a deeper analysis is warranted, which accounts for the hourly and daily variability in HA use. In addition to the amount of HA use, we also assessed patterns of HA use, defined as minutes of HA use per hour throughout the day. That was done by clustering the 453,612 logged days into typical days of HA use (see subsection Data Analysis). Based on the elbow method (**Figure 4A**), a three-cluster solution was chosen, which accounts for almost 50% of the variance among days. The Silhouette analysis (**Figure 4B**) indicated that the three clusters have predominantly positive scores, and there are no clusters with below-average silhouette scores.

Figure 5 displays the 453,612 days of HA use plotted by the two main principal components and colored by the three clusters. The eigenvectors suggest that the first principal component is negatively correlated with HA use in all hours of the day, differentiating between days of heavy use (**Figure 5A**, left) and days of light use (**Figure 5A**, right). The second principal component, instead, is positively correlated with HA use in the morning hours and negatively correlated with HA use in afternoon and evening hours, differentiating between days of morning HA use (**Figure 5A**, top) and days of HA use later in the day (**Figure 5B**, bottom). For each cluster, outliers were removed, resulting in 440,052 observations belonging to the three clusters. Looking at the hourly mean of HA use for each cluster (**Figure 5B**), it is possible to qualitatively evaluate the patterns underlying the clusters. Three distinct types of days of HA use can be identified: a full day of HA use (cluster 1, containing 204,062 days), a day of afternoon HA use (cluster 2, containing 120,810 days), and a day of sporadic evening HA use (cluster 3, containing 115,180 days).

A significant ($p < 0.001$), but negligible (Cramer's V = 0.05) association was found between the type of day of HA use and the day of the week (i.e., weekend vs. weekday). Full days of HA use occurred slightly (6%) more often during weekdays than during the weekend. Conversely, afternoon and sporadic evening days

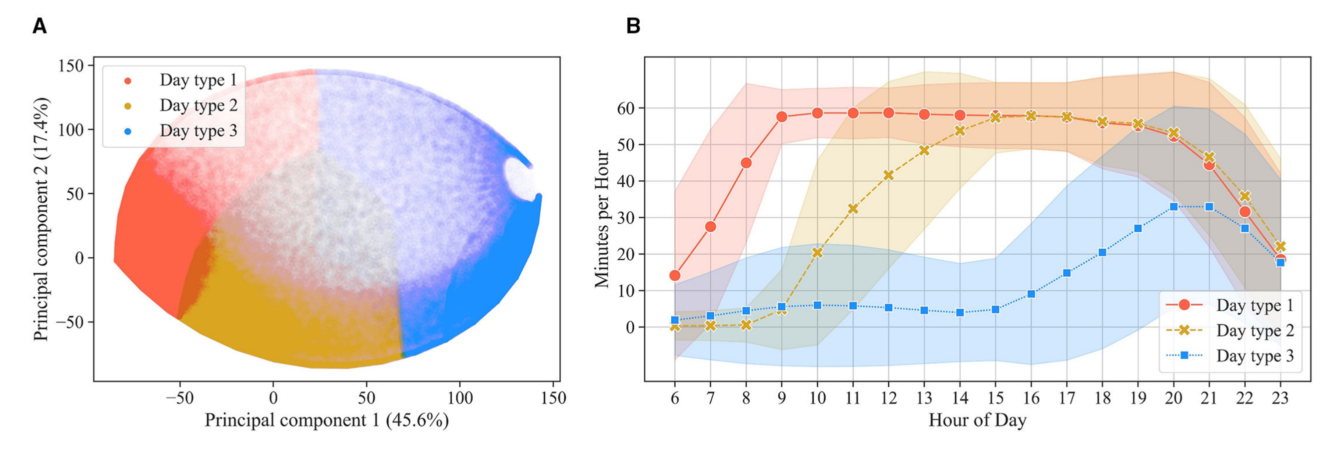


FIGURE 5 | In **(A)**, the days of HA use are displayed as scatterplot against the two main principal components and colored by the three clusters (i.e., three day types). In **(B)**, the mean (±SD) of hourly HA use for each cluster is displayed.

TABLE 1 | Comparison of four different clustering techniques [K-means, HAC (Euclidean distance), HAC (Pearson correlation), and HDBSCAN] based on three internal validation metrics (Silhouette, Davies-Bouldin, and Caliñski-Harabasz).

	K-means	HAC (Euclidean distance and Ward's method)	HAC (Pearson correlation and average linkage)	HDBSCAN (Pearson correlation)
Silhouette (Higher is better)	0.4539	0.4264	0.6400	0.7604
Davies-Bouldin (Lower is better)	0.8267	0.7169	0.6001	0.4176
Caliñski-Harabasz (Higher is better)	18,473	13,732	35,802	70,327

HDBSCAN is the best performing technique according to all three metrics.

occurred slightly (4 and 1%) more often during the weekend than during weekdays.

Clustering Users

Having identified three types of days of HA use enables exploring HA user behavior, thus generating personalized insights, in a way that considers the day-to-day variation of each user. We explored the behavior of HA users by clustering the 15,905 users based on how they used the HAs during the logged days (see subsection Data Analysis). Each user is represented by the proportion of time spent in each of the three types of days of HA use. Four clustering techniques were evaluated. The optimal number of clusters for k-means and both HAC techniques were determined to be three. HDBSCAN also identified three clusters, with the minimum cluster size hyperparameter set to 1,000, in addition to considering some observations as noise. Based on three internal validation metrics, HDBSCAN was chosen (**Table 1**). The Silhouette analysis (**Figure 6**) suggested that the three clusters are of different sizes and have predominantly positive and large scores.

Figure 7A displays the days of HA use experienced by the users belonging to each user group. These plots can be directly compared with **Figure 5A**, which displays all days of HA use from all users. Each user group has a distinctive distribution of days. User group A is the largest cluster (7,862 users) and exhibits a higher density in the left corner of the figure, corresponding with day type 1 (i.e., full day of HA use). User group B (2,442 users) exhibits a higher density in the lower part of the figure, corresponding with day type 2 (i.e., day of afternoon HA use). User group C (3,148 users) has a higher density in the right corner of the figure, corresponding with day type 3 (i.e., day of sporadic evening HA use). Additionally, 2,453 users exhibited atypical behavior and were labeled as noise. The distinctive behavior of the three user groups is confirmed by their average time spent in each of the typical days of HA use (**Figure 7B**). User group A is predominantly having full days of HA usage, user group B is predominantly having days of afternoon HA use, user group C is predominantly having days of sporadic evening HA use. It should be noted that the predominant day of HA use is experienced around 60% of the time by the three user groups.

Validating User Clustering Using Supervised Classifiers

We validated the HA user clustering by training an ensemble of three supervised classifiers (multiclass logistic regression, XGBoost and fully connected neural network) to predict the label of each user (user group A, B, C, or noisy user). The training input was the average day of HA use for each user, defined as minutes of HA use per hour throughout the day (from 6:00 to 23:59).

When evaluating the three individual classifiers based on accuracy and ROC-AUC score (**Table 2**), XGBoost results to

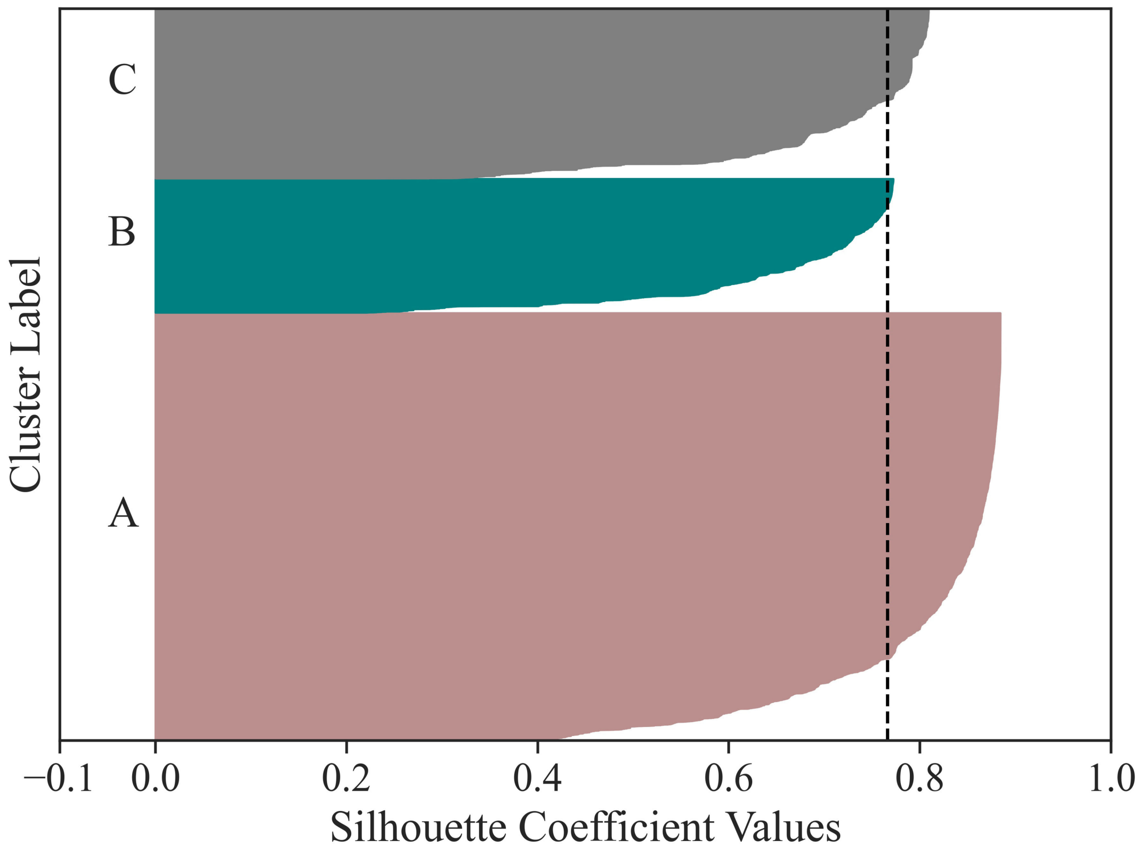


FIGURE 6 | The Silhouette coefficient value of each observation is displayed for the three user clusters (i.e., for each cluster, the observations are ordered by their Silhouette coefficient value and displayed in ascending order as horizontal stacked lines). The average silhouette score is reported (dashed line). The three clusters have predominantly positive and large scores, suggesting valid clustering.

be the best performing classifier. In order to reduce bias, an ensemble of three supervised classifiers was defined. This simulates three artificial experts coming to a decision (40). The ensemble assigns each user to a group by majority voting between the three classifiers. In case where no majority could be defined, the group was decided by the best performing individual classifier (XGBoost). The ensemble accuracy was 86.04%, while the ROC-AUC score was 0.98. While the ensemble has a slightly worse accuracy than XGBoost, relying on classifiers from different classes mitigates the effect of bias that each classifier has. The ROC curves for the ensemble of classifiers (**Figure 8**) show that noisy users exhibiting atypical behavior are the most difficult to classify (i.e., lowest AUC). Conversely, the ensemble of classifiers successfully distinguishes between the three user groups.

It is interesting to inspect the importance attributed by XGBoost to each of the 18 hours considered (**Table 3**). XGBoost values h9 and h15, indicating that these two hours are the ones that mostly differentiate the three user groups. This is consistent with the fact that each user group is characterized by a predominant day of HA use (**Figure 7**), and that h9 and h15 are the most effective hours in differentiating between the three day types (**Figure 5B**).

DISCUSSION

While HA use has been traditionally assessed through subjective self-reports, smartphone-connected HAs enable objective data logging of HA use. This study investigates the objective HA use of a large cohort of HA users. 453,612 days of HA use logged by 15,905 users were analyzed.

The amount of HA use time is informative of how long the HA has been used during a day. On average, the users used the HAs for 10.01 h/day. This value is similar (11, 17) or slightly larger (10, 19) than previous studies objectively measuring HA use. When investigating the variability between users, this study found that 25% of users used the HAs for <8.18 h. This percentage is similar to a study by Laplante-Lévesque et al. (11), but smaller than other studies (12, 18, 19) objectively or subjectively assessing the amount of HA use of several users. The inclusion criteria of this study (i.e., users of the HearingFitness™ feature *via* a smartphone app) and the data cleaning criteria (i.e., days with at least 60 min of HA use) could explain the greater average HA use and the lower percentage of light users. Moreover, a greater average HA use could be explained by the fact that, for binaural HA users this study selected the larger value between the right and left ear. While

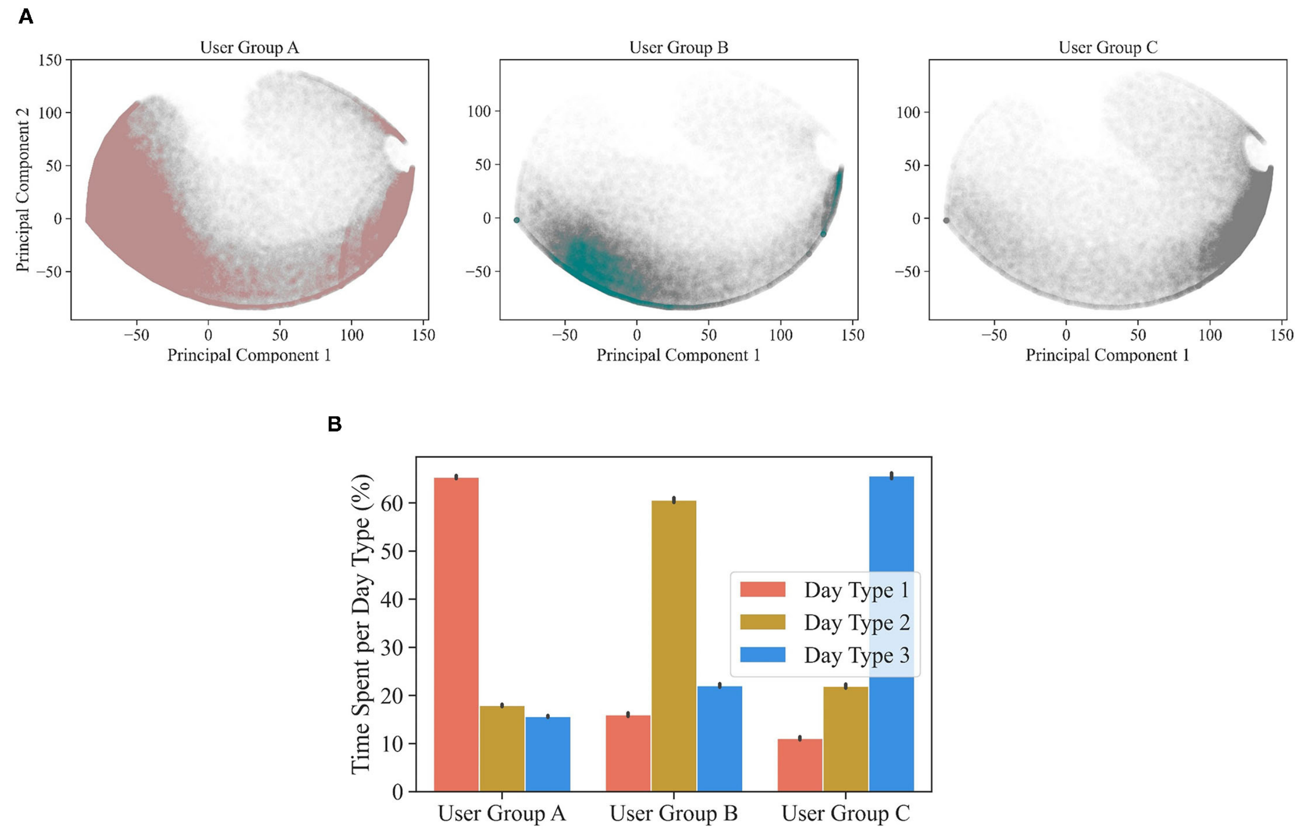


FIGURE 7 | In **(A)**, for each of the three user clusters (i.e., user group A, B, and C), the days of HA use are displayed as scatter plot against the two main principal components. The distinct densities indicate that the three user groups experienced substantially different days of HA use. In **(B)**, the average proportion of time spent (±95% confidence interval) in each day type is displayed for each user cluster.

TABLE 2 | Comparison of three individual classifiers (multiclass logistic regression, XGBoost, and fully connected neural network) and of the classification ensemble based on two performance metrics (accuracy and ROC-AUC score).

Classifier type	Accuracy (0-100 %)	ROC-AUC Score (micro-average)
Logistic regression	81.51	0.97
XGBoost	87.08	0.98
FC neural network	85.56	0.98
Ensemble	86.04	0.98

XGBoost is the best performing individual classifier according to both metrics.

HA users can either exhibit a low or high average amount of daily HA use, their day-to-day fluctuations in HA use provide a deeper understanding of HA use. The fluctuations in day-to-day HA use (i.e., within-user SD) were lower for light and heavy users compared to medium users, proving that a substantial number of users consistently displayed diverse behaviors in terms of HA use.

In addition to the amount of HA use, continuous data logging enables assessing how and when HAs were used during the day. Based on patterns of hourly use, the 453,612 days of HA use were clustered into three typical days. Forty-four percent of days were characterized by full HA use. This indicates that generally, when worn, HAs tend to be turned on in the morning (around 7), used uninterruptedly throughout the day, and turned off in the evening (around 22). Twenty-seven percent of days were characterized by afternoon use. This indicates that HAs are occasionally turned on in the late morning (around 11) and used uninterruptedly until the evening (around 22). This behavior might be due to a different individual daily rhythm or to a day encompassing different activities (e.g., weekend had a significant, but negligible effect on the day type). Twenty-six percent of days were characterized by sporadic evening HA use. This suggests that HAs are sometimes used in isolated occasions and for a limited number of hours. The remaining days (3%) were atypical days of HA use and exhibited infrequent behavior.

Based on the proportion of time spent in each of the typical days of HA use, the 15,905 users were clustered in three user groups. This method allowed to investigate users' behavior while preserving the individual day-to-day variability in HA use. Almost half of the users (group A, 49% of users) predominantly had full days of HA use. This group might include users that have an active life and engage in social interactions starting in the

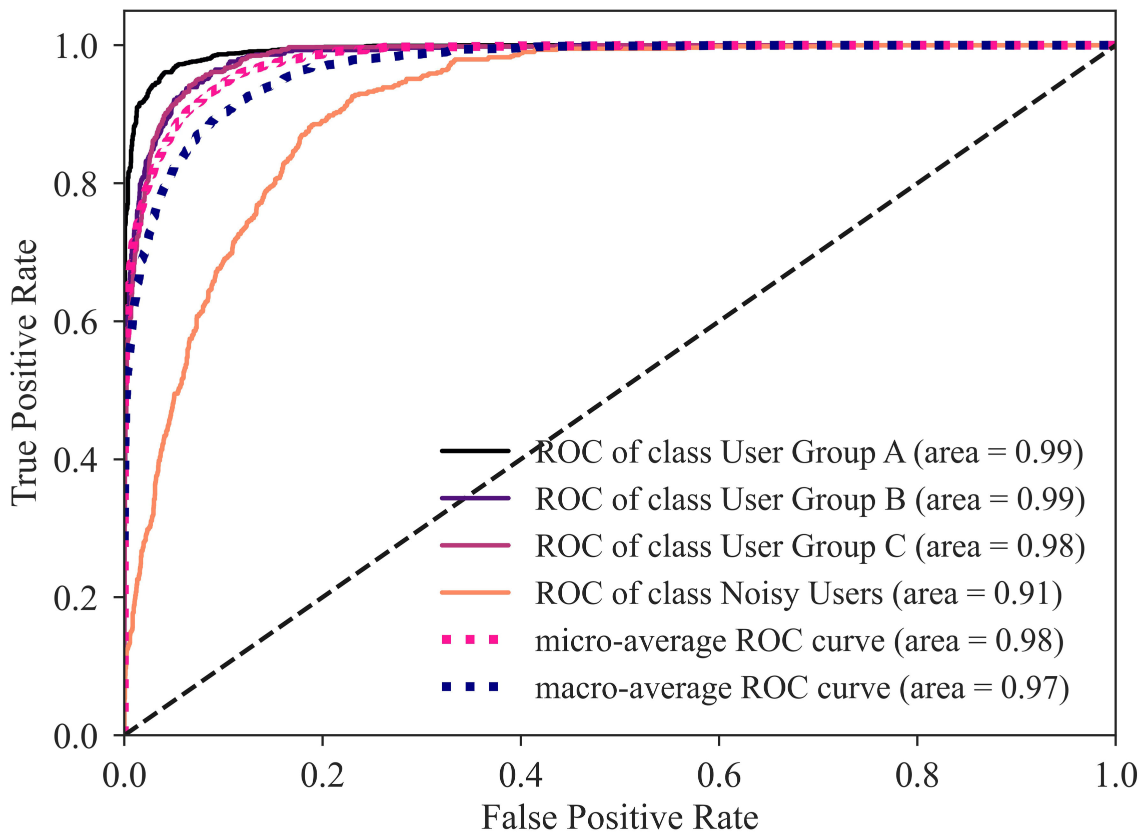


FIGURE 8 | ROC-AUC plot for the ensemble of classifiers, illustrating the tradeoff between sensitivity (True Positive Rate) and specificity (False Positive Rate). The ideal point is the top-left corner, higher AUC is better. In this multiclass scenario, the individual classes are first binarized, the individual scores are computed for each user group, then micro-averages and macro-averages are calculated for each classifier.

morning and throughout the entire day. Because of the inclusion criteria of this study (i.e., users of a smartphone app that tracks HA use), this group might be overrepresented. A smaller portion of users (group B, 15%) predominantly had days of afternoon use. This group might include users that engage in activities and social interactions later in the day. Group A and B, together, indicate that 64% of users tended to use the HAs uninterruptedly, a percentage similar to the 57% found by Laplante-Lévesque et al., (11). Twenty percent of users (Group C) predominantly had days of sporadic evening use. This group might either contain users that are not acclimatized to their HAs or users that do not depend on their HAs and only need them in specific situations (16). The remaining 15% of users were classified as noise, suggesting that some users have an uncommon behavior, more evenly alternating among the typical days of HA use. This percentage is in line with a study by Laplante-Lévesque et al., (11), according to which 23% of the subjects described their HA use to be different from day to day. Interestingly, in all three user groups, we found that the predominant day of HA use accounted for ∼60% of the time, suggesting that users exhibited a substantial within-user variability in terms of day type experienced throughout the logged days. This aspect might not emerge from self-reported assessments that suffer from recall bias, as indicated by a previous study in which most participants (77%) reported their HA use to be the same every day (11).

The user clustering was validated by training a supervised classification ensemble to predict the cluster to which each user belongs. The high accuracy achieved by the supervised classifier ensemble (∼86%) indicates valid user clustering. Indeed, this approach is based on the idea that good clustering should also support good classification, where the better the classification performance the higher the quality of the partition. As such, a high-quality partition is defined by compact clusters separated from each other to the extent that an artificial expert (i.e., a supervised classifier) can distinguish the cluster to which a new user belongs (39). This evaluation was performed to complement internal validation methods (i.e., using information of the clustering process). Internal validation methods attempt to evaluate cluster structure quality, the appropriate clustering algorithm, and the number of clusters without additional information but depend on assumptions such as the presence of underlying structure for each cluster, resulting in weaker results when they do not hold. Alternatively, cluster quality could theoretically be evaluated using external validation, which requires additional, "true" cluster labels to compare against. In real-world scenarios, finding "true" labels is often difficult as

TABLE 3 | Input feature importance returned by XGBoost.

H6	H7	H8	H9	H10	H11	H12	H13	H14	H15	H16	H17	H18	H19	H20	H21	H22	H23
0.013	0.014	0.028	**0.326**	0.017	0.019	0.021	0.039	0.064	**0.256**	0.042	0.037	0.021	0.025	0.024	0.014	0.017	0.015

*The values indicate how valuable each of the 18 features (from H6 in the morning to H23 at night) are in the construction of the boosted decision trees (internal to the model). The values greater than 0.2 are marked in bold. The model values divergence points between the three day types (H9, H15) (see **Figure 5B**). Total value is 1.*

raw data may not have reference labels, thus making external validation methods unusable.

Clustering users based on their HA use patterns provides a deeper insight into the adoption of hearing care treatments and paves the way for more personalized solutions. For instance, users that predominantly have days of sporadic evening HA use might have specific needs compared to the users that uninterruptedly use the HA for the entire day. They might only need the HAs in specific situations and thus benefit from targeted HA settings or features. Additionally, training a supervised classifier based on data labeled by a clustering technique enables future predictions for new users. Based on the average day of HA use of a new user, the classifier can predict her user group, thereby identifying users with similar behaviors and potentially leveraging on the accumulated knowledge of existing users. This can improve the clinical flow by helping audiologists make data-driven decisions.

Looking into the future, a more advanced level of personalization could improve the quality of hearing care solutions and help alleviate major challenges concerning new users, such as the cold start problem. This can be defined as the delay between starting to use the HAs and the moment when enough data was generated locally for meaningful results. Furthermore, an individual's dynamic sound environment, or soundscape, may also be an important factor for personalization. Considering the large number of soundscapes a user may be exposed to throughout the day (public transport, social events, work environments, etc.), additional features can potentially account for both the within-user and the between-user variability. An effective clustering technique for grouping similar users may serve to balance this increase in complexity, especially if advanced privacy-preserving techniques such as federated learning and differential privacy are considered. Federated learning is a machine learning framework where models are trained locally, and afterwards aggregated between participating users. This type of model development could provide access to unrivaled amounts of quality user data, as privacy concerns can only be alleviated if users never have to give away their data. Real-world implementation of such a technique could provide tangible benefits to both existing users, as well as improve the experience of new users, thus enabling next-generation privacy focused personalization systems.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is not publicly available. Requests to access the data should be directed to the corresponding author.

ETHICS STATEMENT

In the sign-up process, the participants actively gave their consent for data to be collected, stored, and used for research purposes on aggregated levels. No personal identifier was collected. No ethical approval was required for this study according to Danish National Scientific Ethical Committee (26).

AUTHOR CONTRIBUTIONS

AP and T-IS conceived and designed the study, organized the database, and performed the statistical analysis. AP wrote the manuscript. JC, KJ, JL, NP, and KS supervised the findings and revised the final manuscript. All authors contributed to the article, read, and approved the submitted version.

REFERENCES

1. World Health Organization. *Deafness and Hearing Loss.* (2021). Available online at: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss (accessed May 25, 2021).
2. Kirkwood DH. Major survey documents negative impact of untreated hearing loss on quality of life. *Hear J.* (1999) 52:32. doi: 10.1097/00025572-199907000-00004
3. Jayakody DMP, Friedland PL, Eikelboom RH, Martins RN, Sohrabi HR. A novel study on association between untreated hearing loss and cognitive functions of older adults: baseline non-verbal cognitive assessment results. *Clin Otolaryngol.* (2018) 43:182–91. doi: 10.1111/coa.12937
4. Reed NS, Altan A, Deal JA, Yeh C, Kravetz AD, Walhagen M, et al. Trends in health care costs and utilization associated with untreated hearing loss over 10 years. *JAMA Otolaryngol Head Neck Surg.* (2019) 145:27–34. doi: 10.1001/jamaoto.2018.2875
5. Kochkin S. The *Impact of Untreated Hearing Loss on Household Income. Better Hearing Institute*. (2005). Available online at: http://ihsinfo.org/IhsV2/News_Info/MarkeTrak7_ImpactUntreatedHLIncome.pdf
6. Kochkin S, Rogin CM. Quantifying the obvious: the impact of hearing instruments on quality of life. *Hear Rev.* (2000) 7:6–34. doi: 10.1097/00025572-200002000-00004
7. Vuorialho A, Karinen P, Sorri M. Effect of hearing aids on hearing disability and quality of life in the elderly. *Int J Audiol.* (2006) 45:400–5. doi: 10.1080/14992020600625007
8. Perez E, Edmonds BA. A systematic review of studies measuring and reporting hearing aid usage in older adults since 1999: a descriptive summary of measurement tools. *PLoS ONE.* (2012) 7:e31831. doi: 10.1371/journal.pone.0031831
9. Mäki-Torkko EM, Sorri MJ, Laukli E. Objective assessment of hearing aid use. *Scand Audiol.* (2001) 30:81–2. doi: 10.1080/010503901300007146
10. Gaffney P. Reported hearing aid use versus datalogging within a VA population. *Hear Rev*. (2008) 15:42–7.

11. Laplante-Lévesque A, Nielsen C, Jensen LD, Naylor G. Patterns of hearing aid usage predict hearing aid use amount (data logged and self-reported) and overreport. *J Am Acad Audiol.* (2014) 25:187–98. doi: 10.3766/jaaa.25.2.7
12. Solheim J, Hickson L. Hearing aid use in the elderly as measured by datalogging and self-report. *Int J Audiol.* (2017) 56:472–9. doi: 10.1080/14992027.2017.1303201
13. Doyle JB, Raghunathan RR, Cellum I, Li G, Golub JS. Longitudinal tracking of sound exposure and hearing aid usage through objective data logs. *Otolaryngol Head Neck Surg.* (2018) 159:110–6. doi: 10.1177/0194599818766056
14. Nelson Kakulla B. *Tech Trends of the 50+*. Washington, DC: AARP Research (2020). doi: 10.26419/res.00329.001
15. Pontoppidan NH, Li X, Bramsløw L, Johansen B, Nielsen C, Hafez A, et al. Data-driven hearing care with time-stamped data-logging. *Proc Int Sympos Audit Audiol Res*. (2018) 6:127–34. Available online at: https://proceedings.isaar.eu/index.php/isaarproc/article/view/2017-15
16. Laplante-Lévesque A, Jensen LD, Dawes P, Nielsen C. Optimal hearing aid use: focus groups with hearing aid clients and audiologists. *Ear Hear.* (2013) 34:193–202. doi: 10.1097/AUD.0b013e31826a8ecd
17. Taubman LB, Palmer CV, Durrant JD, Pratt S. Accuracy of hearing aid use time as reported by experienced hearing aid wearers. *Ear Hear.* (1999) 20:299–305. doi: 10.1097/00003446-199908000-00003
18. Bertoli S, Staehelin K, Zemp E, Schindler C, Bodmer D, Probst R. Survey on hearing aid use and satisfaction in Switzerland and their determinants. *Int J Audiol.* (2009) 48:183–95. doi: 10.1080/14992020802572627
19. Timmer BHB, Hickson L, Launer S. Hearing aid use and mild hearing impairment: learnings from big data. *J Am Acad Audiol.* (2017) 28:731–41. doi: 10.3766/jaaa.16104
20. Wong LLN, Hickson L, McPherson B. Hearing aid satisfaction: what does research from the past 20 years say? *Trends Amplif.* (2003) 7:117–61. doi: 10.1177/108471380300700402
21. Humes LE, Garner CB, Wilson DL, Barlow NN. Hearing-aid outcome measures following one month of hearing aid use by the elderly. *J Speech Lang Hear Res.* (2001) 44:469–86. doi: 10.1044/1092-4388(2001/037)
22. Laplante-Lévesque A, Knudsen LV, Preminger JE, Jones L, Nielsen C, Öberg M, et al. Hearing help-seeking and rehabilitation: perspectives of adults with hearing impairment. *Int J Audiol.* (2012) 51:93–102. doi: 10.3109/14992027.2011.606284
23. Jensen NS, Pedersen AK, Petersen EB, Theill J. When one size does not fit all: special hearing solutions for special occasions. *Hear Rev.* (2019) 26:22–6.
24. Staehelin K, Bertoli S, Probst R, Schindler C, Dratva J, Sutz EZ. Gender and hearing aids: patterns of use and determinants of nonregular use. *Ear Hear.* (2011) 32:e26–37. doi: 10.1097/AUD.0b013e3182291f94
25. Laplante-Lévesque A, Pontoppidan NH, Mazevski A, Schum D, Behrens T, Porsbo M. *Data-Driven Hearing Care With Hearing Fitness^{TM}: Oticon Shares Openly Its Innovative Vision and Roadmap*. Smørum: Oticon A/S (2017).
26. National Videnskabsetisk Komité. *What Should I Report? National Scientific Ethics Committee.* (2020). Available online at: https://www.nvk.dk/forsker/naar-du-anmelder/hvilke-projekter-skal-jeg-anmelde (accessed May 31, 2021).
27. Walker EA, Mccreery RW, Spratford M, Oleson JJ, Buren JV, Bentler R, et al. Trends and predictors of longitudinal hearing aid use for children who are hard of hearing HHS public access. *Ear Hear.* (2015) 36:38–47. doi: 10.1097/AUD.0000000000000208
28. Lloyd S. Least squares quantization in PCM. *IEEE Trans Inf Theory.* (1982) 28:129–37. doi: 10.1109/TIT.1982.1056489
29. Arthur D, Vassilvitskii S. K-means^{++}: the advantages of careful seeding. In: *18th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2007*. New Orleans, LO: Society for Industrial and Applied Mathematics (2007). p. 1027–35.
30. Thorndike RL. Who belongs in the family? *Psychometrika.* (1953) 18:267–76. doi: 10.1007/BF02289263
31. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math.* (1987) 20:53–65. doi: 10.1016/0377-0427(87)90125-7
32. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Prettenhofer P, et al. Scikit-learn: machine learning in python. *J Mach Learn Res.* (2011) 12:2825–30.
33. Rokach L, Maimon O. Clustering methods. In: Maimon O, Rokach L, editors. *Data Mining and Knowledge Discovery Handbook*. Boston, MA: Springer (2005). doi: 10.1007/0-387-25465-X_15
34. Campello RJGB, Moulavi D, Sander J. Density-based clustering based on hierarchical density estimates. In: Pei J, Tseng VS, Cao L, Motoda H, Xu G, editors. *Advances in Knowledge Discovery and Data Mining. PAKDD 2013. Lecture Notes in Computer Science, Vol. 7819*. Berlin; Heidelberg: Springer. doi: 10.1007/978-3-642-37456-2_14
35. Hahsler M, Piekenbrock M. *dbscan: Density Based Clustering of Applications with Noise (DBSCAN) and Related Algorithms*. (2021). Available online at: https://cran.r-project.org/package=dbscan
36. Caliñski T, Harabasz J. A dendrite method for cluster analysis. *Commun Stat*. (1974) 3:1–27. doi: 10.1080/03610917408548446
37. Davies DL, Bouldin DW. A cluster separation measure. IEEE Trans Pattern Anal Mach Intell PAMI (1979). 1:224–7. doi: 10.1109/TPAMI.1979.4766909
38. McInnes L, Healy J, Astels S. hdbscan: Hierarchical density based clustering. *J Open Source Softw.* (2017) 2:205. doi: 10.21105/joss.00205
39. Rodríguez J, Medina-Pérez MA, Gutierrez-Rodríguez, AE, Monroy R, Terashima-Marín H. Cluster validation using an ensemble of supervised classifiers. *Knowl Based Syst*. (2018) 145:134–44. doi: 10.1016/j.knosys.2018.01.010
40. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* New York, NY: ACM (2016) p. 785–94.
41. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys.* (1943) 5:115–33. doi: 10.1007/BF02478259
42. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: an imperative style, high-performance deep learning library. In: Wallach H, Larochelle H, Beygelzimer A, et al., editors. *Advances in Neural Information Processing Systems 32.* Curran Associates, Inc., (2019). p. 8024–35.
43. Bengfort B, Danielsen N, Bilbro R, Gray L, McIntyre K, Richardson G, et al. Yellowbrick v0.6 (0.6). *Zenodo*. (2018). doi: 10.5281/zenodo.1206264
44. Nakano R. reiinakano/scikit-plot: v0.2.1 (v0.2.1). *Zenodo*. (2017). doi: 10.5281/zenodo.293191
45. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Lawrence Erlbaum Associates (1988). doi: 10.4324/9780203771587

Conflict of Interest: AP, T-IS, and KJ are employed by Demant A/S. JC, NP, and KS are employed by Oticon A/S.

The remaining author declares 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.

frontiers in Digital Health

ORIGINAL RESEARCH
published: 13 September 2021
doi: 10.3389/fdgth.2021.724997

Validation of SHOEBOX QuickTest Hearing Loss Screening Tool in Individuals With Cognitive Impairment

Andrew Frank[1*†], Serena Goldlist[1*†], Amy E. Mark Fraser[2] and Matthew Bromwich[2,3,4]

[1] Bruyère Research Institute, Ottawa, ON, Canada, [2] SHOEBOX Ltd, Ottawa, ON, Canada, [3] Department of Otolaryngology-Head and Neck Surgery, CHEO, Ottawa, ON, Canada, [4] Department of Otolaryngology- Head and Neck Surgery, University of Ottawa, Ottawa, ON, Canada

OPEN ACCESS

Edited by:
Dennis L. Barbour,
Washington University in St. Louis,
United States

Reviewed by:
Palash Chandra Banik,
Bangladesh University of Health Sciences, Bangladesh
Hemant Ghayvat,
Linnaeus University, Sweden

***Correspondence:**
Andrew Frank
afrank@bruyère.org
Serena Goldlist
serenagoldlist@gmail.com

† These authors share joint first authorship

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: 14 June 2021
Accepted: 20 August 2021
Published: 13 September 2021

Citation:
Frank A, Goldlist S, Mark Fraser AE and Bromwich M (2021) Validation of SHOEBOX QuickTest Hearing Loss Screening Tool in Individuals With Cognitive Impairment.
Front. Digit. Health 3:724997.
doi: 10.3389/fdgth.2021.724997

Objectives: The aim of this study was to validate a novel iPad-based rapid hearing loss screening tool (SHOEBOX QuickTest) in individuals with cognitive impairment.

Design: Cross-sectional validation study.

Setting: Bruyère Research Institute, Ottawa, Canada.

Subjects and Methods: Twenty-five individuals with mild cognitive impairment (MCI) and mild dementia from the Bruyère Memory Program were included in this study. The study consisted of two components: (1) SHOEBOX QuickTest hearing screener and (2) a conventional hearing test (pure tone audiometry).

Measurements: Hearing was assessed at 1,000, 2,000, and 4,000 Hz separately for each ear. The agreement between hearing ability groupings (good vs. reduced) from conventional hearing test and SHOEBOX QuickTest was determined. Specifically, accuracy, sensitivity, specificity, as well as alignment between conventional thresholds and hearing threshold ranges.

Results: An overall accuracy of 84% was observed for SHOEBOX QuickTest, and a sensitivity and specificity of 100 and 66.7%, respectively. 72% ([95% CI], 60.0–84.1%) of conventional audiometry thresholds were within the pre-established 10 dB SHOEBOX QuickTest.

Conclusion: SHOEBOX QuickTest is a valid hearing loss screening tool for individuals with cognitive impairment. Implementing this iPad-based screening tool in memory clinics could not only aid in the timely diagnosis of hearing loss, but also assist physicians in providing a better assessment of cognitive impairment by ruling out hearing loss as a confounding variable.

Keywords: hearing screening, memory loss, hearing impairment, mild cognitive impairment, aging, mild dementia, audiometry

INTRODUCTION

Hearing loss has been identified as the most prevalent sensory disability in the world impacting approximately 432 million adults worldwide (1, 2). It has been associated with many adverse consequences such as social isolation (3), depression (4, 5), safety issues, decline in independence and reduced quality of life (6, 7). Despite its widespread presence, hearing loss is largely under-recognized and under-treated. It has been estimated that 67–86% of people who experience this disability do not use any form of hearing aid or other assistive technology (8). The World Health Organization (WHO) estimates an annual global cost of US$750 billion due to unaddressed hearing loss (1). This figure is expected to rise with the number of people facing this problem increasing globally. The economic impact is especially dire in countries with aging populations as prevalence of hearing loss increases with age (9). In the United States, it is estimated that two-thirds of people over 70 years old are affected (10). Typically, screening for hearing loss is not included in the battery of tests recommended by physicians for older adults.

In addition to hearing loss, dementia is a one of the major causes of disability among older adults worldwide (11). Dementia is an umbrella term for brain disorders leading to deterioration in cognitive function (11). At any given time, it is estimated that 5–8% of people aged 60 and older are suffering from dementia and 10 million new cases are added each year across the globe (11). The most common cognitive assessments are the Montreal Cognitive Assessment (MoCA), Mini-Mental State Exam (MMSE) and Mini-Cog (12). Currently, screening for hearing loss is not incorporated into these cognitive assessments even though it is estimated that over 60% of adults with cognitive impairment also have a hearing impairment (13). It has been proposed that hearing loss might be a marker for cognitive decline and could be a modifiable risk factor for dementia (14). Thus, ruling out hearing loss could assist physicians in providing a better assessment of cognitive impairment.

The gold standard for assessing hearing loss is pure tone audiometry (PTA) administered by trained audiologists (15). PTA assesses hearing sensitivity by determining hearing thresholds that are required to perceive a tone at least 50% of the time. Hearing thresholds are assessed at different frequencies ranging from 500 to 8,000 Hz and are then plotted on an audiogram to determine if patient's hearing levels are within normal limits (16). The limitation of PTA is that it requires access to specialized medical equipment and staff. However, despite the growing need for audiology professionals in our aging society, work force analyses have indicated that the demand for hearing specialists will outpace available capacity over the next few decades (17, 18). Therefore, there is a growing need to validate a reliable and effective tool to quickly screen older adults (including those who are cognitively impaired) for hearing difficulties to effectively triage them to specialists.

Alternative hearing loss screening methods that do not require specialized health care professionals or expensive medical equipment have been developed recently. These methods are more accessible as they are administered on personal computers (19), tablets (20) or smartphones (21, 22). Although these options provide a potentially more convenient and quicker assessment, issues concerning lack of validation and the effects of environmental noise (i.e., noise limiting) are unaddressed. SHOEBOX Audiometry (SHOEBOX Ltd, Ottawa, Canada) developed an approach to manage background noise levels by utilizing sound-attenuating headphones with their SHOEBOX QuickTest application. However, SHOEBOX QuickTest has yet to be validated in individuals with cognitive impairment.

To date, only one study has assessed hearing loss in cognitively impaired individuals using a screening method not administered by audiology professionals. Pletnikova and colleagues assessed the feasibility of using a tablet-based audiometer in individuals with cognitive impairment (23). Although it could reliably test 59% of the patients, lower cognitive assessment scores (i.e., MMSE) were associated with less reliable results. Furthermore, the study did not compare against a gold standard (i.e., PTA).

Even with the established association between hearing loss and cognitive decline combined with high rates of undiagnosed hearing loss in older adults, no studies have explored the suitability of a rapid hearing loss screening tool to screen for hearing loss in a population of individuals with cognitive impairment. Therefore, the aim of this study was to validate the usefulness of the iPad-based SHOEBOX QuickTest (SHOEBOX Ltd.) hearing screening application in a group of older individuals with cognitive impairment.

MATERIALS AND METHODS

Participants

Twenty-five (25) individuals followed at the Bruyère Memory Program (Bruyère Continuing Care) were recruited into this study. All participants were diagnosed with mild cognitive impairment (MCI) or mild dementia. All patients meeting inclusion criteria were approached by a research staff member who explained the study and if interested, obtained informed consent. The experimental protocol was approved by Bruyère Ethics Review Board and participants were free to withdraw at any point. Demographic information regarding age, gender, and diagnosis, as well as the number of attempts to complete the SHOEBOX QuickTest are displayed in **Table 1**.

Hearing Assessments

The iPad-based SHOEBOX QuickTest application was performed with calibrated sound-attenuating headphones. Testing took place in a quiet office at Élisabeth Bruyère Hospital. The complete test included two main components, a set of four questions followed by a series of pure tone presentations. The tone presentation component was performed separately for each ear (right then left ear) and included frequencies of 1,000, 2,000, and 4,000 Hz. Participants tapped the circle in the middle of the iPad screen (see **Figure 1**) to indicate that they had heard a sound. The presentation level (volume) varied algorithmically depending on the response or lack of response to the previous tone presentation. A starting volume of 70 dB HL (Hearing Level) was used. Tests were completed using the RadioEar DD450 transducers and results were not displayed to the patients. Pure tones (dB HL) were measured at each

TABLE 1 | Characteristics of participants with cognitive impairment.

Characteristic	***N* = 25**
Sex	
Male	14 (56%)
Female	11 (44%)
Age	
50–60	1 (4%)
60–70	4 (16%)
70–80	16 (64%)
>80	4 (16%)
Diagnosis	
Mild dementia	6 (25%)
MCI	19 (76%)
Number of attempts to complete SHOEBOX QuickTest	
1	18 (72%)
2	4 (16%)
3	3 (12%)

Values are reported as number of participants (%).

of these frequencies in each ear and categorized into 10 dB ranges (i.e., 0–10, 10–20, 20–30, 30–40, 40–50, 50–60, 60–70, 70–80). Participants were defined as having hearing loss if ranges obtained at any frequency (i.e., 1,000, 2,000 and 4,000 Hz) exceeded 30 dB HL (i.e., 30–40 dB HL) in any of their ears.

Following completion of the SHOEBOX QuickTest, participants underwent a conventional pure tone audiometry hearing assessment by an audiologist at the Ottawa Audiology Services Clinic, Ottawa, Canada (located at Élisabeth Bruyère Hospital) using circumaural headphones in a sound booth. The audiologist was blinded to the SHOEBOX QuickTest results. The SHOEBOX QuickTest was always conducted first, in order to disadvantage it against any potential sequential learning effect. Assessments included PTA as well as otoscopy performed on both ears. Results from testing at frequencies of 1,000, 2,000, and 4,000 Hz were used to obtain an overall assessment of hearing ability. Similar to SHOEBOX QuickTest, air conduction thresholds (dB HL) were measured at each of these frequencies to determine hearing loss. Participants were defined as having hearing loss from conventional pure tone audiometry if thresholds obtained at any frequency (i.e., 1,000, 2,000, and 4,000 Hz) exceeded 30 dB HL in any of their ears.

Statistical Analysis

Presence of hearing loss was determined by the SHOEBOX QuickTest and conventional audiometry for each participant separately. Reduced hearing ability was defined as >30 dB and was entered in a 2 x 2 contingency table to determine the accuracy of SHOEBOX QuickTest in screening for hearing loss (sensitivity and specificity). Analysis was performed by a at Bruyère Research Institute using Statistical Program for the Social Sciences (SPSS) version 24 (SPSS Inc., Chicago, IL).

The pure tone thresholds obtained from conventional audiometry were compared to the corresponding ranges

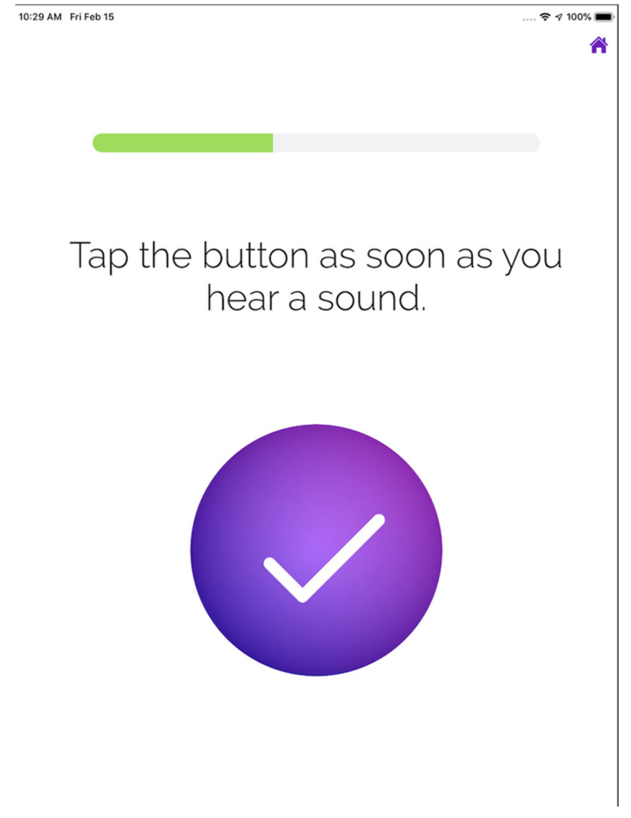


FIGURE 1 | Display on iPad screen for participants to tap.

estimated in SHOEBOX QuickTest at each frequency and each ear. To determine the correlation between the two assessments, the proportion of pure tone thresholds obtained from conventional audiometry that fell within the estimated pure tone ranges from SHOEBOX QuickTest were computed with 95% confidence intervals. We also computed thresholds obtained by conventional audiometry that were within +/– 5 dB the range obtained by SHOEBOX QuickTest.

RESULTS

A total of 25 patients (mean age = 73.84 years) with MCI or mild dementia were tested using SHOEBOX QuickTest and conventional pure tone audiometry. All patients were able to complete the SHOEBOX QuickTest in three attempts without any assistance. Reasons for greater number of attempts required include not responding quickly enough, and not understanding the task on the first attempt (see **Figure 1**). Comparison of the SHOEBOX QuickTest to conventional pure tone audiometry is shown in **Table 2**. The sensitivity and specificity for SHOEBOX Quick Test were 100% (13/13) and 66.7% (8/12) respectively. Of the four patients who had conflicting results between conventional audiometry and the SHOEBOX QuickTest, all were identified as having hearing loss on the SHOEBOX QuickTest

TABLE 2 | Accuracy of SHOEBOX QuickTest in screening for hearing loss (>30 dB HL) in people with cognitive impairment.

	Conventional Audiometry	
SHOEBOX	≤30 dB HL	>30 dB HL
≤30 dB HL	8	0
>30 dB HL	4	13
Total	12	13

but good hearing on conventional audiology exam. The positive predictive value was 76% (CI 59–88%) and the negative predictive value was 100%, with an accuracy of 84%.

The measured pure tone threshold ranges computed using the SHOEBOX QuickTest were compared to the threshold obtained using conventional audiometry to determine correlation between the two assessments. 72% (95% [CI], 60.0%-84.1%) of conventional audiometry thresholds were within the 10 dB range of the SHOEBOX QuickTest. If we expanded the range by +/−5 dB given by SHOEBOX QuickTest, 89.33% (95% [CI], 98.34–80.33) of thresholds obtained by conventional audiometry were included.

DISCUSSION

The primary aim of this study was to validate the usefulness of an iPad-based hearing screener as a screening tool in patients with mild cognitive impairment (MCI) or mild dementia. Overall, the test was accurate with 84% of the tested population having matching results between conventional audiometry and the SHOEBOX QuickTest. SHOEBOX QuickTest was a highly sensitive assessment, correctly identifying every patient who had hearing loss ($n = 13$) as measured by an audiologist (sensitivity 100%). It is important to note that SHOEBOX QuickTest may be overly sensitive, identifying hearing loss in some individuals who do not have hearing loss ($n = 4$). Furthermore, despite the fact that SHOEBOX QuickTest was not intended to give exact thresholds, our results demonstrate that it is still relatively accurate at estimating hearing ranges. Lastly, three of the four patients with unknown cerumen accumulation were correctly identified using SHOEBOX QuickTest. Taking together, although SHOEBOX QuickTest identified some false positives, it can reliably screen for hearing loss in older adults with cognitive impairment.

In our study, all of participants were able to complete the self-administered screening tool within three attempts. This may be because we restricted our participants to those diagnosed with mild-stage cognitive impairment (i.e., mild cognitive impairment and mild dementia) and were likely less impaired than a previous study which included participants with more advanced dementia (23). Future studies should seek to validate the SHOEBOX QuickTest in participants with more advanced cognitive impairment.

There is a plethora of literature detailing the relationship between hearing loss and dementia. A recent article by Griffiths et al. proposed that hearing loss leads to an impoverished sensory environment that decreases stimulation and cognitive processing (24). The impoverished auditory input negatively alters brain structure and function which is a risk factor for the development of dementia. Using hearing aids has been associated with a reduced risk of developing dementia (25). Preliminary results in a recent study have demonstrated significant improvement in cognition associated with hearing aid use in older adults (26). Taken together, these findings suggest that early auditory rehabilitation may prevent cognitive decline. Implementing screening tools such as SHOEBOX QuickTest in memory clinics could be one strategy to increase use of hearing aids in patients with cognitive impairment, which may lessen further cognitive decline.

Despite the association between hearing loss and cognitive decline there is a paucity of literature evaluating the reliability of objective hearing loss screening tools in this population. Our study provides preliminary evidence that rapid, objective hearing screeners such as SHOEBOX QuickTest can reliably be used to screen for hearing loss in individuals with mild cognitive impairment. These types of self-administered objective screeners do not require expertise in hearing testing and provide immediate, actional able results (e.g., individual should be referred for a complete audiological evaluation) reducing some potential barriers to implementation. Further research is needed to explore the implementation of hearing screening within memory clinic programs.

CONCLUSION

SHOEBOX QuickTest is a valid and accurate hearing loss screening tool for individuals with cognitive impairment. Implementing this screening tool in memory clinics can not only aid in a timely diagnosis of hearing loss, but it can also assist physicians in providing a better assessment of cognitive impairment by ruling out hearing loss as a confounding variable.

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 Bruyère Research Institute. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AF designed the methodological approach, collected the data, and supervised the findings and revised the final manuscript. SG performed the data analysis. SG, AMF, MB, and AF wrote the article. All authors contributed to the article and approved the submitted version.

FUNDING

Trillium Health Partners (Trillium) has a vision to develop a pan-Canadian integrated health network (CAN Health) that would facilitate the adoption of Canadian developed technology solutions and break through the historic challenges that health and life sciences companies have faced, enabling them to scale and grow in Canada. Government of Canada funded this project; Ottawa Audiology Services Clinic, Jean Dallaire for conducting the audiology tests. SHOEBOX Inc. CHEO (funds held by MB) covered open access journal costs.

ACKNOWLEDGMENTS

We would like to express our gratitude to:

Funded by the Government of Canada | Financé par le gouvernement du Canada | Canada

- Government of Canada funded this project; Ottawa Audiology Services Clinic, Jean Dallaire for conducting the audiology tests.
- SHOEBOX Ltd. for providing the equipment to conduct the SHOEBOX QuickTest hearing screener, and Vanessa Gauthier-Davidson, In-house Audiologist, for her support.
- Marilyn Reed, Audiology Practice Advisor, Baycrest Health Sciences for her consultation and advice.
- Bruyère Memory Program for working with the research staff on study recruitment.
- Elizabeth Legace for her support with the data analysis.

REFERENCES

1. World Health Organization. *Deafness and hearing loss.* [2021]. Available online at: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss (accessed January 16, 2021).
2. World Health Organization. *World Report on Disability*. Genevia: WHO. (2011).
3. Shukla A, Harper M, Pedersen E, Goman A, Suen JJ, Price C, et al. Hearing loss, loneliness, and social isolation: a systematic review. *Otolaryngol Neck Surg.* (2020) 162:622–33. doi: 10.1177/0194599820910377
4. Mener DJ, Betz J, Genther DJ, Chen D, Lin FR. Hearing loss and depression in older adults. *J Am Geriatr Soc.* (2013) 61:1627–9. doi: 10.1111/jgs.12429
5. Lawrence BJ, Jayakody DMP, Bennett RJ, Eikelboom RH, Gasson N, Friedland PL. Hearing loss and depression in older adults: a systematic review and meta-analysis. *Gerontologist.* (2020) 60:137–54. doi: 10.1093/geront/gnz009
6. Ciorba A, Bianchini C, Pelucchi S, Pastore A. The impact of hearing loss on the quality of life of elderly adults. *Clin Interv Aging.* (2012) 7:159–63. doi: 10.2147/CIA.S26059
7. Cherko M, Hickson L, Bhutta M. Auditory deprivation and health in the elderly. *Maturitas.* (2016) 88:52–7. doi: 10.1016/j.maturitas.2016.03.008
8. National Academies of Sciences, Engineering, and Medicine. Hearing Health Care for Adults: Priorities for Improving Access and Affordability. Washington, DC: The National Academies Press. (2016) p. 1. doi: 10.17226/23446
9. Goman AM, Lin FR. Prevalence of hearing loss by severity in the United States. *Am J Public Health.* (2016) 106:1820–2. doi: 10.2105/AJPH.2016.303299
10. Bainbridge KE, Wallhagen MI. Hearing loss in an aging American population: extent, impact, and management. *Annu Rev Public Health.* (2014) 35:139–52. doi: 10.1146/annurev-publhealth-032013-182510
11. World Health Organization. Dementia (2020). Available online at: https://www.who.int/news-room/fact-sheets/detail/dementia (accessed January 16, 2021).
12. MedlinePlus Medical Test. Cognitive Testing (2020). Available online at: https://medlineplus.gov/lab-tests/cognitive-testing (accessed January 27, 2021).
13. Nirmalasari O, Mamo SK, Nieman CL, Simpson A, Zimmerman J, Nowrangi MA, et al. Age-related hearing loss in older adults with cognitive impairment. *Int Psychogeriatr.* (2017) 29:115–21. doi: 10.1017/S1041610216001459
14. Lin FR. Hearing Loss and Cognition Among Older Adults in the United States. *J Gerontol A Biol Sci Med Sci.* (2011) 66A:1131–6. doi: 10.1093/gerona/glr115
15. Sliwinska-kowalska M. "Chapter 19 - Hearing". In: Lotti M, Bleecker ML, editors. *Handbook of Clinical Neurology*. Elsevier. (2015). p. 341–63. doi: 10.1016/B978-0-444-62627-1.00018-4
16. Mediscape. *Audiology Pure-Tone Testing.* (2018). Available online at: https://emedicine.medscape.com/article/1822962-overview (accessed January 27, 2021).
17. Windmill IM, Freeman BA. Demand for audiology services: 30-yr projections and impact on academic programs. *J Am Acad Audiol.* (2013) 24:407–16. doi: 10.3766/jaaa.24.5.7
18. Goulios H, Patuzzi RB. Audiology education and practice from an international perspective. *Int J Audiol.* [2008] 47:647–64. doi: 10.1080/14992020802203322
19. Folmer RL, Vachhani J, McMillan GP, Watson C, Kidd GR, Feeney MP. Validation of a computer-administered version of the digits-in-noise test for hearing screening in the United States. *J Am Acad Audiol.* (2017) 28:161–9. doi: 10.3766/jaaa.16038
20. Mosley CL, Langley LM, Davis A, McMahon CM, Tremblay KL. Reliability of the home hearing test: implications for public health. *J Am Acad Audiol.* (2019) 30:208–16. doi: 10.3766/jaaa.17092
21. Bright T, Pallawela D. Validated smartphone-based apps for ear and hearing assessments: a review. *JMIR Rehabil Assist Technol.* (2016) 3:13. doi: 10.2196/rehab.6074
22. Masalski M, Grysiński T, Krecicki T. Hearing tests based on biologically calibrated mobile devices: comparison with pure-tone audiometry. *JMIR MHealth UHealth.* (2018) 6:7800. doi: 10.2196/mhealth.7800
23. Alexandra P, Reed Nicholas S, Halima A, Sevil Y, Milap N, Joshua B, et al. Identification of hearing loss in individuals with cognitive impairment using portable tablet audiometer. *Perspect ASHA Spec Interest Groups.* (2019) 4:947–53. doi: 10.1044/2019_PERS-SIG8-2018-0018
24. Griffiths TD, Lad M, Kumar S, Holmes E, McMurray B, Maguire EA, et al. How can hearing loss cause dementia? *Neuron.* (2020) 108:401–12. doi: 10.1016/j.neuron.2020.08.003
25. Mahmoudi E, Basu T, Langa K, McKee MM, Zazove P, Alexander N, et al. Can hearing aids delay time to diagnosis of dementia, depression, or

falls in older adults? *J Am Geriatr Soc.* [2019] 67:2362–9. doi: 10.1111/jgs.16109

26. Sarant J, Harris D, Busby P, Maruff P, Schembri A, Lemke U, et al. The effect of hearing aid use on cognition in older adults: can we delay decline or even improve cognitive function? *J Clin Med.* (2020) 9:254. doi: 10.3390/jcm9010254

Conflict of Interest: MB is the Chief Medical Officer at SHOEBOX Ltd. AMF is the Manager, Research and Data Analytics at SHOEBOX Ltd.

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.

frontiers in Digital Health

ORIGINAL RESEARCH
published: 17 September 2021
doi: 10.3389/fdgth.2021.724714

Check for updates

Activity Tracking Using Ear-Level Accelerometers

Martin A. Skoglund[1,2], Giovanni Balzi[3], Emil Lindegaard Jensen[3], Tanveer A. Bhuiyan[4] and Sergi Rotger-Griful[2]*

[1] Division of Automatic Control, Department of Electrical Engineering, The Institute of Technology, Linköping University, Linkoping, Sweden, [2] Eriksholm Research Centre, Oticon A/S, Snekkersten, Denmark, [3] Department of Electrical Engineering, Technical University of Denmark, Ørsteds Plads, Lyngby, Denmark, [4] Oticon A/S, Smorum, Denmark

Introduction: By means of adding more sensor technology, modern hearing aids (HAs) strive to become better, more personalized, and self-adaptive devices that can handle environmental changes and cope with the day-to-day fitness of the users. The latest HA technology available in the market already combines sound analysis with motion activity classification based on accelerometers to adjust settings. While there is a lot of research in activity tracking using accelerometers in sports applications and consumer electronics, there is not yet much in hearing research.

Objective: This study investigates the feasibility of activity tracking with ear-level accelerometers and how it compares to waist-mounted accelerometers, which is a more common measurement location.

Method: The activity classification methods in this study are based on supervised learning. The experimental set up consisted of 21 subjects, equipped with two XSens MTw Awinda at ear-level and one at waist-level, performing nine different activities.

Results: The highest accuracy on our experimental data as obtained with the combination of Bagging and Classification tree techniques. The total accuracy over all activities and users was 84% (ear-level), 90% (waist-level), and 91% (ear-level + waist-level). Most prominently, the classes, namely, standing, jogging, laying (on one side), laying (face-down), and walking all have an accuracy of above 90%. Furthermore, estimated ear-level step-detection accuracy was 95% in walking and 90% in jogging.

Conclusion: It is demonstrated that several activities can be classified, using ear-level accelerometers, with an accuracy that is on par with waist-level. It is indicated that step-detection accuracy is comparable to a high-performance wrist device. These findings are encouraging for the development of activity applications in hearing healthcare.

Keywords: activity tracking, accelerometer, classification, machine learning, supervised learning, hearing aids, hearing healthcare

OPEN ACCESS

Edited by:
Jing Chen,
Peking University, China

Reviewed by:
Parisis Gallos,
National and Kapodistrian University of Athens, Greece
Ivan Miguel Pires,
Universidade da Beira Interior, Portugal

****Correspondence:***
Martin A. Skoglund
martin.skoglund@liu.se

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: *14 June 2021*
Accepted: *17 August 2021*
Published: *17 September 2021*

Citation:
Skoglund MA, Balzi G, Jensen EL, Bhuiyan TA and Rotger-Griful S (2021) Activity Tracking Using Ear-Level Accelerometers. Front. Digit. Health 3:724714. doi: 10.3389/fdgth.2021.724714

1. INTRODUCTION

A strong trend in modern hearing aid (HA) development and research is the inclusion of more sensing technologies. The driver behind this is the wish for better, more personalized, and self-adaptive (1–3) devices that can handle environmental changes (4–7) and cope with day-to-day fitness of the users. Current HAs usually try to analyze the soundscape and adjust the settings according to a formula. However, recent HAs have advanced further by combining sound analysis with motion activity classification based on accelerometers to adjust settings with the aim of a better user experience. A few other possible uses of accelerometers in HAs are as follows: fall detection (8) to alert caretakers; tap detection for user interfacing (9); and health monitoring based on physical activity (10). The backbone of the above-mentioned applications is accurate and robust activity tracking that can determine and distinguish between several relevant activities, e.g., standing, sitting, walking, running, and more. While a lot of research in activity tracking and classification using accelerometers has been in sports applications (11–14) and general consumer electronics (15–17), such as smart-watches and cell phones, the hearing research body is small. The results in this contribution is based on the work in Balzi (18). The background to this study is that accelerometers are, or are to appear, in hearing devices and that it is of fundamental interest to investigate their usefulness in the activity tracking. The key objective of this study is to investigate the feasibility of activity tracking with ear-level accelerometers and how it compares to waist-mounted accelerometers, which is the more common measurement location in sports and healthcare. The activity classification method is based on supervised learning on experimental data from 21 subjects. The scope of the investigation is limited to 21, normal hearing, healthy subjects, and 9 activities.

2. METHOD

This section outlines the relevant details of the activity tracking methodology based on accelerometer data and machine learning.

2.1. Accelerometer Measurements

It is assumed that the sensors are mounted rigidly onto the users and that any kind of mounting play is negligible. It is further assumed that sensor axes are orthogonal, that the sensitivities are known and linear in the working span, and that sensor biases are negligible. The assumed inertial (fixed) coordinate frame, with axes (XYZ), is a local, right-handed, Euclidean frame with the Z-axis parallel to the local gravity vector. The data from a tri-axial accelerometer are then

$$a = R(a^i - g) + e, \tag{1}$$

where $a = [a_x, a_y, a_z]^T$ is the body referenced measurement for the sensor axes (xyz), R is a rotation matrix relating the orientation of the inertial frame and the body frame, $a^i = [a_X, a_Y, a_Z]^T$ is the acceleration in an inertial frame, the local gravity vector $g \approx [0, 0, 9.81]^T \mathrm{m/s}^2$ is assumed constant, and the noise, e, is assumed Gaussian distributed with the same standard deviation (SD), σ_e, in each axis, $e \sim \mathcal{N}(0, \sigma_e I)$. The measured forces can be divided in to static forces, such as, constant acceleration and gravity, and the dynamic forces that are due to motion of changing rate, e.g., nodding and shaking. Note that even in the ideal case without noise, it is not possible to solve (Equation 1) for a^i and other data, e.g., a magnetometer or a high-grade gyroscope is needed to resolve the rotation R, see Titterton et al. (19) for details. In most situations, the human body accelerations are small compared to gravity, and it is, therefore, possible to estimate the inclination, i.e., each sensor axis angle with respect to the gravity vector, which is related to the orientation in roll and pitch.

2.2. Accelerometer Features

In machine learning, feature extraction is pre-processing of data with the intention of increasing the overall performance of classifiers. The underlying idea is that certain transformations can yield more information, higher independence, and give larger margins for class separability. Feature selection is very much application dependent and usually require domain knowledge, though computationally expensive automated methods exist, [see, e.g., (20)] for an overview. The selected features described below are inspired by the work of Masse et al. (21), Gjoreski et al. (22), and Hua et al. (23) and have been adapted with this application in mind. The features are defined from 13 metrics of which 10 are applied to each axis, resulting in a total of 33 features as described below.

2.2.1. Tilt Angles

The tilt angles, sometimes referred to as inclination, are defined as

$$\phi_k = \arccos(\frac{a_k}{r}), \quad k = \{x, y, z\}, \tag{2}$$

where

$$r = \sqrt{a_x^2 + a_y^2 + a_z^2} \tag{3}$$

and is indicative of each axis angle with respect to the local gravity vector. Errors in the tilt angles arise from the presence of motion and noise.

2.2.2. Acceleration Vector Change

The acceleration vector change (AVC) is a motion-sensitive metric defined by the mean of the absolute value of the differences in the acceleration vector length (Equation 3), and the mean is calculated in a window with size M as

$$AVC = \frac{1}{M} \sum_{i=1}^{M} \frac{|r_{i+1} - r_i|}{T_s} \tag{4}$$

where i denotes samples the i^{th} sample at a given, fixed, sampling frequency f_s with the sampling interval $T_s = 1/f_s$.

2.2.3. Signal Magnitude Area

The signal magnitude area (SMA) is defined over a window with size M as

$$SMA = \frac{1}{M} \sum_{i=1}^{M} |a_{x_i}| + |a_{y_i}| + |a_{z_i}| \tag{5}$$

and it is a measure of the magnitude.

2.2.4. Mean and SD

The mean (Equation 6) and SD (Equation 7) are computed for each axis over a window with size M as

$$\mu_k = \frac{1}{M}\sum_{i=1}^{M} a_{k_i}, \quad k = \{x, y, z\}, \tag{6}$$

$$\sigma_k = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(a_{k_i} - \mu_k)^2}, \quad k = \{x, y, z\}. \tag{7}$$

2.2.5. Root Mean Square

Similar to the SD (Equation 7), the root mean square (RMS) is computed for each axis over a window of size M as

$$RMS_k = \sqrt{\frac{1}{M}\sum_{i=1}^{M} a_{k_i}^2}, \quad k = \{x, y, z\}. \tag{8}$$

2.2.6. Minimum and Maximum

Minimum (MIN) and maximum (MAX) values per axis over a window with size M are defined by

$$MIN_k = \min\{a_{k_i}\}_{i=1}^{M}, \quad k = \{x, y, z\} \tag{9}$$

and

$$MAX_k = \max\{a_{k_i}\}_{i=1}^{M}, \quad k = \{x, y, z\} \tag{10}$$

respectively.

2.2.7. Median

The median is the center value of a size-ordered sample, and it is not skewed by large or small values as the mean is. The median, can e.g., be used to detect burst noise outliers in data and is computed for each axis over a window of size M as

$$MEDIAN_k = \text{median}(\{a_{k_i}\}_{i=1}^{M}) \quad k = \{x, y, z\}. \tag{11}$$

2.2.8. Median Absolute Deviation

The median absolute deviation (MAD) is a measure of sample variability around the median. The MAD is computed for each axis over a window of size M as

$$MAD_k = \text{median}(\{a_{k_i} - MEDIAN_k\}_{i=1}^{M}), \quad k = \{x, y, z\} \tag{12}$$

where the $MEDIAN$ from Equation (11) is used.

2.2.9. Skewness

The skewness (SKW) is the third standardized moment of a sample and is a measure of the asymmetry of a distribution about the mean. Using the previously defined μ_k (Equation 6) and σ_k (Equation 7), the sample SKW is computed for each axis over a window of size M as

$$SKW_k = \frac{\frac{1}{M}\sum_{i=1}^{M}(a_{k_i} - \mu_k)^3}{\sigma_k^3}, \quad k = \{x, y, z\}. \tag{13}$$

Note that other approximations for sample skewness are possible (24).

2.2.10. Counts per Second

Counts per second (counts/s) is a widely used measure in the activity tracking. The computation of counts/s is proprietary of ActiGraph LLC and is usually carried out with ActiGraph accelerometer devices. However, Crouter et al. (25) details on how to derive this measure on a standard accelerometer and for the work here our Matlab implementation is based on Brønd et al. (26).

2.3. Machine Learning

The activity classification is based on a supervised learning to train the classifier. Three classification methods are considered in this study, K-nearest-neighbor (KNN), linear-discriminant-analysis (LDA), and decision tree (DT). Further improvement of the classification can be obtained using ensemble learning methods such as Boosting and Bootstrap aggregation (Bagging), and variations thereof.

2.3.1. Classifiers

In supervised learning, a set of training instances with corresponding class labels is given, and a classifier is trained and used to predict the class of an unseen instance, [see, e.g., (27)] for details. The N samples of training data x and class labels y were ordered in pairs $\{(x_1, y_1), \ldots, (x_N, y_N)\}$ such that the i-th feature vector $x_i \in \mathbb{R}^p$ corresponds to the binary class label vector $y_i \in \mathbb{Z}_2^c$. For the case with a single tri-axial accelerometer and the features described in section 2.2, the feature vector dimension, p, is 33 per sample while the class label vector dimension c is 9.

The three well-known, but rather different, supervised classifiers are considered, and the choice to use these was based on the availability of good implementations. The classifiers are as follows:

- The KNN classifier (28, 29) is here based on the Euclidean distance between the test- and training samples. However, other distance measures can be used.
- The LDA, or Fisher's discriminant (30), is a statistical method to find linear combinations in the feature space to separate the classes, and it is carried out by solving a generalized eigenvalue problem.
- DT has a flow chart-like structure where the root corresponds to feature inputs, the branches of the descending test-nodes represents the outcome of the test, and each leaf-node represents a class label, [see, e.g., (31)].

2.3.2. Ensemble Training

Ensemble training is used to increase the predictive classification (or regression) performance by learning a combination of several classifiers. The two main categories used here are Bagging (31) and Boosting (32) with a few selected variations. In Bagging, the classifiers are trained in parallel on randomly sampled training data while in Boosting the training is carried out sequentially as the classifiers and data are weighted according to their importance. The first, and most well-known, Boosting algorithm is called AdaBoost (short for Adaptive Boosting) and was originally formulated in Freund and Schapire (33).

3. EXPERIMENT

Experimental data were collected from 21 voluntary subjects performing a series of tasks representative of the stipulated activities. A total of three tri-axial accelerometers were used. Experimental data were collected jointly in the projects (18, 34).

3.1. Subjects

Prior to the experiments, the subjects were informed about the experiment procedure and the use of data before deciding on their participation with oral consent. Data were stored and labeled anonymously. The 21 subjects had an age range between 24 and 60 years old, $\mathcal{N}(32.6, 9.6)$, both women and men with a height bracket of $1.55 - 195$ m, $\mathcal{N}(1.78, 0.091)$. None of the subjects had any reported health issues. The subjects did not receive any compensation.

3.2. Data

For all subjects, two of the accelerometers were placed on each side of the head at ear-level, see **Figure 1**, to mimic HA sensors and the third accelerometer was placed at waist-level using an elastic strap, see **Figure 2**, as it is a more common region for activity measurements and it is also representative of an in-pocket smartphone.

The accelerometers are XSens MTw Awinda produced by XSens Technologies B.V. and are battery powered, wireless devices, inertial sensors containing tri-axial accelerometers and gyroscopes and also tri-axial magnetometers enabling accurate orientation estimation when the device is stationary. The data are collected wirelessly using the MT Manager software by XSens on a PC laptop running Microsoft Windows 10 at a sampling rate, f_s, of 100 Hz that was decided to be fast enough for the intended activities. Data are manually labeled based on visual inspection during the experiment to match the activities in section 3.4.

FIGURE 1 | Ear-level accelerometer placement used in the experiment.

FIGURE 2 | Waist-level accelerometer placement used in the experiment.

3.3. Task

The experiment was carried out in a room with a soft carpet at Oticon main offices, Smørum, Denmark. For tasks involving lying down and falling, a mattress was used. Each of the 21 subjects from whom the data have been gathered was asked to perform 6 different tasks while wearing all three accelerometers, and between each task all the data from the accelerometers were saved and anonymously cataloged. Except for the accelerometer data, from each subject, only gender, age, and height were collected. To every subject, the same specific information about which actions to be carried out was given by reading out loud from a manuscript, and no restrictions were communicated regarding how to carry out the various exercises with the intention of increasing the possibility of movement variability in the activities. With more in-class variation used for training, the classifier is less prone to over-fitting at the expense of higher probability of between-class overlap. The mean test duration was about 22 min, including pauses, and generated about 13–14 min of data per subject.

3.4. Physical Activities

The choice of activities to track is a trade-off between how clearly activities can be discerned from each other, the likelihood of activities being present in the daily routines of the subjects, and the intended use of activity tracking. A typical scenario, not addressed here, is that HA users often remove the HAs when lying down to rest. Hence, for a HA applications, in-ear detection, e.g., using accelerometers could be useful. The fidelity of activity categories is chosen as either resting or moving, and no intensity or within-class variation is considered.

The resting activities are as follows:

Act1 :**Standing** in a still position.
Act2 :**Sitting**
Act3 :**Lying face-up (LFU)**
Act4 :**Lying face-down (LFD)**
Act5 :**Lying side (LS)** on either left or right side

and the moving activities are as follows:

Act6 :**Walking, on the floor**
Act7 :**Jogging, moderate pace in circles/square**
Act8 :**Falling** on whichever side, some subjects could not simulate a perfect falling motion and, therefore, have been asked to perform a fast transition from standing to Lying down, at top of their capability.
Act9 :**Transitioning (TRN)** all instances not being any of the other activities, e.g., going from one activity to another.

Note that there is no specific class for head motion as it was predicted being difficult to correctly label and that there is already significant head motion within all the moving activities sections 3.4 to 3.4.With e.g., a waist-level accelerometer, it may be possible to separate head motions from general body motions, but it is beyond this study.

4. RESULTS

As an initial step, 200 min of the dataset in Anguita et al. (35) was used on all combinations of classfiers and ensemble training methods described in Section 2.3. The data are open, pre-labeled, in-pocket cell phone, and it has six activity classes: walking; walking stairs up; walking stairs down; sitting; standing; and laying. For the use here, all the walking classes were considered the same. The best predictive classification accuracy was obtained using the DT and Bagging, and it was furthermore also the case for our experimental data. Consequently, all following results are obtained using DT and Bagging.

4.1. Pre-processing

The data from the 21 subjects were randomly partitioned in to two groups with all activities present in both groups and with 70% used for training and 30% used for testing (validation). Training with cross-validation is carried out for as many times as there are trees, i.e., 100–500 depending on the setup. The tree depth is 480 with 13,121 nodes. Features are computed for each sample at 100 Hz with the window size to one sample, $M = 1$ for the features AVC and SMA, while $M = f_s = 100$, centered at the current sample, for the other applicable features. All results are obtained using the Machine Learning Toolbox in Matlab 2021b and with dependencies to the Optimization Toolbox for certain classifiers.

4.2. Classification

The main performance target here is the accuracy of predicted class labels in data not used for training as it is a common measure in supervised learning, [see, e.g., (11, 12, 14, 36, 37)]. The accuracy is defined as the number of correctly classified labels divided by the total number of labels and simply states how much of the data not used for training that is correctly classified. In **Table 1**, the classification result using both ear-level accelerometers is illustrated based on 500 Decision trees and Bagging trained with a learning rate of 0.1 and showing an overall predicted accuracy of 84.4%. Most prominently, the classes, namely, standing 3.4, jogging 3.4, laying side 3.4, lying face-down 3.4, and, walking 3.4, all have an accuracy of above 90%. The lowest scoring activities are as follows: falling 3.4, sitting 3.4, and transitioning 3.4. The falling activity is often confused with transitioning, which, in turn, generally is confounded with all other activities. The overall accuracy is more than 90% without the sitting activity.

4.3. Feature Evaluation

From a computational perspective, it is good to minimize the number of features needed and, therefore, the relative importance of features for each activity is analyzed using 100 Decision trees and Bagging, with a learning rate of 0.1. In **Table 2**, the total accuracy per feature (or pairs in some cases) for each activity is considered where accelerations only on the top row are considered that base model and the contribution of each additional feature and activity are below. Notably, the tilt angles have only a marginal positive effect on the sitting activity and mostly negative effect on all other activities. Other features with little importance are RMS and counts/s. In **Table 3**, the increase (or regression) per feature and activity, compared to the base level accelertions only, is shown. The overall most important features are: mean, SD, MIN, MAX, median, and MAD. Furthermore, in **Table 4** a summary of the best and worst activity per feature is shown and, as expected, the best ranking features are important for several activities.

4.4. Sensor Combinations

One of the main motivations of this study is to compare the feasibility of ear-level activity tracking compared to waist-level

TABLE 1 | Confusion matrix of the predicted accuracy with Bagging and Decision Tree using both ear-level accelerometers.

	LFD	LFU	Falling	Jogging	LS	Sitting	Standing	TRN	Walking	C	NC
LFD	6,294				1,221			93		82.73	17.27
LFU	1,108	22,754			55			195		94.37	5.63
Falling		87	364		3		16	323	27	44.39	55.61
Jogging			266	43,928			231	932	1,807	93,14	6.86
LS					7,311			26		99.65	0.35
Sitting						2,347	33,213	242	13	6.55	93.45
Standing	8					2,108	1,42,889	1,804	1,496	96.35	3.65
TRN	357	1,082	642	107	116	10	8,135	27,138	4,399	64.64	35.36
Walking			43	28		2	3,505	11,865	1,58,433	91.12	8.88
C	81.04	95.11	27.68	99.69	83.98	52.54	76.01	63.68	95.34		
NC	18.96	4.89	72.32	0.31	16.02	47.46	23.99	36.32	4.66		

The row-normalized row summary on the right displays the percentages of correctly and incorrectly classified observations for each true class, while the column-normalized column summary below the matrix displays the percentages of correctly (C) and incorrectly (NC) classified observations for each predicted class.

TABLE 2 | Total accuracy per feature (or pairs of features) per activity compared to just using accelerations (ACC) only.

	OA	LFD	LFU	Falling	Jogging	LS	Sitting	Standing	TRN	Walking
ACC only	72.17	93.70	85.40	13.70	58.70	71.70	5.80	91.90	39.40	78.10
Tilt Angles	71.67	81.80	79.50	14.00	58.80	69.10	6.30	91.80	39.50	78.10
AVC	75.50	95.00	94.30	16.10	70.30	76.00	5.70	93.10	40.20	81.60
SMA	75.47	95.20	94.20	17.20	70.20	74.90	5.90	93.00	40.20	81.60
Mean + SD	82.71	98.80	94.40	36.00	90.00	99.70	10.70	94.10	60.30	88.40
RMS	72.53	93.70	83.40	14.40	58.70	97.20	5.80	92.00	39.50	78.10
MAX + MIN	82.50	90.30	94.40	33.00	89.50	99.20	10.50	94.10	57.90	89.00
Median + MAD	81.76	99.10	94.60	28.90	89.90	98.80	10.90	94.40	54.00	87.10
Skewness	76.17	95.00	88.70	13.30	52.10	87.30	4.60	95.80	47.40	84.90
Counts/Sec	71.73	78.60	80.80	14.40	58.70	74.30	5.80	91.90	39.50	78.00

The overall accuracy (OA) in the left-most column is the effect when adding each feature.

TABLE 3 | Accuracy improvement per feature and activity.

	OA	LFD	LFU	Falling	Jogging	LS	Sitting	Standing	TRN	Walking
ACC only	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Tilt Angles	−0.50	−11.90	−5.90	0.30	0.10	−2.60	0.50	−0.10	0.10	0.00
AVC	3.33	1.30	8.90	2.40	11.60	4.30	−0.10	1.20	0.80	3.50
SMA	3.30	1.50	8.80	3.50	11.50	3,20	0.10	1.10	0.80	3.50
Mean + SD	10.54	5.10	9.00	22.30	31.30	28.00	4.90	2.20	20.90	10.30
RMS	0.36	0.00	−2.00	0.70	0.00	25.50	0.00	0.10	0.10	0.00
MAX + MIN	10.33	−3.40	9.00	19.30	30.80	27.50	4.70	2.20	18.50	10.90
Median + MAD	9.59	5.40	9.20	15.20	31.20	27.10	5.10	2.50	14.60	9.00
Skewness	4.00	1.30	3.30	−0.40	−6.60	15.60	−1.20	3.90	8.00	6.80
Counts/s	−0.44	−15.10	−4.60	0.70	0.00	2.60	0.00	0.00	0.10	−0.10

*Same setup as in **Table 2** but corrected for baseline accuracy obtained by accelerations only.*

TABLE 4 | Best- and worst-case per feature accuracy using the features listed in the left column based on **Table 3**.

Feature	Best activity	Worst activity
ACC only	X	X
Tilt Angles	X	Lying face-down
AVC	Jogging, Lying face-up	X
SMA	Jogging, Lying face-up	X
Mean + SD	Jogging, Lying face-up, Falling, Transitioning	X
RMS	Lying side	Lying face-up
MAX + MIN	Jogging, Transitioning, Falling, Lying side	Lying face-down
Median + MAD	Jogging, Lying side, Transitioning, Falling	X
Skewness	Lying side	Jogging
Counts/s	Lying side	Lying face-down

Cells marked with X means that there is no improvement or regression in using that feature.

sensoring. For all sensor combinations and data types the same features and training were computed to get the comparable classification results. In **Table 5**, a confusion matrix shows the results of using only the waist-level accelerometer, giving an overall accuracy of 89.6% In **Table 6**, a confusion matrix showing the result from two ear-level accelerometers and the waist-level accelerometer is used together with 500 Decision trees, giving an overall accuracy of 91.6%. One of the reasons for the improvement is that the addition of the waist-level accelerometer makes the sitting activity easier to distinguish with 94.4% correct compared to 52.5% when using only ear-level accelerometers and this performance increase also shows on the standing activity as the previously discussed confusion is decreased. In **Table 7**, the overall accuracy using the various combinations of sensors is shown; for instance, 25 Hz accelerometer data from a wrist-worn Garmin Vivosmart 4 are used with only a 56.3% accuracy. In **Table 7**, 100 Decision trees compared to the previous 500 were chosen to save computations and the accuracy decrease is negligible. The gyroscope and orientation data are obtained from the XSens devices. Note that the orientation data are adapted for stationary orientations and, therefore, of low-pass characteristics and potentially not well suited for all aspects of this application. It can be noted that the waist-level accelerometer alone is rather efficient and, as noted before, the combination with the ear-level accelerometers gives even better performance.

TABLE 5 | Confusion matrix using Bagging and Decision Tree using only the waist accelerometer.

	LFD	LFU	Falling	Jogging	LS	Sitting	Standing	TRN	Walking	C	NC
LFD	6,993							615		91.92	8.08
LFU	1,181	21,752						1,206		90.11	9.89
Falling	18	66	413		22		26	252	23	50.37	49.63
Jogging			23	44,008			360	306	2,467	93.31	6.69
LS	1,471				5664			202		77.20	22.80
Sitting						24,032	9,027	2,744	12	67.10	32.90
Standing			6	5		6,756	1,38,157	1,833	1,548	93.16	6.84
TRN	543	1,065	579	98	64	937	6,211	28,777	3,802	68.39	31.61
Walking			17	168		10	3,696	2,690	1,67,295	96.22	3.78
C	68.52	95.06	39.79	99.39	98.50	75.73	87.73	74.50	95.52		
NC	31.48	4.94	60.21	0.61	1.50	24.27	12.27	25.50	4.48		

The row-normalized row summary on the right displays the percentages of correctly and incorrectly classified observations for each true class, while the column-normalized column summary below the matrix displays the percentages of correctly (C) and incorrectly (NC) classified observations for each predicted class.

TABLE 6 | Confusion matrix using Bagging and Decision Tree using both ear-level accelerometers, and the waist accelerometer.

	LFD	LFU	Falling	Jogging	LS	Sitting	Standing	TRN	Walking	C	NC
LFD	6,274				1,221			113		82.47	17.53
LFU	1,156	22,769			10			177		94.43	5.57
Falling		86	374		8		20	328	4	45.61	54.39
Jogging			91	44,120			245	632	2,076	93.55	6.45
LS					7,324			13		99.82	0.18
Sitting						24,202	10,175	1,436	2	67.58	32.42
Standing				6		844	1,44,503	1,369	1,556	97.45	2.55
TRN	254	1,010	533	78	112	591	5,947	30,698	2,493	73.59	26.41
Walking			33	47			3,339	4,563	1,65,894	95.41	4.59
C	81.65	95.41	36.28	99.70	84.43	94.40	87.99	78.05	96.44		
NC	18.35	4.59	63.72	0.30	15.57	5.60	12.01	21.95	3.56		

The row-normalized row summary on the right displays the percentages of correctly and incorrectly classified observations for each true class, while the column-normalized column summary below the matrix displays the percentages of correctly (C) and incorrectly (NC) classified observations for each predicted class.

TABLE 7 | Accuracy with different sensor combinations using Bagging and 100 Decision trees.

L	L & R	L, R & W	W	L R ACC only	GYR	L, R & GYR	ORI	L, R, & ORI	L, R, & Garmin	Garmin
83,96	84,34	91,57	89,62	74,48	70,93	84,99	47,18	84,76	81,56	56,25

Left (L), Right (R), and Waist (W) are short for the respective accelerometer placements. GYR is short for Gyroscope, and ORI is short for Orientation.

4.5. Step Detection

Another concrete measure that can be useful for activity tracking is step detection, which was also analyzed in Acker (10). For the walking and jogging activities, ear-level step detection was computed based on Bai et al. (38) and Abadleh etal. (39) using the AVC feature resulting in 95% and a 90% accuracy, respectively, using both ear-level accelerometers. This can be compared with the highly optimized Garmin Forerunner 35 giving a 99% (walking) and 95% (jogging) accuracy, respectively.

5. DISCUSSION

The main objective of this study is to compare ear- and waist-level activity tracking performance. Therefore, it is not fundamental to have features and classifiers that could outperform the works of others and such a comparison is beyond this study. While not directly comparing to other methods, the overall ear-level activity classification results are encouraging in the proposed setup.

As noted, it is difficult to separate falling and transitioning with ear-level data only. Possible explanations are that the selected features are not sensitive enough to distinguish between falling and transitioning and that transitioning is too general [or complex as Dernbach et al. (11)] and can, for instance, be confounded with general head movements. More controlled falling experiments, such as, Burwinkel and Xu (8) could provide useful insights. The sitting activity was confounded with the standing activity as they are typically rather similar, and the only potential differences at ear-level between the two may be in postural sway that should be clearer for standing subjects. This difficulty was also found in Parkka et al. (40), and an accelerometer below the waist is particularly useful here. At waist-level it is easier to distinguish sitting and standing and it is possibly explained by the change in the tilt angles for seated subjects.

Designing features sensitive to particular classes is an engineering task requiring expertise. As noted in section 4.3, some features are not that well-suited for any of the activities and would benefit from further tuning, such as, other pre-processing and different window sizes, or should otherwise be omitted. A feature that is sensitive to postural sway (low-frequency component) could potentially support separating standing and sitting at ear-level.

Sensor combinations can improve the results, [see, e.g., (14, 40)], and here the combination of ear- and waist-level data is the overall best. On another positive note, the accuracy difference between one and two ear-level devices is small and this is a good news for HA applications, as single-sided hearing compensation is common. The wrist data, here from the Garmin device, may be difficult in general as arms may do many types of motions not specifically relating to the activities.

The use of gyroscope data is common, [see, e.g., (41–43)], and was expected to improve the results in general. However, all sensor types were processed in the same fashion as accelerometer data with the same features and are a possible explanation of the poor performance of many of the additional data types in **Table 7**.

Wrist data were poor for activity tracking in the setup here, but the output of the proprietary algorithms on these types of devices suggests that a lot more can be achieved on ear-level devices too. The step detection is almost on par with the commercial Garmin device, for the short durations considered here, and these typically utilize additional sensors, e.g., magnetometer and gyroscope, and their algorithms can be considered state-of-the-art.

6. CONCLUSIONS

We investigated the feasibility of ear-level accelerometers for activity tracking in comparison to waist-level accelerometers. Many activities can be classified with an accuracy that is on par with a waist-level accelerometer, and this is particularly encouraging for the development of activity applications in hearing healthcare. Furthermore, we indicate that step-detection accuracy is comparable to a high-performance wrist device. It is also shown that higher predictive performance can be obtained when combining ear- and waist-level accelerometer data, and this could potentially assist in isolating head motion from full body activities, opening for a higher granularity of activity classes.

Noteworthy limitations in this study were as follows: the modest number of test subjects (21); the number of activities (9) per test subject; that the manual data labeling may have errors; and efforts spent on feature design and learning methods. Also, data with a more control and a clearer reference, e.g., a motion capture system, could provide valuable insight at the expense of more costly and complex experiments.

Future directions should consider further feature design with, e.g., multi-tapers and various transforms. The design should start with time-frequency analysis of activities for guidance. Experiments on a larger, more diverse population, with additional knowledge on head motion/orientation throughout, can open up for higher performance and other activity classes. The classification methods could be further improved considering the state-of-the-art in Deep Learning as initially explored by Ronao and Cho (37) and Hammerla et al. (44).

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

The majority of manuscript preparation was carried out by MS with assistance from all authors. GB generated results. GB, TB, and MS developed the methodology. GB and EJ carried out the experiments and data collection.

FUNDING

This work was financially supported by the Swedish Research Council (Vetenskapsrådet, VR 2017-06092 Mekanismer och behandling vid åldersrelaterad hörselnedsättning).

ACKNOWLEDGMENTS

The authors are grateful for the kind support from François Patou at Oticon Medical, Elvira Fischer, and Martha Shiell at Oticon.

REFERENCES

1. Johansen B, Flet-Berliac YPR, Korzepa MJ, Sandholm P, Pontoppidan NH, Petersen MK, et al. Hearables in hearing care: discovering usage patterns through IoT devices. In: Antona M, Stephanidis C, editors. *Universal Access in Human-Computer Interaction*. Human and Technological Environments. Cham: Springer International Publishing (2017). p. 39–49.
2. Korzepa MJ, Johansen B, Petersen MK, Larsen J, Larsen JE, Pontoppidan NH. Modeling user intents as context in smartphone-connected hearing aids. In: *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*. UMAP '18. New York, NY: Association for Computing Machinery (2018). p. 151–5.
3. Wang EK, Liu H, Wang G, Ye Y, Wu TY, Chen CM. Context recognition for adaptive hearing-aids. In: *2015 IEEE 13th International Conference on Industrial Informatics (INDIN)*. Cambridge: IEEE (2015). p. 1102–7.
4. Insignia. *Insio NX* (2021). Available online at: https://www.signia.net/en/hearing-aids/other/insio-nx-bluetooth/; https://www.signia.net/en/hearing-aids/other/insio-nx-bluetooth/.
5. Starkey Livio AI Hearing Aids (2021). Available online at: https://www.starkey.com/hearing-aids/livio-artificial-intelligence-hearing-aids. Available from: https://www.starkey.com/hearing-aids/livio-artificial-intelligence-hearing-aids.
6. Oticon More Hearing Aids (2021). Available online at: https://www.oticon.com/solutions/more-hearing-aids.
7. GN Resound, Resound One (2021). Available online at: https://www.resound.com/en-us/hearing-aids/resound-hearing-aids/resound-one; https://www.resound.com/en-us/hearing-aids/resound-hearing-aids/resound-one.
8. Burwinkel JR, Xu B. Preliminary examination of the accuracy of a fall detection device embedded into hearing instruments. *Am Acad Audiol*. (2020) 31:393–403. doi: 10.3766/jaaa.19056
9. Starkey. Livio Ai Tap Control Feature (2019). Available online at: https://www.starkey.co.uk/blog/articles/2019/04/Livio-AI-hearing-aids-tap-control-featur.
10. Acker K. *The Ear is the New Wrist: Livio AI's Sensors Lead in Step Accuracy* (2018). Available online at: https://starkeypro.com/pdfs/white-papers/The_Ear_Is_The_New_Wrist.pdf.
11. Dernbach S, Das B, Krishnan NC, Thomas BL, Cook DJ. Simple and complex activity recognition through smart phones. In: 2012 *Eighth International Conference on Intelligent Environments*. Guanajuato: IEEE (2012). p. 214–21.
12. Jobanputra C, Bavishi J, Doshi N. Human activity recognition: a survey. *Procedia Comput Sci*. (2019) 155:698–703. doi: 10.1016/j.procs.2019.08.100
13. Brezmes T, Gorricho JL, Cotrina J. Activity recognition from accelerometer data on a mobile phone. In: *Proceedings of the 10th International Work-Conference on Artificial Neural Networks: Part II: Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living*. IWANN '09. Berlin; Heidelberg: Springer-Verlag;. (2009). p. 796–9.
14. Ermes M, Pärkkâ J, Mäntyjärvi J, Korhonen I. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *IEEE Trans Inf Technol Biomed*. (2008) 12:20–26. doi: 10.1109/TITB.2007.899496
15. Kwapisz JR, Weiss GM, Moore SA. Activity recognition using cell phone accelerometers. *ACM SigKDD Explor Newsl*. (2011) 12:74–82. doi: 10.1145/1964897.1964918
16. Atallah L, Lo B, King R, Yang GZ. Sensor positioning for activity recognition using wearable accelerometers. *IEEE Trans Biomed Circ Syst*. (2011) 5:320–9. doi: 10.1109/TBCAS.2011.2160540
17. Khan AM, Lee YK, Lee SY, Kim TS. Human activity recognition via an accelerometer-enabled-smartphone using kernel discriminant analysis. In: *2010 5th International Conference on Future Information Technology*. Busan: IEEE (2010). p. 1–6.
18. Balzi G. *Activity Tracking Using Ear-Level Accelerometers Through Machine Learning*. Technical University of Denmark, Department of Electrical Engineering (2021).
19. Titterton DH, Weston JL. Strapdown inertial navigation technology. In: *IEE Radar, Sonar, Navigation and Avionics Series*. Stevenage: Peter Peregrinus Ltd. (1997).
20. Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. *IEEE Trans Pattern Anal Mach Intell*. (2013) 35:1798–828. doi: 10.1109/TPAMI.2013.50
21. Masse LC, Fuemmeler BF, Anderson CB, Matthews CE, Trost SG, Catellier DJ, et al. Accelerometer data reduction: a comparison of four reduction algorithms on select outcome variables. *Med Sci Sports Exerc*. (2005) 37:S544. doi: 10.1249/01.mss.0000185674.09066.8a
22. Gjoreski H, Lustrek M, Gams M. Accelerometer placement for posture recognition and fall detection. In: *2011 Seventh International Conference on Intelligent Environments*. Nottingham: IEEE (2011). p. 47–54.
23. Hua A, Quicksall Z, Di C, Motl R, LaCroix AZ, Schatz B, et al. Accelerometer-based predictive models of fall risk in older women: a pilot study. *NPJ Digit Med*. (2018) 1:1–8. doi: 10.1038/s41746-018-0033-5
24. Joanes DN, Gill CA. Comparing measures of sample skewness and kurtosis. *J R Stat Soc Ser D*. (1998) 47:183–9. doi: 10.1111/1467-9884.00122
25. Crouter SE, Clowers KG, Bassett Jr DR. A novel method for using accelerometer data to predict energy expenditure. *J Appl Physiol*. (2006) 100:1324–31. doi: 10.1152/japplphysiol.00818.2005
26. Brønd JC, Andersen LB, Arvidsson D. Generating ActiGraph counts from raw acceleration recorded by an alternative monitor. *Med Sci Sports Exerc*. (2017) 49:2351–60. doi: 10.1249/MSS.0000000000001344
27. Bishop C. *Pattern Recognition and Machine Learning*. Springer (2006). Available online at: https://www.microsoft.com/en-us/research/publication/pattern-recognition-machine-learning/.
28. Fix E, Hodges JL. Nonparametric discrimination: consistency properties. In: *USAF School of Aviation Medicine, Randolph Field, Texas* Universal City, TX (1951). p. 21–49.
29. Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Trans Inf Theory*. (1967) 13:21–27. doi: 10.1109/TIT.1967.1053964
30. Fisher RA. The use of multiple measurements in taxonomic problems. *Ann Eugen*. (1936) 7:179–88. doi: 10.1111/j.1469-1809.1936.tb02137.x
31. Breiman L, Friedman JH, Olshen RA, Stone CJ. *Classification and Regression Trees*. Boca Raton, FL: Chapman & Hall (1984).
32. Schapire RE. The strength of weak learnability. *Mach Learn*. (1990) 5:197–227. doi: 10.1007/BF00116037
33. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci*. (1997) 55:119–39. doi: 10.1006/jcss.1997.1504
34. Lindegaard Jensen E. *Head Motion Estimation Using Ear-Worn Binaural Accelerometers*. Technical University of Denmark, Department of Electrical Engineering (2021).
35. Anguita D, Ghio A, Oneto L, Parra X, Reyes-Ortiz JL. A public domain dataset for human activity recognition using smartphones. In: *21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN*. Belgium: Bruges (2013). p. 437–42.
36. Khan AM, Lee YK, Lee SY. Accelerometer's position free human activity recognition using a hierarchical recognition model. In: *The 12th IEEE International Conference on e-Health Networking, Applications and Services*. Lyon: IEEE (2010). p. 296–301. doi: 10.1109/HEALTH.2010.5556553
37. Ronao CA, Cho SB. Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Syst Appl*. (2016) 59:235–44. doi: 10.1016/j.eswa.2016.04.032
38. Bai YW, Yu CH, Wu SC. Using a three-axis accelerometer and GPS module in a smart phone to measure walking steps and distance. In: *2014 IEEE 27th Canadian Conference on Electrical and Computer Engineering (CCECE)*. Toronto, ON: IEEE (2014). p. 1–6.
39. Abadleh A, Al-Hawari E, Alkafaween E, Al-Sawalqah H. Step detection algorithm for accurate distance estimation using dynamic step length. In: *2017 18th IEEE International Conference on Mobile Data Management (MDM)*. Daejeon: IEEE (2017). p. 324–7.
40. Parkka J, Ermes M, Korpipaa P, Mantyjarvi J, Peltola J, Korhonen I. Activity classification using realistic data from wearable sensors. *IEEE Trans Inf Technol Biomed*. (2006) 10:119–28. doi: 10.1109/TITB.2005.856863
41. Anik MAI, Hassan M, Mahmud H, Hasan MK. Activity recognition of a badminton game through accelerometer and gyroscope. In *201619th International Conference on Computer and Information Technology (ICCIT)*. Dhaka: IEEE (2016). p. 213–7.
42. Zhang Y, Markovic S, Sapir I, Wagenaar RC, Little TD. Continuous functional activity monitoring based on wearable tri-axial accelerometer and gyroscope. In: *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*. Dublin: IEEE (2011). p. 370–3.

43. Shoaib M, Bosch S, Incel OD, Scholten H, Havinga PJ. Fusion of smartphone motion sensors for physical activity recognition. *Sensors*. (2014) 14:10146–76. doi: 10.3390/s140610146
44. Hammerla NY, Halloran S, Plötz T. Deep, convolutional, and recurrent models for human activity recognition using wearables. In: *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI'16)*. New York, NY: AAAI Press (2016). p. 1533–40.

Conflict of Interest: MS, TB, and SR-G are 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.

frontiers in Digital Health

ORIGINAL RESEARCH
published: 20 September 2021
doi: 10.3389/fdgth.2021.739370

Check for updates

Connected Hearing Devices and Audiologists: The User-Centered Development of Digital Service Innovations

Marie Luengen[1]*, *Christopher Garrelfs*[1], *Kamil Adiloğlu*[2]*, *Melanie Krueger*[2], *Benjamin Cauchi*[3], *Uwe Markert*[4], *Marei Typlt*[4], *Martin Kinkel*[5] *and Carsten Schultz*[1]

[1] Department for Technology Management, Kiel University, Kiel, Germany, [2] HörTech gGmbH, Oldenburg, Germany, [3] OFFIS e.V., Institute for Information Technology, Oldenburg, Germany, [4] Audifon GmbH & Co. KG, Kölleda, Germany, [5] KIND GmbH & Co. KG, Burgwedel, Germany

OPEN ACCESS

Edited by:
Dennis L. Barbour,
Washington University in St. Louis,
United States

Reviewed by:
Nomfundo Moroe,
University of the Witwatersrand,
South Africa
Gloria Cosoli,
Marche Polytechnic University, Italy

***Correspondence:**
Marie Luengen
luengen@bwl.uni-kiel.de
Kamil Adiloğlu
k.adiloglu@hoertech.de

Specialty section:
This article was submitted to
Connected Health,
a section of the journal
Frontiers in Digital Health

Received: 10 July 2021
Accepted: 27 August 2021
Published: 20 September 2021

Citation:
Luengen M, Garrelfs C, Adiloğlu K, Krueger M, Cauchi B, Markert U, Typlt M, Kinkel M and Schultz C (2021) Connected Hearing Devices and Audiologists: The User-Centered Development of Digital Service Innovations. Front. Digit. Health 3:739370. doi: 10.3389/fdgth.2021.739370

Today, medical technology manufacturers enter the service market through the development of digital service innovations. In the field of audiology, these developments increasingly shift the service capacities from audiologists to manufacturers and technical systems. However, the technology-driven developments of manufacturers lack acceptance of hearing device users and undermine the important role of audiologists within the service provision. By following a user-centered design approach in order to deal with the technological and social challenges of disruptive services, we aim to develop service innovations on an integrated service platform in the field of tele-audiology. To ensure the acceptance of technology-driven service innovations among hearing device users and audiologists, we systematically integrated these actors in a participatory innovation process. With qualitative and quantitative data we identified several requirements and preferences for different service innovations in the field of tele-audiology. According to the preferences of the different actors, we proposed a service platform approach based on a connected hearing device in three pillars of application: 1) one-to-one (1:1) service innovations based on a remote fitting concept directly improve the availability of services offered by audiologists without being physically present. Based on this, 2) one-to-many (1:N) service innovations allow the use of the connected hearing device as an indirect data source for training a machine learning algorithm that empowers users through the automation of service processes. A centralized server system collects the data and performs the training of this algorithm. The optimized algorithm is provided to the connected hearing devices to perform automatic acoustic scene classification. This in turn allows optimization of the hearing devices within each acoustic scene. After the user-centered development of the different service innovations which are designed to converge on an integrated service platform, we experimentally evaluated the functionality and applicability of the system as well as the associated role models between the technical system, the hearing device users and audiologists.

As a future outlook, we show potentials to use the connected hearing device for 3) cross-industry (N:M) service innovations in contexts outside the healthcare domain and give practical implications for the market launch of successful service innovations in the field of tele-audiology.

Keywords: service innovation, user-centered design, integrated service platform, remote fitting, machine learning

INTRODUCTION

Digital Service Innovations in Audiology

According to the World Report on Hearing (1), hearing difficulties are among the most common diseases worldwide. Unaddressed hearing loss is the third largest cause of years lived with disability globally. Over 1.5 billion people currently experience some degree of hearing loss, with over 400 million people living with disabling hearing loss. Only one out of five people suffering from hearing difficulties is using any type of supporting device today. Although hearing difficulties affect all ages, the prevalence increases for elderly people. Due to an aging population, the number of people with at least some degree of hearing difficulties will continuously grow in the future. Through the use of new digital technologies, increasing attention has been paid to the development of service innovations in healthcare (2, 3). Artificial intelligence (AI), big data analysis, Internet of Things (IoT) and the smartification of products into product-service-systems enable healthcare organizations to find new ways of value creation (4).

In contrast to traditional services, digital services are characterized by high levels of re-programmability, homogeneity and self-reference (5, 6). Re-programmability enables the subsequent adaptation and extension of an already implemented service solution (7). Homogeneity includes the possibility to store and transmit digital information and processes which allow for more scalability, broader user targeting, and faster strategic actions (6). Finally, digital services reinforce the diffusion and development of other digital services and technologies which lead to lower entry barriers for future service developments and an even higher diffusion of such service offerings. The field of audiology is predestined for service innovations aiming at increasing the rate of candidates using hearing devices [HDs, e.g., hearing aids (HAs), cochlear implants (CIs)] and improving the quality of current service processes. As such, service organizations within the field of audiology can substantially benefit from service innovations. Especially in the field of tele-audiology, service innovations can increase the efficiency and quality of the fitting process for HDs, enable remote services outside of professional stores and meet the trends for personalized care and user empowerment.

Technological and Social Challenges

As service innovations lead to fundamental changes of existing value creation processes between manufacturers, audiologists, and HD users, the development and implementation is faced with numerous technological as well as social challenges. Technological challenges are related to the integrity of the data and the performance of the delivered services. The data that the considered services will use can be highly sensitive, including personal information, location and audio signals. Consequently, the data should be exchanged safely and stored securely. This concern has been one of the main focuses from the conception to the implementation of the technical architecture (cf. section Technical Realization). Concerns for data integrity are driven by increasingly stringent regulations, such as the general data protection regulation (GDPR) acted by the European Union, as well as by end-users being sensitive to these issues. The main challenges concerning the performance of the delivered services come from the difficulty of using advanced signal processing and machine learning algorithms with the limited computational resources which are typically available in HDs. HDs must perform the signal processing in real-time. The incoming signal is processed faster than a latency, called just noticeable difference, so that the HD users do not suffer under issues like reduced lip synchronization. Therefore, keeping the latency induced by signal processing within HDs below a certain level is crucial for the usability. Due to the limited computational resources designing computationally demanding signal processing algorithms for HDs in terms of required operations as well as memory requirements is a challenging task. Algorithms require many iterations for optimal performance, so called convergence. However, algorithms with high memory requirements are not suitable for HDs. Further, the data processed on a HD have a highly sensitive character. Therefore, preserving the privacy of a HD user is important. For these reasons, HDs are not allowed to save any audio material, which can contain clues about the private life of the HD users. Further, it is not allowed to save data which contain features that can be used to reconstruct the acoustic environment. As a result, machine learning algorithms that require large amounts of data to perform optimally have to be incorporated in a HD with caution.

Next to the technological challenges the implementation of digital service innovations can raise challenges on the side of the different user groups. Service innovations that are based on disruptive technological developments have the potential to fundamentally change current value creation processes of audiologists. Disruptive service innovations challenge existing processes, routines and competencies. As competencies shift to technical systems, service innovations can lack acceptance of audiologists. Further, manufacturers in the field of audiology start to enter the service market through digital services. The service developments of manufacturers are often closely related to their product offerings and commercialized as product-service systems (8). Although current digital service offerings

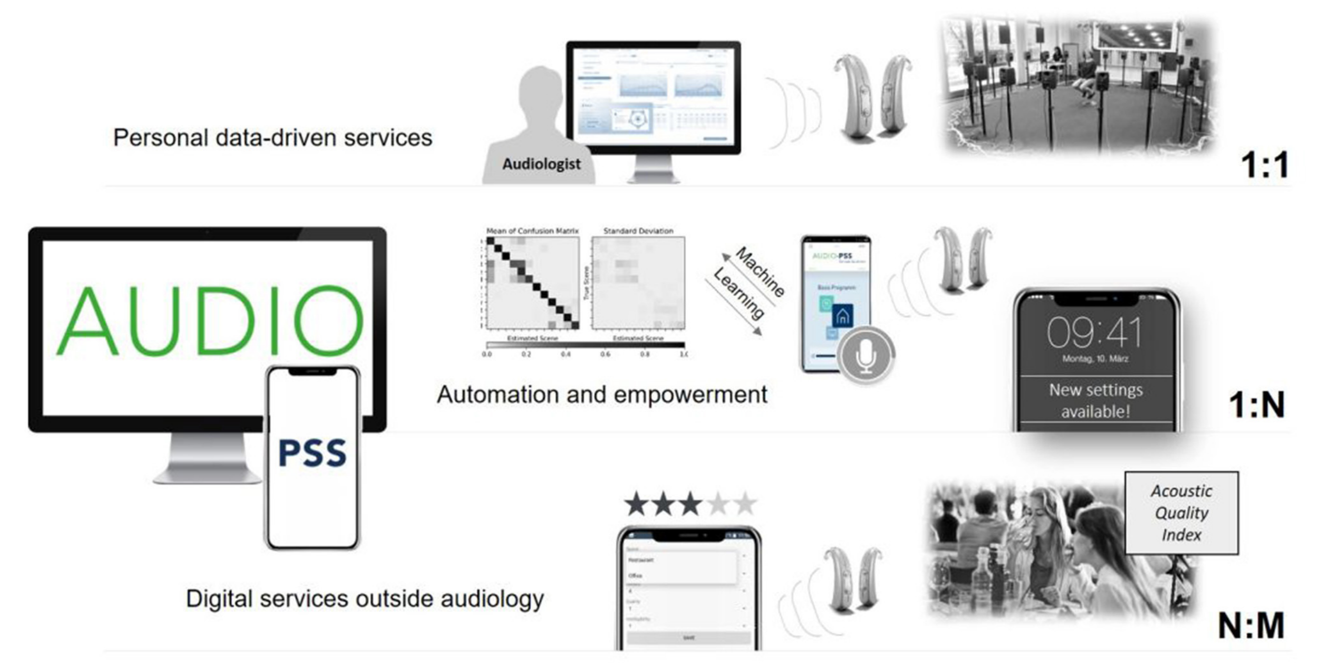


FIGURE 1 | Overview of the Audio-PSS platform approach.

of HD manufacturers are distributed and provided through audiologists, the information and data gathered remain on the side of the manufacturers. The outflow of knowledge, experience and user information can further weaken long-term competences of audiologists which leads to resistance and lack of acceptance. However, service innovations will only be implemented successfully if the complementary interaction between manufacturers, audiologists and HD users is developed and managed with respect to the specific requirements of digital service platforms. As such, service innovations require the close interaction between different actors like manufacturers, audiologists and even HD users (9). The collaboration and integration of all actors of the service ecosystem is necessary to ensure the functionality and applicability of previously individual service components on an integrated service platform.

The "Audio-PSS" Service Platform Approach

With the joint project *Audio-PSS (development of product service systems in tele-audiology)* we aimed for a participatory development of digital service innovations based on a connected HD which are designed to converge on an integrated service platform. Thereby, we combined advanced technological developments with the important role of human interaction between the audiologist and the HD users. **Figure 1** provides an overview of the *Audio-PSS* platform approach that ranges from personal data-driven services (1:1 services) to automation and empowerment (1:N services) and opportunities for new service innovations outside the core field of audiology (N:M services). *1:1 services* are service innovations that aim to improve the relationship between audiologists and HD users, e.g., by allowing an audiologist to continuously monitor HD settings and usage behavior or allow the HD user to contact the audiologist while being in a demanding listening situation for an instantaneous improvement of the settings. The initial fitting, e.g., of HAs, is followed by a familiarization phase in which users collect experiences and impressions of handling HAs in their everyday life. However, the user must memorize these experiences until the next appointment with the audiologist at which they talk about the experiences during the last weeks with the HAs. A common service platform could enable remote online and offline services and could provide documentation of personal experiences for later access by the audiologist. *1:N services* include service innovations that use cloud computing to exchange data, experiences and recommendations of a variety of HD users to automatically fit the HD program through the analysis of aggregated user data. For instance, collective assistance functions can help especially in difficult listening situations by analyzing parameters from the current situation (e.g., location, reverberation, sound pressure level, signal-to-noise ratio) with a machine learning system that was trained offline on aggregated data. *N:M services* are cross-industry innovations that use the data of the connected HD in contexts outside the healthcare domain, e.g., by measuring the acoustic quality of restaurants. For instance, through acoustical monitoring and a simultaneous experience sampling by users; restaurants, cinemas, theaters, educational instructions and other localities as well as public spaces can be recommended with regard to their acoustics, communication experience and well-being, creating an *acoustic quality index*.

In this paper, we provide insights into our project and show the main results of the individual research activities from both the social and technical perspective. From the social perspective, we systematically integrated HD users and audiologists in the innovation process in order to design the service innovations according to their individual needs and preferences. First, we uncovered potentials for service innovations through gaps or specific problems in the current service provision of HD users and audiologists. Further, the user groups contributed to the usability and design of service innovations, e.g., by testing and evaluating different feedback and process scenarios of the remote fitting (RF) concept. With the close integration of audiologists and HD users, we consequently aimed to change the dominating view that digital services substitute personal services. With our audiologist-centered service platform, we show potential pathways how service innovations can create new ways of value creation, improve the quality and interaction within existing service processes and enhance the competencies and opportunities of audiologists in the future service ecosystem. The requirement analysis revealed a great potential for more disruptive service innovations in the context of RF services that audiologists can offer without being physically present. Advanced technological developments like machine learning can increase the efficiency of RF services to the extent that the service can be provided by a service center or by the user alone, independent from personal interaction with an audiologist. As such, there is a great need for a design of the interaction between the HD user, the audiologist, and the technical system that ensures service innovation acceptance.

On the technical side, we show how such service innovations can be technically designed, developed and implemented on one service platform. We designed and implemented a hardware/*s*oftware architecture (10) that allows audiologists to remotely access the HDs of their customers, e.g., for RF, and customers to request support from them. Additionally, this architecture allows to collect a large dataset of audio signals that can be used to improve the performance of HDs. These signals, recorded in a wide range of acoustic scenes, can be used to train advanced acoustic scene classifiers in a remote server and transmit the resulting models from this training to the considered HDs where the recognition can then be achieved in real-time. Though real-time acoustic scene classification in HDs remains a challenging task, the training being done remotely makes it more feasible. As a proof of concept, we developed a machine learning based acoustic scene classifier to automatically recognize acoustic scenes relevant for HD users. For the training of the classifier and the evaluation of the recognition accuracy, we recorded a diversity of acoustic scenes. In order to preserve the privacy of the HD users, we encoded these recordings using sparse features, which cannot be reverse-engineered for reconstructing the original audio. For the training and testing as well as for the real-time evaluation of the acoustic scene classifier, these features were incorporated.

To the best of our knowledge, we are the first who take a participatory innovation development approach to tailor advanced technological developments with specific user-centered preferences within the field of tele-audiology. With our research, we show promising pathways for audiologists to be a major part in the future value chain of service innovations within the HD industry. Beyond that, we show how machine learning can be integrated in an overall architecture and open new opportunities for RF services within audiology. Based on the findings of our experimental evaluation, we demonstrate the benefits of our technical developments regarding their efficacy to enable and improve new ways of RF and the importance of the relationship between the technical system, the HD user and the audiologist.

The remainder of this paper is structured as follows. First, we will present the results of the technology-related requirements for digital service innovations of HD users and audiologists which we collected and analyzed through qualitative (interviews) and quantitative data (surveys). In the subsequent section, we will present the technical infrastructure we realized to meet those requirements. Therefore, we present the system architecture of the connected HD and our core technological developments in the context of RF: acoustic scene classification through machine learning and the overall development and implementation of the RF infrastructure. After the technical realization, we evaluated the technical developments and social aspects with laboratory experiments, which are presented in section Evaluation. We will conclude this article with future directions and mention technical possibilities as well as social implications to increase acceptance among HD users and audiologists in the context of the integrated service platform.

REQUIREMENT ANALYSIS OF HEARING DEVICE USERS AND AUDIOLOGISTS

Innovation through co-creation plays an important role in the healthcare service context. In a collaborative development process, audiologists can be important actors in the development of future innovations as they have broad knowledge and experience in both the product and service world. Additionally, the development of service innovations within tele-audiology also benefits from the integration of HD users as this provides important insights to the user experience during the process of fitting HDs. In the first part of the project, we therefore investigated the requirements of audiologists and HD users to align the technological developments of manufacturers with the needs and preferences of the user groups.

Exploration of User Preferences and Technological Developments

We applied an explorative empirical approach consisting of qualitative interviews with audiologists and HD users and an analysis of secondary information on recent developments of manufacturers to compare the current developments with the user preferences. First, we identified specific preferences for future service innovations along the opportunities of a connected HD based on 22 semi-structured interviews with HD users and audiologists. The interviews were recorded and transcribed. Subsequently we validated the results on the identified user needs for service innovations during a 2-day workshop with both groups.

A total set of 85 requirements for future services based on a connected HD has been identified, clustered in 14 different service functionalities. The HD users highlighted requirements for extended mobile control functions, followed by RF services, and for the consideration of individual hearing preferences through scene classification and machine learning. The audiologists highlighted the development and provision of an interactive hearing training, and the development of a scene classification solution in order to monitor the hearing environment and related issues of HD users. Self-services are merely highlighted by the HD users that formulate preferences for services that represent extended mobile control settings: *"I don't really want to adjust anything and certainly not a second device. If these functions were integrated in the smartphone I would say: Great. So many settings on one or two buttons: I'm freaking out."* Further, a number of HD users and audiologists demand new, audiologist-centered RF solutions: *"Is it possible to bring the audiologist along the various hearing environments I experience? It would be fantastic if the audiologist could adjust my hearing aid remotely, that would be genius."* **Figure 2** shows an overview of the identification and categorization of user preferences. The left side shows the different service categories that have been introduced before. The central column summarizes the coding of the related statements on preferences of the users from the interviews. Finally, the right column categorizes the different user preferences into specific service functionalities.

To identify the current service-related technological developments within the field of tele-audiology, a qualitative patent analysis was conducted (11). The patent analysis was based on the patent database FamPat (Questel). The search period was set from 2008 to 2018. The selected patents included functions about connectivity, use of mobile devices, methods of classification, automation, RF as well as the creation and use of new digital interfaces. In recent years, technological developments focused on the creation of an infrastructure for new digital services from hardware-based rudimentary remote functions in 2008 to machine learning developments in 2018. The patent analysis revealed 21 core technological developments. Within the field of *personal data-driven services* (1:1 services), remote assistance like video chat between audiologists and HD users or chat-based assistance (e.g., WO2013020594) (12) was predominant. More disruptive service developments were reflected by patents on the monitoring of auditory data (e.g., EP3035710) (13) and performance data (e.g., DE102015203288) (14). Further developments enabled intuitive, hands-free interfaces using voice commands (e.g., US2018061411) (15) or physical action based control (e.g., US2014126759) (16) through gestures, eyes, or body movement. With 15 patents, scene classification functions were the most frequently patented technological development in the recent field of automation and empowerment (1:N service, e.g., EP3288293) (17). See **Figure 3** for an overview of the developments in tele-audiology since 2008. The figure shows the evolution of technological trends in the field of tele-audiology which were found out with the patent analysis.

Based on the comparison of the user preferences and the service-related technological developments within the field of tele-audiology, our results indicate that manufacturers, HD users and audiologists focus on different service categories. Audiologists articulate more preferences than HD users but focus primarily on personal data-driven services (1:1). Further, the technological complexity often exceeds the capabilities of most users. Currently, many disruptive service innovations that drive new service opportunities originate predominantly from manufacturers.

User-Centered Design of Different Digital Services

As a result of the explorative pre-study we derived a large set of service innovations expressed by the user groups that are further evaluated within the following two studies. In order to validate the results of the qualitative pre-study we carried out two online surveys with 671 audiologists and 184 HD users. Particularly in the case of 1:N services, different design approaches and beliefs on the future relation between audiologists and HD users were identified, so that we intended to quantitatively determine the specific needs and preferences of audiologists and HD users. Six service innovations were included in both surveys: *Remote fitting* (we distinguished between *time-delayed RF through audiologists* and *instantaneous RF through a service center*), *AI-based assistance function for situation-specific HD optimization*, *Optimized automatic scene and situation classifier*, *Hearing Coach, Hearing training,* and *Linkage to social networks.* In the survey of audiologists, we added six service innovations as a result of the patent analysis so that the audiologists evaluated twelve service innovations in the field of tele-audiology in total. The survey focused on the audiologists' intention to offer these services to the customers in the near future. The HD users evaluated eight service innovations and specific design approaches of more disruptive service innovations, e.g., the design of RF services. In this context they had to choose between different specifications of the services, for example between (1) *no remote fitting*, (2) *time-delayed remote fitting by their attending audiologist* and (3) *instantaneous remote fitting by a service center*. Due to the different requirements and preferences expressed in the explorative study, we have adapted the service innovations somewhat to the target group in the quantitative survey.

Service Innovation Acceptance of Audiologists

As established healthcare organizations are characterized by strong path dependencies, they typically show a strong dependency upon their mainstream users. While the current service processes typically fit the needs and preferences of existing users, they do not match the increasing demand for digital services and online offerings of emerging target users. As such, service organizations face highly uncertain demands in the future (18) and struggle to balance the ambidexterity between routine and innovation (19). At the same time, service organizations face institutional tensions that hamper the development of future innovation (20). For instance, innovative service organizations face missing reimbursement opportunities for digital services by health insurances, a lack of telemedical regulations, or strict process standards of professional associations. We aimed to integrate the audiologists as providers of service innovations

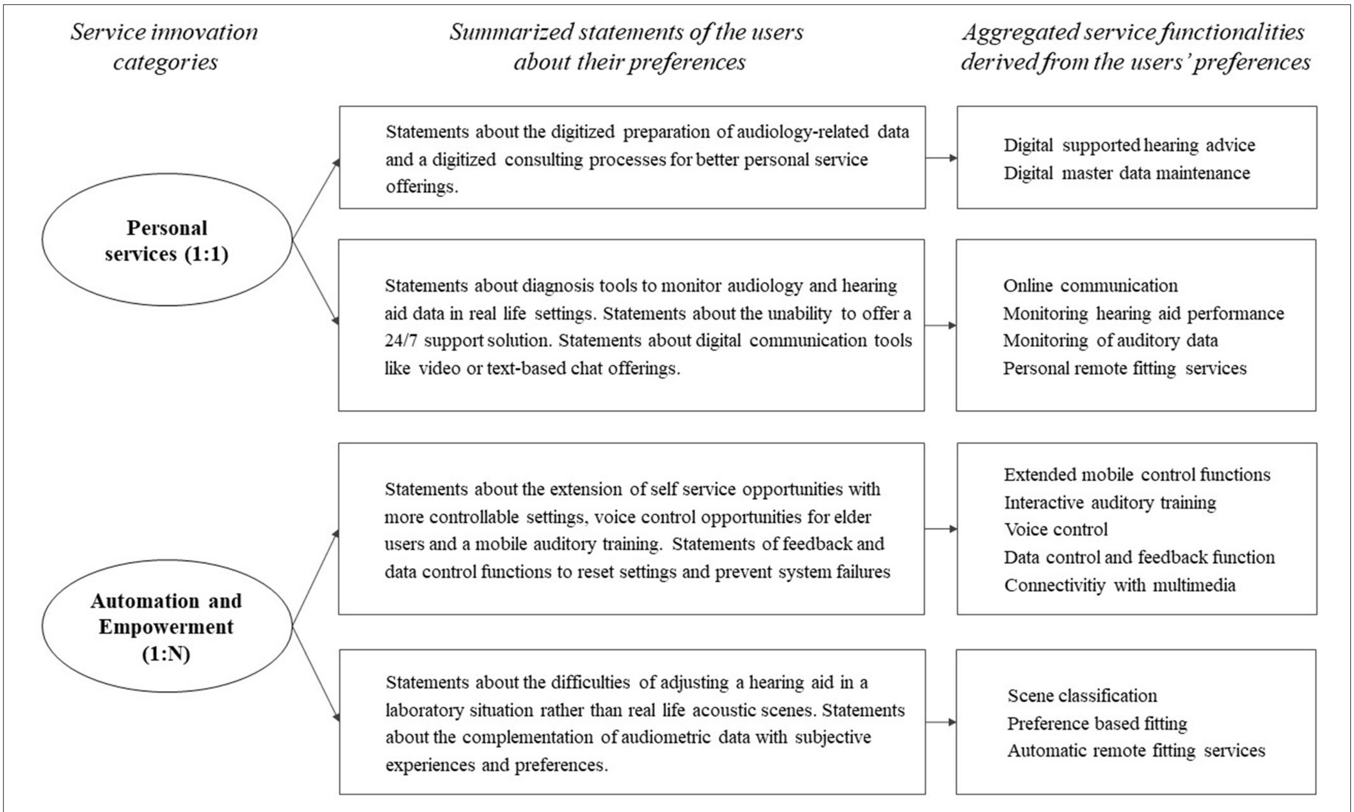


FIGURE 2 | Identification and categorization of user preferences according to their statements in the interview series.

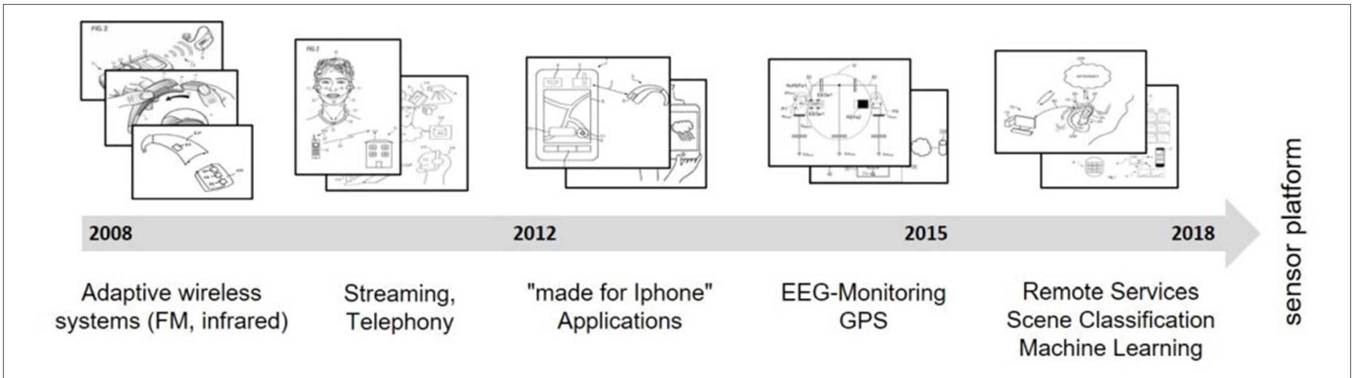


FIGURE 3 | Evolution of the technological trends in the field of audiology.

to analyze these challenges further. Therefore, the objective of the study was to find out about audiologists' intention to offer different digital service innovations and their general evaluation of the derived twelve service innovations. The primary research question in this study is: *How does the disruptive potential of service innovations impact innovation acceptance of frontline employees in audiology?*

We surveyed 671 audiologists from different companies and in different positions in Germany. The online survey included questions regarding the acceptance of different service innovations as well as demographic characteristics, digital competence and innovativeness. We used established scales that originate from literature on research of acceptance and innovation management. Within our survey, the audiologists (63% female; mean age = 20–29 years, 56% younger than 30) were employed for 11 years on average (mean = 10.89, SD = 9.41). 14.8% of them were apprentices, 48.6% journeymen/journeywomen, 32.5% master craftsmen/craftswomen and 2.5% others. From a theoretical viewpoint, we distinguished between incremental

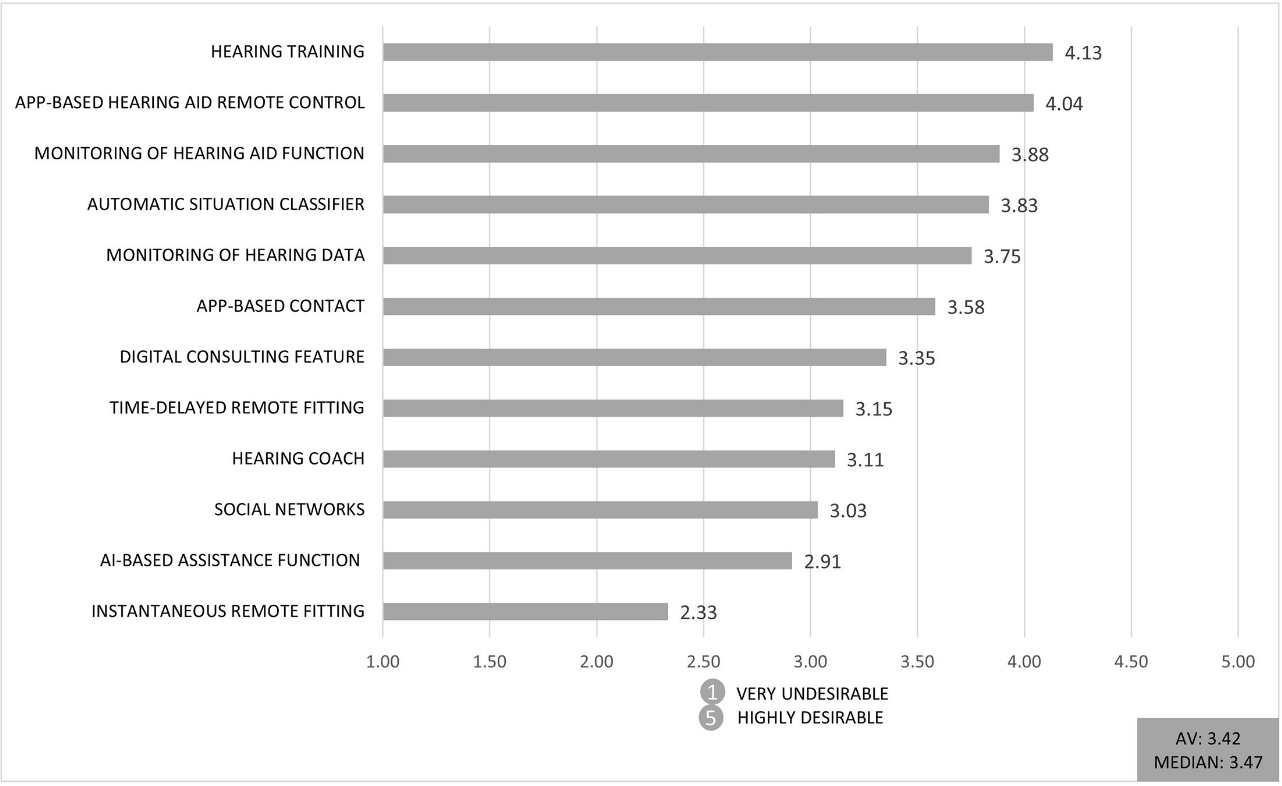


FIGURE 4 | Mean response to the question *"To what extent do you wish to offer the digital service within the next 5 years?"* (1 = very undesirable to 5 = highly desirable; AV = Average).

and disruptive digital service innovations. We classified *six* service innovations as incremental, e.g., *app-based contact, time-delayed remote fitting through audiologist* and *monitoring of hearing aid function*. The remaining six service innovations were classified as disruptive, e.g., *automatic situation classifier, instantaneous remote fitting through service center* and *AI-based assistance function*.

Only 11% of respondents have a high level of digital competence, and feel confident in using digital tools. This can possibly have a negative impact on the evaluation of disruptive service innovations. But on the other hand 44 % of respondents rate their company as highly innovative. In **Figure 4**, the twelve digital service innovations are ranked by the intention of audiologists to offer each digital service to their customers within the next 5 years, from *1 = very undesirable* to *5 = highly desirable*.

Hearing training was rated best among the presented services (4.13 on a 5-point Likert scale). Whereas, *instantaneous remote fitting through a service center* was rated the lowest (2.33 on a 5-point Likert scale). The first service is a complement to the audiologists' work, and therefore can facilitate their daily work, whereas the lowest rated service is a potential threat to their current economic position if the fitting of HDs is taken over by an external service provider. As we have previously classified the services according to their assumed threat potential, this result supports our theoretical perspective that more disruptive service innovations are declined by the audiologists.

As an overarching result, incremental innovations are rated 12% better on average than more disruptive innovations. **Figure 5** depicts that more incremental service innovations are rated higher than disruptive service innovations regarding the audiologists' technology acceptance. The acceptance of the service innovations was measured by a multi item scale, containing questions like *"I rate the digital service as useful for my field of work." "I think my customers would respond positively to the digital service."* or *"I think that colleagues in my company would advise me to use the digital service."* The graph shows a summarized factor of the used items for *acceptance of service innovations*. As mentioned above the expectation of change in competence plays a decisive role in acceptance of audiologists.

Further, we compared the time-delayed RF with instantaneous RF. The RF services differ by (1) a time dimension and (2) a personal dimension. The time-delayed RF was conducted by the personal audiologist whereas the instantaneous RF was conducted by a service center. We hypothesized that for audiologists the fitting by an external service center would be perceived as a threat to their current competences. For this reason we also investigated a contrary RF service in which the audiologist participates strongly. Consequently we defined the

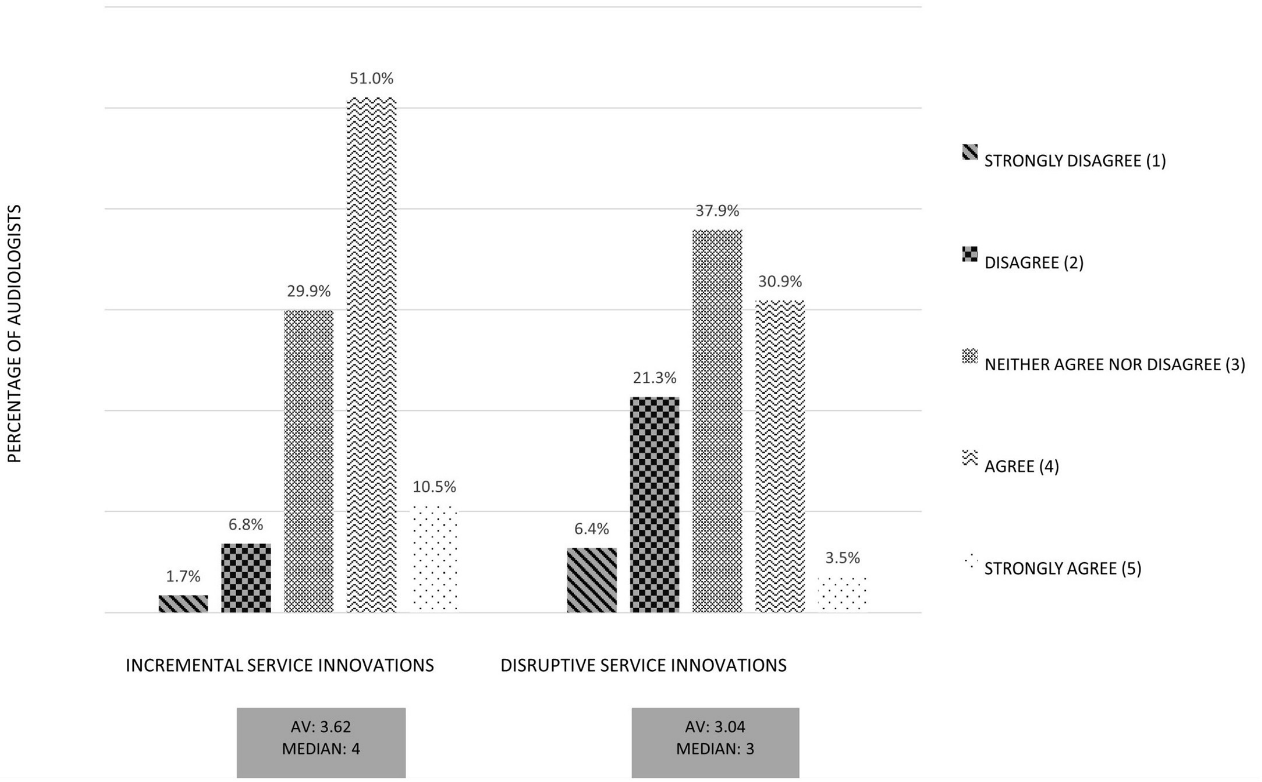


FIGURE 5 | Comparison of incremental service innovations and disruptive service innovations: Aggregated factor Technology Acceptance. Exemplary Item: *"I would like to work with the digital service"*.

time-delayed RF as an incremental service innovation and the instantaneous RF as a disruptive service innovation, from the viewpoint of audiologists. By comparing the two RF services with regard to technology acceptance it showed that the time-delayed RF was evaluated substantially better (see **Figure 6**). Again, the figure shows a summarized factor of the used items for *acceptance of service innovations*. In particular, the fear of quality losses in the case of RF by a service center had a negative influence on the evaluation of this digital service. Moreover, the service possibly was perceived as a threat for the area of competence by audiologists.

Service Innovation Acceptance of Hearing Device Users

In addition to the audiologists, the potential users of these digital services were surveyed. Hearing loss is a sensory disorder that greatly affects the life and social interaction of the affected person. Therefore, many HD users would like more assistance from their HDs in acoustically difficult listening situations, in the handling and adjustment of HDs, as well as in the rehabilitation and acceptance of their hearing impairment. To better understand user preferences, an online survey with HD users was conducted with the goal to identify different service innovations and the participants' willingness to use them in everyday life. The primary research question here was: *Which service innovations and which specifications of service innovations are HD users most likely to use?*

For this purpose, an online survey with 184 HD users (67% female, 85% older than 60 years, median: 70 years) was conducted. All participants were experienced HD users and 78% of them wore their HDs more than 8 h a day. The HD classes that the participants currently use covered basic, mid-range and high-end HDs, which meant that the HDs differed in terms of both technical and comfort features (e.g., noise reduction, situation detection, microphone directionality).

Many modern HDs already offer the possibility to be controlled via smartphone and app. In addition to the technology class of the HDs, the use of smartphones in everyday life is also potentially decisive for the acceptance of this new technology. Eighty percentage of the participants stated that they do not use an app to operate the HD (e.g., program change, volume change) and 14% do not use a smartphone. Regardless of HD control, a total of 54% use their smartphone one to several times a day and 18% one to several times a week, which is an important prerequisite for the use of future digital services. Furthermore, there are technical requirements for HDs to be compatible with digital services (e.g., Bluetooth capability) which are currently not fulfilled by all HD classes. Depending on the HD class and the fixed amount covered by the health insurance, the amount of the private co-payment that HD users have to pay for a technically

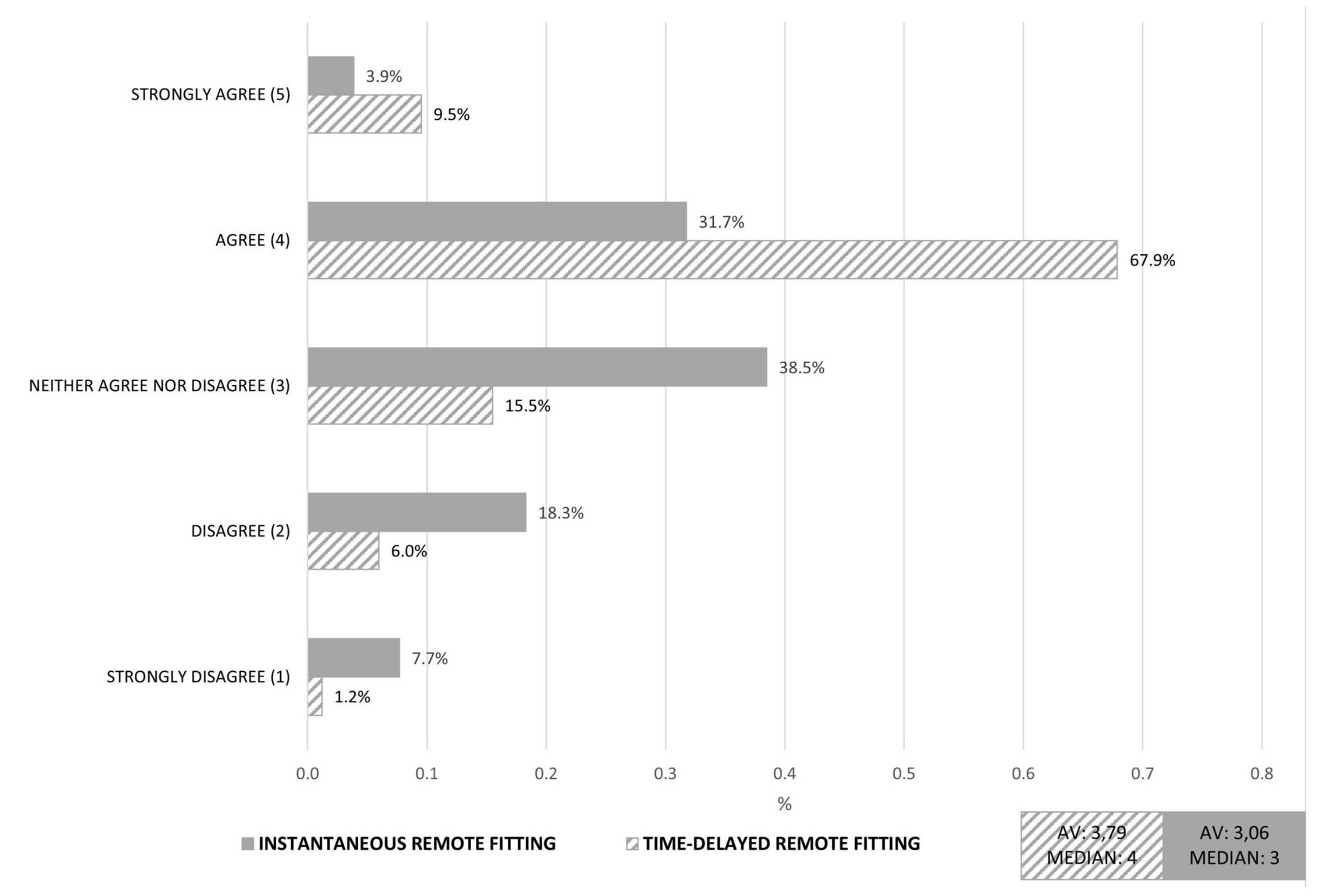


FIGURE 6 | Comparison of instantaneous and time-delayed remote fitting: Aggregated factor Technology Acceptance. Exemplary Item: *"I would like to work with the digital service"*.

more advanced HD provision varies. Twenty-eight percentage of participants with public health insurance reported that they paid between 0 and 500 € per ear for their HDs. Thirty-three percentage invested between 500 and 1,500 €, 16% between 1,501 and 2,000 € and 18% more than 2,001 € per ear for a high-class HD. Overall, technology acceptance and personal interest in new technology were rated on average as "partially agree" (3) and "fairly agree" (4) (mean age = 3.43, SD = 0.91), with their own technology competence rated as "less agree" (mean age = 2.03, SD = 0.90). Nevertheless, the participants assess that learning the technique is under their control [M = 3.82 ("fairly true"), SD = 0.90].

- In the questionnaire, the users were introduced to eight different digital service technologies, each in three separate configurations: *Remote fitting* (no remote fitting, time-delayed remote fitting through audiologist, instantaneous remote fitting through service center),
- *AI-based assistance function for situation-specific HD optimization* (disabled assistance function, user-defined adjustment of the HD setting (limited assistance function), automatic situation-dependent optimization of the HD setting),
- *Optimized automatic scene and situation classifier* (disabled automatic scene classifier, user preference-based scene classifier, automatic scene classifier of any acoustic listening situation),
- *Smart home connectivity* [no smart home connectivity, smart home connection with household appliances (e.g., washing machine), smart home connection with household appliances and home technologies (e.g., doorbell)],
- *Hearing Coach* (no hearing coach necessary, hearing coach only virtual in the app, hearing coach virtual in the app and meetings in person),
- *Hearing training* (no hearing training necessary, hearing training only virtual in the app, hearing training virtual in the app and meetings in person),
- *Connection to social networks* (no connection, connection only virtual in the app, connection virtual in the app and meetings in person),
- *Recording of vital parameters* (no recording, recording of vital parameters without exchange with doctor or audiologist,

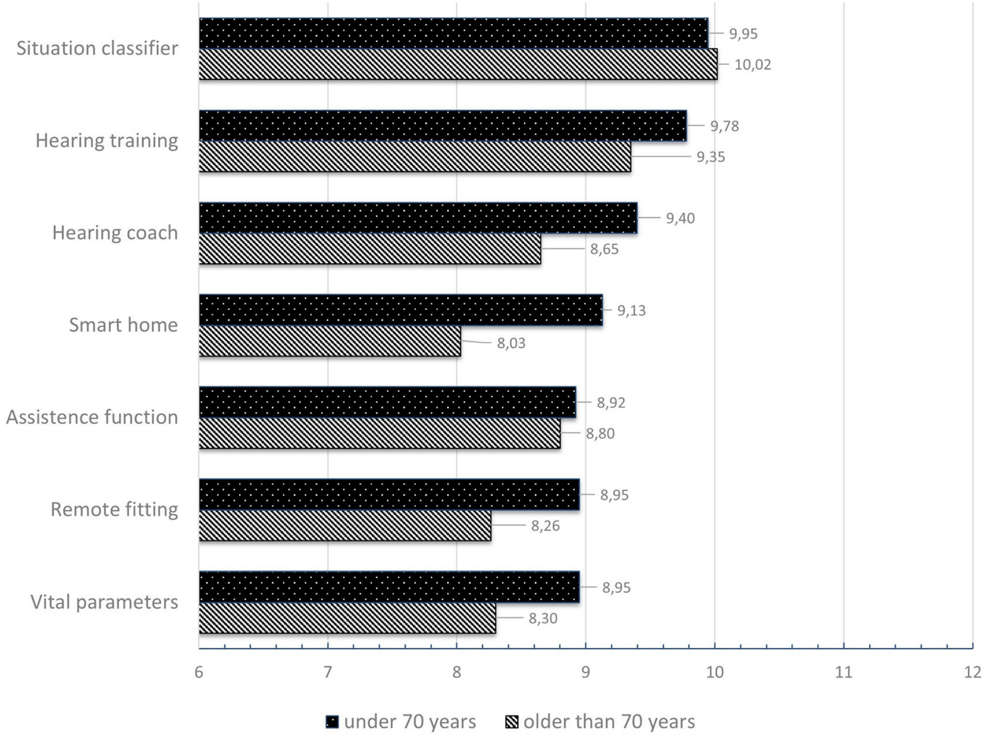


FIGURE 7 | Ranking of the preferred service technologies classified by age.

recording of vital parameters with exchange with doctor or audiologist and comparison with other users).

For the evaluation of the preferred service innovations, the individual scores for the three alternatives were summed (maximum rating = 12, minimum rating = 6). Results were analyzed and ranked for two age groups (above and below 70 years, see **Figure 7**). Overall, the service *optimized automatic scene and situation classifier* is most desired, followed by *hearing training*. Regarding the *remote fitting services* alone, it can be noted that it is ranked in a split 5th place. However, participants under 70 years are noticeably more open toward the service than those over 70.

The rating of each of the three different alternatives for service innovations: remote *fitting*, *AI-based assistance function for situation-specific HD optimization*, and o*ptimized automatic scene and situation classifier* can briefly be summarized: The respondents under 70 years of age preferred the function *instantaneous remote fitting through service center*. The older respondents also preferred this type of RF, but also indicated that they actually prefer the face-to-face fitting interview. In terms of technology *AI-based assistance function for situation-specific HD optimization*, both age groups preferred the limited assistance function. They would like to be supported by an AI system, but retain the ability to decide about the HD setting and fitting. In the under-70 group, the *user preference-based scene classifier* specification was clearly preferred over the *automatic scene classifier of any acoustic listening situation*. The participants in the group over 70 years of age, on the other hand, preferred the *automatic classifier*, closely followed by *user preference-based scene classifier*.

Summary of the Requirement Analysis

The exploration of user preferences and technological developments led to a diverse set of service innovations which we quantitatively analyzed through two surveys with audiologists and HD users. We investigated which service innovations and their specifications are most desired by both groups. It was particularly noticeable that the RF services were evaluated quite differently by HD users and audiologists. Although many HD users preferred the instantaneous RF, the personal component still plays a role, as evidenced by the decision of those over 70 against a RF at all. On the other hand, the audiologists were clearly in favor of time-delayed RF by an audiologist, because they feared a loss of competencies and decreasing quality of customer consultation. For this reason we decided to further clarify the preferences for the different RF specifications and the impact on the acceptance of the adjusted HD fitting. Consequently we followed up the requirement analysis with another evaluation study (cf. section Evaluation). In the social evaluation of RF we gave the HD users three variants to test: RF through audiologist, RF through AI, and a mixed form: RF through AI with approval by audiologists. In

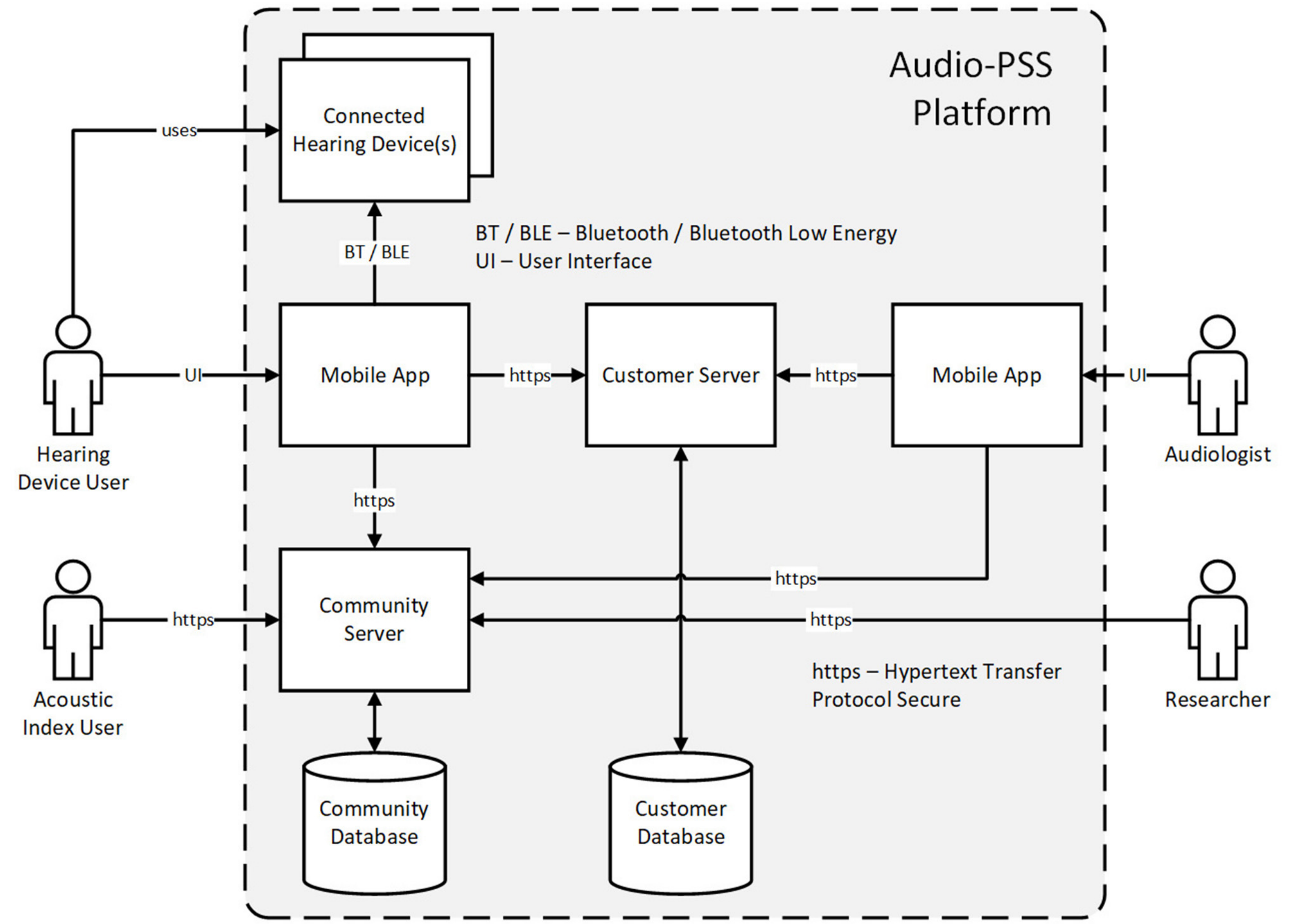


FIGURE 8 | Overview of the platform architecture to support 1:1, 1:N and N:M services.

technological development, we focused on the mixed form to consider a system that combines both advantages of personal and digital services to meet the preferences of both audiologists and HD users.

TECHNICAL REALIZATION

In this section, the technical realization based on the results of the requirement analysis is presented. We decided for a platform design as HD users and audiologists favor a wide range of functionalities. However, both sides emphasized the importance of a personal relationship, which is why we aimed to offer an integrative platform where all stakeholders are joined. While many functionalities are incremental and can be fed directly into market development, we focused on RF and scene classification, as they require a high level of initial research and development effort. Developing a platform also had the advantage to build and test components for several services independently from each other. The modularity of a platform enables platform providers to adapt or expand their service offerings depending on the market needs as well as technological opportunities. As service platforms offer great interoperability, the HD as well as the collected data can be used for several service solutions provided by the integration of different stakeholders (6). Furthermore, new service innovations can be introduced continuously after the launch of the platform. The overall architecture for the RF approach and the respective subsystems will be described in the following sections.

System Architecture of the Connected Hearing Device

Overview

In order to fulfill the needs expressed by HD users and audiologists during the field study described in section Requirement Analysis of Hearing Device Users and Audiologists, a dedicated hardware/software architecture had to be designed and implemented. We took advantage of the Bluetooth connectivity available in modern HDs to design an architecture that is able to support the one-to-one (1:1), one-to-many (1:N) and cross-industry (N:M) services considered in this paper. **Figure 8** depicts an overview of the complete architecture (10). All use cases rely on users accessing one or both of the two used databases, referred to as the customer database and community database.

The customer database is used to store personalized, and hence potentially sensitive, data and can be accessed by HD users and by audiologists i.e., to support 1:1 service innovations. The community database contains anonymized data used to support the 1:N service innovations and cross-industry (N:M) service innovations. Different software clients are provided for each type of users, to visualize and, if relevant, edit the content of the databases. Both, HD users and audiologists, can use mobile applications to access the data. HD users can edit the information about themselves, send messages to audiologists and select new available HD settings. HD users can as well view the acoustic index and make new entries to it. Audiologists can see the information about HD users under their care, answer their messages and change their HD settings during the RF. Acoustic index users can visualize the existing entries using a web browser. Those entries are accessed through a computing cluster that is also used by researchers to improve HD processing, more specifically, to train performant acoustic scenes classifiers. As this paper focuses on the RF application, the next subsections describe the elements of the architecture that support RF, namely the customer server and database as well as the mobile application used by the HD users. The development of acoustic scene classifiers as well as efforts to develop privacy preserving features is described in section Machine Learning for Acoustic Scene Classification.

Customer Server and Database

As mentioned in the previous section the customer server is required to store the HD user related data that could be associated with audiological data, such as audiograms or the history of fitting sessions by the audiologist. These data are sensitive with regard to the data protection regulation as they link individuals with their health data. Hence precautions have to be taken to ensure that data is processed in conformity with regulations. In our setup, an SQL server was used. The customer server can be accessed from the app through a secured cloud authentication process (OAuth 2) using encrypted communication (TLS 1.2).

The audiologist uses the customer server to manage the individual history of fittings. Together with the feedback from the user's app (evaluated listening situation as well as sound recordings) the audiologist can create an updated set of HD parameters to improve the HD settings (1:1 service) and send them back to the user. Based on individual proposals and fitting sessions stored and managed from the customer server, the entire content of the customer database can be used to feed the community server, hence supporting 1:N services.

User Interface for Remote Audiological Services

As a user interface for the patients as well as for the audiologists we provide a mobile smart phone application. It is a core component of the mobile part of the system. It connects the HDs on the one end with servers in the backend and serves as a user interface to gather feedback and display data. By interacting with this app, both stakeholders can perform many tasks, which would usually require a personal meeting, without such a meeting necessary.

Having several use cases and functions in mind the AudioPSS app is designed to be a platform. Designing the app as a platform means on the one hand to create a solution that can be used on different target systems (iOS, Android) without implementing it multiple times. Platform specific concerns like HA drivers and notification systems are dedicated to the particular target system. The other aspect refers to the app as a host for different functions and services. **Figure 9** depicts the architecture.

Remote Measurement of Hearing Loss

For the characterization of the hearing loss, typically a pure tone audiogram is measured for both ears. For these measurements calibrated hardware is required. By using the technical infrastructure, we set up a combined solution, which integrates a calibrated hardware/software system with the remote AudioPSS-app. The calibrated hardware is called the portable hearing lab (21). An open source signal processing platform for hearing research called the *open Master Hearing Aid* (22) is running on this platform. This solution provides the audiometer backend of the measurement set up. The architecture is shown in **Figure 10**.

The remote AudioPSS-app provides two functionalities for measuring the audiogram: by using the first functionality, which is called the remote audiogram app, the patient can perform the audiogram measurement remotely without visiting the audiologist. The second functionality (remote controlling app) is designed for the audiologist to observe the measurement. The audiologist receives the measurement results once the measurement has been completed and can thereby judge the quality of the measurement.

Remote Fitting

There are two basic remote fitting scenarios: (a) perform a first fit and (b) execute a follow-up fit. Based on the audiogram data gathered from the remote measurement above the remote audiogram app provides additionally the possibility to perform a first fit using the portable hearing lab device and a selected fitting rule. Hence the patient can get an immediate benefit from the remote measurement.

As for the follow-up fit workflow, a connected user can send his feedback about the current listening situation using categorized symbols and an attached sound snippet record to the backend. This message including the attachments is stored on the customer server. The audiologist, located either in his office or operating from a service center, analyzes the current settings, the listening situation and the user's feedback resulting in a new fitting proposal. The proposal is sent back to the user's smartphone where the user will be notified about it. Then the updated settings can be downloaded, installed and tested. Depending on the testing the new settings can be kept or discarded resulting in another feedback loop with the audiologist.

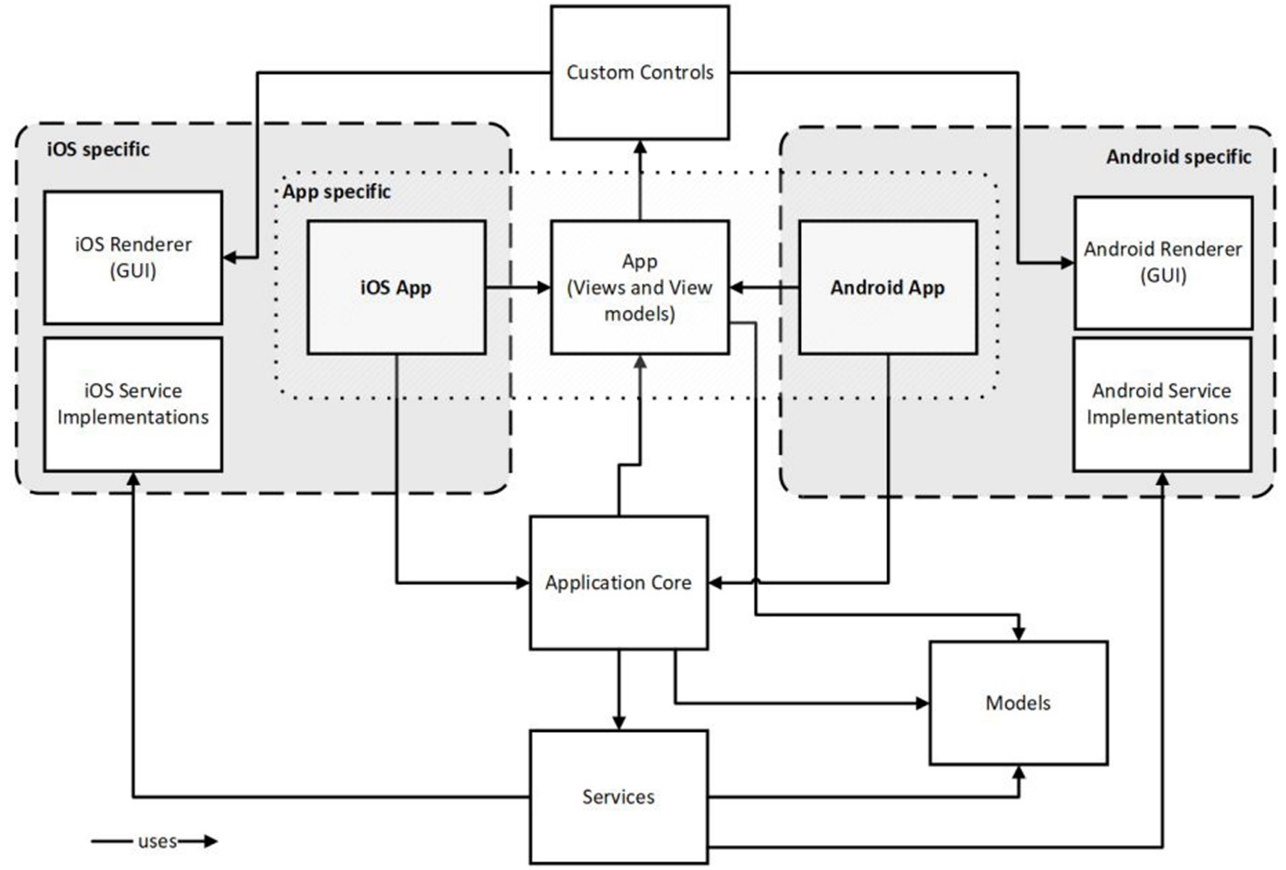


FIGURE 9 | Software architecture of the mobile app.

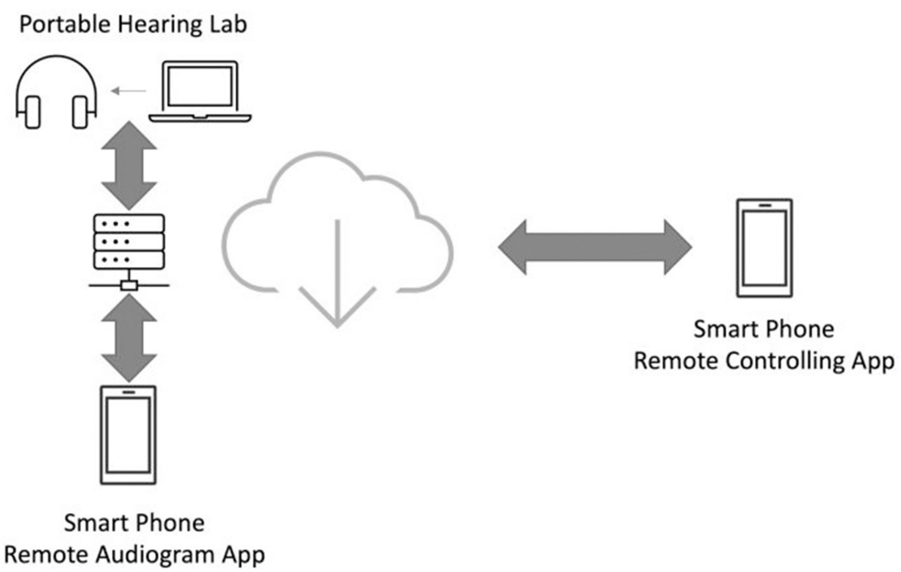


FIGURE 10 | The architecture of the remote audiogram measurement performed using the remote AudioPSS-app.

Machine Learning for Acoustic Scene Classification

Due to the limited computational resources of HDs the range of acoustic scenes that can be recognized and programmatically treated inside the HD remains limited. In this paper, we introduce a machine learning approach for the recognition of acoustic scenes by incorporating the provided infrastructure for connected HDs to continuously optimize a machine learning model.

In order to train such models, a large training data set is required. In the proposed approach based on the IoT, these sound samples will directly be recorded by the HD users, which raises the question about how to respect the privacy of the HD users. Therefore, we cannot use the sound samples directly to train a machine learning system. State-of-the-art machine learning approaches extract some lower dimensional features (23), which carry the essence of the underlying sound sample. By doing this a simpler machine learning architecture with a smaller number of parameters can be defined. Moreover, we have to make sure that these features cannot be used to reconstruct the underlying sound samples. For this reason, we computed a set of privacy preserving features for training and testing of the machine learning system.

Privacy Preserving Audio Features

State-of-the-art scene classification approaches use variants of spectral features in the logarithmic domain, e.g., log-mel spectra or mel frequency cepstral coefficients, for acoustic scene classification (24). However, the underlying audio signals can be reconstructed from these features. For this reason, we decided to use sparse decomposition techniques (25, 26) for encoding the audio signals as sparsely as possible, which account for two important aspects. On the one hand, the characteristics of the underlying acoustic scenes, where these audio signals were recorded, are preserved. On the other hand, an intelligible reconstruction of the underlying audio signals is not possible.

In order to decompose a given audio signal, a Gammatone filterbank (27) is constructed, which is composed of a given number of M Gammatone filters. A Gammatone filter is characterized by its center frequency and bandwidth. It has been shown that the impulse response of the Gammatone filters is similar to the basilar membrane measured in cats (28). Therefore, Gammatone filters are frequently used for modeling the human auditory system.

A sparse signal representation of a given audio signal x at time t can be approximated as a linear superposition of K Gammatone filters γ_{f_k,t_k} called atoms with center frequency f_k and time offset t_k, selected from a Gammatone filterbank with M filters is defined as follows:

$$x(t) = \Sigma_{k=1}\, a_k \gamma_{f_k,t_k}(t) + \varepsilon(t), \tag{1}$$

where a_k is the amplitude of the corresponding Gammatone filter and $\varepsilon(t)$ is the residual. As a decomposition algorithm, we employ the matching pursuit algorithm proposed by Mallat and Zhang (29). The amplitude, center frequency and time offset of the K Gammatone filters $\gamma_{f_m,\,t^*}$ are defined to be the ones, which maximally correlate with the residual:

$$\left(f_k, t_k\right) = argmax < \varepsilon_k(t), \gamma_{f_m,t^*} > \tag{2}$$

The sparse decomposition of a given audio signal in comparison with itself is shown in **Figure 11**.

For the sparse decomposition of the audio signals, we generated a Gammatone filterbank of M = 64 filters and computed K = 1,024 Gammatone atoms for each audio signal of 10 s.

Privacy Preserving Acoustic Scene Classification

In order to recognize a diversity of acoustic scenes automatically, we defined a deep learning architecture using convolutional neural networks (CNNs). In the proposed approach, we aim to train the CNNs on a cloud server in an offline mode and download the pre-trained network to each HD connected to the proposed system.

The defined network architecture consists of four convolutional layers followed by two fully connected layers. The first convolutional layer consists of 32 filters of size 3 × 3. In the following layers the number of filters is doubled with the filter size kept constant. Between each convolutional layer batch normalization and max pooling of size 2 × 2 is applied. After the batch normalization, the rectified linear unit is applied as an activation function. The first fully connected layer consists of 280 neurons. After the first fully connected layer, we apply another rectified linear unit. In order to prevent the network from overfitting, there is a dropout layer with 50% dropout probability between the two fully connected layers. The second fully connected layer is at the same time the output unit consisting of 14 neurons, which is equal to the number of acoustic scenes within the database used for training and testing the system.

EVALUATION

In the following section we present the evaluation of the RF service from a technical as well as from a social viewpoint. In addition we technically evaluated the Acoustic Scene Classification.

Technical Evaluation

Remote Fitting Evaluation (Audiometry App)

In order to evaluate the performance of the RF concept, we performed subjective measurement experiments with 18 normal hearing listeners aged between 20 and 54 years. An audiologist performed three pure tone audiogram measurements with each test subject. Subsequent to these measurements, each test subject performed three audiogram measurements using the remote audiogram app. For both measurement paradigms five frequencies (500, 1,000, 2,000, 4,000, 6,000 Hz) were measured using the audiometry headphones (Sennheiser HDA 200). Both paradigms have a measurement resolution of 5 dBs, which means that in both paradigms the hearing thresholds are increased with 5 dBs steps.

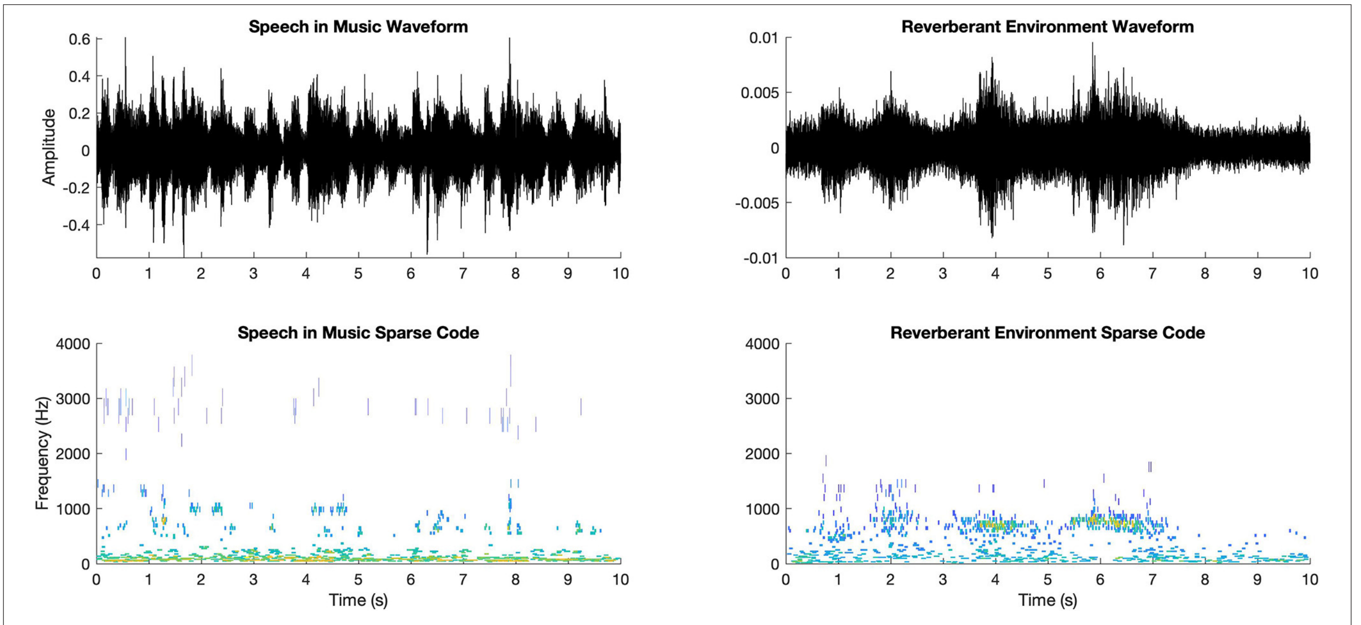


FIGURE 11 | Two sample waveforms (upper row) and the corresponding sparse decompositions (lower row).

In **Figure 12**, the results of the audiogram measurements as a difference between the pure tone audiogram and remote audiogram are shown. The zero line indicates that the measurement results of the pure tone audiogram and the remote audiogram are equal. Negative deviation means that the results obtained with the standard pure tone audiogram indicate higher hearing thresholds, e.g., less hearing loss than those obtained with the remote application.

As one can easily see, the remote audiogram app on average measures hearing thresholds around 5 dB lower than the corresponding pure tone audiogram measurements. The reason for this discrepancy can be explained by the differences between the reaction times in hitting a button or clicking on a smartphone screen. The test-retest reliability in clinical audiogram measurements also typically is in the 5 dB-range. Therefore, the results of the remote audiogram app approximate the pure tone measurements with sufficient accuracy.

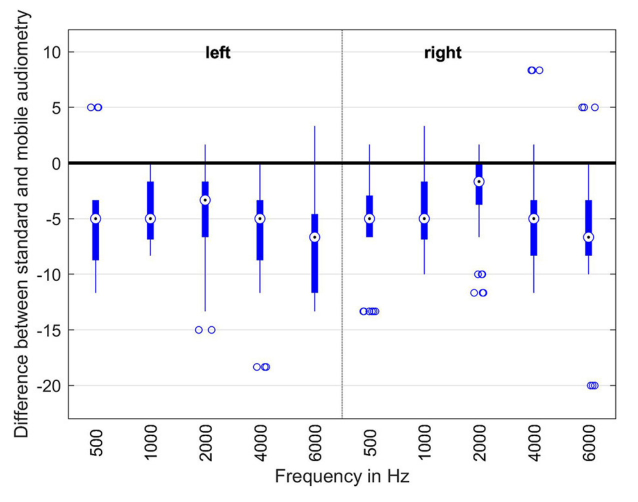


FIGURE 12 | Evaluation of the different audiogram measurement methods. The boxes indicate the measurement points within the 25 and 75% percentiles. The dots within the boxes are the median of all the measurement points for this particular frequency. The thin lines indicate the minimum and maximum measurement points. The dots below and above the thin lines are the outliers.

Evaluation of the Acoustic Scene Classification

In order to train supervised machine learning models, a large training data set is required. While the technical infrastructure for the connected HDs was still in the making, we recorded a binaural data set Hearing Aid Research Data Set (HeAR-DS) (30) to provide a reliable basis to train a supervised machine learning model to perform acoustic scene classification for HDs.

HeAR-DS contains fourteen acoustic scenes as shown in **Figure 13** that were defined in close cooperation with audiologists and with the German hearing aid manufacturer and project partner *audifon* to cover acoustic scenes relevant in everyday life of HD users. These acoustic scenes can be categorized in three groups. The *speech* group consists inherently of speech, where the target speaker changes simultaneously and hence is hard to determine. Therefore, *Cocktail party* and *Interfering speakers* belong to that group. The acoustic scenes in the *background* group contain pure background noises (i.e., no significant speech components). Finally, the *speech in background* group consists of acoustic scenes with a target speaker embedded in one of those backgrounds. In order to create the acoustic

Speech			
Cocktail Party	667		
Interfering Speakers	1481		
Background		**Speech in Background**	
In Traffic	530	Speech in Traffic	470
In Vehicle	584	Speech in Vehicle	511
Music	1496	Speech in Music	1495
Quiet Indoors	525	Speech in Quiet Indoors	426
Reverberant Environment	315	Speech in Reverberant Environment	692
Wind Turbulence	595	Speech in Wind Turbulence	439

FIGURE 13 | The acoustic scenes of the HeaR-DS research data set is shown. The number next to each acoustic scene indicates the number of samples in the corresponding acoustic scene. Each sample is 10 s. long.

scenes in this group, background signals were manually mixed with speech signals.

The proposed CNN was randomly initialized, trained and tested 10 times. For the optimization of the weights the Adam optimizer was used. As the size of the training data set is large, we use mini-batches of size 24 samples to train the network. For each training session, a maximum of 250 iterations was performed. An early stopping mechanism was also integrated so that after 10 iterations without any improvement in the loss, the training was terminated. We randomly split the HeAR-DS into training (70%) and test data sets (30%) so that there was no overlap between the sets during each training and test phase. The class wise test accuracies, the confusion matrix and its standard deviation are shown in **Figure 14**.

The average overall accuracy of all 10 training sessions is 87.26%. All acoustic scenes were successfully classified with high classification accuracies except two acoustic scenes. *Cocktail party* and *reverberant environments* were frequently mixed up with one another. One can easily recognize this phenomenon on the confusion matrix. We hypothesize that the reason for this result is caused by the fact that the sparse codes of these two acoustic scenes are highly similar to one another. There is another slight confusion between the acoustic scenes *speech in traffic* and *speech in vehicle*. As the recordings for the acoustic scene *In Vehicle* were made within different cars in traffic, this confusion is acceptable. The other acoustic scenes were classified with very high accuracies and with very little variance among different training and test sessions.

Social Evaluation of Remote Fitting

Additionally to the successful evaluation of the technical implementation of RF we also investigated the user preferences and usability regarding the different RF scenarios. The findings of the requirement analysis reveal that the design and interaction between audiologists, HD users and the technical system can be shaped in different ways. The preferences of both audiologists and users suggest a hybrid service process where audiologists, HD users and the technical system interact closely. However, it remains unclear whether such service design exceeds other service interaction processes in terms of user acceptance as well as subjective and even objective hearing quality. Therefore, we experimentally evaluate different RF scenarios with 18 experienced and inexperienced HD users between 59 and 80 years of age (M = 73 years, SD = 5.4 years, six female) with a mild to moderate hearing loss. The objectives of this study were to investigate the preferences of hearing-impaired participants for their HD fitting and whether these preferences influence the perception, measured by objective and subjective methods, when HD fittings are identical for each scenario.

In this study, participants were fitted with audifon HAs lewi R. Two programs were implemented into the HAs: the standard setting program 1 and a "comfort in noise"-setting as program 2. Three different fitting scenarios were investigated in this study: fitting adjustments by (1) an audiologist, (2) fitting by the AI system, and (3) fitting by both the audiologist and AI system. To ensure comparability of the measurement results, the modification of the fitting in all three scenarios was realized by switching from preset program 1 to program 2. For the experiment, three different acoustically complex listening situations were set up in the laboratory using eight circularly arranged loudspeakers, in which the subject had the task to repeat the sentences presented and to fill in a questionnaire afterwards. For the fitting adjustment, the probands were asked to either verbally describe their hearing problems in the listening situation (scenario 1) or enter them into the app for the AI system (scenarios 2 and 3). Subsequently, the setting in the HAs was changed by switching programs unnoticed by the participant and the test was repeated.

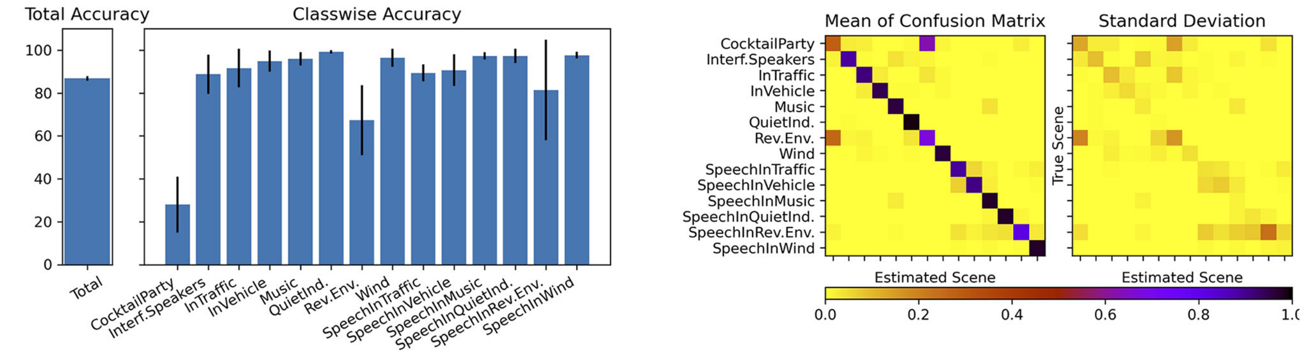


FIGURE 14 | Class wise test accuracies **(left)** and matrix plots showing mean and standard deviation of the confusion matrix **(right)**.

The objective speech intelligibility tests performed in each scenario turned out very similarly for all three fitting scenarios (*mean word scoring in %:* scenario 1: 69%, scenario 2: 70%, scenario 3: 72%). The subjective questionnaires also show that the study participants did not notice any perceptible differences in, among other things, speech understanding, listening effort, or sound based on the different fitting scenarios. Furthermore, participants were asked to rate *"Which type of remote fitting would you prefer?"* (ranking from 1 to 3).

The results show that 38.9% of the participants prefer the mixed option (scenario 3: AI + audiologist) for adjusting their HAs, and 38.9% state that the audiologist-based fitting is their favorite. Only 22.3% preferred the scenario with an AI-only automated fitting. Overall, 72.2% of participants ranked the mixed option first or second, whereas only 66.7% said this was the case for the audiologist-only option. **Figure 15** depicts the preference ranking of the three fitting scenarios. In conclusion, the HD users did not notice any difference in the objective parameters but have a clear preference for a technical solution in which the audiologist is still actively involved. This is seemingly in contrast to the data from the requirements analysis, where the HD users preferred the instantaneous RF through a service center (see section Service Innovation Acceptance of Hearing Device Users). However, those over 70 years old rated the classical consulting on site (no RF) as the best alternative. And in the explorative interviews, HD users stated that the personal component should not be replaced. Since the objective evaluation of the RF scenarios did not differ, the tendency toward a technical solution, supported by an audiologist, was strengthened.

CONCLUSION AND FUTURE OUTLOOK

In this paper, we presented our concept of an integrated service platform for digital service innovations in audiology. At the beginning of the project, we identified the requirements of HD users and of audiologists on different digital services. Based on the user needs we focused on the technical infrastructure needed to meet those requirements. In this context we presented the system architecture of the connected HD and the acoustic scene classification through machine learning as well as the overall development and implementation of the RF infrastructure. To conclude the user-centered developments, we evaluated the acoustic scene classification from a technical standpoint and the RF solution from both technical and social standpoints. With our interdisciplinary study, we developed digital service innovations based on connected HDs which converge on an integrated service platform. With our platform approach we focused on advanced 1:1 and 1:N services as RF and acoustic scene classification. For the participatory innovation process we combined advanced technology developments with the important role of human interaction between audiologists and HD users.

We found out that the preferences of HD users, audiologists and manufacturers differ with regard to potential digital services. The HD users highlighted preferences for self-services, followed by RF services, and for services on the 1:N domain like the consideration of individual hearing preferences through scene classification and machine learning, whereas the audiologists approved particularly an interactive auditory training, and the development of an acoustic scene classification solution. An audiologist-centered RF solution was mentioned by both user groups. Based on the comparison of the user preferences and the patent analysis of service-related technological developments, we found that more disruptive service innovations mainly originate from manufacturers. With the subsequent user surveys the trends of the explorative analysis were supported: The audiologists preferred more personal data-driven services, whereas the HD users preferred self-services and services on the 1:N domain. For RF services it was particularly noticeable that they were rated quite differently by the user groups. Audiologists clearly preferred time-delayed RF by an audiologist, whereas many HD users preferred the instantaneous RF. Nevertheless, the personal component still plays a role, as evidenced by the decision of those over 70 against RF.

We developed a hardware/software architecture that can support all service dimensions (1:1, 1:N and N:M) underlying the project. Relying on state-of-the-art database development and communication protocols, we ensure that this architecture is both safe and scalable. Indeed, the architecture design allows to easily increase the number of users, both HD users and audiologists, as well as the connections between them, e.g.,

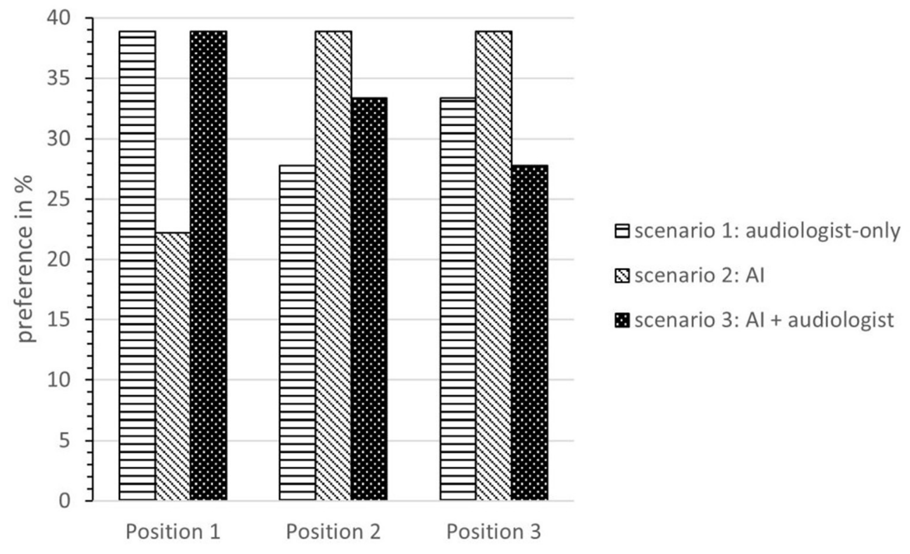


FIGURE 15 | Preference ranking of the three fitting scenarios.

allowing each audiologist to support a larger number of HD users. Consequently, though we tested this technical solution only on a limited number of participants, it could easily be applied to a much larger user group. One of the main applications of this architecture is RF, whose proof of concept has been evaluated in the form of a remote audiogram measurement. Measuring the pure tone audiogram remotely was a key step for realizing the RF approach. The evaluation studies have shown that our proposed remote audiogram app is able to measure the hearing thresholds within the normal test-retest reliability of audiogram measurements. It also enables a first fit of the HDs remotely and is therefore a convenient tool for the RF process.

For the social evaluation we experimentally evaluated three different fitting scenarios: fitting adjustments (1) by an audiologist, (2) by the AI system, and (3) by both the audiologist and AI system. The objective speech intelligibility tests performed in each scenario turned out very similarly for all three fitting scenarios. In the subjective evaluation of the different RF scenarios, the mixed scenario scored the best, which shows that HD users are open toward a technical solution but still prefer the integration of a personal component.

As our results show, the development of a service platform for digital services in healthcare requires interdisciplinary collaboration. With our project team we combined expert knowledge from audiology, information technology, and service research. With the combination of social and engineering science, we made valuable implications for the further developments of digital services in the field of tele-audiology. Moreover, with multiple stakeholder integration, we showed how a service platform can connect advanced technological developments to users and audiologists. For manufacturers, we highlighted the necessity of integrating audiologists and HD users in the development of future service processes and offered a system architecture that integrates all relevant stakeholders on the same service platform. For audiologists, we raise awareness for the topic of digitalization in audiology and show that audiologists should take an active part in the development of digital service innovations. The early alignment and participation in service innovation processes ensures the sovereignty and importance of audiologists in future service processes. But most of all, the platform approach opens up new alternative ways of value creation. The re-programmability and homogeneity of digital innovations can be used to create service innovations beyond the core field of audiology. For instance, cross-industry innovations on the basis of the hearing data can be used for an acoustic index by evaluating the acoustic quality of restaurants.

The connected HDs paradigm technically allows to perform many measurements, which typically take place on site remotely without the physical presence of the HD user. These kinds of measurements offer a large flexibility to the HD users as well as to the audiologists. The results of the remote audiogram study showed that HD users can measure audiograms remotely and perform the first fit for their HDs based on the measurement results themselves. The possibility to monitor and accompany the measurement by an audiologist remotely reduces the risk and gives the HD user more confidence. Currently, the remote audiogram app depicts only one possible usage of the technology of connected HDs concerning the 1:1 services scenario. However, many other measurements can be handled using the same infrastructure in the future.

Acoustic scene classification indicates a possible application of the 1:N services scenario using the connected HA technology. While increasing the speech intelligibility, preserving the privacy of the HD users has the highest priority in the HA signal processing. The sparse features encode the essence of an audio signal, while preventing a reconstruction of the audio. Hence, these features are very well-suited for tasks, which require collection of data for post-processing from individual HD users.

The results (87%) of the conducted acoustic scene classification experiments showed that the proposed sparse features with deep learning architectures turned out to perform in the same range as the state-of-the-art log-mel features (85%) (30), for this task. A slim CNN architecture provides sufficient accuracy for recognizing an acoustic scene. The recognition happens every 10 s, which is frequent enough for a smooth performance of a HA, because the acoustic scene typically does not change more frequently.

The hardware/software architecture should be further evaluated in a wider field study taking advantage of its scalability. This would allow us to confirm the robustness of the developed architecture as well as to confirm or infirm the conclusion made in this paper. Additionally, such a study would allow us to prove the feasibility of the acoustic index and to measure the interest of the end users for this service. Moreover, a usability study could allow us to improve the user interface of the mobile application to make it easier to use. Exemplary usability evaluation methods are a "cognitive walkthrough" where experts evaluate the ease of use as well as a "usability testing" of the demonstrator by potential end users. The importance of user friendliness of such an application on the acceptance of the provided services indeed remains to be quantified.

To further quantify the results of the first RF evaluation we currently survey HD users toward their intention to use different RF options, with regard to the status quo during the Covid-19 pandemic but also in perspective after the pandemic. Future research can also focus on the use and development of sensor data, to further optimize connected digital services beyond the scope of audiology, e.g., wearable devices like fitness tracker. In this context future research can extend the core field of hearing improvement to other use cases in the context of telemedicine, which has become increasingly important since the Covid-19 pandemic.

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://www.hoertech.de/en/research/open-tools-for-science/417-hoertech/englisch-ht/projects-ht/open-mha/770-hear-ds.html.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Research Ethics Committee of the Carl von Ossietzky University of Oldenburg for non-medical research projects. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ML, CG, and MKr collected the data and performed the data analysis for the requirement analysis. BC and UM implemented the architecture used for the project. BC focused on the backend while UM focused on the development of the mobile application. KA collected the data and performed the data analysis for the technical implementation and evaluation of Remote Fitting and Acoustic Scene Classification. MKr collected the data and performed the data analysis for the social evaluation. ML coordinated the manuscript draft and the service sections. KA coordinated the technical sections. ML, CG, KA, MKr, BC, and UM wrote the main parts of the article. MT and MKi revised and completed the article. CS designed the methodological approach, supervised the project and revised the final manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was part of the Audio-PSS project (www.audio-pss.de) funded by the German Federal Ministry of Education and Research (BMBF) within the Program Innovations for Tomorrow's Production, Services, and Work and managed by the Project Management Agency Karlsruhe (PTKA). The grant numbers are: 02K16C200, 02K16C201, 02K16C202, 02K16C203, and 02K16C204.

ACKNOWLEDGMENTS

We thank the *Hörzentrum Oldenburg gGmbH* for providing participants and for their support during the evaluation measurements. Furthermore, we would like to thank our project partners, who did not explicitly contribute to the article, but made valuable contributions to the research project: Jörg-Hendrik Bach, Marco Eichelberg, Andreas Hüwel, Heiner Lütjen, Markus Meis, Frerk Müller- von Aschwege, and Hartmut Richter.

REFERENCES

1. *WHO: World Report on Hearing*. Geneva: World Health Organization. Licence: CC NY-NC-SA 3.0 IGO. (2021).
2. Searchfield GD, Sanders PJ, Doborjeh Z, Doborjeh M, Boldu R, Sun K, et al. A state-of-art review of digital technologies for the next generation of tinnitus therapeutics. *Front Digit Health.* (2021) 3:724370. doi: 10.3389/fdgth.2021.724370
3. Christensen JH, Saunders GH, Havtorn L, Pontoppidan NH. Real-world hearing aid usage patterns and smartphone connectivity. *Front Digit Health.* (2021) 3:722186. doi: 10.3389/fdgth.2021.722186
4. Verganti R, Vendraminelli L, Iansiti M. Innovation and design in the age of artificial intelligence. *J Prod Innov Manag.* (2020) 37:212–27. doi: 10.1111/jpim.12523
5. Hinings B, Gegenhuber T, Greenwood R. Digital innovation and transformation: an institutional perspective. *Inform Org.* (2018) 28:52–61. doi: 10.1016/j.infoandorg.2018.02.004
6. Yoo Y, Henfridsson O, Lyytinen K. Research commentary —the new organizing logic of digital innovation: an agenda for information systems research. *Inform Syst Res.* (2010) 21:724–35. doi: 10.1287/isre.1100.0322
7. Kallinikos J, Aaltonen A, Marton A. The ambivalent ontology of digital artifacts. *MISQ.* (2013) 37:357–70. doi: 10.25300/MISQ/2013/37.2.02

8. Matzner M, Büttgen M, Demirkan H, Spohrer J, Alter S, Fritzsche A, et al. Digital transformation in service management. *SMR.* (2018) 2:3–21. doi: 10.15358/2511-8676-2018-2-3
9. Vargo SL, Lusch RF. Service-dominant logic: continuing the evolution. *J Acad Mark Sci.* (2008) 36:1–10. doi: 10.1007/s11747-007-0069-6
10. Cauchi B, Eichelberg M, Hüwel A, Adiloglu K, Richter H, Typlt M. Hardware/software architecture for services in the hearing aid industry. In: *IEEE 20th International Conference on e-Health Networking, Applications and Services* (2018). p. 1–6. doi: 10.1109/HealthCom.2018.8531138
11. Faix A. *Patente im strategischen Marketing: Sicherung der Wettbewerbsfähigkeit durch systematische Patentanalyse und Patentnutzung*. Berlin: E. Schmidt (1998). 466 p.
12. Flueckiger M, Phonak AG. *Method for Providing Distant Support to a Personal Hearing System User and System for Implementing Such a Method*. WO2013020594A1. Geneva: World Intellectual Property Organization (2011).
13. Bjelosevic A, Oticon AS. *Monitoring System for a Hearing Device*. EP3035710A2. Munich: European Patent Office (2016).
14. Reinlein C, Sivantos PTE. *Verfahren zur Ermittlung von trägerspezifischen Nutzungsdaten eines Hörgeräts, Verfahren zur Anpassung von Hörgeräteeinstellungen eines Hörgeräts, Hörgerätesystem und Einstelleinheit für ein Hörgerätesystem*. DE102015203288. Munich: Deutsches Patent- und Markenamt (2015).
15. Bhat VK, Oticon AS. *Hearing Aid Device With Speech Control Functionality*. US2018061411A1. Alexandria, VA: United States Patent (2017).
16. Hauschultz SI, Rasmussen KBO, Oticon AS. *Body-Worn Control Apparatus for Hearing Devices*. US2014126759A1. Alexandria, VA: United States Patent (2013).
17. Aschoff S, Sivantos PTE. *Verfahren und Einrichtung zur Einstellung eines Hörhilfegeräts*. EP3288293A1. Munich: European Patent Office (2018).
18. Adner R. When are technologies disruptive? A demand-based view of the emergence of competition. *Strat Mgmt J.* (2002) 23:667–88. doi: 10.1002/smj.246
19. Cummins N, Schuller BW. Five crucial challenges in digital health. *Front Digi. Health.* (2020) 2:536203. doi: 10.3389/fdgth.2020.536203
20. Markides C. Disruptive innovation: in need of better theory. *J Prod Innov Manag.* (2006) 23:19–25. doi: 10.1111/j.1540-5885.2005.00177.x
21. Pavlovic C, Kassayan R, Prakash SR, Kayser H, Hohmann V, Atamaniuk A. A high-fidelity multi-channel portable platform for development of novel algorithms for assistive listening wearables. *J Acoust Soc Am.* (2019) 146:2878. doi: 10.1121/1.5136986
22. Kayser H, Herzke T, Maanen P, Pavlovic C, Hohmann V. Open master hearing aid (openMHA)—An integrated platform for hearing aid research. *J Acoust Soc Am.* (2019) 146:2879. doi: 10.1121/1.5136988
23. Mesaros A, Heittola T, Virtanen T. "A multi-device dataset for urban acoustic scene classification," in: *Detection and Classification of Acoustic Scenes and Events Workshop (DCASE 2018)*.
24. Barchiesi D, Giannoulis D, Stowell D, Plumbley MD. Acoustic scene classification: classifying environments from the sounds they produce. *IEEE Signal Process Mag.* (2015) 32:16–34. doi: 10.1109/MSP.2014.2326181
25. Smith E, Lewicki MS. Efficient coding of time-relative structure using spikes. *Neural Comput.* (2005) 17:19–45. doi: 10.1162/0899766052530839
26. Adiloglu K, Annies A, Wahlen E, Purwins H, Obermayer K. A graphical representation and dissimilarity measure for basic everyday sound events. *IEEE Trans Audio Speech Lang Process.* (2012) 20:1542–52. doi: 10.1109/TASL.2012.2184752
27. Patterson RD, Holdsworth J. A functional model of neural activity patterns and auditory images. *Adv Speech Hearing Lang Process.* (1996) 3:547–63.
28. Carney LH, Yin TC. Temporal coding of resonances by low-frequency auditory nerve fibers: single-fiber responses and a population model. *J Neurophysiol.* (1988) 60:1653–77. doi: 10.1152/jn.1988.60.5.1653
29. Mallat SG, Zhang Z. Matching pursuits with time-frequency dictionaries. *IEEE Trans Signal Process.* (1993) 41:3397–415. doi: 10.1109/78.258082
30. Hüwel A, Adiloglu K, Bach J-H. Hearing aid research data set for acoustic environment recognition. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2020)*. p. 706–10.

Author Disclaimer: The authors are responsible for the contents of this publication.

Conflict of Interest: MKi was employed by the company KIND GmbH & Co. KG. UM and MT were employed by the company audifon GmbH & Co. KG.

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.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 29 September 2021
doi: 10.3389/fnins.2021.724007

Auditory Tests for Characterizing Hearing Deficits in Listeners With Various Hearing Abilities: The BEAR Test Battery

Raul Sanchez-Lopez[1,2*], Silje Grini Nielsen[1], Mouhamad El-Haj-Ali[3], Federica Bianchi[1,4], Michal Fereczkowski[1,3,5], Oscar M. Cañete[1], Mengfan Wu[3,5], Tobias Neher[3,5], Torsten Dau[1*†] and Sébastien Santurette[1,6*†]

[1] Hearing Systems Section, Department of Health Technology, Technical University of Denmark, Kgs. Lyngby, Denmark, [2] Interacoustics Research Unit, Kgs. Lyngby, Denmark, [3] Institute of Clinical Research, University of Southern Denmark, Odense, Denmark, [4] Oticon Medical, Smørum, Denmark, [5] Research Unit for ORL-Head & Neck Surgery and Audiology, Odense University Hospital & University of Southern Denmark, Odense, Denmark, [6] Centre for Applied Audiology Research, Oticon A/S, Smørum, Denmark

OPEN ACCESS

Edited by:
Dennis L. Barbour,
Washington University in St. Louis,
United States

Reviewed by:
Anna Warzybok,
University of Oldenburg, Germany
Ken W. Grant,
Walter Reed National Military Medical
Center, United States

***Correspondence:**
Raul Sanchez-Lopez
rsalo@dtu.dk
Torsten Dau
tdau@dtu.dk
Sébastien Santurette
sesn@oticon.com

† These authors have contributed equally to this work and share last authorship

Specialty section:
This article was submitted to
Auditory Cognitive Neuroscience,
a section of the journal
Frontiers in Neuroscience

Received: 11 June 2021
Accepted: 11 August 2021
Published: 29 September 2021

Citation:
Sanchez-Lopez R, Nielsen SG, El-Haj-Ali M, Bianchi F, Fereczkowski M, Cañete OM, Wu M, Neher T, Dau T and Santurette S (2021) Auditory Tests for Characterizing Hearing Deficits in Listeners With Various Hearing Abilities: The BEAR Test Battery. Front. Neurosci. 15:724007. doi: 10.3389/fnins.2021.724007

The Better hEAring Rehabilitation (BEAR) project aims to provide a new clinical profiling tool—a test battery—for hearing loss characterization. Although the loss of sensitivity can be efficiently measured using pure-tone audiometry, the assessment of supra-threshold hearing deficits remains a challenge. In contrast to the classical "attenuation-distortion" model, the proposed BEAR approach is based on the hypothesis that the hearing abilities of a given listener can be characterized along two dimensions, reflecting independent types of perceptual deficits (distortions). A data-driven approach provided evidence for the existence of different auditory profiles with different degrees of distortions. Ten tests were included in a test battery, based on their clinical feasibility, time efficiency, and related evidence from the literature. The tests were divided into six categories: audibility, speech perception, binaural processing abilities, loudness perception, spectro-temporal modulation sensitivity, and spectro-temporal resolution. Seventy-five listeners with symmetric, mild-to-severe sensorineural hearing loss were selected from a clinical population. The analysis of the results showed interrelations among outcomes related to high-frequency processing and outcome measures related to low-frequency processing abilities. The results showed the ability of the tests to reveal differences among individuals and their potential use in clinical settings.

Keywords: audiology, hearing loss, loudness, binaural processing, speech perception, spectro-temporal resolution, auditory profile

1. INTRODUCTION

In current clinical practice, hearing loss is diagnosed mainly on the basis of pure-tone audiometry (ISO 8253-1, 2010). The audiogram helps differentiate between conductive and sensorineural hearing losses and can characterize the severity of the hearing loss from mild to profound. However, the pure-tone audiogram only assesses the sensitivity to simple sounds, which is not necessarily related to listening abilities at supra-threshold sound pressure levels (e.g., a person's ability to discriminate speech in noise).

Pure-tone audiometry is often complemented by speech audiometry (ISO 8253-3, 2012), which is a test typically performed in the form of word recognition performance in quiet (Anderson et al., 2018). Although this test can provide information about supra-threshold deficits (Gelfand, 2009), measurements of speech understanding in noise have been found more informative (Nilsson et al., 1994; Killion et al., 2004). Since improving speech intelligibility is usually the main goal of successful hearing rehabilitation, several auditory factors affecting speech intelligibility in noise have been investigated (e.g., Glasberg and Moore, 1989; Houtgast and Festen, 2008; Strelcyk and Dau, 2009). Audibility (in conditions with fluctuating maskers), frequency selectivity (in conditions with stationary noise), and temporal processing acuity (in conditions with speech interferers) have been identified as important factors affecting speech reception thresholds in noise when using meaningful sentences as speech material (e.g., Rhebergen et al., 2006; Oxenham and Simonson, 2009; Johannesen et al., 2016; Desloge et al., 2017)[1]. Thus, a hearing evaluation that goes beyond pure-tone sensitivity and speech intelligibility in quiet would be expected to provide a more accurate characterization of a listener's hearing deficits.

In Denmark, the Better hEAring Rehabilitation (BEAR) project was initiated with the aim of developing new diagnostic tests and hearing-aid compensation strategies for audiological practice. Although the assessment of individual hearing deficits can be complex, new evidence suggests that the perceptual consequences of a hearing loss can be characterized effectively by two types of hearing deficits, defined as "auditory distortions" (Sanchez-Lopez et al., 2018). By analyzing the outcomes of two previous studies (Johannesen et al., 2016; Thorup et al., 2016) with a data-driven approach, Sanchez-Lopez et al. (2018) identified high-frequency (HF) hearing loss as the main predictor of one of the distortions, whereas the definition of the second type of distortion was inconclusive. The inconclusiveness in the prediction of the second distortion was most likely due to differences between the two studies in terms of hearing loss profiles and outcome measures. Here, a new dataset was therefore collected based on a heterogeneous group of listeners with audiometric hearing losses ranging from very mild to severe and with a large range of audiometric profiles. To that end, the most informative tests resulting from the analysis of Sanchez-Lopez et al. (2018) were included, together with additional auditory tests that had shown potential for hearing profiling in other previous studies. The tests included in the current study are referred to as the BEAR test battery.

The characterization of hearing deficits beyond the audiogram was considered in several earlier studies (e.g., Saunders et al., 1992; Santurette and Dau, 2012; Lecluyse et al., 2013; Brungart et al., 2014; Esch and Dreschler, 2015; Rönnberg et al., 2016). Among them, the HEARCOM project (Vlaming et al., 2011) proposed an extended hearing profile formed by the results of several behavioral tests. These tests targeted various auditory domains, such as audibility, loudness perception, speech perception, binaural processing, and spectro-temporal resolution, as well as a test of cognitive abilities. Importantly, while the auditory domains considered in the BEAR test battery are similar to the ones considered in the HEARCOM project, the BEAR project aims to additionally classify the patients in subcategories and to create a link between hearing capacities and hearing-aid parameter settings.

The tests included in the BEAR test battery were chosen based on the following criteria: (1) There is evidence from the hearing research literature that the considered test is informative (i.e., it provides information about the individual hearing deficits) and reliable (i.e., the result of the test does not vary over time). (2) The outcomes of the test may be linked to a hearing-aid fitting strategy. (3) The outcome measures are easy to interpret and to explain to the patient. (4) The task is reasonably time-efficient or can be suitably modified to meet this requirement (e.g., by changing the test paradigm or developing an out-of-clinic solution). (5) The test implementation can be done with equipment available in clinics. (6) The tasks are not too demanding for patients and clinicians. (7) Tests with several outcome measures are prioritized. (8) The language-independent tests are also prioritized. Although a large list of tests candidates was considered in an early stage, discussions among the authors and other BEAR partners to shorten the list led to the current proposal. Some classical tests were discarded because a suitable alternative was more promising. For example, the short-increment sensitivity index (SISI test) was discarded since there was a more informative candidate for measuring loudness perception.

The selected test battery included measures of audibility, loudness perception, speech perception, binaural processing abilities, spectro-temporal modulation (STM) sensitivity, and spectro-temporal resolution. It was implemented and tested in older listeners with different hearing abilities (from mild to severe hearing losses) as a representative sample of the population of hearing-aid user candidates mainly affected by age-related hearing loss. The goals of the study were as follows: (1) To collect reference data from a representative sample of HI listeners for each of the selected tests, (2) to analyze the test–retest reliability of these tests, (3) to analyze the relationships between the different outcome measures, and (4) to propose a version of the test battery that can be implemented in hearing clinics.

2. OVERVIEW OF THE TEST BATTERY

The test battery consisted of ten tests (9 tests besides the pure-tone audiometry). The outcomes of the proposed tests are divided into six categories. **Table 1** shows the tests and the corresponding auditory domains or categories. For convenience, the domains spectro-temporal modulation sensitivity and spectro-temporal resolution are presented together in the category spectro-temporal processing. The following sections

[1]The factors identified correspond to the authors' conclusions based on cited references. For example, Johannesen et al. (2016) identified the basilar membrane compession as a predictor of speech intelligibility in stationary noise and temporal processing as a predictor of speech-in-speech intelligibility. Rhebergen et al. (2006), Oxenham and Simonson (2009), and Desloge et al. (2017) identified the audibility of the soft speech sounds in the presence of fluctuating maskers as a crucial factor for speech intelligibility.

TABLE 1 | List of the tests included in the BEAR test battery and their corresponding auditory domains.

Test name	Category	Variables
Pure-tone audiometry[a]	Audibility	AUD_{LF}, AUD_{HF}
Extended audiometry at high frequencies [b]		FLFT
Adaptive categorical		HTL_{LF}, HTL_{HF}
loudness scaling[c]	Loudness	MCL_{LF}, MCL_{HF}
	perception	$DynR_{LF}$, $DynR_{HF}$
		$Slope_{LF}$, $Slope_{HF}$
Word recognition scores[d]	Speech	SRT_{Q}, maxDS
Hearing in noise test[e]	Perception	SRT_{N}, $SScore^{+4dB}$
Spectro-temporal modulation test[i]		$sSTM_{8}$, $sSTM_{4k}$
	Spectro-	$fSTM_{8}$, $fSTM_{4k}$
Extended audiometry in noise [j,k,l]	temporal	TiN_{LF},TiN_{HF}
	processing	SMR_{LF}, SMR_{HF},
		TMR_{LF}, TMR_{HF}
Maximum frequency for IPD detection[f]	Binaural	IPD_{fmax}
Binaural pitch[g]	processing	BP_{20}
Extended binaural audiometry in noise[h]	abilities	BMR

LF, Lower-frequencies, HF, higher-frequencies.
AUDx, Pure-tone average at low (x=LF; f $\leq$ 1kHz) or high (x = HF; f > 1kHz) frequencies. // FLFT: Fixed-level frequency threshold at 80 dB SPL. ACALOS outcome variabless are averaged for low (x=LF; f $\leq$ 1kHz) and high (x = HF; f > 1kHz) frequencies. // HTL, Hearing threshold level MCL: Most comfortable level / Slope: Slope of the loudness function / DynR: Dynamic range // SRT_Q: Speech reception threshold in quiet / maxDS: Maximum speech discrimination score. // SRT_N: Speech reception threshold in noise / $SScore^{+4dB}$: Sentence recognition score at +4 dB SNR // sSTM: Sensitivity for detecting a spectro-temporally modulated noise at 20log(m) = -3 dB, where m is the modulation depth / fSTM: Fast version of the STM test (Bernstein et al., 2016) // Extended audiometry outcome measures were measured at 0.5 kHz (x=LF) and at 2 kHz (x=HF) // eAUD-N: Tone detection in TEN noise // TMR: Temporal masking release // SMR: Spectral masking release // IPD_{fmax}: Frequency threshold for detecting an interaural phase difference of 180°. // BP_{20}: Binaural pitch detection scores for 20 presentations // BMR: Binaural masking release.
[a]*ISO 8253-1, 2010;* [b]*Rieke et al., 2017;* [c]*Brand and Hohmann, 2002;* [d]*ISO 8253-3, 2012;* [e]*Nielsen and Dau, 2011;* [f]*Füllgrabe and Moore, 2017;* [g]*Santurette and Dau, 2012;* [h]*Durlach, 1963;* [i]*Bernstein et al., 2016;* [j]*Moore et al., 2000;* [k]*Schorn and Zwicker, 1990;* [l]*van Esch and Dreschler, 2011.*

introduce the experimental methods and present all tests individually. The dataset is publicly available in a Zenodo repository (Sanchez-Lopez et al., 2021b). More details about the method can be found in the **Supplementary Material** in the data repository.

2.1. Reference Data From Younger Normal-Hearing Listeners

Although many of the tests included in the test battery are based on previous studies with normative data, a group of 11 young normal-hearing (NH) listeners were tested in the facilities of the Technical University of Denmark (DTU) and the University of Southern Denmark (SDU) to obtain reference data for this specific implementation of each of the tests. The summary statistics of the outcome variables from **Table 1** for these NH listeners are shown in **Table 2**.

2.2. Time Efficiency of the Test Battery

The examiners kept track of the time used by each of the participants in completing the test battery. In the case of unexpected events (e.g., unexpected or incongruent results), these events were cautiously annotated for later investigation. Regarding the test procedure, additional repetitions of the threshold estimations were needed if: (1) a repetition was considered as an outlier (i.e., a given threshold was greater than three scaled median absolute deviations of the two repetitions); or (2) the responses of the listeners during the tracking procedure were inconsistent or reached the maximum or minimum possible values. In that case, the measurement was considered an invalid or "missing" data point.
The timing of the individual tests is shown in **Figure 1**. Besides, the probability of needing an additional measurement and mean number of extra repetitions per listener are shown in **Table 3**. The repetitions were only suggested when the test was done using the alternative-forced choice (AFC) framework (i.e., the IPD test, the STM test, and the eAUD test in all the conditions). The total testing time was approximately 2.5 h excluding the initial interview, information about the study, and preparations.

3. GENERAL METHODS

3.1. Participants and General Setup

Seventy-five listeners (38 of them females) participated in the study, who were aged between 59 and 82 years (median: 71 years). Five participants were considered older normal hearing (ONH) with thresholds below 25 dB hearing level (HL) in the frequency range between 0.25 and 4 kHz in both ears and no larger than 40 dB HL at 8 kHz (PTA $\leq$ 22 dB HL)[2]. PTA was defined as the pure-tone average between 0.5, 1, and 2 kHz as is typically reported (Vermiglio et al., 2020). Two of these participants were not usual hearing-aid users. The hearing-impaired listeners (HI) group consisted of 70 participants with symmetric sensorineural hearing losses. Symmetric sensorineural hearing loss was defined as an interaural difference (ID) $\leq$ 15 dB HL at frequencies below 8 kHz and ID $\leq$ 25 dB HL at 8 kHz and air-bone gap $\leq$ 10 dB HL. The pure-tone audiograms of the participants are shown in **Figure 2**. The participants eligible for the present study had audiometric thresholds $\leq$ 55 dB HL (pure-tone audiometry not older than 1 year) in the range between 125 and 1,000 Hz. Participants with a pure tone threshold $\geq$ 75 dB HL at 2 kHz were excluded from the study as it was unlikely that it would be feasible to perform all the tests due to audibility issues.

The participants were recruited from the BEAR database (Houmoller et al., 2021) at Odense University Hospital (OUH), from the patient database at Bispebjerg Hospital (BBH), and from the database at the Hearing Systems Section at the Technical University of Denmark (DTU). The study was approved by the Science-Ethics Committee for the Capital Region of Denmark, H-16036391. All participants gave written informed consent and some of them received economical compensation for their participation, depending on each test

[2]While other listeners presented PTA $\leq$ 22 dB, the individual thresholds did not fulfill this criteria.

TABLE 2 | Reference data of the young normal-hearing group.

Outcome	5th	25th	50th	75th	95th	Unit
AUD_{LF}	–5	0	0	5	10	dB HL
AUD_{HF}	–10	0	5	10	10	dB HL
FLFT	14.42	15.60	17.88	18.02	19.31	kHz
HTL_{LF}	–7.5	–2.5	1.25	2.5	7.5	dB HL
HTL_{HF}	–5.5	2.5	8.75	15	23.5	dB HL
MCL_{LF}	55	62.5	70	72.5	78	dB HL
MCL_{HF}	62.5	70	75	80	87.5	dB HL
$DynR_{LF}$	87.5	92.5	97.5	105	117.5	dB
$DynR_{HF}$	74.5	85	92.5	100	105	dB
$SLope_{LF}$	0.27	0.30	0.33	0.37	0.42	CU/dB
$Slope_{HF}$	0.27	0.30	0.33	0.37	0.44	CU/dB
SRT_{Q}	6	10	14	18	21	dB SPL
maxDS	96.7	100	100	100	100	% Corr
SRT_{N}	–3.48	–1.84	–0.85	–0.19	1	dB SNR
$SScore^{+4dB}$	80	90	95	100	100	% Corr
$sSTM_{8}$	2.40	3.07	3.07	3.07	3.07	d'
$sSTM_{4k}$	0.30	2.40	3.07	3.07	3.07	d'
$fSTM_{8}$	–13.5	–12.7	–12.3	–10.5	–6.1	dB ML
$fSTM_{4k}$	–9.1	–6.5	–4.375	–3.375	–2.7	dB ML
TiN_{LF}	64.25	66.5	68.25	69.5	73.6	dB HL
TiN_{HF}	64.1	68	69.375	70	73.85	dB HL
TMR_{LF}	3.55	7.4167	9.375	13	15.95	dB
TMR_{HF}	3.1	8.25	10.875	13	18.15	dB
SMR_{LF}	16.55	19.75	21.75	23.25	26.15	dB
SMR_{HF}	19.6	26	28.5833	30.75	31.25	dB
IPD	0.86	1.20	1.22	1.29	1.40	kHz
BP_{20}	50	98.75	100	100	100	% Corr
BMR	14.3	15.875	17	18.4375	22.1	dB

The data are shown as the 5th, 25th, 50th, 75th, and 95th percentiles. In tests performed monoaurally, the summary corresponds to the data of both ears merged together (i.e., 22 ears).

site's regulations and whether the participant chose to participate without compensation.

3.2. Equipment

The basic audiological assessment consisted of pure-tone audiometry, wideband tympanometry (Rosowski et al., 2013) and middle ear muscle reflex, and was conducted in the facilities of OUH, BBH, and DTU. The rest of the tests were performed via PC in a double-walled sound-insulated booth (BBH and DTU) or in a small anechoic chamber (OUH). The tests were implemented in Matlab with a graphical user interface (GUI) that the examiner could operate without programming experience. Most of the tests were implemented using a modular framework for psychoacoustic experiments (AFC; Ewert, 2013), except for HINT, provided by Jens Bo Nielsen and Binaural Pitch test which was a reimplementation of the Binaural Pitch Test v1.0, Bispebjerg hospital, 2008. The participants were seated in the room and the stimuli were presented through headphones (Sennheiser HDA200) connected to a headphone-amplifier (SPL phonic) and an audio interface (RME Surface 24-bit). The equipment was calibrated using an artificial ear according to IEC 60318-1:2009. The tests consisting of threshold estimation using the AFC framework were repeated at least two times and the mean of the two measurements was considered as the final value. To ensure the quality of the data collected, additional repetitions were suggested by the framework until a certain standard deviation across measures was achieved.

3.3. Analysis of Test Reliability

The test–retest reliability of the test battery was assessed using intraclass correlation coefficients (ICC; Koo and Li, 2016), and the standard error of measurement (SEM; Stratford and Goldsmith, 1997). Since the ICC can be prone to misleading results in the case of systematic biases, the differences in

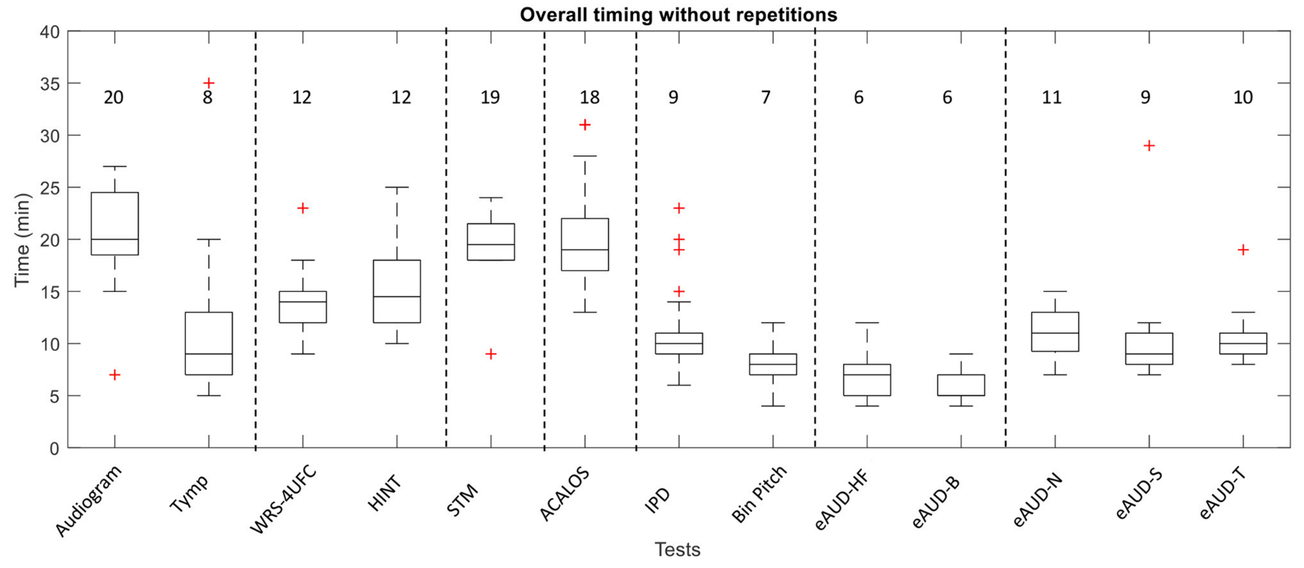


FIGURE 1 | The overall time of the different tests in the test battery including the instructions. The data correspond to the annotations of the examiners. The basic examination with the audiometry and the tympanometry (Tymp) are included. The numbers represent the rounded median in minutes.

TABLE 3 | Table with the probability of needing repetitions (PR), and the probability of having missing values (PM).

Test	PR (%)	PM (%)	PT (%)	E.Rep.
STM	42.86	90.79	88.16	4.32
IPD	10.77	10.97	20.55	1.87
eAUD-HF	5.63	4.05	9.46	1.85
eAUD-N	66.67	46.58	82.19	3.00
eAUD-S	48.57	52.70	75.68	3.07
eAUD-T	53.85	46.58	75.34	3.27
S_0N_0	42.59	27.03	58.11	2.00
$S_\pi N_0$	20.59	9.11	27.03	1.85

The total probability of repetitions (PT). The mean number of extra repetitions (E. Rep).

the means were evaluated as well. The only tests with a likely bias were the individual measures of the extended audiometry at lower frequencies. More details can be found in **Supplementary Table 1**. It was of special interest to test the reliability in older listeners with different hearing abilities. Therefore, test–retest measurements were performed with a subgroup consisting of 11 participants for all tests of the test battery (excluding the screening spectro-temporal modulation test). According to the hearing thresholds estimated with ACALOS, two of those listeners had near-normal hearing thresholds at all tested frequencies, three had average hearing thresholds > 30 dB HL, and the remaining 6 listeners presented gently sloping hearing thresholds < 55 dB HL at all frequencies. The participants were aged between 59 and 82 years (median 69 years). The retest session was conducted within 4 months after the first visit.

4. HIGH-FREQUENCY AUDIBILITY

Recently, elevated thresholds at HFs (> 8 kHz) have been linked to the concept of "hidden hearing loss" and synaptopathy (Liberman et al., 2016). However, the measurement of audiometric thresholds above 8 kHz is not part of the current clinical practice. The fixed-level frequency threshold (FLFT) has been proposed as a quick and efficient alternative to HF audiometry (Rieke et al., 2017; Prendergast et al., 2020). The test is based on the detection of a tone presented at a fixed level. The frequency of the tone is varied toward HFs and the maximum audible frequency at the given level is estimated in an adaptive procedure. Here, a modified version of FLFT, using warble tones presented at 80 dB SPL, was used as the extended audiometry at high frequencies (eAUD-HF).

4.1. Method

The procedure used here was a yes/no task using a single-interval adjustment matrix (SIAM) yes-no task procedure (Kaernbach, 1990). As in traditional up-down procedures, the target can be presented in a given trial or not. If the target was detected, the frequency of the warble tone was increased according to a given step size; if it was not detected, the frequency was decreased. However, the adjustment of the target depends on the participant's behavior across trials, characterized by the false alarm rate (catch trials) and the hit rate. If the participant gives unbiased responses and keeps the criterion, the tests is an adaptive procedure similar to a 1-up 1-down. When the participant is caught, the step size becomes double. For each run, the first four reversals were discarded, and the threshold of each

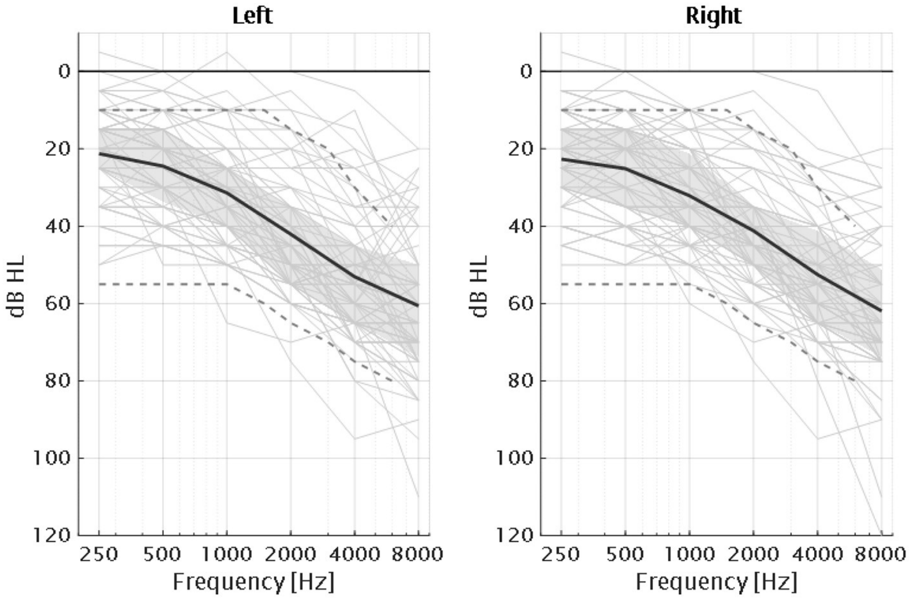


FIGURE 2 | Audiograms of the 75 participants of the study together with the average for each ear (dark solid lines) and interquartile ranges (gray areas). The gray dashed lines correspond to the standard audiograms N1 and N4 according to the IEC60118-15 (Bisgaard et al., 2010).

TABLE 4 | Summary of the results of the extended audiometry for high frequencies eAUD-HF.

			ONH			HI		
Outcome measure	**Level**	**Ear**	**Mean (SD)**	**Q1**	**Q3**	**Mean (SD)**	**Q1**	**Q3**
eAUD-HF	80	LE	10.9 (1.2)	10.2	11.9	7.57 (2.7)	5.3	10
FLFT (kHz)	dB SPL	RE	11.7 (1.1)	10.9	12.5	8.12 (2.3)	6.7	10.2

The results are shown as the mean, standard deviation, first and third quantiles (25 and 75th percentiles) for each ear.

trial was calculated as the average of the two subsequent reversals. In the catch trials, no sound was presented. The warble tone $wt(f_c, t)$ was defined by the expression 1:

$$wt(f_c, t) = \sin\left(2\pi f_c t + \frac{f_c f_e}{f_r}\sin(2\pi f_c t)\right), \tag{1}$$

where f_c is the stimulus frequency, $f_r = 4Hz$ is the frequency rate, and $f_e = 4.3\%$ is the frequency excursion of the frequency modulation.

4.2. Results and Discussion

The results of the FLFT measured at 80 dB SPL are shown in **Table 4**.

The maximum frequency threshold for a tone presented at 80 dB SPL (eAUD-HF) was 11 kHz for the ONH listeners and 8 kHz for the HI listeners. The HI group showed larger variability compared to the ONH group (interquartile range: 6 vs. 10 kHz). The eAUD-HF test showed very good reliability (ICC = 0.89; SEM = 495 Hz). These results suggest that the FLFT paradigm might be a good time-efficient alternative to the traditional audiometry for measuring HF sensitivity. A recent study pointed out the importance of off-frequency listening and the role of the excitation of the basal cochlea when presenting narrow-band stimuli in high levels (Encina-Llamas et al., 2019). Knowing the hearing sensitivity at HFs of a given patient might be crucial for better understanding of their supra-threshold deficits. Moreover, the eAUD-HF can include different levels and be useful not only for ototoxicity monitoring but also in association with other supra-threshold measures. For example, if FLFT is measured at a conversational level (i.e., 65 dB SPL), or at frequency-dependant levels corresponding to the speech spectrum, this measure could help to estimate the contribution of audible off-frequency listening to speech intelligibility or loudness perception.

5. LOUDNESS PERCEPTION

Loudness perception can substantially differ between NH and HI listeners and has been connected to the peripheral non-linearity (e.g., Jürgens et al., 2011). While the growth of loudness shows a non-linear behavior in a healthy ear, the results from HI listeners suggest that loudness perception becomes linear when outer-hair

TABLE 5 | Summary of the results of the adaptive categorial loudness scaling test (ACALOS).

Outcome measure	Freq. Range	Ear	ONH			HI		
			Mean (SD)	Q1	Q3	Mean (SD)	Q1	Q3
MCL (dB HL)	LF	LE	81.5 (14.8)	73.3	84.1	80.6 (8.4)	76.4	85.8
		RE	76.5 (13.2)	70	80	79.1 (7.9)	74.7	84.1
	HF	LE	79.0 (17.6)	66.6	90.8	82.7 (12.3)	75.8	90
		RE	73.8 (17.2)	65	80	80.3 (9.9)	74.7	87.5
Slope (CU/dB)	LF	LE	0.35 (0.1)	0.3	0.4	0.45 (0.1)	0.3	0.5
		RE	0.36 (0.1)	0.3	0.4	0.48 (0.2)	0.3	0.5
	HF	LE	0.45 (0.1)	0.3	0.4	0.84 (0.5)	0.5	0.9
		RE	*0.41 (0.1)*	0.3	0.4	0.81 (0.4)	0.5	0.9
DynR (dB HL)	LF	LE	91.5 (16.8)	78.3	97.5	*76.7 (15.8)*	*64.5*	88.3
		RE	91.1 (18.8)	79.1	100	73.9 (16.0)	61.6	86.8
	HF	LE	77.6 (18.2)	72.5	85.8	50.8 (15.1)	*40.6*	60.2
		RE	78.6 (17.9)	67.5	*90.8*	50.7 (15.5)	38.9	60.4

The results of the most comfortable level (MCL), slope of the growth of loudness (Slope), and dynamic range (DynR) are shown for low frequency (LF) and high frequency (HF) as the mean, standard deviation, first, and third quantiles (25 and 75th percentiles) for each ear.

cell (OHC) function is affected (e.g., Moore, 2007). Besides, the possibilities of characterizing hearing deficits, the loudness function can be used for fitting hearing aids (e.g., Oetting et al., 2018). Adaptive categorical loudness scaling (ACALOS; Brand and Hohmann, 2002) is the reference method for the current standard (ISO 16832, 2006) for loudness measurements.

5.1. Methods

According to the ACALOS method, 1/3-octave band noise was presented sequentially and the participant had to judge the perceived loudness using a 11-category scale ranging from "not heard" to "extremely loud." The presentation level of the next stimulus was calculated based on the previous trials. The raw results, which correspond to categorical units (CU) spanned between 0 and 50, were fitted to a model of loudness as described in Oetting et al. (2014). The outcome measures of the ACALOS presented here are the most comfortable level (MCL), the slope of the loudness function (Slope), and the dynamic range (DynR) defined as the difference between the uncomfortable level (50 CU) and the hearing threshold (0.5 CU). Low-frequency (LF) average corresponds to the stimuli centered at 250, 500, and 1,000 Hz, and HF average corresponds to the stimuli centered at 2, 4, and 6 mkHz.

5.2. Results and Discussion

The results of the ACALOS outcome measures are shown in **Table 5**. The hearing thresholds (HTL) estimated by ACALOS were significantly correlated with the pure-tone audiometric thresholds ($\rho = 0.88$; $p << 0.0001$) even when looking at the HI group alone ($\rho = 0.83$; $p << 0.0001$) despite the use of different stimuli and procedure.

The average MCL estimate ranged between 74 and 83 dB HL in both groups and for both frequency ranges. The only appreciable difference between the two groups in terms of MCL was found at HFs and only in the right ear. The average slope of the loudness growth was slightly steeper for the HI listeners in the LF range (0.45 CU/dB for HI vs. 0.35 CU/dB for ONH) and substantially steeper in the HF range (0.8 CU/dB for HI vs. 0.45 CU/dB for NH). The average dynamic range was between 80 and 90 dB HL for the ONH listeners, and smaller for the HI listeners, especially at HFs (50.8 dB). Regarding the test–retest reliability, ACALOS showed an excellent reliability for estimating the hearing thresholds (ICC = 0.94; SEM = 4.5 dB), good reliability for estimating the MCL (ICC = 0.68, SEM = 6.5 dB), and very good reliability for estimating the slope (ICC = 0.82; SEM = 0.07 CU/dB). Overall, these results supported the inclusion of ACALOS in a clinical test battery, as it provides several outcomes (hearing thresholds, growth of loudness, MCL and dynamic range). ACALOS also showed a high time efficiency (around 10 min per ear).

6. SPEECH PERCEPTION IN QUIET

6.1. Method

The word recognition score with four unforced choices (WRS-4UFC) test was proposed as a systematic and self-administered procedure that allows the estimation of supra-threshold deficits in speech perception in quiet. The speech material was the same as the one used for standard speech audiometry (Dantale I; Elberling et al., 1989) in Danish. The self-administered procedure consisted of the presentation of one word where the participant has to answer in a 4-unforced-choice paradigm (4UFC). After the acoustical presentation of each word, the target written word was assigned randomly to one of four buttons shown to the participant. The other three buttons contained words that were also taken from the Dantale-I corpus. They were chosen based on the lowest Levenshtein phonetic distance (Sanders and Chin,

TABLE 6 | Summary of the results of the word recognition scores (WRS-4UFC) test.

Outcome measure	Ear	ONH			HI		
		Mean (SD)	Q1	Q3	Mean (SD)	Q1	Q3
SRT_Q (dB HL)	LE	19.9 (7.1)	16.5	*19.2*	41.5 (13.5)	*31.8*	50.6
	RE	23.3 (8.9)	*17.2*	29.0	42.7 (12.6)	33.9	51.1
maxDS (%)	LE	99.2 (1.6)	100	100	97.2 (4.1)	*95.3*	100
	RE	97.2 (1.8)	95.5	97.6	93.9 (6.4)	92.1	98.4

The results of the speech recognition threshold (SRT_Q), and maximum discrimination score (maxDS) are shown as the mean, standard deviation, first, and third quantiles (25 and 75th percentiles) for each ear.

2009) from the target. The term "unforced" corresponds to an additional choice, a question mark, that the listener can press if none of the four options are considered the right answer. Four lists of 25 words were presented at 40, 30, 20, and 10 dB above the individual PTA, in this order. A logistic function with two independent free asymptotes was fitted to the results from each individual ear and the speech reception threshold (SRT_Q), at which 50% of the words were recognized, and maximum speech discrimination score (maxDS), which was the maximum value of the function (i.e., the upper asymptote). Both outcomes were estimated using psignifit 4 software (Schütt et al., 2016).

6.2. Results and Discussion

The results of the WRS-4UAFC outcome measures are shown in **Table 6**.

The HI listeners' SRT_Q were, on average, 20 dB higher than the ones of the ONH group. The interquartile range for the HI group was about 19 dB, whereas for the ONH group it was 3 dB for the left ear (LE) and 11.8 dB for the right ear (RE). The maxDS for both groups was close to 100%. However, the HI listeners showed larger variability, especially in the right ear (SD= 6.42%). In the analysis of the test–retest variability, the WRS-4UFC test showed poor to moderate reliability, especially at low levels (PTA + 10 dB; ICC = 0.25). However, at the higher presentation levels (i.e., individual PTA + 40 dB) the standard error of the measurement was only 4% (1 word). Regarding clinical applicability, the WRS-4UFC needs to be compared to traditional speech audiometry to explore the influence of using closed- vs. open-set and forced- vs. unforced-choice test procedures on the results.

7. SPEECH PERCEPTION IN NOISE

The Hearing in Noise Test (HINT; Nilsson et al., 1994) is an adaptive sentence recognition test carried out with speech-shaped noise. The following assumptions are considered in HINT (based on Plomp, 1978): (1) Speech materials made up of meaningful sentences yield a steep psychometric function; (2) stationary noise with the same spectral shape as the average spectrum of the speech material makes the speech reception threshold in noise (SRT_N) less dependent on the spectral characteristics of the speaker's voice. Furthermore, the signal-to-noise ratio (SNR) between the target and masker is better defined across the frequency range. (3) The SRT_N is independent of the absolute noise level as long as the noise level is above the "internal noise" level. Therefore, it is recommended to present the noise at least 30 dB above the "internal noise." The internal noise is defined as the sum of the SRT in quiet of the tested listener and the SRT in noise for NH listeners, for a given speech material.

7.1. Methods

The Danish HINT was used as in Nielsen and Dau (2011) to obtain the SRT_N but in a monoaural presentation. Additionally, a 20-sentence list was presented at a fixed SNR of +4 dB and scored to obtain a sentence recognition score ($SScore^{+4dB}$). The presentation level of the noise was set between 65 and 85 dB SPL to ensure that the noise was always presented 30 dB above the individual PTA. Each ear was tested individually. All participants were tested using the same list with the same ear. Since small differences across lists were found in Nielsen and Dau (2011), this was done to ensure that all the listeners were tested with an equally difficult list. However, for the test–retest reliability study, the list and ear presented were randomized, only using lists 6–10. The listeners did not report recalling sentences from the test.

7.2. Results and Discussion

The results of the HINT outcome measures are shown in **Table 7**.

The SRT_N for ONH listeners were, on average, 2 dB higher than the ones reported in Nielsen and Dau (2011). This bias was also observed in the YNH listeners. However, this might be explained by the fact that they used diotic presentation, which can lead to a 1.5 dB improvement as reported by Plomp and Mimpen (1979). The results also showed a lower SRT_N (1.5 dB) and higher $SScore^{+4dB}$ (4%) for the right ear in both groups of listeners. According to Nielsen and Dau (2011), there was a significant main effect of test list. Such differences are seen mainly for lists 1–4, which were the lists used here. Therefore, the observed interaural difference can be ascribed to a list effect. However, it might be ascribed to other factors as, for example, a right-ear advantage as the one observed in NH listeners with tinnitus (Tai and Husain, 2018). Unfortunately, the difference across lists was not taken into account in the experimental design so that we cannot conclude that the difference is only due to the list effect. The ICC values (SRT_N: ICC = 0.61; $SScore^{+4dB}$: ICC = 0.57) indicated only moderate reliability of the HINT.

TABLE 7 | Summary of the results of the hearing in noise test (HINT).

Outcome measure	Ear	ONH			HI		
		Mean (SD)	Q1	Q3	Mean (SD)	Q1	Q3
SRT_N (dB SNR)	LE	1.0 (0.7)	0.4	1.5	4.1 (3.4)	1.4	6.7
	RE	–0.5 (1.1)	–1.0	0.0	2.6 (3.8)	0.0	4.2
$SScore^{+4dB}$ (%)	LE	85.0 (11.7)	85	90	60.0 (26.6)	40	85
	RE	91.0 (9.6)	90	95	*62.3 (24.0)*	48.7	80

The results of the speech recognition threshold in noise (SRT_N), and sentence recognition score at +4 dB SNR ($SScore^{+4dB}$) are shown as the mean, standard deviation, first, and third quantiles (25 and 75th percentiles) for each ear.

The SRT_N showed an SEM = 1.02 dB, which is below the step size of the test (2 dB). The $SScore^{+4dB}$ showed an SEM value of 7.94%, which corresponds to an error in one of the sentences. However, the reliability of the test can be improved by using an adaptive method as the one described in Wagener et al. (2003), Rønne et al. (2017) where the SRT_N estimation was optimized using a combination of word scoring and a maximum likelihood procedure.

8. SPECTRO-TEMPORAL MODULATION SENSITIVITY

A speech signal can be decomposed into spectral and temporal modulations. While speech-in-noise perception assessment leads to some confounds due to the variety of speech corpora, noise maskers, and test procedures that can all affect the results, the assessment of the contrast sensitivity of simpler sounds might be of interest for characterizing a listener's spectro-temporal processing abilities. Bernstein et al. (2013) showed significant differences between NH and HI listeners for detecting STM in random noise. These differences corresponded to specific conditions that were also useful for the prediction of speech-in-noise performance in the same listeners. Lately, the assessment of STM sensitivity in these specific conditions gained an increasing interest due to its potential for predicting speech intelligibility (Bernstein et al., 2016; Zaar et al., 2020) and for assessing cochlear-implant candidacy (Choi et al., 2016). Here, STM sensitivity was assessed using a new test paradigm that may be more suitable for a clinical implementation than the previous psychoacoustic versions of the test. The test was performed in two conditions: an LF condition (similar to the one previously used in Bernstein et al., 2016) and an HF condition (Mehraei et al., 2014).

8.1. Methods

The stimuli were similar to those of Bernstein et al. (2016) and Mehraei et al. (2014), but a different presentation paradigm was employed. A sequence of four noises was presented in each trial. The first and third stimulus always contained unmodulated noise, whereas the second and fourth stimuli could be either modulated or unmodulated. The test was performed with a low-frequency (LF) 3-octaves wide stimulus centered at 800 Hz ($sSTM_8$ and $fSTM_8$), and a 1-octave wide stimulus centered at 4 kHz ($sSTM_{4k}$ and $fSTM_{4k}$). The stimuli were presented at 75 dB sound pressure level (SPL). After the sequence was presented, the listener had to respond whether the four sounds were different ("yes") or the same ("no"). Two procedures involving catch trials were evaluated. The first test the screening spectro-temporal sensitivity (sSTM), a test consisting of 10 stimuli modulated at 20log(m) = –3 dB modulation level (ML), where m is the modulation depth, and five unmodulated ones presented in random order. The two runs of the two conditions could be completed in approximately 4 min. The outcome measure was the listener's contrast sensitivity (d')[3] in the task. The second test was the "fast" spectro-temporal sensitivity (fSTM), which tracked the 80% point of the psychometric function using a yes/no task and the single-interval adjusted matrix (SIAM; Kaernbach, 1990) paradigm. The first two reversals were discarded and the thresholds were the average of the last four reversals. A negative response increased the modulation by 4 times the step size and by 5 times when there was a "caught." These parameters were chosen for maximizing the attainability of the test after a pilot investigation. For the sSTM test, the stimulus was presented diotically, whereas for the fSTM the test was presented in each ear individually in a monoaural presentation.

8.2. Results and Discussion

The results of the spectro-temporal modulation sensitivity tests outcomes are shown in **Table 8**.

The screening STM test (sSTM) shows the sensitivity in terms of d', where the maximum value is d' = 3, (i.e., 10 modulated and 5 unmodulated stimuli correctly detected). In the hypothetical case when all the catch trials are detected, the lowest d' value can be –0.3. The ONH listeners showed a high sensitivity in the LF condition (d' = 2.6) and a somewhat lower sensitivity in the HF condition (d' = 1.63) corresponding to 65% correct responses. The HI listeners showed a higher variability and a lower sensitivity in the LF condition (≈ 70% correct) and substantially lower sensitivity in the HF condition (0–50% correct responses). The threshold-tracking procedure (fSTM) showed results between –9 and –1.6 dB ML in the ONH group, whereas the HI listeners showed thresholds between –3.50 and –0.6 dB

[3] d' was defined as $Z(\frac{N_H+0.5}{H+1}) - Z(\frac{N_{FA}+0.5}{FA+1})$, where Z refers to the z-score transformation, H is the total number of target presentations, and FA is the total number of catch trials.

TABLE 8 | Summary of the results of the spectro-temporal modulation sensitivity tests.

			ONH			HI		
Outcome measure	**Freq. Range**	**Ear**	**Mean (SD)**	**Q1**	**Q3**	**Mean (SD)**	**Q1**	**Q3**
sSTM -3dB (d')	LF	Bin	2.6 (0.6)	2.4	3	1.7 (1.3)	0.4	3
	HF		1.6 (0.8)	1.1	2.4	0.6 (1.1)	–0.3	1.4
fSTM (dB ML)	LF	LE	–7.7 (1.8)	–9	–7.6	–2.8 (2.1)	–3.5	–0.8
		RE	–5.1 (3.1)	–7.2	–1.6	–1.6 (1.3)	–2	–0.6
	HF	LE	–8.0 (2.0)	–8.6	–6.2	–2.6 (2.4)	–3.8	–0.6
		RE	–5.6 (3.6)	–8.6	–2.1	–1.9 (1.5)	–2	–1

The results of the screening STM sensitivity test (sSTM), and the threshold of the fast STM sensitivity test (fSTM) are shown for the low-frequency stimulus (8, because it is centered at 800 Hz) and high-frequency (4 k, centered at 4 kHz) as the mean, standard deviation, first, and third quantiles (25 and 75th percentiles) for each ear.

ML in the LF condition. Although the results of the fSTM LF condition were consistent with Bernstein et al. (2016), the results in the HF condition showed higher thresholds than the ones in Mehraei et al. (2014). This can be ascribed to the higher presentation level used in Mehraei et al. (2014) than in the current test procedure. According to Mehraei et al. (2014) fSTM_{4k} could be a good predictor of frequency selectivity but, in the present study, the majority of the listeners presented elevated thresholds or could not complete the test. Therefore, this condition was excluded for further analysis. The fSTM showed an excellent reliability (ICC = 0.91; SEM = 0.93 dB ML) in the LF condition. However, several HI listeners were not able to complete the procedure for the HF condition. Overall, the use of the SIAM tracking procedure allowed us to obtain accurate thresholds, although additional repetitions were required, especially in the HF condition. This might be because the psychometric function for detecting the stimulus can be shallower in this condition, or because the 100% detection could not be reached even in the fully modulated trials. Therefore, a Bayesian procedure being able to estimate the threshold and slope of the psychometric function, such as the Bayes Fisher information gain (Figure: Remus and Collins, 2008), might be more suitable for this type of test. Another reason explaining the inability of the listeners to perform the test can be ascribed to the stimulus. Zaar et al. (2018) used a longer stimulus (1s), a diotic presentation, and a hearing loss compensation that ensured the audibility of the stimulus in all its frequency range. In their study, all the listeners were able to perform the tests and their sensitivity thresholds were well below the maximum value.

9. EXTENDED AUDIOMETRY IN NOISE

The extended audiometry in noise (eAUD) is a tone detection test intended to assess different aspects of auditory processing by means of a task similar to pure-tone audiometry. The tone is presented in the presence of noise and the listener has to indicate whether the tone was perceived or not. The aspects of auditory processing assessed here are (1) tone-in-noise detection and (2) spectral and temporal resolution.

9.1. Tone-in-Noise Detection

In pure-tone audiometry, a given patient has to detect the simple stimulus (e.g., sinusioids) in quiet aiming at estimating the hearing thresholds of the listener. A simple way to explore the supra-threshold performance is to perform a tone-in-noise detection test by presenting noise at supra-threshold levels and obtaining the masked thresholds. However, the characteristics of the noise such as bandwidth, level, or inherit modulations can affect the results. Moore et al. (2000) proposed a test paradigm using a special type of noise, which is able to provide the same masking in the entire frequency range, so the hearing thresholds of an NH listener would raise according to the level of the noise. This is the so-called threshold-equalizing noise (TEN). The advantage of the TEN test is that the expected masked thresholds are similar to the level of the noise (i.e., as TEN is played at 70 dB per equal rectangular bandwidth (ERB), the masked threshold is expected to be at 70 dB SPL). Although this test was originally design to detect dead cochlear regions, recent evidence suggests that tone-in-noise detection can be representative of supra-threshold deficits beyond the audiogram (Schädler et al., 2020).

9.2. Spectro-Temporal Resolution

Frequency and temporal resolution are aspects of hearing that are fundamental for the analysis of perceived sounds. While NH listeners exhibit a frequency selectivity on the order of one-third of an octave when using isoinput levels (from Glasberg and Moore, 1990; Eustaquio-Martín and Lopez-Poveda, 2011), HI listeners typically have broader auditory filters, leading to impaired frequency selectivity (Moore, 2007). Temporal resolution can be characterized by the ability to "listen in the dips" when the background noise is fluctuating based on the so-called masking release (Festen and Plomp, 1990). Schorn and Zwicker (1990) proposed an elaborated technique for assessing both spectral and temporal resolution using two tests: (1) Psychoacoustical tuning curves and (2) temporal resolution curves. In both cases, the task consists of detecting a pure tone that is masked by noise or another tone while the spectral or temporal characteristics of the masker are varied. Later, Larsby and Arlinger (1998) proposed a similar paradigm, the F-T test,

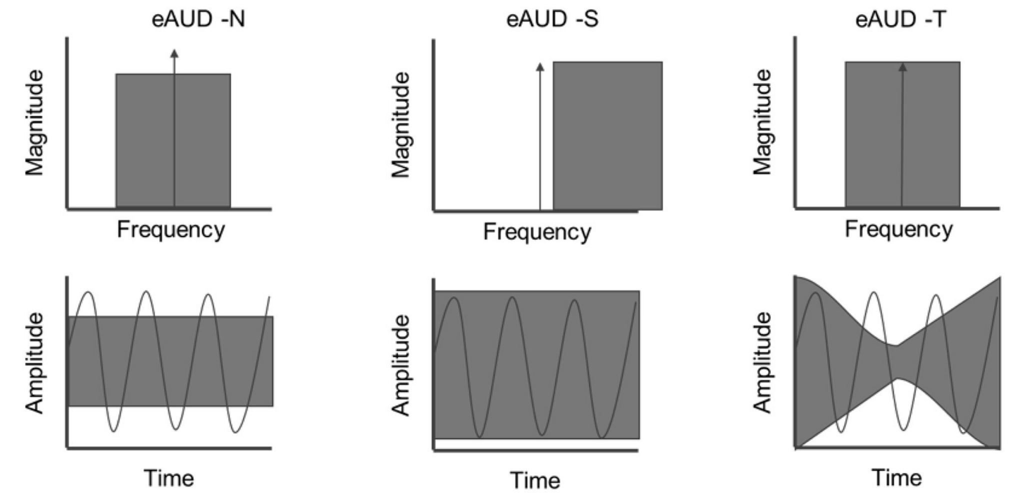


FIGURE 3 | Sketch of the conditions of the spectro-temporal resolution measures of the extended audiometry in noise (eAUD). The top panel shows the spectrum of the noise and target pure-tone (delta), the bottom panel shows both signals in the time domain. Left panel: Tone in noise condition (eAUD-N). Middle panel: Spectral condition (eAUD-S). Right panel: Temporal condition (eAUD-T).

which was successfully tested in HI listeners (van Esch and Dreschler, 2011). However, the F-T test is based on a Bekesy-tracking procedure, which can be demanding and less reliable for some listeners than an adaptive procedure (Rhebergen et al., 2015). Here, the spectro-temporal resolution was assessed using a new test. This test is a tone-in-noise detection task consisting of three conditions as sketched in **Figure 3**.

1. eAUD-N: The tone is embedded in a 1-octave-wide threshold equalizing noise (TEN-HL; Moore, 2001). Because of the properties of the TEN-HL, the tone detection threshold is comparable to the level of the noise in dB HL.
2. eAUD-S: The tone is embedded in a TEN that has been shifted up in frequency. In the spectral domain, this yields spectral unmasking of the tone, so the detection threshold is lower than in eAUD-N.
3. eAUD-T: The tone is embedded in a temporally modulated noise with the same spectral properties as the one in eAUD-N. In the temporal domain, the modulations of the noise yield temporal unmasking, so the tone can be detected in the dips.

The outcome measures were focused on the temporal and spectral benefits expected in the eAUD-S and eAUD-T conditions compared to the eAUD-N condition. While in the noise condition (eAUD-N) the threshold is expected to be approximately at the level of the noise, in the temporal and spectral conditions the thresholds should be obtained at a lower level, showing temporal masking release (TMR) and spectral masking release (SMR).

9.3. Methods

The procedure used here was a yes/no task using a SIAM procedure (Kaernbach, 1990) similar to the one used in the eAUD-HF. Here, a TEN was presented together with a warble tone. If the target was detected, the target-presentation level is decreased according to a given step size; if it was not detected, the level is increased. If the stimulus was not presented (catch trial) but the listener provided a positive response, the level is decreased compared to the previous trial. As in the eAUD-HF, for each run, the threshold of each trial was calculated as the average of the last four reversals. The noise was presented at 70 dB HL. The LF condition corresponded to the detection of a 0.5 kHz warble tone, whereas the HF condition corresponded to a 2 kHz warble tone. The final threshold was calculated as the mean threshold of two repetitions. In the eAUD-S condition, the center frequency of the noise was $f_{c,\text{noise}} = 1.1 f_{\text{tone}}$. In the eAUD-T condition, the modulation frequency of the noise was set to, $f_m = 4$ Hz. The outcome measures of the eAUD are 1) the tone-in-noise threshold (TiN), (2) the temporal masking release (TMR), and (3) the spectral masking release (SMR). The first two reversals were discarded and the thresholds were the average of the last four‘reversals.

9.4. Results and Discussion

The results of the extended audiometry in noise outcomes are shown in **Table 9**.

The TiN showed a larger variance for the ONH group (SD = 4.5 dB HL) at LFs. The detection thresholds were in line with previous work with thresholds close to the noise presentation level (70 dB HL) (Vinay et al., 2017). The TMR shown by the NH group was larger at HFs (10 dB) than at LFs (7 dB). The HI group showed, on average, similar TMR only at LFs. The SMR shown by the ONH listeners was 19 dB for LFs and 26 dB for HFs. In contrast, for the HI listeners, the SMR was 7 dB lower only in the HF condition. The reliability of the eAUD was moderate for most of the conditions (ICC $\leq$ 0.75). The eAUD-S at LFs showed good reliability (ICC = 0.85; SEM = 1.78 dB). The masking release estimates showed good reliability only for the HF condition. The reason for this might be

TABLE 9 | Summary of the results of the extended audiometry in noise (eAUD).

Outcome measure	Freq. Range	Ear	ONH Mean (SD)	ONH Q1	ONH Q3	HI Mean (SD)	HI Q1	HI Q3
TiN (dB HL)	LF	LE	70.4 (4.5)	68	71.5	71.8 (2.6)	70.2	73.2
eAUD-N		RE	69.2 (4.6)	65.2	72.5	72.0 (2.8)	69.6	74.3
	HF	LE	71.1 (2.5)	69.7	72.7	74.7 (3.4)	72.5	76.1
		RE	70.8 (3.6)	70.5	71.7	74.2 (3.1)	72	76.2
TMR (dB)	LF	LE	7.5 (3.4)	6	7.5	7.7 (4.0)	6.1	10.1
eAUD (N -T)		RE	5.2 (3.3)	4	7.6	8.3 (2.7)	6.5	10.3
	HF	LE	13.0 (0.6)	12.7	13.2	7.9 (5.0)	5	11.6
		RE	10.7 (3.1)	9.1	10.2	8.1 (5.2)	5.1	10.7
SMR (dB)	LF	LE	19.3 (3.6)	16.5	21.7	19.6 (17.7)	17.7	23.2
eAUD (N -S)		RE	18.8 (4.6)	17	21.2	20.0 (5.2)	16.5	23.8
	HF	LE	26.8 (4.5)	27.5	29	19.3 (9.5)	12.1	26.3
		RE	27.2 (3.7)	26.2	29.5	19.5 (9.9)	12	26.8

The results of the tone-in-noise (TiN), temporal masking release (TMR), and spectral masking release (SMR) are shown for the low-frequency condition (LF; 500 Hz) and high-frequency condition (HF; 2 kHz) as the mean, standard deviation, first, and third quantiles (25 and 75th percentiles) for each ear.

that masking release is a differential measure, and the cumulative error is, therefore, higher than that of each individual measure. The reduced reliability can be explained to some extent by the method used. To have a similar procedure as in pure-tone audiometry, the parameters of the SIAM tracking procedure were set accordingly. However, this made the test challenging and the listeners consistently missed several catch trials. Thus, extra trials were required to improve measurement accuracy, especially in the eAUD-N condition. Furthermore, the standard error of the measurement was in most cases larger than the final step size (2 dB). As in the case of the fSTM, a different procedure, such as Bayesian adaptive methods, might increase measurement reliability.

10. BINAURAL PROCESSING ABILITIES

Binaural hearing is useful for sound localization and the segregation of complex sounds (Darwin, 1997). Interaural differences in level or timing are processed for spatial hearing purposes in the auditory system. In the case of hearing loss, the neural signal at the output of the cochlea can be degraded, which may lead to reduced binaural abilities typically connected to temporal fine structure (TFS) processing. Based on a method of estimating the upper-frequency limit for detecting an interaural phase difference (IPD) of 180° (IPD_{fmax} Ross et al., 2007; Neher et al., 2011; Santurette and Dau, 2012), Füllgrabe and Moore (2017) recently proposed a refined test as a feasible way to evaluate TFS sensitivity. This paradigm was used in recent research that suggested that IPD_{fmax} might be related to non-auditory factors (Strelcyk et al., 2019) and affected by factors beyond hearing loss, such as musical training (Bianchi et al., 2019). Therefore, the IPD_{fmax} might be a task that requires auditory and non-auditory processing abilities beyond TFS sensitivity.

In contrast, binaural pitch detection assesses binaural processing abilities in a different manner. This test requires the detection of pitch contours embedded in noise, which are diotically or dichotically evoked. While the diotic condition can be resolved monoaurally, the dichotic condition requires the binaural processing abilities to be sufficiently intact to detect the contour. Previous studies showed that some listeners were unable to detect binaural pitch, regardless of the audiometric configuration (Santurette and Dau, 2012; Sanchez-Lopez et al., 2018). Therefore, it was of interest to compare the results of these two binaural processing tests.

Another approach for evaluating the binaural processing abilities is assessing binaural masking release (Durlach, 1963), which has been used in several studies (e.g., Strelcyk and Dau, 2009; Neher, 2017) and implemented in some commercial audiometers (Brown and Musiek, 2013). In this paradigm, a tone-in-noise stimulus is presented in two conditions: (1) a diotic condition where the tone is in phase in the two ears, and (2) a dichotic condition where the tone is in antiphase in the two ears. The difference between the two yields the benefit for tone detection due to binaural processing, the so-called binaural masking release (BMR).

10.1. Methods

The maximum frequency for detecting an IPD of 180° with pure-tones was obtained using a 2-AFC tracking procedure similar to the one used in Füllgrabe and Moore (2017). The stimuli were presented bilaterally in both ears as two sequencies of four tones. One sequence contained an ABAB sequence, where A means a diotic presentation and B an IPD of 180° between the tones presented to each ear, and the other an AAAA sequence. A

TABLE 10 | Summary of the results of the binaural processing abilities tests.

Outcome measure	Ear	ONH			HI		
		Mean (SD)	Q1	*Q3*	Mean (SD)	Q1	Q3
IPD_{fmax} (kHz)	Bin	0.76 (0.26)	0.59	0.98	0.69 (0.27)	0.52	0.88
BP_{20} (%)	Bin	87.5 (25.0)	87.5	100	80.7 (30.9)	70	100
BMR (dB) ($S_0N_0 - S_\pi N_0$)	Bin	16.5 (4.7)	13.5	17.5	14.7 (4.6)	12.2	*17.5*

The results of the maximum frequency for IPD detection (IPD_{fmax}), binaural pitch detection scores (BP_{20}), and the binaural masking release (BMR) are shown as the mean, standard deviation, and first and third quantiles (25 and 75th percentiles) for each ear.

positive response (detection) increased the frequency of the tone, and a negative response a decrease of the frequency. Although the stimuli duration and procedure were similar, the step size used here was slightly different, starting with steps of 2/3 octave and decreasing to a final step size of 1/6 octave in each reversal. The last six reversals were used for estimating the threshold. The frequency threshold (IPD_{fmax}) was obtained from the average of two runs.

Binaural pitch detection scores were obtained using a clinical implementation of the test proposed by Santurette and Dau (2012). A 3-min sequence of noise was presented bilaterally. Ten diotic and ten dichotic pitch contours, embedded in the noise, had to be detected by the listener. The tones forming the pitch contours were generated by adding frequency-specific IPDs to the presented noise (Cramer and Huggins, 1958). The outcome measure of the binaural pitch test was the percentage score of detecting the dichotic pitch contours only, averaged across two repetitions (BP20).

The BMR was assessed using the same method as the extended audiometry. Two measurements were required: (1) tone-in-noise detection presented diotically (S_0N_0) and tone-in-noise detection presented dichotically, i.e., with the tone in anti-phase across the two ears ($S_\pi N_0$).

10.2. Results and Discussion

The results of the tests assessing binaural processing abilities are shown in **Table 10**.

The listeners in the ONH and HI groups showed IPD_{fmax} thresholds around 700 Hz with a standard deviation ($\approx$ 270 Hz) and interquartile range ($\approx$ 370 Hz) similarly in both groups. These results are in line with the ones reported in Füllgrabe and Moore (2017). The IPD_{fmax} test showed excellent reliability (ICC = 0.95; SEM = 65.4 Hz), and the median time needed for two repetitions was 10 min. This suggests that IPD_{fmax} is a reliable measure of binaural processing abilities that can reveal substantial variability among both NH and HI listeners, which is valuable for highlighting individual differences among patients.

The overall results from the binaural pitch test for the NH listeners showed > 87.5% correct detection, whereas the HI listeners' results showed a higher variability, with an interquartile range from 70 to 100%. The test showed excellent reliability (ICC = 0.98; SEM = 4%), which may be influenced by ceiling effects since 6 participants got 100% correct responses. Listeners reported a positive experience due to the test being short and easy to understand.

The BMR shown by both groups was around 15 dB, as expected from previous studies (Durlach, 1963). This BMR is, in essence, similar to the binaural masking level differences that are available in some clinical devices. However, in this test battery, the use of threshold equalizing noise provoked the S_0N_0 condition thresholds to be similar to the noise presentation level (i.e., 70 dB HL). The variability of this condition was substantially lower than the $S_\pi N_0$, suggesting that the use of $S_\pi N_0$ can be sufficient and more informative. A similar reasoning has been recently reported in Grant et al. (2021), where the dichotic condition alone was a sensitive auditory measure associated with the effects of noise exposure.

11. EXPLORATORY ANALYSIS

The collection of tests included in the test battery was intended to explore different and potentially independent aspects of hearing to obtain an auditory profile with controlled interrelations among the tests. A factor analysis performed in the HEARCOM study (Vlaming et al., 2011) based on data from 72 HI subjects revealed auditory dimensions: (1) HF processing, (2) audibility, (3) LF processing, and (4) recruitment. In the current study, the results of the behavioral tests were analyzed further in order to explore possible interrelations between the various outcome measures.

11.1. Methods

First, the data were pre-processed as in Sanchez-Lopez et al. (2018) to reduce the number of variables. The outcome variables of the frequency-specific tests were divided into LF ($\leq$1 kHz) and HF (> 1 kHz) variables. This decision was supported by a correlation analysis performed on the complete set of outcome variables, where the outcomes corresponding to 2, 4, and 6 kHz as well as the ones corresponding to 0.25, 0.5, and 1 kHz were highly intercorrelated. For the tests performed monaurally, the mean of the two ears was taken as the resulting outcome variable. The resulting dataset (BEAR3 dataset[4])

[4]The BEAR3 available at Zenodo contains an observation labeled "0," which corresponds to the results of one of the examiners and it is not used in the present analysis.

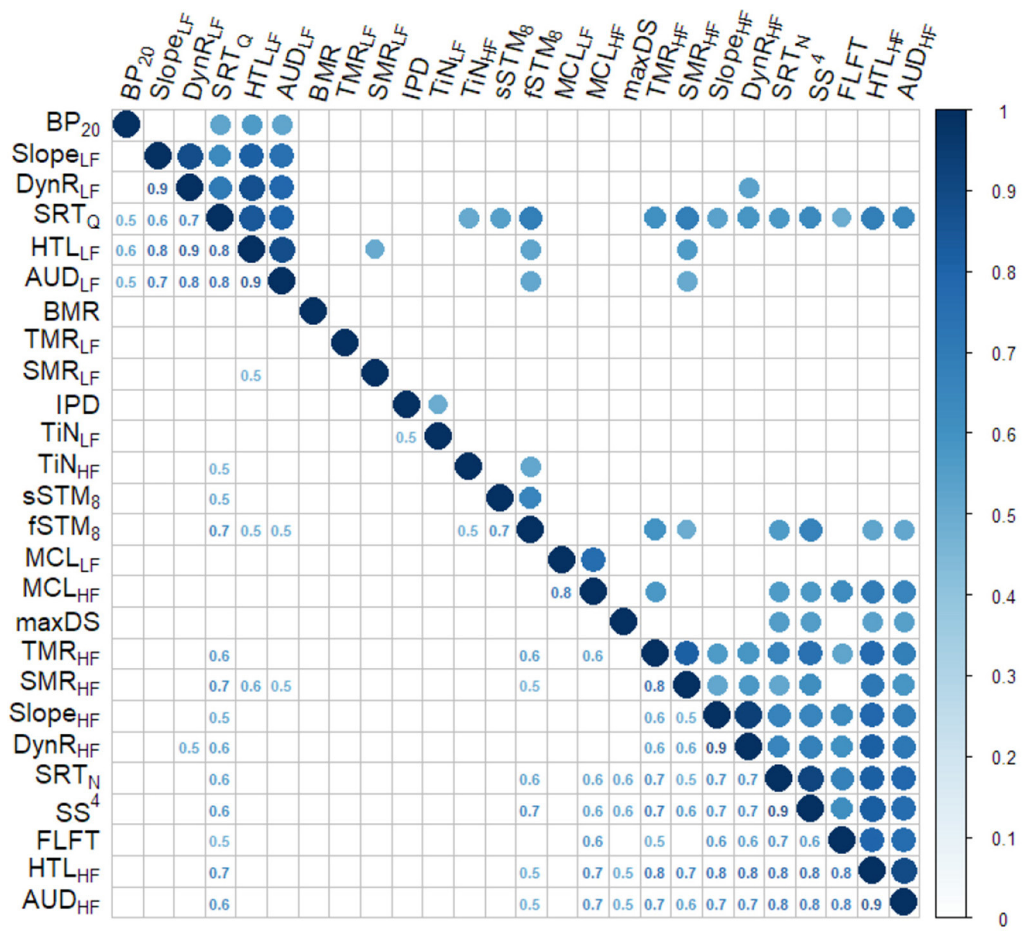


FIGURE 4 | Correlation plot of the data set BEAR3. The upper part shows the significantly correlated variables as colored circles. The lower panel shows the numeric correlation value.

contained 26 variables, divided into six groups corresponding to the six aspects of auditory processing considered here. The exploratory analysis consisted of a correlation analysis using Spearman correlations and factor analysis. The factor analysis was performed using an orthogonal rotation ("varimax") and the method of maximum likelihood. The number of components was chosen to use parallel analysis, the resulting number of components was four.

11.2. Results

Figure 4 shows the results from the correlation analysis performed on the BEAR3 dataset. For convenience, the absolute value of the correlation was used when visualizing the data to show the strength of the correlation. The circles on the right-hand side of the figure depict significant correlations ($p < 0.00001$), and the correlation values are presented on the left-hand side of the figure. Two groups of correlated variables can be observed. The upper-left corner shows variables related to LF processing (dynamic range, the slope of the loudness function, and hearing thresholds) and speech intelligibility in quiet. The bottom-right corner shows a larger group of correlated variables including HF processing, speech intelligibility in noise, and spectro-temporal resolution at HFs. The variables that are not significantly interrelated are shown in the middle part of **Figure 3**, including the three variables related to binaural processing abilities (IPD_{fmax}, BP20 and BMR), which were not significantly correlated to each other. The speech reception threshold in quiet (SRT_Q) and the STM detection were correlated to various variables such as tone-in-noise detection, HF spectro-temporal resolution, LF hearing thresholds, and speech-in-noise perception.

The four factors resulting from the factor analysis showed 63% of explained cumulative variance. The variables with higher loadings (> 0.65) for each of the factors are shown in **Table 11**. The first factor, in terms of the amount of variance explained (19%), was associated with LF loudness perception and speech intelligibility in quiet, whereas the second factor (18% of variance explained) was associated with HF loudness perception. Despite loudness perception being associated with the first and second factor, the MCL was associated, both at HF and LF, with the third factor, while the fourth factor was associated with speech intelligibility in noise.

TABLE 11 | Variables correlated to the four latent orthogonal factors resulting from the factor analysis with the method of maximum likelihood (ML).

	ML2 (19%)	ML1 (18%)	ML3 (14%)	ML4 (12%)
HTL_{LF}	0.93			
$DynR_{LF}$	–0.9			
AUD_{LF}	0.82			
$Slope_{LF}$	0.81			
SRT_{Q}	0.67			
$DynR_{HF}$		–0.93		
$Slope_{HF}$		0.82		
HTL_{HF}		0.79		
AUD_{HF}		0.73		
MCL_{HF}			0.92	
MCL_{LF}			0.85	
SRT_{N}				0.77
$SScore^{+4dB}$				–0.78

Columns are sorted in terms of the variance explained by each factor.

12. GENERAL DISCUSSION

The first goal of the present study was to collect data from a heterogeneous population of HI listeners, reflecting their hearing abilities in different aspects of auditory processing. The current study was motivated by the need for a new dataset to refine the data-driven approach for auditory profiling. The dataset should contain a representative population of listeners and outcome measures (Sanchez-Lopez et al., 2018) to allow a refined definition of the two types of auditory distortions and to identify subgroups of listeners with clinical relevance. To refine the data-driven auditory profiling, the BEAR3 dataset fulfills all the requirements discussed in Sanchez-Lopez et al. (2018). Other datasets containing a large number of listeners (e.g., Rönnberg et al., 2016; Gieseler et al., 2017) or physiological measures (e.g., Kamerer et al., 2019) could also be interesting for complementing the auditory profiling beyond auditory perceptual measures.

12.1. Relationships Across Different Aspects of Auditory Processing

The proposed test battery considers outcomes divided into six dimensions of auditory processing. One of the objectives of the study was to investigate the interrelations of different dimensions and measures. The present analysis showed two interesting findings. First, the correlation analysis showed two clusters of variables related to either LF or HF audiometric thresholds. Speech-in-noise perception was associated with HF sensitivity loss, temporal, and spectral masking release, whereas speech-in-quiet was correlated with both LF and HF hearing loss. Several outcomes were not interrelated, especially the outcomes associated with binaural processing abilities. Second, factor analysis yielded latent factors related to LF and HF processing, most comfortable level and speech in noise. Vlaming et al. (2011) showed four dimensions in the factor analysis of the HEARCOM project data corresponding to HF and LF spectro-temporal processing, MCL, and recruitment. In contrast, the current study showed that the slopes of the loudness growth, both at LF and HF, were not interrelated and contributed to the first and second latent factors. Additionally, the speech-in-noise test performed in HEARCOM was associated with the LF processing, whereas, in the present study, speech-in-noise dominates the fourth factor and is significantly correlated with HFs. The reason for this discrepancy might be the use of different types of noise (fluctuating masker) and test procedures in the two studies. Furthermore, in the HEARCOM study, the group participants included some younger hearing-impaired listeners and also participants with asymmetric or mixed hearing losses.

Overall, the data of the present study seem to be dominated by the audiometric profiles, with LF and HF processing reflecting the main sources of variability in the data. However, binaural processing abilities, loudness perception, and speech-in-noise outcomes showed a greater contribution to the variability of the supra-threshold measures than spectro-temporal processing outcomes.

12.2. Effects of the Participant's Cognitive Abilities

In this study, only auditory tests were considered. Indeed, some of the tests were quite demanding, which might have affected the results of listeners with reduced cognitive abilities. Here, only the age could be indicative of a likely cognitive decline, but it cannot confirm or deny this. In terms of age, we observed a significant effect on the results of the IPD and tone-in-noise tests. However, a thorough analysis on the effect of age was not carried out with the existing data. A more interesting approach would be to replicate this study including cognitive tests assessing executive functions, working memory or attention span with the aim of including a heterogeneous group of listeners with various cognitive abilities. Such a study could potentially assess both hearing and cognitive abilities toward a feasible test battery that can be included in the audiological assessment, either in or out of the clinic.

12.3. Extending the Test Battery to Other Clinical Populations

The proposed test battery was tested in a population of hearing-aid user candidates with various hearing abilities. However, the inclusion criteria left out the hearing-impaired listeners that were not likely to suffer from nonsyndromic presbycusis or noise-induced hearing loss. This means that adults with asymmetric hearing losses, severe-to-profound hearing loss, younger people with hearing deficits result of ototoxicity, or genetic conditions could potentially be included in a future study. Nevertheless, some tests might be affected by audibility or other aspects that have to be taken into account first. Furthermore, the test battery could be adapted to the pediatric clinical population, although that may be challenging for some tests.

12.4. Toward Clinical Feasibility of the Tests

The test–retest reliability of the test battery was investigated based on the results of a subset of listeners who participated 2–5 months after the first visit. The analysis was based on the ICC and the

SEM. Some of the tests, such as IPD_{fmax}, binaural pitch, and eAUD-HF (FLFT), showed good to excellent test–retest reliability with all ICC values above 0.9, while other tests, such as the extended audiometry in noise and speech intelligibility in quiet, showed poor reliability. The selected tests were conducted in two sessions and the total time was, on average, 3 h including the instructions and interview. In realistic clinical setups, a subset of tests with high reliability and a reasonably low difficulty would need to be prioritized. For a clinical version of the test battery, other tracking procedures such as Bayesian functional information (Remus and Collins, 2008) might be adopted to improve the reliability and time efficiency in some tasks such as STM and tone detection in noise. Moreover, if time-efficiency is crucial, testing some aspects of auditory processing out of the clinic, as other proposed test batteries for auditory research (Lelo De Larrea-Mancera et al., 2020), might be a solution for completing the patient's hearing profile. The use of speech-in-noise tests can be a useful tool for the characterization of the listener's hearing deficits that can be performed under different conditions, including monaural, binaural, unaided, and aided stimuli presentations. While here the tests were performed monaurally and unaided, a binaural condition as well as at least one aided measure (i.e., with hearing aids) could also be included in clinical practice. A clinical test battery with the subset of tests that showed a good or excellent test–retest reliability should be evaluated in a large scale study. In this study, we explored the use of an extended audiometry using the same test procedure for assessing high-frequency audibility (eAUD-HF), tone-in-noise detection (eAUD-N), spectro-temporal resolution (eAUD-S and eAUD-T), and binaural processing abilities. This procedure can be further explored and be performed by a hearing-care professional rather than in the current experimental setup. However, if the goal is to accurately estimate the hearing deficits of the patient, the test battery should include several aspects of auditory processing and provide detailed information on the supra-threshold deficits of the patient. The tests that showed potential for the clinical implementation were ACALOS, HINT, fSTM (only the LF condition), Binaural Pitch, and IPD_{fmax}. Such a test battery could serve to identify a clinically relevant subset of patients (auditory profiles) that may benefit from specific types of hearing rehabilitation toward a "stratified approach" (Trusheim et al., 2007) for audiology practice.

12.5. Toward Personalized Rehabilitation Based on Hearing Deficits

The present study was motivated by a novel approach for hearing loss characterization, recently proposed in Sanchez-Lopez et al. (2018). In their study, a data-driven profiling method was able to identify four relevant groups of listeners with a large similarity within each group and a substantial dissimilarity across groups. This stratification was possible by using two abstract dimensions (distortion type-I and distortion type-II) that can characterize each individual's hearing deficits. The dataset obtained in the present study was analyzed using the same approach in Sanchez Lopez et al. (2020) showing that speech intelligibility-related deficits and loudness-related deficits were associated with the two abstract and orthogonal dimensions. Moreover, four relevant subpopulations were proposed as the auditory profiles. Although the proposed data-driven method is constrained and other relevant subpopulations may be found using a less restrictive approach, the current definition of the four auditory profiles allows a necessary simplification of the variety of hearing impairments than enables a meaningful stratification.

Following the principles of stratified medicine (Trusheim et al., 2007), the first criterion for implementing personalized treatments is that the identification of the patient subpopulations must be technically feasible. In stratified medicine, the patient's phenotype is usually obtained by clinical biomarkers, which are measurable characteristics that associate the optimal treatment to a patient subpopulation. In the present study, different perceptual measurements are proposed as candidates for establishing the association between the heterogeneous hearing deficits and potential target treatments. However, before the clinical biomarkers can be established, patient subpopulations with a likely different response to different treatments must be identified. This has been explored in Wu et al. (2020, 2021) and Sanchez Lopez et al. (2021). A profile-based hearing-aid fitting process has also been implemented in a large field study within the BEAR project. In that study, the patients will be tested with a subset of the tests presented here, fitted with hearing aids, and their aided performance and self-perceived benefit were evaluated in and out of the clinic. This is expected to set the basis for targeted interventions involving not only hearing-aid fitting, but also the use of new tools for evaluating the experiences of hearing-aid users (Lund et al., 2020) and recommendations for individualized pathways (Sanchez-Lopez et al., 2021a).

13. CONCLUSION

The current study has shown the rationale behind the BEAR test battery and the selected tests for characterizing hearing deficits in listeners with various hearing abilities. The analysis of the data showed that a reduced BEAR test battery has potential for clinical implementation, providing relevant and reliable information reflecting several auditory domains. The proposed test battery showed good reliability, was reasonably time-efficient, and easy to perform. The implementation of a clinical version of the test battery is to be evaluated in future research, e.g., in a larger field study to further refine the auditory profiling approach. Moreover, the current data have been already analyzed for the purpose of auditory profiling (Sanchez Lopez et al., 2020), showing the potential of this test battery for hearing rehabilitation.

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/10.5281/zenodo.3459579.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Science-Ethics Committee for the Capital Region of Denmark H-16036391. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Author contributions according to CRediT (Contributor Roles Taxonomy). RS-L: conceptualization, methodology, software, validation, formal analysis, investigation, data curation, visualization, and writing-original draft and editing. SN: investigation, validation, resources, formal analysis, visualization, and writing-original draft. ME-H-A: methodology, investigation, resources, and writing-review. MF and FB: conceptualization, methodology, supervision, and writing-editing. MW and OC: investigation, resources, and writing-review. TN: methodology, supervision, project administration, and writing-review. SS and TD: conceptualization, methodology, supervision, project administration, funding acquisition, and writing-review. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Innovation Fund Denmark Grand Solutions 5164-00011B (Better hEAring Rehabilitation project), Oticon, GN Resound, Widex, and other partners (Aalborg University, University of Southern Denmark, the Technical University of Denmark, Force Technology, Aalborg, Odense, and Copenhagen University Hospitals).

ACKNOWLEDGMENTS

We thank the staff from OUH, BBH, and HEA, especially JH Schmidt, SS Houmøller, E Kjærbøl, RS Sørensen, and the student helpers from the MSc of Audiology at SDU. We also want to show our gratitude to all the participants in the study. We want to thank I. Holube and two reviewers for their comments in an earlier version of this manuscript, as well as SD Ewert for the AFC framework. The funding and collaboration of all BEAR partners are sincerely acknowledged.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.724007/full#supplementary-material

REFERENCES

Anderson, M. C., Arehart, K. H., and Souza, P. E. (2018). Survey of current practice in the fitting and fine-tuning of common signal-processing features in hearing aids for adults. *J. Am. Acad. Audiol.* 29, 118–124. doi: 10.3766/jaaa.16107

Bernstein, J. G., Mehraei, G., Shamma, S., Gallun, F. J., Theodoroff, S. M., and Leek, M. R. (2013). Spectrotemporal modulation sensitivity as a predictor of speech intelligibility for hearing-impaired listeners. *J. Am. Acad. Audiol.* 24, 293–306. doi: 10.3766/jaaa.24.4.5

Bernstein, J. G. W., Danielsson, H., Hällgren, M., Stenfelt, S., Rönnberg, J., and Lunner, T. (2016). Spectrotemporal modulation sensitivity as a predictor of speech-reception performance in noise with hearing aids. *Trends Hear.* 20, 1–17. doi: 10.1177/2331216516670387

Bianchi, F., Carney, L. H., Dau, T., and Santurette, S. (2019). Effects of musical training and hearing loss on fundamental frequency discrimination and temporal fine structure processing: Psychophysics and modeling. *J. Assoc. Res. Otolaryngol.* 20, 263–277. doi: 10.1007/s10162-018-00710-2

Bisgaard, N., Vlaming, M. S. M. G., and Dahlquist, M. (2010). Standard audiograms for the iec 60118-15 measurement procedure. *Trends Amplif.* 14, 113–120. doi: 10.1177/1084713810379609

Brand, T., and Hohmann, V. (2002). An adaptive procedure for categorical loudness scaling. *J. Acoust. Soc. Am.* 112, 1597–1604. doi: 10.1121/1.1502902

Brown, M., and Musiek, F. (2013). Pathways: The fundamentals of masking level differences for assessing auditory function. *Hear J.* 66, 16. doi: 10.1097/01.HJ.0000425772.41884.1d

Brungart, D. S., Sheffield, B. M., and Kubli, L. R. (2014). Development of a test battery for evaluating speech perception in complex listening environments. *J. Acoust. Soc. Am.* 136, 777–790. doi: 10.1121/1.4887440

Choi, J. E., Hong, S. H., Won, J. H., Park, H.-S., Cho, Y. S., Chung, W.-H., et al. (2016). Evaluation of cochlear implant candidates using a non-linguistic spectrotemporal modulation detection test. *Sci. Rep.* 6, 35235. doi: 10.1038/srep35235

Cramer, E. M., and Huggins, W. H. (1958). Creation of pitch through binaural interaction. *J. Acoust. Soc. Am.* 30, 413–417. doi: 10.1121/1.1909628

Darwin, C. J. (1997). Auditory grouping. *Trends Cogn. Sci.* 1, 327–333. doi: 10.1016/S1364-6613(97)01097-8

Desloge, J. G., Reed, C. M., Braida, L. D., Perez, Z. D., and D'Aquila, L. A. (2017). Masking release for hearing-impaired listeners: the effect of increased audibility through reduction of amplitude variability. *J. Acoust. Soc. Am.* 141, 4452–4465. doi: 10.1121/1.4985186

Durlach, N. I. (1963). Equalization and cancellation theory of binaural masking-level differences. *J. Acoust. Soc. Am.* 35, 1206–1218. doi: 10.1121/1.1918675

Elberling, C., Ludvigsen, C., and Lyregaard, P. E. (1989). Dantale: a new danish speech material. *Scand. Audiol.* 18, 169–175. doi: 10.3109/01050398909070742

Encina-Llamas, G., Harte, J., Dau, T., Shinn-Cunningham, B., and Epp, B. (2019). Investigating the effect of cochlear synaptopathy on envelope following responses using a model of the auditory nerve. *J. Assoc. Res. Otolaryngol.* 20, 363–382. doi: 10.1007/s10162-019-00721-7

Esch, T. E. M. V., and Dreschler, W. A. (2015). Relations between the intelligibility of speech in noise and psychophysical measures of hearing measured in four languages using the auditory profile test battery. *Trends Hear.* 19:2331216515618902. doi: 10.1177/2331216515618902

Eustaquio-Martín, A., and Lopez-Poveda, E. A. (2011). Isoresponse versus isoinput estimates of cochlear filter tuning. *J. Assoc. Res. Otolaryngol.* 12, 281–299. doi: 10.1007/s10162-010-0252-1

Ewert, S. (2013). "Afc-a modular framework for running psychoacoustic experiments and computational perception models," in *Proceedings of the International Conference on Acoustics AIA-DAGA 2013* (Merano), 1326–1329.

Festen, J. M., and Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J. Acoust. Soc. Am.* 88, 1725–1736. doi: 10.1121/1.400247

Füllgrabe, C., and Moore, B. C. J. (2017). Evaluation of a method for determining binaural sensitivity to temporal fine structure (tfs-af test) for older listeners with normal and impaired low-frequency hearing. *Trends Hear.* 21, 233121651773723. doi: 10.1177/2331216517737230

Gelfand, S. A. (2009). *Essentials of Audiology*. 3rd Edn. New York, NY: Thieme.

Gieseler, A., Tahden, M. A. S., Thiel, C. M., Wagener, K. C., Meis, M., and Colonius, H. (2017). Auditory and non-auditory contributions for unaided speech recognition in noise as a function of hearing aid use. *Front. Psychol.* 8:219. doi: 10.3389/fpsyg.2017.00219

Glasberg, B. R., and Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hear. Res*. 47, 103–138. doi: 10.1016/0378-5955(90)90170-T

Glasberg, B. R., and Moore, B. J. (1989). Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech. Scandinavian audiology. *Supplementum* 32, 1–25.

Grant, K. W., Kubli, L. R., Phatak, S. A., Galloza, H., and Brungart, D. S. (2021). Estimated prevalence of functional hearing difficulties in blast-exposed service members with normal to near-normal-hearing thresholds. *Ear Hear*. doi: 10.1097/AUD.0000000000001067. [Epub ahead of print].

Houmoller, S. S., Wolff, A., Moller, S., Narne, V. K., Narayanan, S. K., Godballe, C., et al. (2021). Prediction of successful hearing aid treatment in first-time and experienced hearing aid users: using the international outcome inventory for hearing aids. *Int. J. Audiol*. doi: 10.1080/14992027.2021.1916632. [Epub ahead of print].

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

ISO 16832 (2006). Acoustics-loudness scaling by means of categories. *Int. Organ. Standard*. 2006, 1–12. Available online at: https://www.iso.org/standard/32442.html

ISO 8253-1 (2010). Acoustics-audiometric test methods-part 1: pure-tone air and bone conduction audiometry. *Int. Organ. Standard*.

ISO 8253-3 (2012). Acoustics. audiometric test methods-part 3: speech audiometry. *Int. Organ. Standard*.

Johannesen, P. T., Pérez-González, P., Kalluri, S., Blanco, J. L., and Lopez-Poveda, E. A. (2016). The influence of cochlear mechanical dysfunction, temporal processing deficits, and age on the intelligibility of audible speech in noise for hearing-impaired listeners. *Trends Hear*. 20, 233121651664105. doi: 10.1177/2331216516641055

Jürgens, T., Kollmeier, B., Brand, T., and Ewert, S. D. (2011). Assessment of auditory nonlinearity for listeners with different hearing losses using temporal masking and categorical loudness scaling. *Hear. Res*. 280, 177–191. doi: 10.1016/j.heares.2011.05.016

Kaernbach, C. (1990). A single-interval adjustment-matrix (siam) procedure for unbiased adaptive testing. *J. Acoust. Soc. Am*. 88, 2645–2655. doi: 10.1121/1.399985

Kamerer, A. M., Kopun, J. G., Fultz, S. E., Neely, S. T., and Rasetshwane, D. M. (2019). Reliability of measures intended to assess threshold-independent hearing disorders. *Ear. Hear*. 40, 1267–1279. doi: 10.1097/AUD.0000000000000711

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. Am*. 116, 2395–2405. doi: 10.1121/1.1784440

Koo, T. K., and Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med*. 15, 155–163. doi: 10.1016/j.jcm.2016.02.012

Larsby, B., and Arlinger, S. (1998). A method for evaluating temporal, spectral and combined temporal-spectral resolution of hearing. *Scand. Audiol*. 27, 3–12. doi: 10.1080/010503998419641

Lecluyse, W., Tan, C. M., McFerran, D., and Meddis, R. (2013). Acquisition of auditory profiles for good and impaired hearing. *Int. J. Audiol*. 52, 596–605. doi: 10.3109/14992027.2013.796530

Lelo De Larrea-Mancera, E. S., Stavropoulos, T., Hoover, E. C., Eddins, D. A., Gallun, F. J., and Seitz, A. R. (2020). Portable automated rapid testing (part) for auditory assessment: validation in a young adult normal-hearing population. *J. Acoust. Soc. Ame*. 148, 1831–1851. doi: 10.1121/10.0002108

Liberman, M. C., Epstein, M. J., Cleveland, S. S., Wang, H., and Maison, S. F. (2016). Toward a differential diagnosis of hidden hearing loss in humans. *PLoS ONE* 11:e0162726. doi: 10.1371/journal.pone.0162726

Lund, K., Ordoñez, R., Nielsen, J. B., and Hammershøi, D. (2020). Sentence-based experience logging in new hearing aid users. *Am. J. Audiol*. 29, 631–637. doi: 10.1044/2020_AJA-19-00077

Mehraei, G., Gallun, F. J., Leek, M. R., and Bernstein, J. G. W. (2014). Spectrotemporal modulation sensitivity for hearing-impaired listeners: dependence on carrier center frequency and the relationship to speech intelligibility. *J. Acoust. Soc. Am*. 136, 301–316. doi: 10.1121/1.4881918

Moore, B. C., Huss, M., Vickers, D. A., Glasberg, B. R., and Alcantara, J. I. (2000). A test for the diagnosis of dead regions in the cochlea. *Br. J. Audiol*. 34, 205–224. doi: 10.3109/03005364000000131

Moore, B. C. J. (2001). Dead regions in the cochlea: Diagnosis, perceptual consequences, and implications for the fitting of hearing aids. *Trends Amplif*. 5, 1–34. doi: 10.1177/108471380100500102

Moore, B. C. J. (2007). *Cochlear Hearing Loss*. West Sussex: John Wiley & Sons, Ltd.

Neher, T. (2017). Characterizing the binaural contribution to speech-in-noise reception in elderly hearing-impaired listeners. *J. Acoust. Soc. Am*. 141:EL159. doi: 10.1121/1.4976327

Neher, T., Laugesen, S., Søgaard Jensen, N., and Kragelund, L. (2011). Can basic auditory and cognitive measures predict hearing-impaired listeners' localization and spatial speech recognition abilities? *J. Acoust. Soc. Am*. 130, 1542–1558. doi: 10.1121/1.3608122

Nielsen, J. B., and Dau, T. (2011). The danish hearing in noise test. *Int. J. Audiol*. 50, 202–208. doi: 10.3109/14992027.2010.524254

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

Oetting, D., Brand, T., and Ewert, S. D. (2014). Optimized loudness-function estimation for categorical loudness scaling data. *Hear. Res*. 316, 16–27. doi: 10.1016/j.heares.2014.07.003

Oetting, D., Hohmann, V., Appell, J. E., Kollmeier, B., and Ewert, S. D. (2018). Restoring perceived loudness for listeners with hearing loss. *Ear. Hear*. 39, 664–78. doi: 10.1097/AUD.0000000000000521

Oxenham, A. J., and Simonson, A. M. (2009). Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference. *J. Acoust. Soc. Am*. 125, 457–468. doi: 10.1121/1.3021299

Plomp, R. (1978). Auditory handicap of hearing impairment and the limited benefit of hearing aids. *J. Acoust. Soc. Am*. 63, 533–549. doi: 10.1121/1.381753

Plomp, R., and Mimpen, A. M. (1979). Speech-reception threshold for sentences as a function of age and noise level. *J. Acoust. Soc. Am*. 66, 1333–1342. doi: 10.1121/1.383554

Prendergast, G., Hymers, M., and Lee, A. (2020). A quick and reliable estimate of extended high-frequency hearing. *Int. J. Audiol*. 59, 823–827. doi: 10.1080/14992027.2020.1767810

Remus, J. J., and Collins, L. M. (2008). Comparison of adaptive psychometric procedures motivated by the theory of optimal experiments: Simulated and experimental results. *J. Acoust. Soc. Am*. 123, 315–326. doi: 10.1121/1.2816567

Rhebergen, K. S., Van Esch, T. E., and Dreschler, W. A. (2015). Measuring temporal resolution (release of masking) with a hughson-westlake up-down instead of a békèsy-tracking procedure. *J. Am. Acad. Audiol*. 26, 563–571. doi: 10.3766/jaaa.14087

Rhebergen, K. S., Versfeld, N. J., and Dreschler, W., a. (2006). Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise. *J. Acoust. Soc. Am*. 120, 3988–3997. doi: 10.1121/1.2358008

Rieke, C. C., Clavier, O. H., Allen, L. V., Anderson, A. P., Brooks, C. A., Fellows, A. M., et al. (2017). Fixed-level frequency threshold testing for ototoxicity monitoring. *Ear. Hear*. 38, e369–e375. doi: 10.1097/AUD.0000000000000433

Rönnberg, J., Lunner, T., Ng, E. H. N., Lidestam, B., Zekveld, A. A., Sörqvist, P., et al. (2016). Hearing impairment, cognition and speech understanding: exploratory factor analyses of a comprehensive test battery for a group of hearing aid users, the n200 study. *Int. J. Audiol*. 55, 623–642. doi: 10.1080/14992027.2016.1219775

Rønne, F. M., Laugesen, S., and Jensen, N. S. (2017). Selection of test-setup parameters to target specific signal-to-noise regions in speech-on-speech intelligibility testing. *Int. J. Audiol*. 56, 559–567. doi: 10.1080/14992027.2017.1300349

Rosowski, J. J., Stenfelt, S., and Lilly, D. (2013). An overview of wideband immittance measurements techniques and terminology. *Ear Hear*. 34, 9s–16s. doi: 10.1097/AUD.0b013e31829d5a14

Ross, B., Tremblay, K. L., and Picton, T. W. (2007). Physiological detection of interaural phase differences. *J. Acoust. Soc. Am*. 121, 1017–1027. doi: 10.1121/1.2404915

Sanchez Lopez, R., Fereczkowski, M., Neher, T., Santurette, S., and Dau, T. (2020). Robust data-driven auditory profiling towards precision audiology. *Trends Hear*. 24:2331216520973539. doi: 10.1177/2331216520973539

Sanchez Lopez, R., Fereczkowski, M., Santurette, S., Dau, T., and Neher, T. (2021). Towards auditory profile-based hearing-aid fitting: fitting rationale and pilot evaluation. *Audiol. Res*. 11, 10–21. doi: 10.3390/audiolres11010002

Sanchez-Lopez, R., Bianchi, F., Fereczkowski, M., Santurette, S., and Dau, T. (2018). Data-driven approach for auditory profiling and characterization of individual hearing loss. *Trends Hear*. 22, 233121651880740. doi: 10.1177/2331216518807400

Sanchez-Lopez, R., Dau, T., and Whitmer, W. M. (2021a). Audiometric profiles and patterns of benefit: a data-driven analysis of subjective hearing difficulties and handicaps. *Int. J. Audiol*. doi: 10.1080/14992027.2021.1905890. [Epub ahead of print].

Sanchez-Lopez, R., Nielsen, S. G., El-Haj-Ali, M., Bianchi, F., Fereczkowski, M., Cañete, O., et al. (2021b). Data from "Auditory tests for characterizing hearing deficits in listeners with various hearing abilities: The BEAR test battery" (v1.1) [Data set]. Zenodo. doi: 10.5281/zenodo.4923009

Sanders, N. C., and Chin, S. B. (2009). Phonological distance measures. *J. Quant. Linguist*. 16, 96–114. doi: 10.1080/09296170802514138

Santurette, S., and Dau, T. (2012). Relating binaural pitch perception to the individual listener's auditory profile. *J. Acoust. Soc. Am*. 131, 2968–2986. doi: 10.1121/1.3689554

Saunders, G. H., Field, D. L., and Haggard, M. P. (1992). A clinical test battery for obscure auditory dysfunction (oad): development, selection and use of tests. *Br. J. Audiol*. 26, 33–42. doi: 10.3109/03005369209077869

Schädler, M. R., Hüdlsmeier, D., Warzybok, A., and Kollmeier, B. (2020). Individual aided speech-recognition performance and predictions of benefit for listeners with impaired hearing employing fade. *Trends Hear*. 24:2331216520938929. doi: 10.1177/2331216520938929

Schorn, K., and Zwicker, E. (1990). Frequency selectivity and temporal resolution in patients with various inner ear disorders. *Audiology* 29, 8–20. doi: 10.3109/00206099009081641

Schütt, H. H., Harmeling, S., Macke, J. H., and Wichmann, F. A. (2016). Painfree and accurate bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vision Res*. 122, 105–123. doi: 10.1016/j.visres.2016.02.002

Stratford, P. W., and Goldsmith, C. H. (1997). Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data. *Phys. Ther*. 77, 745–750. doi: 10.1093/ptj/77.7.745

Strelcyk, O., and Dau, T. (2009). Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing. *J. Acoust. Soc. Am*. 125, 3328. doi: 10.1121/1.3097469

Strelcyk, O., Zahorik, P., Shehorn, J., Patro, C., and Derleth, R. P. (2019). Sensitivity to interaural phase in older hearing-impaired listeners correlates with nonauditory trail making scores and with a spatial auditory task of unrelated peripheral origin. *Trends Hear*. 23:233121651986449. doi: 10.1177/2331216519864499

Tai, Y., and Husain, F. T. (2018). Right-ear advantage for speech-in-noise recognition in patients with nonlateralized tinnitus and normal hearing sensitivity. *J. Assoc. Res. Otolaryngol*. 19, 211–221. doi: 10.1007/s10162-017-0647-3

Thorup, N., Santurette, S., Jørgensen, S., Kjærbøl, E., Dau, T., and Friis, M. (2016). Auditory profiling and hearing-aid satisfaction in hearing-aid candidates. *Dan. Med. J*. 63, 1–5. Available online at: https://ugeskriftet.dk/dmj/auditory-profiling-and-hearing-aid-satisfaction-hearing-aid-candidates

Trusheim, M. R., Berndt, E. R., and Douglas, F. L. (2007). Stratified medicine: strategic and economic implications of combining drugs and clinical biomarkers. *Nat. Rev. Drug Discov*. 6, 287–293. doi: 10.1038/nrd2251

van Esch, T. E. M., and Dreschler, W. A. (2011). Measuring spectral and temporal resolution simultaneously: a comparison between two tests. *Int. J. Audiol*. 50, 477–490. doi: 10.3109/14992027.2011.572083

Vermiglio, A. J., Soli, S. D., Freed, D. J., and Fang, X. (2020). The effect of stimulus audibility on the relationship between pure-tone average and speech recognition in noise ability. *J. Am. Acad. Audiol*. 31, 224–232. doi: 10.3766/jaaa.19031

Vinay, H.ansen, A. S., Raen, Ø., and Moore, B. C. J. (2017). Reference thresholds for the ten(hl) test for people with normal hearing. *Int. J. Audiol*. 56, 672–676. doi: 10.1080/14992027.2017.1307531

Vlaming, M. S. M. G., Kollmeier, B., Dreschler, W. A., Martin, R., Wouters, J., Grover, B., et al. (2011). Hearcom: hearing in the communication society. *Acta Acustica United Acustica* 97, 175–192. doi: 10.3813/AAA.918397

Wagener, K., Josvassen, J. L., and Ardenkjær, R. (2003). Design, optimization and evaluation of a danish sentence test in noise. *Int. J. Audiol*. 42, 10–17. doi: 10.3109/14992020309056080

Wu, M., Cañete, O. M., Schmidt, J. H., Fereczkowski, M., and Neher, T. (2021). Influence of three auditory profiles on aided speech perception in different noise scenarios. *Trends Hear*. 25:23312165211023709. doi: 10.1177/23312165211023709

Wu, M., Sanchez-Lopez, R., El-Haj-Ali, M., Nielsen, S. G., Fereczkowski, M., Dau, T., et al. (2020). Investigating the effects of four auditory profiles on speech recognition, overall quality, and noise annoyance with simulated hearing-aid processing strategies. *Trends Hear*. 24:2331216520960861. doi: 10.1177/2331216520960861

Zaar, J., Simonsen, L., Behrens, T., and Laugesen, S. (2018). "Towards a clinically viable spectro-temporal modulation test," in *International Hearing Aid Conference 2018, IHCON 2018* (Lake Tahoe).

Zaar, J., Simonsen, L. B., Behrens, T., Dau, T., and Laugesen, S. (2020). "Investigating the relationship between spectro-temporal modulation detection, aided speech perception, and directional noise reduction preference in hearing-impaired listeners," in *Proceedings of the International Symposium on Auditory and Audiological Research, Vol. 7* (Nyborg), 181–188. Available online at: http://proceedings.isaar.eu/index.php/isaarproc/article/view/2019-22

Conflict of Interest: RS-L was employed by the company Interacoustics A/S, Denmark. FB was employed by the company Oticon Medical, Denmark. SB was employed by company 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.

frontiers in Digital Health

MINI REVIEW
published: 07 October 2021
doi: 10.3389/fdgth.2021.724748

User-Operated Audiometry Project (UAud) – Introducing an Automated User-Operated System for Audiometric Testing Into Everyday Clinic Practice

Christos Sidiras[1], Raul Sanchez-Lopez[2,3], Ellen Raben Pedersen[1], Chris Bang Sørensen[1], Jacob Nielsen[1] and Jesper Hvass Schmidt[4,5,6]*

[1] *Faculty of Engineering, The Maersk Mc-Kinney Møller Institute, University of Southern Denmark, Odense, Denmark,* [2] *Interacoustics Research Unit, Kongens Lyngby, Denmark,* [3] *Hearing Systems Section, Department of Health Technology, Technical University of Denmark, Kongens Lyngby, Denmark,* [4] *Department of Clinical Research, Faculty of Health Science, University of Southern Denmark, Odense, Denmark,* [5] *OPEN, Open Patient Data Explorative Network, Odense University Hospital, Odense, Denmark,* [6] *Research Unit for ORL-Head and Neck Surgery and Audiology, Odense University Hospital and University of Southern Denmark, Odense, Denmark*

OPEN ACCESS

Edited by:
Dennis L. Barbour, Washington University in St. Louis, United States

Reviewed by:
Robert Eikelboom, Ear Science Institute Australia, Australia
Piotr Henryk Skarzynski, Institute of Physiology and Pathology of Hearing (IFPS), Poland

***Correspondence:**
Christos Sidiras
chsi@mmmi.sdu.dk

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: *14 June 2021*
Accepted: *13 September 2021*
Published: *07 October 2021*

Citation:
Sidiras C, Sanchez-Lopez R, Pedersen ER, Sørensen CB, Nielsen J and Schmidt JH (2021) User-Operated Audiometry Project (UAud) – Introducing an Automated User-Operated System for Audiometric Testing Into Everyday Clinic Practice. Front. Digit. Health 3:724748. doi: 10.3389/fdgth.2021.724748

Hearing loss is the third leading cause of years lived with disability. It is estimated that 430 million people worldwide are affected, and the number of cases is expected to increase in the future. There is therefore increased pressure on hearing health systems around the world to improve efficiency and reduce costs to ensure increased access to quality hearing health care. Here, we describe the User-Operated Audiometry project, the goal of which is to introduce an automated system for user-operated audiometric testing into everyday clinic practice as a means to relieve part of this pressure. The alternative to the existing referral route is presented in which examination is executed via the user-operated system. This route is conceptualized as an interaction between the patient, the system, and the hearing care professional (HCP). Technological requirements of the system and challenges that are related to the interaction between patients, the user-operated system, and the HCPs within the specific medical setting are discussed. Lastly, a strategy for the development and implementation of user-operated audiometry is presented, which includes initial investigations, a validation study, and implementation in a real-life clinical situation.

Keywords: audiometry, hearing loss, user-operated, automated, hearing aid, hearing test, telehealth

INTRODUCTION

Hearing loss is the third leading cause of years lived with disability (1). The World Health Organization estimates that 430 million people worldwide live with a disabling hearing loss and one-third of older adults (>65 years) are affected by this condition (2). The annual cost for untreated hearing loss is estimated by the WHO to be 980 billion USD globally. Recent studies have also highlighted hearing loss as being one of the greatest modifiable risk factors for cognitive decline, dementia, and depression later in life (3–6).

Since the most common type of hearing loss is associated with age and given that the percentage of people above the age of 65 is increasing (2), it is expected that the number of people with hearing loss will also increase (1). There is therefore increased pressure on hearing health systems around the world to improve efficiency and reduce costs to ensure increased access to quality hearing health care.

Treatment of hearing loss often involves the examination of hearing function, the selection and fitting of hearing aids, and the evaluation of hearing-aid performance. As part of the initial assessment, the pure-tone audiometry is time-consuming, as testing time may exceed 20 min (7, 8) and seems to be the bottleneck for further clinical decisions in the current health care system. According to data from the Danish Health and Medicine Authority (9), the current waiting time for the examination in the public system is up to 86 weeks. Currently, this assessment requires the presence of a hearing care professional (HCP). Although pure-tone audiometry is not the only examination of the initial assessment, it is probably the most time-consuming part. The development of a user-operated audiometry system, that is, one that does not require the presence of a HCP, will arguably help to free resources needed in the initial assessment procedure.

Here, we describe the User-Operated Audiometry (UAud) project, the goal of which is to introduce an automated system for user-operated audiometric testing into everyday clinical practice. Importantly, the user-operated system is not intended to replace the audiological assessment done by an HCP, but rather to offer an alternative to manual audiometric testing when applicable. Furthermore, the UAud project focuses on the user-operated diagnostic examination, in the clinical environment and with calibrated clinical devices. Although these approaches, such as asynchronous tele-audiology (10), have shown their potential as screening tools and likely impact for reducing the global burden of hearing loss, the UAud scope is on the more efficient use of human resources in the hearing healthcare services (11). It is therefore expected that the user-operated hearing assessments will reduce the clinical hours spent on air conducted pure-tone tests by a significant amount, freeing up time that can be spent better by the HCP on counseling the patient on using hearing aids and/or to see more patients throughout the day.

The purpose of this perspective article is (a) to briefly review previous relevant related work about user-operated audiometry, (b) to describe the scope and focus of the UAud project, (c) to create a perspective on the challenges and possible barriers related to the inclusion of the new examination paradigm, and (d) to present a strategy for addressing these challenges and effectively implement user-operated audiometry in the daily practice.

Previous Research

The opportunities of automated audiometry have been explored since the beginning of the computer age (12) and has been widely used for research purposes (13). Further, automated audiometry is implemented for medical purposes, though limited to screening (14, 15). Recently, research efforts have shown promising results toward its potential diagnostic use (16–18), both in the clinic and as an opportunity for implementing tele-audiology (19, 20). In a systematic review, Mahomed et al. (21) suggested that automated pure-tone audiometry provides an accurate measure, but validation is still needed for specific cases such as difficult-to-test populations. In a more recent review, Shojaeemend and Ayatollahi (22) concluded that automated audiometry produces clinically acceptable results compared with traditional audiometry.

Considerable research contributions have been focused on the Automated Method for Testing Auditory Sensitivity (AMTAS®) test, a single-interval, forced choice method (yes-no paradigm) task with an adaptive algorithm for pure-tone detection thresholds (23). This test is intended to be used in-the-clinic with standardized diagnostic equipment but in an asynchronous user-operated approach. The HCP instructs the patient and supervises the accuracy of the results, but they do not need to be present while the test is being carried out. A series of studies has explored its accuracy and validity in clinical settings in children, adults, and elderly populations (23–28). Overall, the use of user-operated audiometry using medical equipment and a controlled environment shows promise for the implementation of AMTAS® in the current clinical practice.

Automated Audiometric Tests Beyond Pure-Tone Audiometry

The automatization of presenting pure tones to a patient while concurrently analyzing the patient's responses does not present many technical challenges and has the potential to be included in a user-operated version suitable for a broad part of the population. However, the implementation of other user-operated audiological tests such as speech audiometry is more challenging. Research efforts have been made in the direction of automatizing speech-in-noise recognition (29, 30), and the development of self-scoring multilingual speech tests (31, 32). These tests still need a careful selection and validation of the speech material, and their ecological validity is limited. Either the so-called sentence-matrix tests [first introduced by Hagerman (33)] or the digit-triplet test [first introduced by Smits et al. (34) as a screening speech intelligibility test by telephone] are both affected by the same drawbacks. On one hand, the development and validation of the speech material in different languages is easier than for other tests (e.g., hearing in noise test) but it still requires substantial efforts. On the other hand, are the speech reception thresholds obtained using these tests are unrealistic since the speech stimulus is cognitively undemanding (fixed structured sentences or digits) and it is presented as a closed set (the patient has only a limited number of possible responses). Therefore, this argues against the introduction of a standardized test for assessing speech recognition abilities in noise for worldwide implementation.

The assessment of the patient's ability to extract the essential features of the speech signal may lend itself to a more practical solution. The Audible Contrast Threshold (ACT™) is a new clinical test measuring spectro-temporal modulation sensitivity, in which the subject's task is to discriminate between spectro-temporally modulated noise and non-modulated noise. The idea

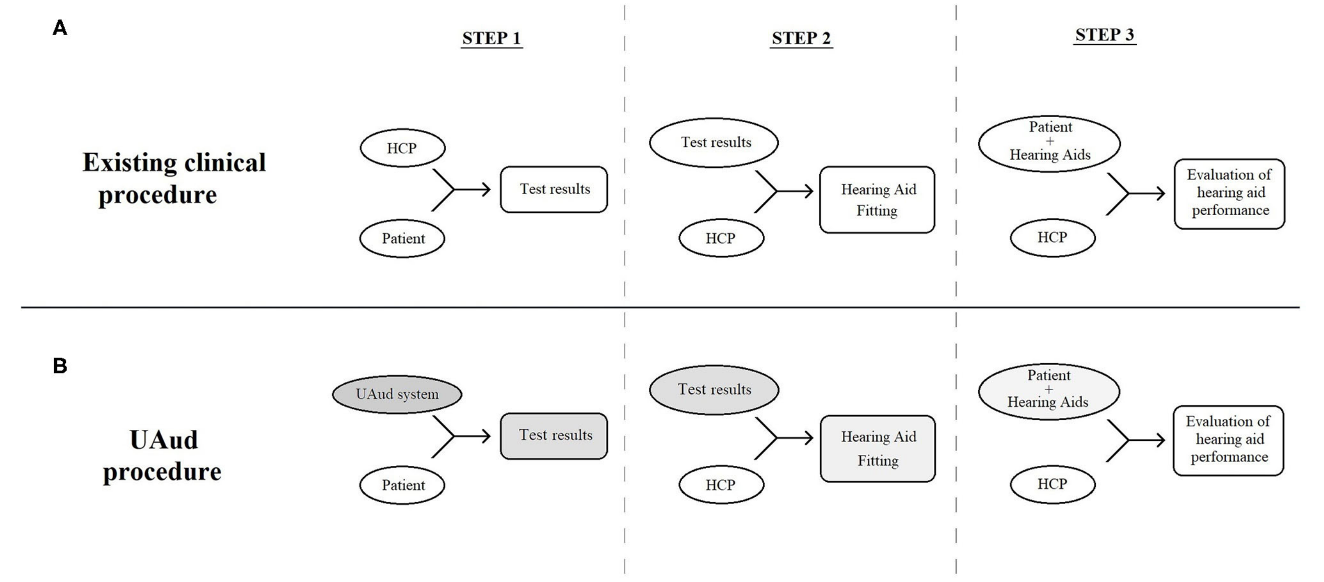


FIGURE 1 | Schematic representation of the 3-step existing **(A)** and user-operated audiometry **(B)** procedure, consisting of three parties, i.e., the patient, the user-operated system and the HCP. The user-operated examination affects step 1, and consequently steps 2 and 3, as medical decisions are based on the results of the former.

that the similarity of the spectro-temporal characteristics of the modulated stimuli and real speech would make sensitivity to spectro-temporal modulations a good predictor of speech-in-noise performance has been investigated by several research groups (35–37) with promising results. Subsequent research (38) indicated that the spectro-temporal modulation test used was too difficult for about 1/3 of the large clinical population tested. This problem was eventually solved by Zaar et al. (39), who further went on to show how spectro-temporal modulation detection thresholds predict aided speech-in-noise recognition in an ecologically valid scenario, as well as the benefit from using noise reduction in hearing aids (40). This research (39, 40) forms the basis of the clinical ACT™ test, which thus is a suprathreshold audiometric proxy for speech-in-noise testing with hearing aids measured with a language-independent stimulus. A further advantage of the ACT™ test is that it lends itself well to automatic user-operated implementation.

INCORPORATING USER-OPERATED AUDIOMETRY WITHIN THE EXISTING CLINICAL PROCEDURE

The existing clinical procedure for hearing-aid fitting consists of three main steps. Step 1: examination of hearing–audiometry, step 2: initial fitting of the hearing aids, and step 3: evaluation of hearing aid performance and re-adjustment if needed (see **Figure 1A**). Normally, all three steps are carried out by an HCP. The whole procedure requires the involvement of two parties (patient and HCP) who interact throughout all three steps.

The UAud project explores an alternative procedure by freeing the HCP from the major part of the first step (i.e., the audiometric tests, excluding the otoscopy and anamnesis) and instead introducing the user-operated system, while all the tasks and responsibilities of the HCP in steps 2 and 3 remain unchanged. This new procedure can be conceptualized as a human-digital system-human interaction (i.e., the patient, the user-operated system and the HCP, respectively). The three parties interact more than once and in more than one way (see **Figure 1B**), but the interaction is more complicated compared to the existing clinical procedure. The user-operated examination affects step 1, and consequently steps 2 and 3, as the practical hearing-aid fitting is based on the results of the former. Further and beyond these effects, there are less obvious ones. The HCP's reservations and skepticism may negatively affect both hearing-aid fitting and evaluation/re-adjustment (13). This may happen if either the patient or the HCP is skeptical of the accuracy of the hearing examination.

In order to better understand the dynamics of the 3-parties system (patient, user-operated system, and HCP), it seems useful to have a detailed perspective of the relevant characteristics of the parts, (i.e., humans and digital technology). Humans are goal-oriented and goal-directed (41). They learn quickly, have powerful selective attention, can be comparatively easily excited and get focused on something they find interesting, either feature or process (42). On the other side, humans are easily distracted, lose their interest quickly, or even give up if they get disappointed, confused, or tired. They have limited cognitive resources [e.g., working memory and attention; (43)], and they are often cognitively or emotionally biased (44).

Digital technology comes with its own pros and cons. Compared to humans, it is typically governed by clearer and known rules, so it can be standardized comparatively easily, and still offers some versatility through settings in the system. On the other hand, digital technology tends to become less adaptable after its release, and paradigm changes may require completely new technology to be developed, whereas a human might only need short retraining. Thus, every detail on every process and all potential pitfalls/bugs, must be thought of and addressed in advance before implementation. Excluding an arguably significant, still not relevant to the present study, part of current technology, i.e., artificial intelligence, one could argue that humans learn while technology is set.

Bringing together humans and digital technology is a challenge on its own, as large differences between them may negatively affect their interaction. Making digital technology more user-friendly and measuring the quality of the interaction is a wide field of study of its own and has a fast-growing body of research output in the last few decades. Indicatively, a search in Scopus with keywords ["usability" AND "technology"] yields 717 review studies alone (searched on 8-March-2021), while there is an increasing interest on usability for health evaluation and intervention tools [for reviews see (45–47)]. In the UAud project, the quality of the human-system interaction is crucially important, as deviations from optimum may affect not only the examination, i.e., the first step, but also the whole procedure.

Insights on the effects of the introduction of user-operated audiometry can be gained by comparing with audiometry operated by an HCP. Focus should be given on those aspects of human and digital technology that are crucial for the quality of the hearing examination results. In **Table 1** the main advantages of the HCP-operated test are in line with the disadvantages of the user-operated audiometry and vice versa. While the user-operated system can make use of the computation intelligence for analyzing the quality of the test results and apply a well-defined protocol to obtain the patient's hearing thresholds, the HCP-operated examination has the advantage that the professional can supervise and adapt the procedure to patient if needed (23), which is particularly important in certain populations (e.g., children and people with mild to moderate cognitive impairments).

TABLE 1 | Primary examples of the advantages and drawbacks of manually- and user-operated audiometry procedures identified in the UAud project.

	HCP operated audiometry	**User-operated audiometry**
Advantages	- The HCP can start the test when the listener has understood the task - Adapt the procedure and instructions to the individual - Constant supervision and observation - HCPs can trust their own actions	- Indirect measures gathered during the test: Reaction time - The procedure is well-defined, and all listeners are tested identically - Objective metric of the quality of the measurement
Drawbacks	- The protocol and criteria can vary from one HCP to other - The experience of the HCP may affect the measurement	- The user has to learn the task alone - The procedure cannot (a priori) be adapted to the individual - The HCP requires evidence to trust that patients have consistent response criteria

The table is the result of discussions among some of the researchers involved in the initial investigations of the assumption that the user-operated test will be performed by the patient alone.

SCOPE OF THE PROJECT

The UAud project aims to explore the possibilities of implementing user-operated audiometry in everyday clinical practice. The technological progress on automated audiometry of the past decades will be further developed, evaluated, and implemented in the clinic in the form of a system for user-operated testing of air-conduction pure-tone audiogram and ACT^{TM}. As the motivation behind the project is to free work hours from HCPs dedicated to the hearing examination, the system must be handled by the patient ideally without any supervision. All actions taken by the HCP (11, 23) such as preparation for and carrying out the examination and monitoring the behavior of the patient, must be executed in the absence of the HCP.

The requirements for successful implementation of the user-operated audiometry system are:

- The establishment of the necessary software, hardware, and testing environment.
- Acceptable usability and user experience of the system's software and hardware.
- Strategies to deal with situations such as patients needing further instructions or other considerations (patient's presenting tinnitus).
- The system must be designed to ensure that the patient's attention is adequately focused on the task during the examination.
- Optimal time efficiency, that is, the test must be accurate and reliable while avoiding causing patient fatigue due to extensive test time.
- Effective supervision, that is, the system must assess the quality of the examination, and give recommendation for further testing if necessary.

Further, the UAud project must account for the specific setting in which the system will be implemented. The setting includes (a) the medical context, that is, hearing health care and specifically the audiological examination (step 1 in **Figure 1**), (b) the specific use of the test results, that is, informing hearing intervention (as depicted in steps 2 and 3), (c) the people involved, that is, audiological patients, many of which are elderly, and HCPs, and (d) the fact that the patient will be examined alone, i.e., without the supervision of a HCP. Factors that must be accounted for include:

- Emotional factors of the patient at different stages of the process: (a) before the examination (e.g., are his/her feelings positive or negative toward the use of user-operated tests?, is he/she confident that the examination will go well?), (b) during the examination (e.g., does he/she feel confident with this approach or does he/she experience frustrations?), and (c) after the examination (e.g., does he/she feel that the examination went well and the results are accurate?). These factors may affect the examination itself, and further influence the HCP's confidence in the validity of the examination and consequently his/her decisions in the rehabilitation process.
- The lack of the positive effect on the patient due to presence of a human medical expert during the examination. This may partially affect the trust, knowledge, regard, and loyalty that characterizes the patient-HCP relationship (48).
- Intrinsic factors of the patient: computer skills, biases against technology, cognitive decline, and comorbidities.
- HCPs, as all clinicians, are very cautious and demand high quality and evidence before accepting a new examination paradigm as part of their everyday clinical practice (13). It is reasonable to assume that HCPs' skepticism will be pronounced in the case of a user-operated examination.
- HCPs work as a community, at least to a degree (13). Experiences shared among colleagues concerning clinical practices and interventions affect how new practices are accepted (13), and this may play a role in how user-operated audiometry will be received.
- The potential disruption to standard operating procedures, e.g., changing appointment times, effects on allocation of human and physical resources.

TOWARD THE USER-OPERATED AUDIOMETRY IMPLEMENTATION

The success of the UAud project relies on the extent to which the user-operated audiometry system meets the aforementioned requirements and is accepted by both patients and HCPs, enabling wide adoption in audiological practice. Here, a research plan is presented which includes (1) initial investigations, (2) a validation study where this system will be evaluated in a randomized clinical trial, and (3) implementation in a semi-large scale within real-life clinical conditions.

Initial Investigations

The initial investigations consist of (a) prototyping and user-interaction design of the user-operated solutions, (b) evaluation of the equipment and testing environment toward a usable system, (c) implementation and validation of a user-operated ACT™ test, and (d) identifying barriers on implementation. At this early stage, a detailed record of patients' and HCPs' concerns that will inform the whole project's research plan is needed. This will include semi-structured interviews (49) and questionnaires delivered to patients and HCPs. Further, specific research questions will be addressed by running "proof of concept" studies that will help preparations of the validation and implementation and address concerns raised by patients and HCPs. Some research questions will be part of these studies:

- How can the design of the user interface maximize the internal/external motivation and minimize distractions?
- How do the patients interact with the system when doing a new audiometric test?
- Which is the preferred paradigm for assessing the ACT™ in a user-operated test?
- What is the effect of the user (instead of expert) placing of headphones on audiometry and ACT™? How to ensure its quality?
- What is the most effective design for the set of instructions required for handling both software and hardware? How can the system be designed in terms of both content and mode (e.g., text, audio, video, images, pictograms) to guarantee that a majority of patients can successfully complete the test?

Validation Study

Hearing rehabilitation entails different types of interventions. Part of these interventions include the hearing-aid selection, provision, and fitting, which make use of the individual audiometric thresholds. The purpose of the validation study is to demonstrate that the quality of the intervention is not influenced by the modality of the hearing test (i.e., user-operated vs. manual audiometry). A randomized double blind clinical trial will be conducted on two groups of adult patients with a treatable hearing loss (125 in each group). HCPs will provide the intervention based on audiometric thresholds without knowledge of the modality used to obtain them. Two months after the hearing aid fitting, patients will be examined with a test battery which will include self-report outcome measures and an examination of speech intelligibility.

Implementation

The UAud system will be implemented in a real-life situation, involving the allocation of resources for a new workflow in the selected clinics. The implementation of user-operated audiometry will be carried out at Odense University Hospital in Odense and Svendborg, as well as at two private ENT clinics. It will include a total of at least 250 patients. This is probably the most challenging part of the project, as there will be a need for clinicians adopt the user-operated audiometry as part of their daily work. Therefore, an effective change in the hearing healthcare service should involve different phases (50). First, the hearing care professionals have to be aware of the innovation. Second, there should be an understanding that the user-operated audiometry is indeed an opportunity to improve the person-centered service. Third, the barriers explored in the initial investigation have to be addressed so the HCPs can accept the change with a positive attitude. And forth, the actual implementation of the user-operated audiometry in the clinical practice, the confirmation of its benefits and the integration of the service in the clinical setup. An implementation manual will be published describing the user-operated audiometry as well as the limitations and criteria for its correct and efficient use.

AUTHOR CONTRIBUTIONS

All authors have made contributions to the manuscript, participated in the conceptualization, reviewed, and agreed with the final version of the manuscript.

FUNDING

This work was supported by Innovation Fund Denmark Grand Solutions 9090-00089B and the William Demant Foundation.

ACKNOWLEDGMENTS

We thank all partners in the UAud project: University of Southern Denmark, Odense University Hospital, Interacoustics and Oticon for their participation in the conceptualization of the project and this article. A special thanks goes to Søren Laugesen, James Michael Harte, and Louise Thygesen Smidt for their contribution to the article preparation. We also thank the reviewers for their valuable comments.

REFERENCES

1. GBD, 2019 HLC. Hearing loss prevalence and years lived with disability, 1990-2019: findings from the Global Burden of Disease Study 2019. *Lancet*. (2021) 397:996–1009. doi: 10.1016/S0140-6736(21)00516-X
2. World Health Organization. *World Report on Hearing*. Geneva: World Health Organization (2021).
3. Amieva H, Ouvrard C, Giulioli C, Meillon C, Rullier L, Dartigues JF. Self-reported hearing loss, hearing aids, and cognitive decline in elderly adults: a 25-year study. *J Am Geriatr Soc.* (2015) 63:2099–104. doi: 10.1111/jgs.13649
4. Hardy CJD, Marshall CR, Golden HL, Clark CN, Mummery CJ, Griffiths TD, et al. Hearing and dementia. *J Neurol*. (2016) 263:2339–54. doi: 10.1007/s00415-016-8208-y
5. Livingston G, Sommerlad A, Orgeta V, Costafreda SG, Huntley J, Ames D, et al. (2017). Dementia prevention, intervention, and care. *Lancet.* (2017) 390:2673–734. doi: 10.1016/S0140-6736(17)31363-6
6. Griffiths TD, Lad M, Kumar S, Holmes E, McMurray B, Maguire EA, et al. How can hearing loss cause dementia? *Neuron*. (2020) 108:401–12. doi: 10.1016/j.neuron.2020.08.003
7. International Organization for Standardization. *Acoustics. Audiometric Test Methods – Part 1: Pure-Tone Air and Bone Conduction Audiometry*. Geneva: International Organization for Standardization (2010).
8. British Society of Audiology (2017). *Pure Tone Air and Bone Conduction Threshold Audiometry With and Without* Masking. Bathgate, UK. Available online at: https://www.thebsa.org.uk/wp-content/uploads/2018/11/OD104-32-Recommended-Procedure-Pure-Tone-Audiometry-August-2018-FINAL.pdf
9. The Danish Health and Medicines Authority (2021). Available online at: https://www.esundhed.dk/Emner/Patienter-og-sygehuse/MitSygehusvalgProfessionel?treatment_ids=336&organizationGroupId=undefined&regionId=1,2,3,4,5&postalFrom$=$undefined (accessed April 2021).
10. Muñoz K, Nagaraj NK, Nichols N. Applied tele-audiology research in clinical practice during the past decade: a scoping review. *Int J Audiol.* (2020) 60:S4–12. doi: 10.1080/14992027.2020.1817994
11. Margolis RH, Morgan DE. Automated pure-tone audiometry: An analysis of capacity, need, and benefit. *Am J Audiol.* (2008) 17:109–13. doi: 10.1044/1059-0889(2008/07-0047)
12. Békésy GV. A new audiometer. *Acta Oto-Laryngol*. (1947) 35:3411–28. doi: 10.3109/00016484709123756
13. Margolis RH. Automated audiometry then and now. *Audiol Pract.* (2012) 4:9–15. Available online at: https://web.archive.org/web/20210920143938/https://www.audiologypractices.org/automated-audiometry-then-and-now (accessed September 20, 2021).
14. Swanepoel DW, Myburgh HC, Howe DM, Mahomed F, Eikelboom RH. Smartphone hearing screening with integrated quality control and data management. *Int J Audiol.* (2014) 53:841–9. doi: 10.3109/14992027.2014.920965
15. Dillon H, Mee C, Moreno JC, Seymour J. Hearing tests are just child's play: The sound scouts game for children entering school. *Int J Audiol.* (2018) 57:529–37. doi: 10.1080/14992027.2018.1463464
16. Schmidt JH, Brandt C, Pedersen ER, Christensen-Dalsgaard J, Andersen T, Poulsen T, et al. A user-operated audiometry method based on the maximum likelihood principle and the two-alternative forced-choice paradigm. *Int J Audiol.* (2014) 53:383–91. doi: 10.3109/14992027.2013.879339
17. Schlittenlacher J, Turner RE, Moore BCJ. Audiogram estimation using Bayesian active learning. *J Acoust Soc Am.* (2018) 144:421–30. doi: 10.1121/1.5047436
18. Barbour DL, Howard RT, Song XD, Metzger N, Sukesan KA, DiLorenzo JC, et al. Online machine learning audiometry. *Ear Hear.* (2019) 40:918–26. doi: 10.1097/AUD.0000000000000669
19. Brennan-Jones CG, Eikelboom RH, Swanepoel DW. Diagnosis of hearing loss using automated audiometry in an asynchronous telehealth model: a pilot accuracy study. *J Telemed Telecare.* (2017) 23:256–62. doi: 10.1177/1357633X16641552
20. Sandström J, Swanepoel D, Laurent C, Umefjord G, Lundberg T. Accuracy and reliability of smartphone self-test audiometry in community clinics in low income settings: a comparative study. *Ann Otol Rhinol Laryngol.* (2020) 129:578–84. doi: 10.1177/0003489420902162
21. Mahomed F, Swanepoel DW, Eikelboom RH, Soer M. Validity of automated threshold audiometry: a systematic review and meta-analysis. *Ear Hear.* (2013) 34:745–52. doi: 10.1097/01.aud.0000436255.53747.a4
22. Shojaeemend H, Ayatollahi H. Automated audiometry: a review of the implementation and evaluation methods. *Healthc Inform Res.* (2018) 24:263–75. doi: 10.4258/hir.2018.24.4.263
23. Margolis RH, Saly GL, Le C, Laurence J. Qualind^{TM}: a method for assessing the accuracy of automated tests. *J Am Acad Audiol.* (2007) 18:78–89. doi: 10.3766/jaaa.18.1.7
24. Margolis RH, Glasberg BR, Creeke S, Moore BCJ. AMTAS$^{®}$: automated method for testing auditory sensitivity: validation studies. *Int J Audiol.* (2010) 49:185–94. doi: 10.3109/14992020903092608
25. Margolis RH, Frisina R, Walton JP. AMTAS$^{®}$: Automated method for testing auditory sensitivity: II. air conduction audiograms in children and adults. *Int J Audiol.* (2011) 50:434–9. doi: 10.3109/14992027.2011.553206
26. Eikelboom RH, Swanepoel DW, Motakef S, Upson GS. Clinical validation of the AMTAS automated audiometer. *Int J Audiol.* (2013) 52:342–9. doi: 10.3109/14992027.2013.769065
27. Margolis RH, Killion MC, Bratt GW, Saly GL. Validation of the home hearing testa^{TM}. *J Am Acad Audiol.* (2016) 27:416–20. doi: 10.3766/jaaa.15102
28. Margolis RH, Moore BCJ. AMTAS$^{®}$: Automated method for testing auditory sensitivity: III. sensorineural hearing loss and air-bone gaps. *Int J Audiol.* (2011) 50:440–7. doi: 10.3109/14992027.2011.575085
29. Paglialonga A, Polo EM, Zanet M, Rocco G, van Waterschoot T, Barbieri R. An automated speech-in-noise test for remote testing: development and preliminary evaluation. *Am J Audiol.* (2020) 29:564–76. doi: 10.1044/2020_AJA-19-00071
30. Ooster J, Krueger M, Bach J-H, Wagener KC, Kollmeier B, Meyer BT. Speech audiometry at home: automated listening tests via smart speakers with normal-hearing and hearing-impaired listeners. *Trends Hearing.* (2020) 24:2331216520970011. doi: 10.1177/2331216520970011
31. Zokoll MA, Hochmuth S, Warzybok A, Wagener KC, Buschermöhle M, Kollmeier B. Speech-in-noise tests for multilingual hearing screening and diagnostics. *Am J Audiol.* (2013) 22:175–8. doi: 10.1044/1059-0889(2013/12-0061)
32. Akeroyd M, Arlinger S, Bentler R, Boothroyd A, Dillier N, Dreschler W, et al. International Collegium of Rehabilitative Audiology (ICRA)

recommendations for the construction of multilingual speech testsICRA Working Group on Multilingual Speech Tests. *Int J Audiol*. (2015) 54:1–6. doi: 10.3109/14992027.2015.1030513
33. Hagerman B. Sentences for testing speech intelligibility in noise. *Scand Audiol*. (1982) 11:79–87. doi: 10.3109/01050398209076203
34. Smits C, Kapteyn TS, Houtgast T. Development and validation of an automatic speech-in-noise screening test by telephone. *Int J Audiol*. (2004) 43:15–28. doi: 10.1080/14992020400050004
35. Chi T, Gao Y, Guyton MC, Ru P, Shamma S. Spectro-temporal modulation transfer functions and speech intelligibility. *J Acoust Soc Am*. (1999) 106:2719–32. doi: 10.1121/1.428100
36. Bernstein JGW, Mehraei G, Shamma S, Gallun FJ, Theodoroff SM, Leek MR. Spectrotemporal modulation sensitivity as a predictor of speech intelligibility for hearing-impaired listeners. *J Am Acad Audiol*. (2013) 24:293–306. doi: 10.3766/jaaa.24.4.5
37. Mehraei G, Gallun FJ, Leek MR, Bernstein JG. Spectrotemporal modulation sensitivity for hearing-impaired listeners: dependence on carrier center frequency and the relationship to speech intelligibility. *J Acoust Soc Am*. (2014) 136:301–16. doi: 10.1121/1.4881918
38. Bernstein JGW, Danielsson H, Hällgren M, Stenfelt S, Rönnberg J, Lunner T. Spectrotemporal modulation sensitivity as a predictor of speech-reception performance in noise with hearing aids. *Trends Hearing*. (2016) 20:1–17. doi: 10.1177/2331216516670387
39. Zaar J, Simonsen LB, Behrens T, Laugesen S. Towards a clinically viable spectro-temporal modulation test. In: *Poster Session Presented at International Hearing Aid Conference 2018*. Tahoe, CA. (2018).
40. Zaar J, Simonsen L, Behrens T, Dau T, Laugesen S. Investigating the relationship between spectro-temporal modulation detection, aided speech perception, and directional noise reduction preference in hearing-impaired listeners. In: *Proceedings of the International Symposium on Auditory and Audiological Research*, Vol. 7. Nyborg (2020). p. 181–8.
41. Daw ND, O'Doherty JP. Multiple systems for value learning. In: *Neuroeconomics: Decision Making and the Brain: Second Edition*. London, UK:Academic Press (Elsevier) (2014). p. 393–410. doi: 10.1016/B978-0-12-416008-8.00021-8
42. Johnson A, Proctor RW. *Attention: Theory and Practice*. Thousand oaks, CA: Sage Publications, Inc. doi: 10.4135/9781483328768
43. Diamond A. Executive functions. *Annu Rev Psychol*. (2013) 64:135–68. doi: 10.1146/annurev-psych-113011-143750
44. Blanchette I, Richards A. The influence of affect on higher level cognition: a review of research on interpretation, judgement, decision making and reasoning. *Cogn Emot*. (2010) 24:561–95. doi: 10.1080/02699930903132496
45. Contreras-Somoza LM, Irazoki E, Toribio-Guzmán JM, de la Torre-Díez I, Diaz-Baquero AA, Parra-Vidales E, et al. Usability and user experience of cognitive intervention technologies for elderly people with MCI or dementia: a systematic review. *Front Psychol*. (2021) 12:636116. doi: 10.3389/fpsyg.2021.636116
46. LaMonica HM, Roberts AE, Davenport TA, Hickie IB. Evaluation of the usability and acceptability of the InnoWell platform as rated by older adults: survey study. *JMIR Aging*. (2021) 4:e25928. doi: 10.2196/2592
47. Adamou M, Jones SL, Fullen T, Galab N, Abbott K, Yasmeen S. Remote assessment in adults with autism or ADHD: a service user satisfaction survey. *PLoS ONE*. (2021) 16:e0249237. doi: 10.1371/journal.pone.0249237
48. Chipidza FE, Wallwork RS, Stern TA. Impact of the doctor-patient relationship. *Prim Care Companion J Clin Psychiatry*. (2015) 17:360. doi: 10.4088/PCC.15f01840
49. Tanggaard L, Brinkmann S, editors. Interviewet: Samtalen som forskningsmetode. In: *Kvalitative metoder: En grundbog*. 2nd ed. Copenhagen: Hans Reitzels Forlag (2015). p. 29–53.
50. Grol R, Wensing M. Effective implementation of change in healthcare. In: Wensing M, Grol R, Grimshaw J, editors. *Improving Patient Care*. Hoboken, NJ: John Wiley & Sons (2020) 45–71. doi: 10.1002/9781119488620.ch3

Conflict of Interest: RS-L is employed by Interacoustics 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.

frontiers in Public Health

ORIGINAL RESEARCH
published: 14 October 2021
doi: 10.3389/fpubh.2021.725080

Check for updates

French Version of the Antiphasic Digits-in-Noise Test for Smartphone Hearing Screening

Jean-Charles Ceccato[1,2]*, *Marie-Josée Duran*[3], *De Wet Swanepoel*[4], *Cas Smits*[5], *Karina C. De Sousa*[4], *Lewis Gledhill*[1], *Frédéric Venail*[1,6] *and Jean-Luc Puel*[1,2]

[1] INM, Univ Montpellier, Inserm, Montpellier, France, [2] Audiocampus, UFR Pharmacie, Univ Montpellier, Montpellier, France, [3] Fondation Pour l'Audition, Paris, France, [4] Department of Speech-Language Pathology and Audiology, Univ Pretoria, Pretoria, South Africa, [5] Department of Otolaryngology-Head and Neck Surgery, Amsterdam UMC, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam Ear & Hearing, Amsterdam, Netherlands, [6] Otology and Neurotology Unit, Univ Montpellier, CHU Montpellier, Montpellier, France

In France 58% of persons with hearing loss still do not wear hearing aids. Pure-tone audiometry is the traditional gold standard in assessment and screening of hearing impairment, but it requires the use of calibrated devices and soundproof booth. The antiphasic digits-in-noise (DIN) test does not require calibrated material and can run on a standard headset or earbuds connected to a smartphone or a computer. The DIN test is highly correlated with pure tone audiometry and has already shown to be effective in hearing loss screening in its English version promoted by the WHO. The aim of the present study was to develop and validate a French version of the antiphasic DIN test for implementation on a national screening test offered as a smartphone app. The audio files recorded from a French native female speaker were selected and normalized in intensity according to their recognition probability. The French DIN test application was then tested on normal hearing- and hearing-impaired subjects. Based on the strong correlation between pure tone audiometry (PTA) and DIN SRT, we calculated ROC curves and Z-score. For PTA > 20 dB HL, a SNR cutoff of 12.9 dB corresponds to a sensitivity and specificity of 0.96 and 0.93, respectively. To detect moderate and more severe hearing loss (PTA > 40 dB HL), the SNR cutoff was −10.9 dB, corresponding to a sensitivity and specificity of 0.99 and 0.83, respectively. The Z-score was calculated to define statistical criteria of normality for speech-in-noise evaluation. While a score of 0 roughly corresponds to the normality (DIN SRT = −15.4 dB SNR), a subject with DIN SRT > −12.2 (Z-score > 2) is ranked in the hearing loss population. Next, the French antiphasic DIN test was implemented in the Höra iOS and Android apps. In total, 19,545 Höra tests were completed and analyzed. Three quarters of them were classified as normal (74 %) and one quarter presented mild (9%) or more severe loss (17%). Together, results argue for the use of the French version of antiphasic DIN test in the general population to improve the screening of hearing-impaired individuals.

Keywords: hearing loss, French antiphasic DIN test, smartphone, Höra app, speech-in-noise (SIN)

OPEN ACCESS

Edited by:
Changxin Zhang,
East China Normal University, China

Reviewed by:
Jan de Laat,
Leiden University Medical Center, Netherlands
Zilong Xie,
University of Kansas Medical Center, United States

****Correspondence:***
Jean-Charles Ceccato
jean-charles.ceccato@umontpellier.fr

Specialty section:
This article was submitted to Family Medicine and Primary Care, a section of the journal Frontiers in Public Health

Received: *14 June 2021*
Accepted: *13 September 2021*
Published: *14 October 2021*

Citation:
Ceccato JC, Duran MJ, Swanepoel DW, Smits C, De Sousa KC, Gledhill L, Venail F and Puel JL (2021) French Version of the Antiphasic Digits-in-Noise Test for Smartphone Hearing Screening. Front. Public Health 9:725080. doi: 10.3389/fpubh.2021.725080

INTRODUCTION

Hearing loss burden on health and quality of life is often underestimated by public authorities, health care professionals and the public. In the world, hearing loss affects around 20% (1.5 billion people) of the population with an estimated 5.5% (430 million) experiencing significant hearing loss. In France as in other high-income countries, the prevalence is expected to increase rapidly over the next three decades due to an aging population and noise exposure of younger people (1).

The impact of untreated hearing loss is far reaching with economic cost estimates of ∼22.5 billion Euros in France, and 225 billion in Europe (1). On an individual level, untreated hearing loss is associated with social isolation, increased risk of depression, cognitive decline, dementia, and hospitalization (2). On the other hand, hearing aid use is associated with significant improvement in social, psychologic, emotional, and physical aspects of the lives of persons with hearing loss with all degrees of hearing loss (3–5). Hearing loss has recently been identified as the most significant modifiable risk factor in mid-life for dementia (6), which emphasizes the importance of early detection and timely intervention.

In France 58% of persons with hearing loss still do not wear hearing aids (7). For one third of them, the cause relies on the lack of screening since most of cases are referred after self-declaration to general practitioners or targeted hearing loss screening campaigns. For the remainder, either no medical recommendation has been made or the cost of hearing aids has been a barrier (3, 7). Although French Health Insurance recently adopted regulation for full reimbursement of hearing aids (8), the absence of a national screening strategy for hearing loss limits widespread uptake.

Pure-tone audiometry is the traditional gold standard in assessment and screening of hearing. However, it includes inherent limitations that make large-scale population-based screening programs difficult. Firstly, pure tone audiometry requires the use of calibrated devices that are typically set in a sound treated environments and require trained professionals to operate. Secondly, it is insensitive to the early and specific patient complaints such as difficulty with speech understanding in noise (9). Although some efforts in the development of self-testing applications for pure tone hearing thresholds on laptops or mobile devices have been attempted, accurate testing is difficult due to the challenge of uncalibrated testing across various digital devices and headphones (10–17). An alternative to pure tone audiometry is the use of speech-in-noise testing as a screening tool. The suprathreshold and relative measure of a signal-to-noise ratio (SNR) does not require absolute calibrated levels of noise or speech. As a result hearing screening is made possible using non-calibrated devices such as smartphones coupled with generic headphones (18).

Pioneering experiments from Wilson and Smits teams (19–23) investigated the possibility to use digit pairs or digit triplets for speech-in-noise testing. First developed to screen the Dutch population using landline phone, digits and noise were presented in one ear (monaurally) (22, 23). The digits-in-noise test (DIN test, or "digit-triplet test") was then adapted in different languages [e.g., Dutch, German, British, Australian, Polish, Swiss, and French, see Van de Borre et al. for review (24)]. Different platforms have been used for hearing screening (telephone, internet, tablets, smartphones) and different populations have been targeted, mainly adults, but also young school children (25–28). The presentation method of the stimuli differs as well. In most of the studies, both ears are tested separately. In others, stimuli are presented binaurally, for kids for example (18, 28–32). Presenting speech materials binaurally involves more central auditory processes than monaural presentation (33) and halves test duration which reduces task dropout compared to sequentially testing each ear. Since the diotic test strongly relies on the better ear, asymmetric, or unilateral hearing losses are easily missed (31, 34). Moreover, both monaural and binaural/diotic speech-in-noise tests are insensitive to conductive hearing loss (CHL) (31). Recently, the use of antiphasic stimuli has been explored and standardized (31). Presenting stimuli in opposite phase between ears with masking noise interaurally in-phase enable binaural masking release that improve stimuli perception (35, 36). This phenomenon, called binaural masking level difference is frequency dependant and relies on well-preserved and symmetric hearing to be effective. Compared to binaural presentation, speech reception thresholds (SRT) of the antiphasic DIN test are more strongly correlated with the worse ear pure tone average (PTA) across 0.5, 1, 2, and kHz than with the better ear PTA. The SRT distribution range is also wider for the antiphasic DIN SRTs than for diotic DIN SRTs, and differences in SRT are larger between normal hearing and hearing loss persons (31). As a result, antiphasic digits presentation markedly improved the specificity (0.8 vs. 0.71) and the sensitivity (0.9 vs. 0.75) to symmetric and asymmetric sensorineural hearing loss as well as conductive hearing loss compared to diotic presentation (31, 36, 37).

Combined with smartphone technology the antiphasic DIN screening is accessible to a large global audience. For example, the HearZA app was used successfully in a national hearing screening campaign in South Africa, with a binaural test. The World Health Organization (WHO) has adopted this antiphasic DIN screening approach in the use of their hearWHO smartphone application available in English, Spanish and Mandarin (38). To date no antiphasic DIN test has yet been developed in French for a digital platform like smartphone. The purpose of the present study was to develop and to validate a French version of the antiphasic DIN for a smartphone app.

MATERIALS AND METHODS

According to Jardé's law regulating biomedical research in France, this type of evaluation is considered as non-interventional research and did not require institutional review board approval. All eligible participants were informed of the study aims and procedures and provided consent before participation.

Digits Recording and Level Normalization

Recording and Processing the Speech Material

French mono- and bi-syllabic digits (0–9) were selected as speech material. Single-digit recordings were made from a native French

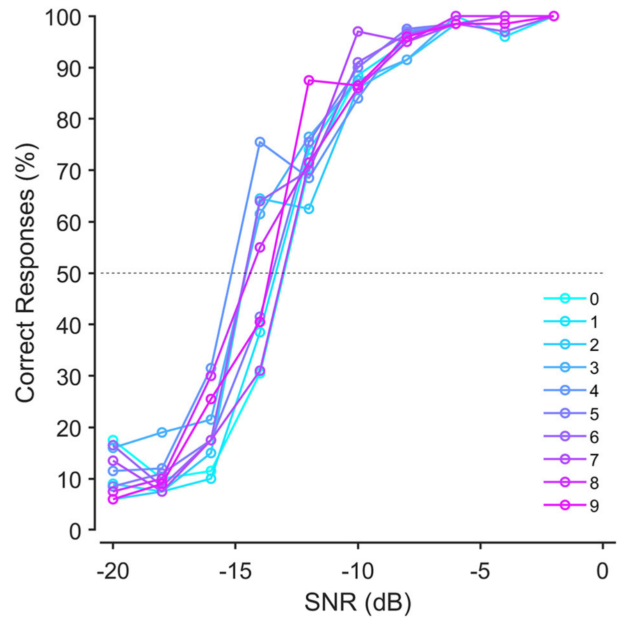


FIGURE 1 | Average speech recognition probabilities for single digits-in-noise before equalization.

female speaker in a sound-proof booth. A carrier sentence "le chiffre" was uttered before pronouncing each digit to allow natural intonation. A microphone (Blue Yeti Microphone) was held ∼5 cm from the speakers' mouth during recordings. The speaker was asked to read four lists of randomly ordered digits. Thus, each digit appeared four times. The recordings were digitalized at 44,1 kHz with floating 32 bits resolution using Audacity software (Audacity). Each digit was then cropped manually from the list and stored separately in WAV files using Audacity software. One speech-language therapist and one phonetic academic teacher rated the four recordings of each digit, according to the naturalness, articulation, voice quality, intonation, and speed of production. The final list of digits was compiled using the best rated digits for digits 0–9.The masking noise was generated by shaping a white noise (using FIR filter in Matlab) with the long-term average spectrum of the 10 selected digits (**Figure 1**). The recording level (RMS in Volt) of the masking noise was set to the average recording level of the 10 digits without any silences (18, 39).

Equalization

Digits were equalized according to their recognition probability. Equalizing digits by applying level corrections to the digits ensured that each digit had a 50% chance of being recognized correctly at the same SNR. To do so, we recruited 20 normal-hearing (both ears) participants aged 20–33 years, with pure-tone thresholds <20 dB HL from 250 to 8,000 Hz. A custom Matlab script was used to generate the sequences of digits superimposed with noise on a laptop (MacBook Pro) that were presented monaurally through circumaural headphones (Sony WH1000XM3). Four lists of 10 digits were successively presented at 10 different SNRs decreasing from −2 to −20 dB in 2 dB steps. For each SNR level the 10 digits were presented randomly. The noise started 500 ms before and finished 500 ms after each digit. The participant was forced to choose a digit, even if it was not recognized (forced-choice procedure). The psychometric curves of recognition for each digit were fitted with a logistic function to determine the speech reception threshold (SRT, i.e., the SNR corresponding to a 50% recognition probability (**Figure 1**). Each digit's recording level was then adjusted using the difference between the SRT of each digit and the average SRT of all the digits (±0.4 dB maximum).

Validation of the French Antiphasic DIN Test

Before administering the DIN test, pure-tone audiometry was completed for all the participants. The two ears were evaluated with air and bone conduction audiometry across the frequencies of 500, 1,000, 2,000, and 4,000 Hz.

Study Design

Pure tone audiometry was performed in a sound booth with a digital audiometer (Audyx) equipped with a supra-aural TDH39 headphone for air conduction and a bone vibrator B71 for bone conduction. The hearing status was determined according to the pure tone threshold average at 500, 1,000, 2,000, and 4,000 Hz. The normal-hearing participants with PTA ≤ 20 dBHL in both ears were students at the University of Montpellier and of the Institute for Neuroscience of Montpellier, relatives of the authors or accompanying hearing loss people. The participants with hearing loss came from the University Hospital of Montpellier or private clinics.

Population of Reference

The beta version of the French antiphasic DIN test was evaluated in a normative reference population, including normal and hearing loss subjects of different ages ($n = 167$). The participants (77 women and 90 men) were aged from 19 to 90 years (mean age 56 years of age ±22). The hearing status of the subject was classified according to the recommendation of the International Bureau for Audiophonology (BIAP) (40). Among 167 subjects tested, 66 had normal-hearing (PTA ≤ 20 dB HL, 32.5 years of age ±11.5), 75 symmetric sensorineural hearing loss (PTA > 20 dBHL, 71.28 years of age ±10.6), 19 unilateral or asymmetric hearing loss (PTA difference between both ears >10 dB, 72.7 years of age ±10). In addition, 7 mixed hearing loss based on air bone gap criteria (PTA difference between bone and air conduction > 20 dB and bone conduction PTA > 20 dB, 74.8 years of age ±8.5) and other test results such as tympanometry and otoscopy.

DIN Test Procedure and Equipment

DIN test was developed in a webapp working on Google Chrome running on a laptop connected to a commercial headphone (Sony WH 1000 XM3). The selected digits were organized in 120 triplets stored in stereo files. The participants were informed on how to enter the digit responses on the computer keyboard. Before starting the test, they were asked to adjust the loudness of the digit-triplets to a comfortable listening level. The noise was presented in-phase (diotic) in both ears, and the digits were antiphasic between the two ears. The noise was present during the entire digit triplet sequence and started 500 ms before the first digit. Successive digits within a triplet were separated by 200 ms intervals. The initial SNR was fixed at 0 dB and the SNR change was obtained by varying the noise level when the SNR is positive, or by varying the digits level when the SNR is negative (31). To prevent possible learning of the masking noise (41), noise refreshment was ensured for each trial by creating a long noise file where different fragments were randomly selected. For each presentation, the application randomly selected 3 different digits to produce the presented triplet superimposed with noise. The subject was required to enter the digits they recognized (or guessed) directly on the laptop as they would perform on the smartphone application. Depending on the answers, the signal-to-noise ratio was adjusted following a 1-up, 1-down adaptive procedure using step size of 4 dB SNR for the first 3 steps, thereafter continuing in 2 dB steps. After 23 triplets presentations, the test stops and the DIN SRT was calculated as the average of the last 19 SNRs (18, 22). Subjects were tested only once without training to obtain scores that reflects more closely the results that could be obtained on the smartphone application for naïve listeners. Adding a training phase in the smartphone application may be counterproductive because the increased test duration will reduce uptake and completion.

The French smartphone-based antiphasic digits-in-noise hearing application (Höra) was developed on iOS and Android and is available for free in the Apple store and Google Play Store. The mobile app can be used with standard headphones or earphones. When the application is launched, a tutorial screen appears to inform the subject how to use the application. The subject is instructed to enter a name or nickname (use is fully anonymous), gender and birthdate. Next, the subject is invited to put on the smartphone headset and adjust the intensity of the continuously presented digit-triplets to a comfortable listening level using a scroll bar. When ready, a "Start Test" button allows the subject to begin testing. A pop-up number pad appears after each presentation of the digits to allow the subject to enter his/her responses. A personalized score out of 100 was deduced from the range of SNRs of our reference population. At the end, the subject is categorized as "normal" (score between 70 and 100), "mild" (score between 50 and 70), or "moderate and worse" (score < 50). For "mild" and "moderate" expected hearing loss, the subject is advised to see a doctor.

Data and Statistical Analysis

Matlab was used for data processing, statistical analyses and for creating figures. Audiometric data and DIN test data, loaded from the text files generated by the web app, were stored in Microsoft Excel. For each subject, the status of hearing (i.e., "normal hearing" or "hearing loss") was determined from the poorer ear PTA of each subject. Pearson correlation coefficient and linear or multilinear regressions were used to assess coevolution of different parameters like age, hearing loss and SNR score of DIN (Matlab functions: corrcoef, fit, confint, predint). Receiver operating characteristics (ROC) curves were calculated (Matlab functions: fitgml, perfcurve) to determine the sensitivity and specificity of the DIN tests for different cutoff values, to detect mild hearing loss and worse (PTA > 20 dB HL or PTA > 25 dB HL depending on the French or international standard, respectively) and moderate hearing loss and worse (PTA > 40 dB HL). Distribution of results were calculated and fitted with unimodal (one) or bimodal (two) Gaussian models (Matlab functions: hist, fitgmdist). Gaussian distribution function is $f(x) = 1/(\sigma\sqrt{2\pi})\ e\hat{}(-[(x-\mu)]\hat{}2/(2\sigma\hat{}2))$, with μ the mean and σ the standard deviation).

RESULTS

Pure tone average (PTA) of the poorer ear was determined by averaging the thresholds at 0.5, 1, 2, 4 kHz to obtain a single value for each ear and each participant (**Figure 2A**). A broken-stick model [$y = max\ (a,\ cAge + a - bc)$] was used to characterize the time course of PTA or SRT, in which "a" represents the normal hearing, "b" the cut-off age and "c" the slope. All parameters are given with 95% confidence interval. In our reference normative population ($n = 167$), the mean PTA was 11.6 dB HL (±3.7 dB) until a cut-off age of 40-years of age (±6 years) where the mean thresholds increase at a pace of 1.18 dB/year (±0.18 dB/year). The time course of DIN SRT was similar to PTA (**Figure 2B**) with a mean score of −15.4 dB SNR (±1.3 dB) up to a cut-off age of 44-years of age (±6 years) where the mean DIN SRT increase at a pace of 0.43 dB SNR/year (±0.07 dB/year). Cut-off ages of PTA and DIN SRT did not differ statistically. Despite speech perception in noise requiring more attention and auditory processing than in quiet, the similarity of PTA and DIN SRT results with ages suggest that both tests are related to peripheral hearing loss.

Sensitivity and Specificity of the DIN Test

Poorer ear PTA was significantly correlated with DIN SRT ($r = 0.82$, $p < 0.001$, **Figure 3A**) across all types and the degrees of hearing loss. ROC curves were calculated to determine the sensitivity (true-positive rate) and specificity (false-positive rate) of the DIN tests for different cut-off values. The optimal SNR cut-off values were chosen using a cost function that optimized the Youden index (lowest misclassification) favoring sensitivity over specificity (**Figure 3B**). According to French regulation, hearing is considered normal when PTA is below 20 dB (BIAP Recommendation 02.1) (40). To detect all hearing losses (PTA > 20 dB HL), a SRT cutoff value of −12.9 dB SNR corresponded to a sensitivity and a specificity of 0.92 [0.84; 0.95] and 0.86 [0.77; 0.93], respectively. For PTA > 25 dB HL, which is the international standard, a SNR cutoff of −11.7 dB corresponded to sensitivity and a specificity of 0.96 [0.9; 0.98] and 0.93 [0.84; 0.97], respectively. To detect moderate hearing loss and worse (PTA > 40 dB HL), the SNR cutoff was −10.9 dB, corresponding to a sensitivity and the specificity of 0.99 [0.92; 1] and 0.83 [0.75; 0.9], respectively.

DIN SRT Distribution

To establish a reference range of DIN SRT, we selected a population composed of people 25-years of age and normal

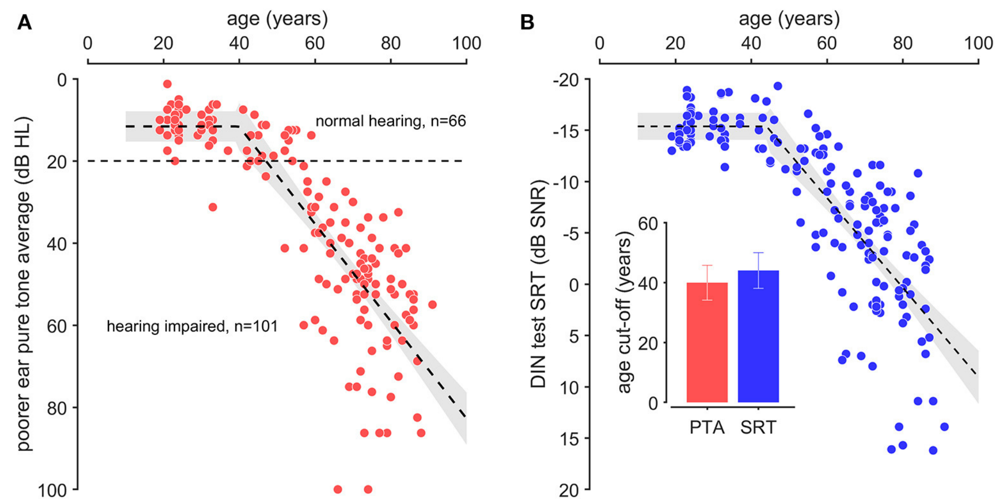


FIGURE 2 | Pure tone average thresholds and digit in noise speech reception threshold as a function of age. **(A)** Pure tone average (PTA) was determined by averaging pure tone thresholds at 0.5, 1, 2, and 4 kHz from the poorer ear of each subject (red dots). **(B)** Digits-in-noise speech reception threshold (DIN SRT, blue dot). The broken-stick regression (dotted line) is presented with confidence interval (gray area). The inserted graph presents the cut-off values of broken-stick regression for PTA (red bar) and DIN SRT (blue bar). Note that the cut-off age for PTA (40 years-old) and DIN SRT (44 years-old) are not statistically different.

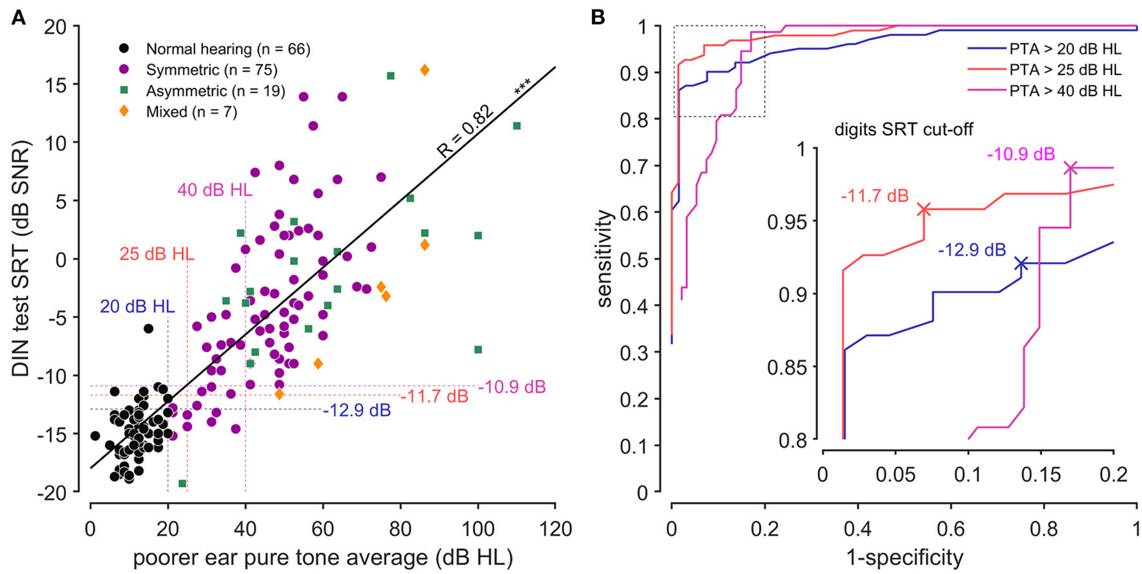


FIGURE 3 | Selectivity and specificity of the DIN test. **(A)** Correlation between antiphasic digit speech reception threshold (DIN SRT) and pure tone average (PTA) from the poorer ear of each subject ($R = 0.82$, $p < 0.001$). The different symbols represent normal hearing (black dots), symmetric sensorineural loss (purple dots), unilateral or dissymmetric loss (green square) and conductive loss (orange diamonds). Vertical dotted lines represent the French standard for mild (and worse) hearing loss (PTA > 20 dB HL, blue), the international standard for mild hearing loss (PTA > 25 dB HL, red) and the moderate (and worse) hearing loss (PTA > 40 dB HL, pink). Horizontal dotted lines represent the corresponding DIN SRT cut-offs. **(B)** Shown are the ROC curves corresponding to PTA > 20 dB HL (blue line), PTA > 25 dB HL (red line) and PTA > 40 dB (pink line). The insert graph emphasizes the DIN SRT cut-offs (sensitivity and specificity).

PTA ($n = 30$), from which we determined the mean (−15.3 dB SNR) and standard deviation (1.6 dB) of the DIN SRT. The Z-score was calculated to associate a value of normality (**Figure 4**). Because the DIN SRT follows a standard normal distribution (Kolmogorov-Smirnov test, $p < 0.001$) for this subgroup of participants, a Z-score of 1.5 (DIN SRT ≤−12.9 dB SNR) corresponds to the 95th percentile, which is verified with the direct evaluation of percentile on data of the subgroup with exactly −13 dB SNR. In other words, a subject with a DIN SRT of −12.1 dB has a Z-score of 2 (97.5% confidence interval), and a subject with a DIN SRT of −10.5 dB has a Z-score of 3 (99.5% confidence interval).

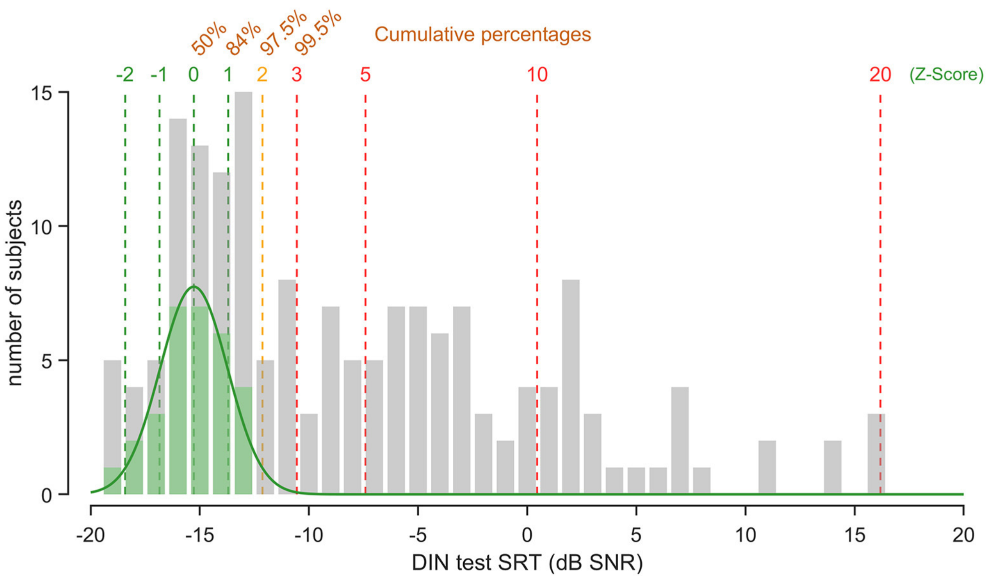


FIGURE 4 | DIN SRT distribution. Shown is the DIN SRT distribution across all the subjects participating to the experiment (gray). Subjects under 25 years of age with normal-hearing (PTA < 20 dB HL) constitutes the normative population (green). Note that the DIN SRT of the normative population follows a standard normal distribution (green line, $R^2 = 0.97$). The Z-scores are shown (dotted lines) with corresponding percentile values.

Mobile Höra Application

We subsequently integrated the French antiphasic DIN test in a mobile app (iOS and Android) called Höra. The use of the Höra smartphone app allowed us to reach a large population of subjects (n = 19,545 from March 26 to April 26, 2021). Overall, the proportion of male users (55.7 %) was more common than the females (45.3 %). Compared against our reference normative population, the smartphones users were younger (median age 55 and 33 years of age, respectively). Both populations showed a bimodal gaussian distribution across ages (**Figure 5A**). The first mode was around 26 years of age ±12 for the reference normative population versus 20 year of age ±10 for smartphones users, and the second mode was 68 year of age ±34 for the reference normative population vs. 75 year of age ±20.41 for smartphone users. For both populations, the normal SRT (−15.4 dB SNR vs. −14.9 dB SNR) and the cut-off age was similar (44 vs. 47 years). In contrast, the slope in the Höra population was less steep (0.16 dB/year ± 0.016 for smartphone users vs. 0.44 dB/year ± 0.09 for reference normative population) attesting to the proportion of hearing loss subjects being less in the Höra app users (**Figure 5B**). Accordingly, 74% of the subjects were classified as "normal" with DIN SRT ≤ 12.9 dB, corresponding to a PTA < 20 dBHL or a Z-score < 1.5 (**Figure 5C**), whereas 9% presented with a suspicion of mild hearing loss (predicted PTA between 20 and 40 dBHL or Z-score between 1.5 and 3) and 20 % had a moderate and worse hearing loss categorization (predicted PTA > 40 dBHL or Z-score > 3) (**Figure 5C**).

DISCUSSION

Our results support the use of the French version of the antiphasic DIN test to detect all hearing losses (symmetric and asymmetric sensorineural hearing loss and mixed hearing loss) with high sensitivity (0.92) and specificity (0.86). We further propose a standard score to compare speech in-noise results to a normal-hearing population.

The onset of the decline of DIN SRT and PTA with age was similar (44 ± 6 dB SNR and 40 ± 6 dB HL for DIN SRT and PTA, respectively). When DIN SRT was expressed as the function of the poorer ear PTA, the correlation was highly significant (R = 0.82, p < 0.001) and comparable with the English antiphasic DIN test (R = 0.82) (31) using the same protocol on a different population with different language. Comparing with the Jansen and colleagues study using 20 subjects (40 ears), the PTA-SRT correlation reported from our reference normative population (n = 167) is higher (0.82 vs. 0.77), probably because of the antiphasic presentation in our study. These results support the conclusion of Van den Borre et al. (24) suggesting that the mode of recording, the vocal material, the construction of the speech-weighted noise seems to have minor effects on the results of the DIN test, and only presentation monaural/binaural vs. antiphasic seems induce significant changes. To evaluate the impact of binaural presentation, we performed an additional study in 19 young adults (mean age 22.2 ± 3.4 years) with normal-hearing (unpublished). Using this binaural protocol, we found a mean SRT value of −10.7 ± 1.3 dB SNR, which closely match with the French test version reported by Jansen using a standard monaural DIN test (−10.2 ± 0.5 dB SNR) via a broadband headphone (42). Therefore, the more favorable DIN SRT measured in our reference population (n = 167, mean SNR= −15.4 ± 1.3 dB) is most likely due to the antiphasic presentation of the digits. This is consistent with the studies of Smits et al. (36) and de Sousa et al. (31, 34) in which antiphasic presentation of digits improved the average SRT with ∼5–7 dB for normal hearing listeners. This improvement in SRT shift is the result of binaural masking level difference enabled using antiphasic presentation (35).

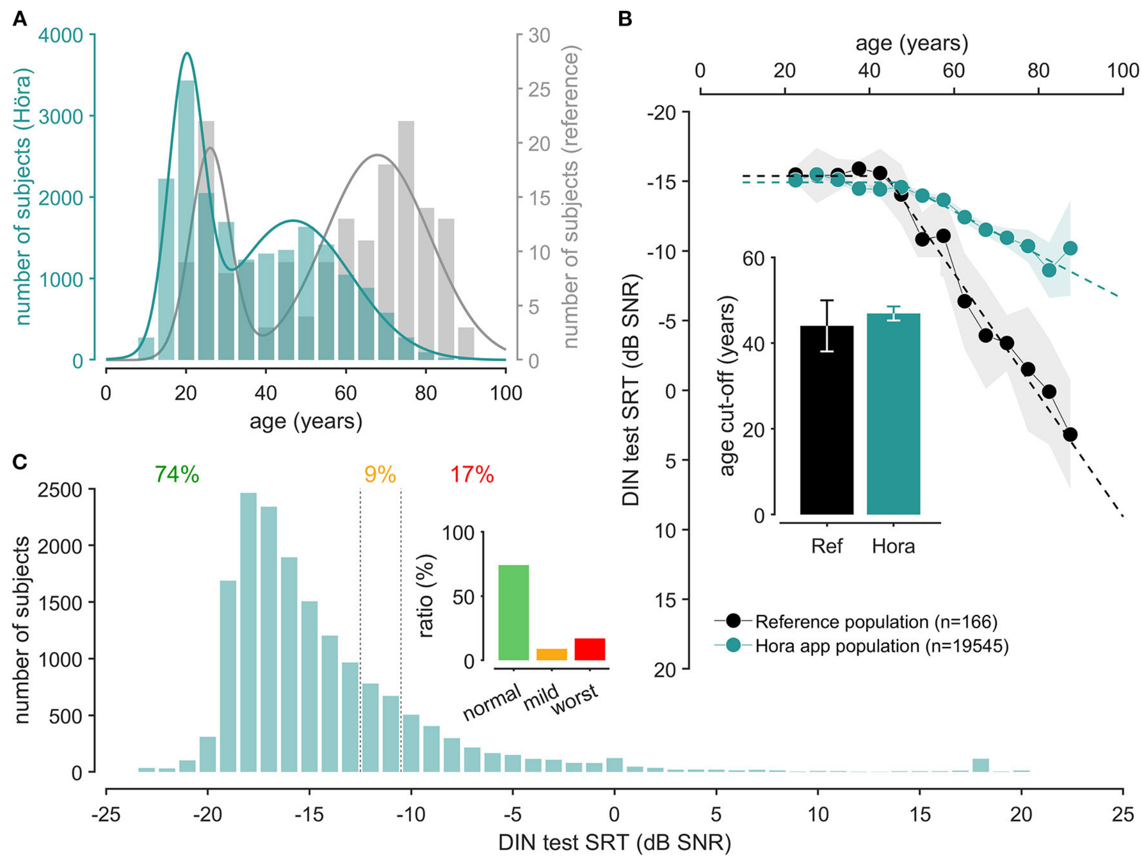


FIGURE 5 | Höra screening campaign results. **(A)** Shown is the age repartition of the Höra population (green, left axis) compared to the reference population (gray, right axis). Curves correspond to the sum of two gaussian model (reference population, gray, $f_{(x)} = 3.76 \times exp(-((x-26)/6.94)^2) + 3.66 \times exp(-((x-67.8)/18.5)^2)$, $R^2 = 0.72$, Höra population, green, $f_{(x)} = 668.46 \times exp(-((x-20)/6.24)^2) + 332 \times exp(-((x-78.6)/20.6)^2$, $R^2 = 0.95$). **(B)** Mean ± SEM of digits-in-noise speech reception thresholds for reference (gray) and Höra (green) population relative to age and the corresponding broken-stick regression. The inserted graph presents the cut-off values of broken-stick regression for reference (gray bar) and Höra (blue bar). Note that the cut-off age for PTA (44 years-old) and DIN SRT (46 years-old) are not statistically different. **(C)** Shown is the DIN SRT distribution across all the subjects participating to the Höra screening campaign (green). The inserted graph presents the distribution of Höra tests as normal hearing (74%), mild hearing loss (9%), or moderate or worst hearing loss (17%).

Based on the strong correlation between PTA and DIN SRT, we calculated ROC curves to determine the best DIN cut-offs predicting hearing impairment.The ROC curve depends on the correlation between the predictor and predicted values. The prediction for a mild hearing loss on the poorer ear according to the international standard (PTA > 25 dB HL) showed a sensitivity and a specificity of 0.96 and 0.93, respectively, compared to 0.95 and 0.73 for the English version (31). The test performs equally well for the prediction of moderate hearing loss (PTA > 40 dB HL) with sensitivity and specificity of 0.99 and 0.93, respectively. These values are better than the values reported for the English version (31) with 0.95 and 0.75 for sensitivity and specificity, respectively. This may be due to some differences in the representation of sub-populations like normal hearing older people or conductive loss. As highlighted by the study on English antiphasic digits in noise test (31), the test is very efficient to detect hearing losses of any type but does not give more information on their characteristics, like the degree of hearing difference for unilateral or asymmetrical losses. This would require to perform either two monaural tests like in older version of the test or a combination of binaural and antiphasic sequential testing that would increase test duration.

The French antiphasic DIN test then demonstrates very high quality of hearing loss prediction, but some outliers can be seen, i.e., false negatives and false positives, and may need some explanations. It is important to keep in mind that PTA and DIN test are subject to measurement errors that are potentialized when compared for screening. In our ROC analysis false negatives values correspond to subjects with hearing loss (PTA $\geq$ 20 dB HL) with a good SRT in noise (DIN SRT < -12.9 dB SNR). This phenomenon can be explained by the antiphasic stimulus presentation, known to improve speech perception in noise for people with well-preserved symmetric low frequencies hearing thresholds (35). On the contrary, false positives correspond to normal-hearing subjects with poor score in DIN test. A first explanation may be a difficulty to maintain attention through the test or a miscomprehension of the task that can occur in non-guided smartphone applications. This can also be explained by "hidden hearing loss," defined as selective reduction of the cochlear nerve synapse number associated with noise exposure or aging (43–45) while outer hair cells remain well-preserved. In other words, false positive may present normal pure tone audiometry while being classified as hearing loss by the application. In any case, they need to be taken care of and advised to see an ENT doctor.

Next, we used the Z-score method to define statistical criteria of normality for speech in noise evaluation. The Z-score is very useful because it allows calculation of the probability of a score occurring within a normal-hearing population and compare two scores that come from different populations with the normal distribution. In our study, a score of 0 roughly corresponds to the normality (DIN SRT $= -15.4$ dB SNR). Note that DIN SRT ≤ -13 dB SNR value (Z-score $= 1.5$, percentile 95) roughly corresponds to the cut-off inferred from the ROC curves (-12.9 dB SNR). Consequently, a subject with DIN SRT of -12.2 (Z-score $= 2$, confidence interval: 97.5%) will be ranked in the hearing loss population. A Z-score of 2 also fits with the recommendation of the French ENT society (46) (3 dB SNR above the norm) estimate from different tests of speech perception in noise (Hint, FrBio, French DIN test or Framatrix). Finally, Z-score calculation may also provide unified basis for inter-language comparison of DIN tests while SRT values ranges are different.

After implementation of the French antiphasic DIN test on iOS and Android mobile apps, it was launched as Höra. In total, 19,545 completed tests were registered and analyzed. Age distributions showed a first mode around 25 years of age and a second mode around 50 years of age. Although the smartphone users are younger than our reference normative population (median age: 33 vs. 55 years of age, respectively), the normal SNR and the cut-off age was similar in both populations, which supports the reliably of the test to predict hearing loss. Three quarters of them were classified as normal (74%) and one quarter presented mild (9%) or more severe loss (17%), which is also consistent with English version (HearZA) realized in 2018 by De Sousa et al. (47). When we compared the Höra study with the French screening test using digit triplet SRTs (42), the number of subjects who fell in the 50–70 year range was lower (23 vs. 60%, respectively) and the percentage of people with a "good" SRT outcome was higher (74 vs. 46%, respectively). The discrepancies are probably due to lower average age in the Höra population (median age: 33 vs. 58 years of age, respectively).

In summary, the present results validate the efficiency of the Höra application in its purpose of screening and raising awareness on hearing loss. Since the application counsels a confirmation by a professional, listeners with hearing loss will benefit from a follow-up diagnostic assessment by a hearing specialist who will confirm the smartphone test result. The French Ministry of Health recently outlined the full reimbursement of hearing aids (8). In addition to the implementation of this reform, screening the French population through an accessible free mobile app has significant potential to increase access to hearing aids and improve the subsequent quality of life in a larger proportion of the population.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JCC, DS, CS, JLP, and FV designed and conceived the study. JCC and MJD coordinated the study. JCC and LG collected the data. JCC, DS, and CS performed the data and statistical analysis. JCC and JLP designed and realized the figures and wrote the first draft. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work was supported by Fondation Pour l'Audition, non-profit foundation recognized as a public-interest since 2015.

ACKNOWLEDGMENTS

We would like to thank Dr. Jérôme Bourien for his valuable advices on statistical analysis, Willy Nguyen and Noelia G. Best who participated to the data collection.

REFERENCES

1. World Health Organization. *World Report on Hearing* (2021). Disponible sur. Available online at: https://www.who.int/publications-detail-redirect/world-report-on-hearing (accessed 20 avr 2021).
2. Loughrey DG, Kelly ME, Kelley GA, Brennan S, Lawlor BA. Association of age-related hearing loss with cognitive function, cognitive impairment, and dementia: a systematic review and meta-analysis. *JAMA Otolaryngol Head Neck Surg*. (2018) 144:115–26. doi: 10.1001/jamaoto.2017.2513
3. ehima. *EuroTrak-Trends-2009–2020* (2020). Disponible sur. Available online at: https://www.ehima.com/wp-content/uploads/2020/07/EuroTrak-Trends-2009-2020-June-2020.pdf (accessed 14 déc 2020).
4. Shield B. *Evaluation of the Social and Economic Costs of Hearing Imparment*. Hear-It AISBL (2019). p. 250.
5. Wilson BS, Tucci DL, Merson MH, O'Donoghue GM. Global hearing health care: new findings and perspectives. *Lancet*. (2017) 390:2503–15. doi: 10.1016/S0140-6736(17)31073-5
6. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet Lond Engl*. (2020) 396:413–46. doi: 10.1016/S0140-6736(20)30367-6
7. ehima. *EuroTrak France* (2018). Available online at: https://www.ehima.com/wp-content/uploads/2018/07/EuroTrak_2018_FRANCE.pdf (accessed 14 déc 2020).
8. Journal officiel. *Décret n° 2019-21 du 11 janvier* (2019). 2019-21 janv 11 (2019).
9. Pienkowski M. On the etiology of listening difficulties in noise despite clinically normal audiograms. *Ear Hear*. (2017) 38:135–48. doi: 10.1097/AUD.0000000000000388
10. Bauer MA, Sales A, Teixeira AR, Morsch P, Lessa AH, Bós ÂJG. Development and accuracy of a hearing screening application. *Braz J Otorhinolaryngol*. (2020) S1808-8694(20)30056-2. doi: 10.1016/j.bjorl.2020.03.009
11. Corona AP, Ferrite S, Bright T, Polack S. Validity of hearing screening using hearTest smartphone-based audiometry: performance

evaluation of different response modes. *Int J Audiol*. (2020) 59:666–73. doi: 10.1080/14992027.2020.1731767
12. Barczik J, Serpanos YC. Accuracy of smartphone self-hearing test applications across frequencies and earphone styles in adults. *Am J Audiol*. (2018) 27:570–80. doi: 10.1044/2018_AJA-17-0070
13. Szudek J, Ostevik A, Dziegielewski P, Robinson-Anagor J, Gomaa N, Hodgetts B, et al. Can Uhear me now? Validation of an iPod-based hearing loss screening test. *J Otolaryngol Head Neck Surg*. (2012) 41:S78–84.
14. Kelly EA, Stadler ME, Nelson S, Runge CL, Friedland DR. Tablet-based screening for hearing loss: feasibility of testing in nonspecialty locations. *Otol Neurotol*. (2018) 39:410–6. doi: 10.1097/MAO.0000000000001752
15. Mahomed-Asmail F, Swanepoel DW, Eikelboom RH, Myburgh HC, Hall J. Clinical validity of hearScreen^{TM} smartphone hearing screening for school children. *Ear Hear*. (2016) 37:e11–17. doi: 10.1097/AUD.0000000000000223
16. Wu W, Lü J, Li Y, Kam ACS, Fai Tong MC, Huang Z, et al. A new hearing screening system for preschool children. *Int J Pediatr Otorhinolaryngol*. (2014) 78:290–5. doi: 10.1016/j.ijporl.2013.11.026
17. Yeung JC, Heley S, Beauregard Y, Champagne S, Bromwich MA. Self-administered hearing loss screening using an interactive, tablet play audiometer with ear bud headphones. *Int J Pediatr Otorhinolaryngol*. (2015) 79:1248–52. doi: 10.1016/j.ijporl.2015.05.021
18. Potgieter J-M, Swanepoel DW, Myburgh HC, Hopper TC, Smits C. Development and validation of a smartphone-based digits-in-noise hearing test in South African English. *Int J Audiol*. (2015) 55:405–11. doi: 10.3109/14992027.2016.1172269
19. Wilson RH, Burks CA, Weakley DG. A comparison of word-recognition abilities assessed with digit pairs and digit triplets in multitalker babble. *J Rehabil Res Dev*. (2005) 42:499–510. doi: 10.1682/JRRD.2004.10.0134
20. McArdle RA, Wilson RH, Burks CA. Speech recognition in multitalker babble using digits, words, and sentences. *J Am Acad Audiol*. (2005) 16:726–39; quiz 763–4. doi: 10.3766/jaaa.16.9.9
21. Wilson RH, Burks CA, Weakley DG. Word recognition of digit triplets and monosyllabic words in multitalker babble by listeners with sensorineural hearing loss. *J Am Acad Audiol*. (2006) 17:385–97. doi: 10.3766/jaaa.17.6.2
22. Smits C, Kapteyn TS, Houtgast T. Development and validation of an automatic speech-in-noise screening test by telephone. *Int J Audiol*. (2004). 43:15–28. doi: 10.1080/14992020400050004
23. Smits C, Houtgast T. Results from the Dutch speech-in-noise screening test by telephone. *Ear Hear*. (2005) 26:89–95. doi: 10.1097/00003446-200502000-00008
24. Van den Borre E, Denys S, van Wieringen A, Wouters J. The digit triplet test: a scoping review. *Int J Audiol*. (2021) 1—18. doi: 10.1080/14992027.2021.1902579
25. Denys S, Hofmann M, van Wieringen A, Wouters J. Improving the efficiency of the digit triplet test using digit scoring with variable adaptive step sizes. *Int J Audiol*. (2019) 58:670–7. doi: 10.1080/14992027.2019.1622042
26. Denys S, Hofmann M, Luts H, Guérin C, Keymeulen A, Van Hoeck K, et al. School-age hearing screening based on speech-in-noise perception using the digit triplet test. *Ear Hear*. (2018) 39:1104–15. doi: 10.1097/AUD.0000000000000563
27. Denys S, Wouters J, van Wieringen A. The digit triplet test as a self-test for hearing screening at the age of school-entry. *Int J Audiol*. (2021) 1–8. doi: 10.1080/14992027.2021.1930204
28. Wolmarans J, De Sousa KC, Frisby C, Mahomed-Asmail F, Smits C, Moore DR, et al. Speech recognition in noise using binaural diotic and antiphasic digits-in-noise in children: maturation and self-test validity. *J Am Acad Audiol*. (2021) 32:315–23. doi: 10.1055/s-0041-1727274
29. Vlaming MSMG, MacKinnon RC, Jansen M, Moore DR. Automated screening for high-frequency hearing loss. *Ear Hear*. (2014) 35:667–79. doi: 10.1097/AUD.0000000000000073
30. Ozimek E, Kutzner D, Sek A, Wicher A. Development and evaluation of Polish digit triplet test for auditory screening. *Speech Commun*. (2009) 51:307–16. doi: 10.1016/j.specom.2008.09.007
31. De Sousa KC, Swanepoel DW, Moore DR, Myburgh HC, Smits C. Improving sensitivity of the digits-in-noise test using antiphasic stimuli. *Ear Hear*. (2020) 41:442–50. doi: 10.1097/AUD.0000000000000775
32. Koopmans WJA, Goverts ST, Smits C. Speech recognition abilities in normal-hearing children 4 to 12 years of age in stationary and interrupted noise. *Ear Hear*. (2018) 39:1091–103. doi: 10.1097/AUD.0000000000000569
33. Cameron S, Dillon H, Glyde H, Kanthan S, Kania A. Prevalence and remediation of spatial processing disorder (SPD) in Indigenous children in regional Australia. *Int J Audiol*. (2014) 53:326–35. doi: 10.3109/14992027.2013.871388
34. De Sousa KC, Smits C, Moore DR, Myburgh HC, Swanepoel DW. Diotic and antiphasic digits-in-noise testing as a hearing screening and triage tool to classify type of hearing loss. *Ear Hear*. (2021).
35. Culling JF, Lavandier M. Binaural Unmasking and Spatial Release from Masking. In: Litovsky RY, Goupell MJ, Fay RR, Popper AN, éditeurs. *Binaural Hearing: With 93 Illustrations*. Cham: Springer International Publishing; (2021). p. 209–41. (Springer Handbook of Auditory Research). doi: 10.1007/978-3-030-57100-9_8
36. Smits C, Watson CS, Kidd GR, Moore DR, Goverts ST. A comparison between the Dutch and American-English digits-in-noise (DIN) tests in normal-hearing listeners. *Int J Audiol*. (2016) 55:358–65. doi: 10.3109/14992027.2015.1137362
37. Smits C. Improving the efficiency of speech-in-noise hearing screening tests. *Ear Hear*. (2017) 38:e385–8. doi: 10.1097/AUD.0000000000000446
38. Swanepoel DW, De Sousa KC, Smits C, Moore DR. Mobile applications to detect hearing impairment: opportunities and challenges. *Bull World Health Organ*. (2019) 97:717–8. doi: 10.2471/BLT.18.227728
39. Smits C, Theo Goverts S, Festen JM. The digits-in-noise test: assessing auditory speech recognition abilities in noise. *J Acoust Soc Am*. (2013) 133:1693–706. doi: 10.1121/1.4789933
40. BIAP - Bureau International d'Audiophonologie. Available online at: https://www.biap.org/en/component/content/article/65-recommendations/ct-2-classification/5-biap-recommendation-021-bis (accessed 23 août 2021).
41. Lyzenga J, Smits C. Effects of coarticulation, prosody, and noise freshness on the intelligibility of digit triplets in noise. *J Am Acad Audiol*. (2011) 22:215–21. doi: 10.3766/jaaa.22.4.4
42. Jansen S, Luts H, Wagener KC, Frachet B, Wouters J. The French digit triplet test: a hearing screening tool for speech intelligibility in noise. *Int J Audiol*. (2010) 49:378–87. doi: 10.3109/14992020903431272
43. Kujawa SG, Liberman MC. Adding insult to injury: cochlear nerve degeneration after ≪ temporary ≫ noise-induced hearing loss. *J Neurosci*. (2009) 29:14077–85. doi: 10.1523/JNEUROSCI.2845-09.2009
44. Liberman MC. Noise-induced and age-related hearing loss: new perspectives and potential therapies. *F1000Research* (2017). 6. Available online at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5482333/ (accessed 31 mars 2021).
45. Monaghan JJM, Garcia-Lazaro JA, McAlpine D, Schaette R. Hidden hearing loss impacts the neural representation of speech in background noise. *Curr Biol*. (2020) 30:4710–21.e4. doi: 10.1016/j.cub.2020.09.046
46. Joly C-A, Reynard P, Mezzi K, Bakhos D, Bergeron F, Bonnard D, et al. *Recommandation de la société française d'audiologie (SFA) et de la société française d'ORL et de chirurgie cervico-faciale (SFORL) pour de l'audiométrie vocale dans le bruit chez l'adulte*. (2020).
47. De SKC, Swanepoel DW, Moore DR, Smits C. A smartphone national hearing test: performance and characteristics of users. *Am J Audiol*. (2018) 27:448–54. doi: 10.1044/2018_AJA-IMIA3-18-0016

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.

ORIGINAL RESEARCH
published: 08 November 2021
doi: 10.3389/fmed.2021.740123

Deep Learning-Based Speech Enhancement With a Loss Trading Off the Speech Distortion and the Noise Residue for Cochlear Implants

Yuyong Kang[1], *Nengheng Zheng*[1,2]* *and Qinglin Meng*[3]

[1] Guangdong Key Laboratory of Intelligent Information Processing, College of Electronics and Information Engineering, Shenzhen University, Shenzhen, China, [2] Pengcheng Laboratory, Shenzhen, China, [3] Acoustics Laboratory, School of Physics and Optoelectronics, South China University of Technology, Guangzhou, China

The cochlea plays a key role in the transmission from acoustic vibration to neural stimulation upon which the brain perceives the sound. A cochlear implant (CI) is an auditory prosthesis to replace the damaged cochlear hair cells to achieve acoustic-to-neural conversion. However, the CI is a very coarse bionic imitation of the normal cochlea. The highly resolved time-frequency-intensity information transmitted by the normal cochlea, which is vital to high-quality auditory perception such as speech perception in challenging environments, cannot be guaranteed by CIs. Although CI recipients with state-of-the-art commercial CI devices achieve good speech perception in quiet backgrounds, they usually suffer from poor speech perception in noisy environments. Therefore, noise suppression or speech enhancement (SE) is one of the most important technologies for CI. In this study, we introduce recent progress in deep learning (DL), mostly neural networks (NN)-based SE front ends to CI, and discuss how the hearing properties of the CI recipients could be utilized to optimize the DL-based SE. In particular, different loss functions are introduced to supervise the NN training, and a set of objective and subjective experiments is presented. Results verify that the CI recipients are more sensitive to the residual noise than the SE-induced speech distortion, which has been common knowledge in CI research. Furthermore, speech reception threshold (SRT) in noise tests demonstrates that the intelligibility of the denoised speech can be significantly improved when the NN is trained with a loss function bias to more noise suppression than that with equal attention on noise residue and speech distortion.

Keywords: cochlear implant, speech enhancement, perceptual property, deep learning, loss function

Edited by:
Alexandros-Apostolos A. Boulogeorgos, University of Piraeus, Greece

Reviewed by:
Stylianos E. Trevlakis, Aristotle University of Thessaloniki, Greece
Ilias Siniosoglou, University of Western Macedonia, Greece
Tobias Goehring, University of Cambridge, United Kingdom

***Correspondence:**
Nengheng Zheng
nhzheng@szu.edu.cn

Specialty section:
This article was submitted to Translational Medicine, a section of the journal Frontiers in Medicine

Received: *13 August 2021*
Accepted: *04 October 2021*
Published: *08 November 2021*

Citation:
Kang Y, Zheng N and Meng Q (2021) Deep Learning-Based Speech Enhancement With a Loss Trading Off the Speech Distortion and the Noise Residue for Cochlear Implants. Front. Med. 8:740123. doi: 10.3389/fmed.2021.740123

INTRODUCTION

A cochlear implant (CI) is an auditory prosthesis playing an essential role in restoring hearing ability for patients with severe-to-profound sensorineural hearing impairment (1, 2). CI recipients can achieve good speech understanding ability in quiet environments. However, their hearing ability degrades dramatically in noisy backgrounds (3, 4). The main reason is that the signal processing in CIs is a very coarse imitation of the sound coding in a healthy cochlea (5). The inner hair cells (around 3,500), in charge of transforming sound vibrations in the cochlea into electrical signals, are replaced by only 12–26 implanted intracochlear electrodes. Signal processing strategies in CIs can only transmit coarsely frequency-resolved temporal envelopes of the speech to stimulate the auditory nerves. Therefore, the information conveyed by the spectro-temporal fine structures, which are very important for speech understanding in noise, are not effectively represented in CI (6, 7). Therefore, speech enhancement (SE) algorithms have been developed to improve speech intelligibility in noisy environments for CI recipients (8–11). Unfortunately, this is still a pending problem.

Even for modern CIs with multiple microphones, single-channel SE algorithms are mostly implemented after a directional processing stage. Typical single-channel SE algorithms for CIs include spectral subtraction (SS), subspace projection (SP), Wiener filtering (WF), time-frequency (T-F) masking, etc. Yang et al. (12) implemented an SS-based SE as a front end to CI, which improved the speech understanding ability of CI recipients significantly in speech-shape noise (SSN) but not significantly in babble noise (Babble). Loizou et al. (13) proposed an SP-based SE where CI recipients received intelligibility improvement in stationary noise. Guevara et al. (14) proposed a multiband single-channel WF SE for CI recipients, where subjects achieved significant intelligibility gain in SSN but slightly improved in cocktail party noise. Koning et al. (15) investigated two T-F masking-based SEs: the ideal binary masking (IBM) and the ideal ratio masking (IRM), on their effectiveness in SE for CIs. Vocoder simulated tests showed that both maskings worked well given known *a priori* signal-to-noise ratio (SNR), but the performance cannot be guaranteed in real conditions due to the unavoidable SNR estimation error. Most of these traditional SE methods rely on an estimate of the noise (or SNR) and a prerequisite on noise stationarity. Therefore, their performances in nonstationary noise are usually not as convincing as in stationary noise.

Data-driven models, particularly the deep-learning (DL) ones, have been applied for SE with promising results, especially in nonstationary noisy environments where most conventional SEs fail. A well-known example is the spectral mapping-based SE, which uses the clean speech spectra as the training targets such that a noisy-to-clean spectral mapping network can be obtained (16). Another example is the masking-based SE, which is similar to the traditional IBM/IRM ones except that a network is trained to estimate the masking gain from the noisy input such that no explicit noise/SNR estimation is required (17). Model-based mapping or masking methods have also been adopted for SE in CI. In SE for CI, enhancement processing can be done on either the acoustic or electric signals. For the acoustic SE, Lai et al. (18, 19) proposed deep neural network (DNN)-based spectral mapping as an SE front end to CI processor. Both objective and subjective evaluations showed superior performance over traditional SEs. Goehring et al. (20) implemented the recurrent neural networks (RNN)-based T-F masking method to enhance the acoustic signal. Results indicated that both objective and subjective evaluations achieved significant improvement. For electric SE, Hu et al. (21) used a Gaussian mixture model (GMM) as the binary classifier to estimate the IBM gains for each electrode channel. Results demonstrated that CI subjects obtained significant improvement on speech understanding in Babble, Train, and Hall noises. Mamun etal. (22) proposed a convolutional neural network (CNN)-based IRM gain estimator to enhance the temporal envelopes (TEs) of each channel. Objective evaluations showed a significant improvement in speech intelligibility in noisy ambiance. Bolner et al. (23) and Goehring et al. (24) used DNN to estimate the electrode-dependent channel gains with which noise components in the TEs can be suppressed. Results showed that DNN-based IRM performed better than WF in both vocoder-simulated and CI-subjective tests. Zheng et al. (25) presented a DL-based CI strategy. Instead of serving as a front end or a built-in module in CI strategy, the NN was built and trained to simulate a specific strategy of a clinical CI device. The NN output was compatible with the clinical device, and the noise robustness of the NN was obtained through data-driven network optimization.

Most of the abovementioned DL-based SEs focus on minimizing the overall difference between the target speech and its denoised estimate, and, usually, the mean-square-error (MSE)-based loss functions are adopted for NN training. NNs trained with separate speech and noise losses have been demonstrated to be beneficial for SE. For example, Xu et al. (26) proposed a masking-based SE, in which the NN to estimate the masking gain was trained with a loss function containing separately computed speech distortion and residual noise. Objective evaluations demonstrated that NN trained with the new loss outperformed the one trained with traditional MSE loss, and the best results were attained when the speech and noise losses were equally combined.

As for CI, due to its coarse imitation of the normal auditory system, the recipients obtain an electric hearing much different from the acoustic hearing of NH people. A well-recognized property of electric hearing is that the recipients are more tolerant of speech distortion but very sensitive to noise (10, 27–29). In contrast, NH people are more sensitive to distortion than noise (27, 30). In addition, different CI recipients have noticeable individual differences due to hearing experience, devices, surgery, physiological conditions, etc. Therefore, the individualized perceptual sensitivity to noise and distortion should be considered in designing SE front ends for patients with CI, and a more sophisticated combination of the two losses should be investigated.

This study aims (1) to investigate perceptual sensitivities of the CI recipients to noise and distortion, and will such sensitivities vary across different noise conditions? and (2) to design an effective SE front end with the knowledge of such sensitivities of CI recipients. We developed a DL-based SE as a front end to the signal processing strategy of CIs. A long-short term memory (LSTM) network was trained to estimate the T-F masking gains. Instead of the MSE, a loss function similar to that in Xu et al. (26) was adopted for network training. By adjusting the weights for trading off the speech distortion and the noise residue, their contributions to speech intelligibility for CI recipients were investigated, upon which an LSTM trained with preference-biased-loss was developed. Finally, a set of subjective experiments was conducted to evaluate the system performance.

ALGORITHM DESCRIPTION

SE Based on Time-Frequency Masking

Assuming that speech and noise are additive in the time domain, i.e.,

$$y(n) = s(n) + d(n) \tag{1}$$

where $y(n)$, $s(n)$, and $d(n)$ denote noisy speech, clean speech, and noise, respectively. Since speech is a short-term stationary signal, the frequency domain representation of (Equation 1) can be obtained by applying short-time Fourier transform (STFT) to the time signals, i.e.,

$$Y(t, f) = S(t, f) + D(t, f) \tag{2}$$

where t and f denote the index of the time frames and the frequency bins for each T-F unit.

Wiener filtering (WF) has been one of the most widely implemented SE methods to estimate $S(t,f)$ from $Y(t,f)$. Given speech and noise uncorrelated, a gain function $G(t,f)$ to suppress the noise can be written as

$$G(t, f) = \left(\frac{|S(t, f)|^2}{|S(t, f)|^2 + |D(t, f)|^2}\right)^{1/2} = \left(\frac{|S(t, f)|^2}{|Y(t, f)|^2}\right)^{1/2} \tag{3}$$

Assuming the effect of phase distortion is negligible, the target speech spectra can be estimated by

$$\hat{S}(t,f) = G(t,f) \cdot Y(t,f) = G(t,f) \cdot |Y(t,f)| \cdot e^{j\varphi_Y(t,f)} \tag{4}$$

where $\varphi_Y(t,f)$ is the phase of the noisy speech.

Time-frequency masking, first proposed for speech separation in computational auditory scene analysis tasks, has been demonstrated to be the most successful in SE tasks. WF can be regarded as T-F masking for noise suppression. Essentially, the masking gain as in (Equation 3) provides the optimal filtering in the sense of minimized MSE, given an accurate estimation of noise or SNR. Unfortunately, such an accurate noise/SNR estimation is usually not an easy task.

Figure 1 shows the diagram for a WF-based SE. As illustrated, a noise estimation from the noisy speech spectra is required for computing the masking gain.

Deep Learning-Based T-F Masking for SE

In the DL-based SE, the masking gain is computed from a pre-trained NN. NNs have been known for their powerful learning ability, given enough training data. Therefore, given a well-trained NN, the gain can be reliably computed from the noisy input without an explicit noise/SNR estimation.

Figure 2 shows the diagram for DL-based SE in which the masking gain is computed from the noisy input by a pre-trained LSTM. Here, LSTM is adopted for its superiority in modeling sequential signals like speech over other networks. As shown, the pre-trained LSTM takes the noisy spectral magnitude, $|Y(t,f)|$, as input and output of the masking gain, $\hat{G}(t,f)$, which multiplies $|Y(t,f)|$ to generate a denoised spectral magnitude, $|\hat{S}(t,f)|$. Finally, the inverse STFT (ISTFT) is employed to recover the time-domain signal from the enhanced magnitude spectra and noisy phase spectra. Unlike the WF, no explicit noise or SNR estimate is required, as the gain is directly estimated by the LSTM.

Loss Functions for NN Training

Many factors affect the performance of an NN, including the network structure, training strategy, optimization method, etc. This study investigates the effect of different loss functions, i.e., the way measuring the difference between NN output and the target signal, on their performance on NN training.

The most adopted loss is the MSE given by

$$J_{MSE} = \frac{1}{T \cdot F} \sum_t \sum_f (|\hat{S}(t,f)| - |S(t,f)|)^2 \tag{5}$$

where T and F are the total numbers of time frames and frequency bins, respectively.

As known, noise suppression in any SE may induce an inevitable distortion to the target speech. Usually, the more noise is suppressed, the more speech gets distorted. The MSE loss in (Equation 5) computes the overall errors, including both speech distortion and noise residue. It forces the DL-based SE system to output estimated speech that is statistically and objectively *optimal* with respect to the data. However, speech perception is subjective, and the data-level objective optimum might not necessarily result in a perceptual optimum. Specifically, perceptual sensitivity to noise and speech distortion varies across different noise conditions and different subjects. Therefore, the NN may benefit from being trained with a loss function trading off the speech distortion and the noise residue.

We introduce a new loss function combining weighted speech distortion and noise residue, noted as weighting loss (WL), to train the LSTM. The WL is given as (26).

$$\begin{aligned} J_{WL} &= \alpha \frac{1}{T \cdot F} \sum_t \sum_f (|\tilde{S}(t,f)| - |S(t,f)|)^2 \\ &+ (1-\alpha) \frac{1}{T \cdot F} \sum_t \sum_f (|\tilde{D}(t,f)|)^2 \end{aligned} \tag{6}$$

where $|\tilde{S}(t,f)| = \hat{G}(t,f) \cdot |S(t,f)|$ is the distorted speech spectrum, i.e., the remaining target speech components after masking, $|\tilde{D}(t,f)| = \hat{G}(t,f) \cdot |D(t,f)|$ is the noise residue, and α is the weighting factor. Given $\alpha = 1$, the loss J_{WL} forces the SE system to retain the target speech components as much as possible, regardless of whether it suppresses the noise components. On the other hand, if $\alpha = 0$, the system suppresses the noise as much as possible, regardless of the speech distortion. That is, the remaining noise residue and the induced speech distortion can be traded off by adjusting the parameter α. From (Equation 5) and (Equation 6), it is easy to infer that $\alpha = 0.5$ does not give J_{WL} identical to J_{MSE} in general, and only when the clean speech could be perfectly estimated (i.e., $|\hat{S}(t,f)| = |S(t,f)|$), $\alpha = 0.5$ gives $J_{WL} = J_{MSE} = 0$.

Figures 3, **4** give the diagrams for the training of the LSTM with the respective MSE- and WL-based loss functions.

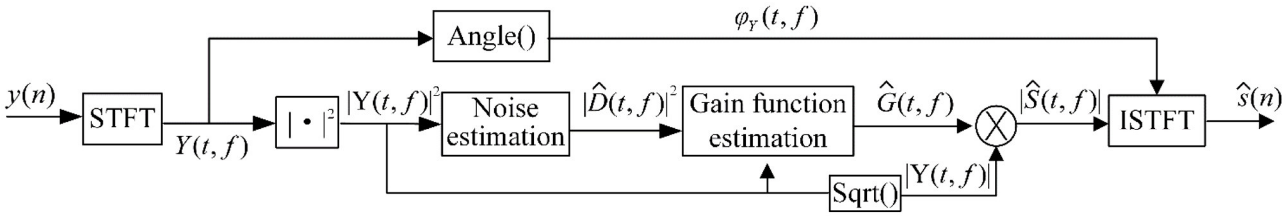


FIGURE 1 | The diagram for WF-based SE.

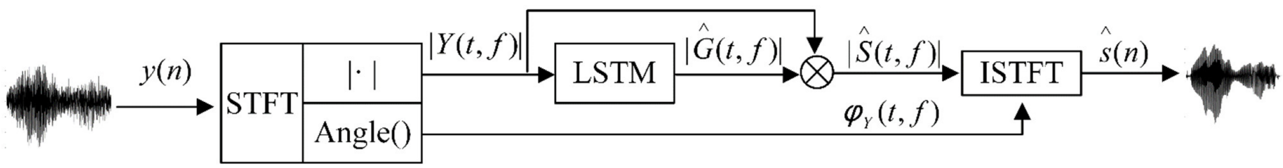


FIGURE 2 | The diagram for SE with DL-estimated T-F masking gain.

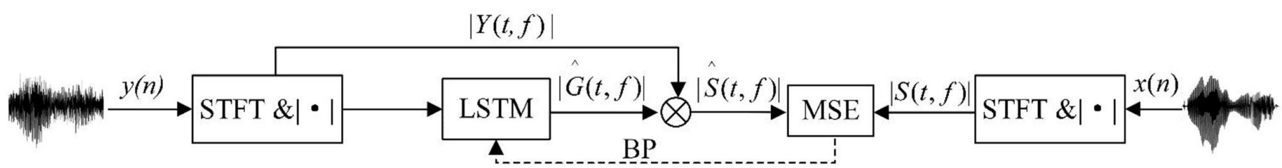


FIGURE 3 | The diagram for training the LSTM with MSE loss.

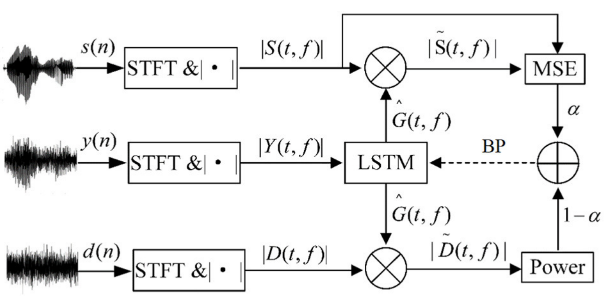


FIGURE 4 | The diagram for the training of the LSTM with weighting loss.

As shown, the LSTMs are optimized iteratively by the backpropagation (BP) algorithm with the respective losses.

EXPERIMENTAL SETTING

Speech Materials

Two speech corpora, i.e., an open-access Chinese speech database built by Tsinghua University, THCHS-30 (31) and the Mandarin hearing in noise test, MHINT-M (32), were adopted for the experiments. THCHS-30 is a Mandarin speech database widely used to develop DL-based speech systems. It contains three subsets, i.e., training, development, and test sets, consisting of 10,000, 893, and 2,495 utterances, respectively. MHINT-M is a Mandarin speech database designed for the subjective listening test. It contains 14 lists, 12 for formal tests and two for practice. There are 20 utterances in each list and 10 Mandarin syllables in each utterance. The noisy speeches were generated by additively mixing the clean ones with two noise samples, SSN and Babble. The SSN noise was generated by shaping (multiplying) a white noise spectrum with an averaged speech envelope. The Babble noise was taken from the NOISE-92 database (33). The duration of SSN and Babble noises are about 10 and 4 min, respectively. In this study, the training and validation of NNs in the training stage used the training and the development sets of THCHS-30, and all tests used the speech signals in MHINT-M. All speech and noise signals were downsampled to 16 kHz for the experiments.

There were, in total, 10 noisy conditions for the training, i.e., two noises (SSN and Babble), each at five SNRs (from 0 to 20 dB in a step of 5 dB), in generating the noisy speech. Each utterance in the training and development sets of THCHS-30 was randomly mixed with a noise segment (randomly picked out from the whole noise recording) in one of the 10 conditions such that there are 10,000 and 839 noisy utterances used for NN training and validation, respectively.

There were 62 noisy conditions for the test, i.e., two noises, each at 31 SNRs (from −10 to 20 dB in a step of 1 dB), for noisy speech generation. Each utterance in MHINT-M was mixed with a noise segment in all 62 conditions. Note that the whole set of noisy speech was used for subjective evaluations, but only a subset, with SNRs from −5 to 15 dB in a step of 5 dB, was selected for objective evaluations.

SE Systems to Be Evaluated

Several DL-based SE systems were developed to examine how the cost function could affect the DL-based SE for CI. Two loss functions introduced in section Loss Functions For NN Training, i.e., the MSE loss and the weighting loss (with different weights), were used for network training. In addition, Wiener filtering-based SE was also developed for comparison, from which the performance gap between the traditional and the DL-based SEs can be shown.

Wiener filtering (WF): instead of the traditional WF, a parametric WF (34) was adopted as the SE front end to CI. The gain function is given as

$$G\left(t,f\right) = \max\left(\frac{|Y\left(t,\ f\right)|^2 - \alpha(t)|\hat{D}\left(t,\ f\right)|^2}{|Y(t,\ f)|^2}, 0.01\right) \tag{7}$$

where $\left|\hat{D}\left(t,\ f\right)\right|^2$ is estimated by an energy-based voice activity detector, the floor parameter 0.01 is set to avoid negative or very small gain, $\alpha(t)$ is a factor to avoid the overestimation of noise and is computed based on the local *a posterior* SNR $[(SNR_{post}(t))]$, i.e.,

$$\alpha(t) = \begin{cases} 3.125, & SNR_{post}(t) < 0dB \\ \frac{-1.875}{20}SNR_{post}(t) + 3.125, & others \\ 1.25, & SNR_{post}(t) > 20dB \end{cases} \tag{8}$$

where $SNR_{post}(t) = 10\log_{10}\frac{\sum_f |Y(t,f)|^2}{\sum_f |\hat{D}(t,f)|^2}$. In this experiment, WF was implemented with the source code of the parametric WF download from https://github.com/dpwe/pitchfilter. More details of the baseline can be referred to the webpage.

T-F masking with gains computed by MSE-trained LSTM (MSE-MASK) The LSTM consisted of three layers, i.e., an input layer with 256 LSTM units, a hidden layer with two fully connected (FC) layers (512 neural units per layer), and an output layer with 256 FC units such that the output has the same dimension as the input. LeakyReLU activation function was applied to the input and hidden layers. Sigmoid activation function was applied to the output layer. The parameters of networks were optimized by Adam optimizer with an initial learning rate of 0.005. When the loss did not decline for two consecutive epochs, the learning rate was reduced to half until < 0.00001. The model was trained for 60 epochs. The validation was implemented after each training epoch. Finally, the best model, i.e., the one with the minimum loss among all the validated ones, was selected for tests.

The long-short term memory was trained with all noisy speech covering all the 10 noise conditions mentioned in section Speech Materials. To train the LSTM, T-F spectra of noisy speech, $|Y(t,f)|$, and their corresponding clean spectra, $|S(t,f)|$, were served as the input features and the training labels, respectively. To generate the feature, each speech signal was first segmented into short frames by a Hanning window with a 32-ms length and 16-ms shift. Then, a 512-point fast Fourier transform was applied to each frame, and a 256-dimensional feature was constructed with the magnitude spectra of nonnegative frequency components.

T-F masking with gains computed by WL-trained LSTM (WL-MASK): The training and validation processes for LSTM were the same as in MSE-MASK, except that the weighting loss J_{WL}, instead of the MSE loss, was used. To investigate the effect of the weighting parameter α, we repeated the training process nine times, each with a specific α from 0.1 to 0.9 in a step of 0.1. That is, there were, in total, nine NNs trained with different α. For each α, an LSTM was trained with all noisy speech covering all the 10 noise conditions mentioned in section Speech Materials.

Figure 5 shows the electrodograms extracted from speech signals generated from the same speech utterance with different noisy processing. The electrodograms were generated by processing the acoustical signal with the CCi-Mobile, a CI research development platform developed by CI-Lab at the University of Texas at Dallas (35). **Figures 5A,B** are for clean speech, and that corrupt by SSN at 0 dB SNR (**Figure 5C**) is for the denoised speech with MSE-MASK DL SE, and (**Figures 5D–L**) are for the denoised speeches with WL-MASK DL SEs with $\alpha = 0.1, 0.2, \cdots, 0.9$, respectively. Two spectro-temporal regions in the electrodograms are marked with red boxes and blue boxes for better illustration. As shown, the noise seriously corrupts the electrodogram. All the SE processings suppress the noise to a certain degree and, at the same time, introduce some speech distortion. The MSE-MASK seems to have a balanced speech distortion and noise residue. As for WL-MASK, noise is mostly suppressed, and a large number of speech components are deleted at small α; as α increase, speech components are mostly retained, so as the noise components. Therefore, user-preference-dependent noise-distortion tradeoff could be achieved by properly selected α.

To further investigate the effect of α in trading off the speech distortion and residual noise in the electrodograms, we computed and compared the current units of the enhanced electrograms and the clean ones. The CI speech processor maps the subband envelopes into currents from 0 unit to 255 units. We consider distortion happens when the current unit of the enhanced electrodogram is lower than that of the clean one; otherwise, there exists residual noise. The degree of speech distortion and noise residue is computed as,

$$C_{dis} = \frac{1}{I \cdot T}\sum_i\sum_t \max\left\{0, C_{ref}\ (i,t) - C(i,t)\right\} \tag{9}$$

$$C_{res} = \frac{1}{I \cdot T}\sum_i\sum_t \max\left\{0, C\left(i,t\right) - C_{ref}\ (i,t)\right\} \tag{10}$$

where i and t represent the indices for electrode channels and time frames, I and T are the numbers of electrode channels and time frames, *and* $C_{ref}\ (i,t)$ and $C\left(i,t\right)$ are the current units of the clean electrodograms and the enhanced ones.

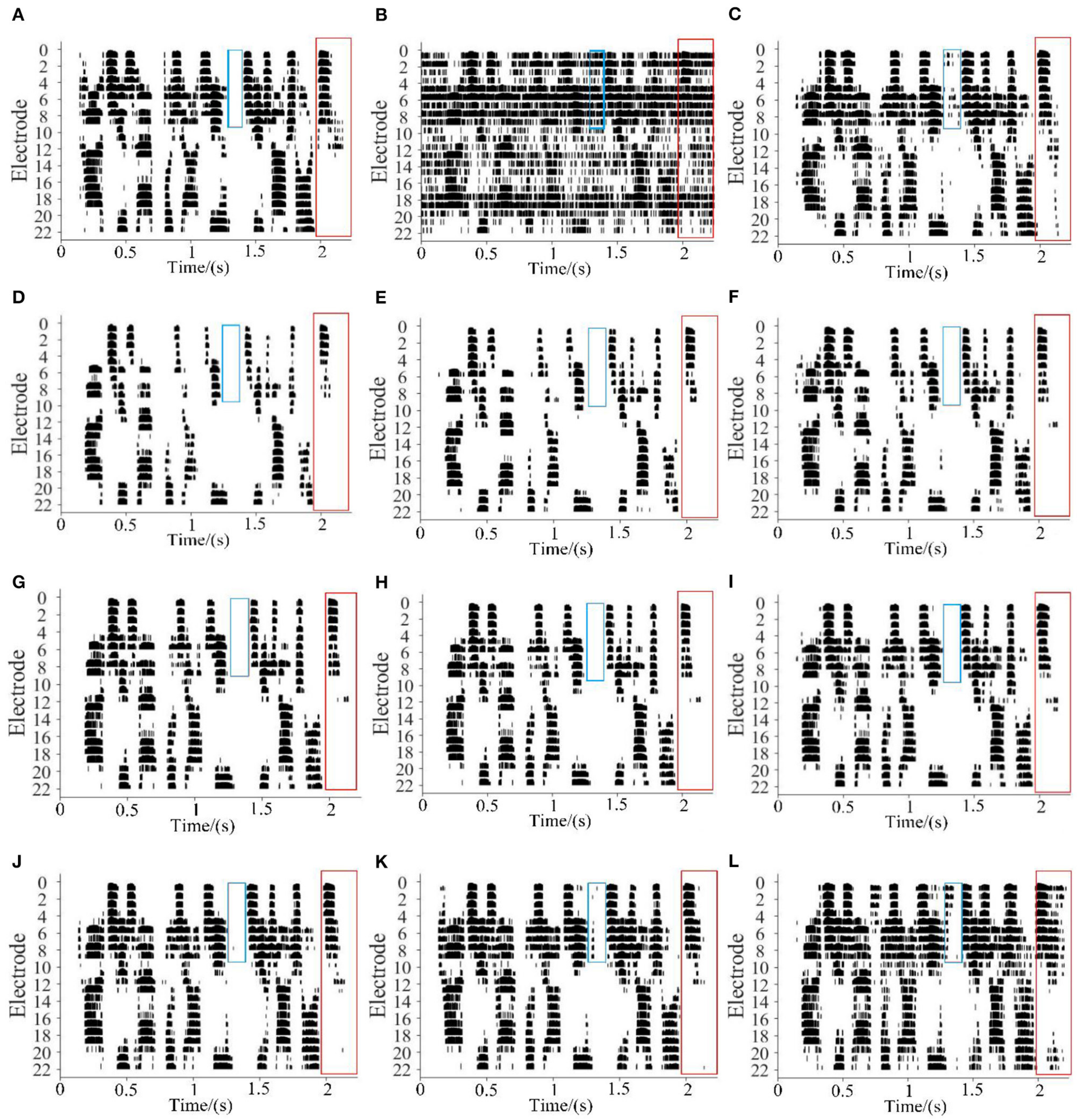


FIGURE 5 | The electrodogram examples of different systems, **(A)** clean speech, **(B)** noisy speech, **(C)** MSE-MASK enhanced, and **(D–L)** WL-MASK with α across from 0.1 to 0.9.

Figure 6 shows the C_{dis} and C_{res} at different α. Ten noise conditions, i.e., two noise types (SSN and Babble), each at five SNRs (−5, 0, 5, 10, and 15 dB), were evaluated. It is clear that, as α increases, the distortion decreases monotonically, and the noise residue increases monotonically in most noise conditions. The only exceptions happen at C_{dis} in −5 dB SSN, C_{dis} in −5 dB Babble, and C_{res} in 0 dB SSN, where some fluctuations appear at around $\alpha = 0.6$. The fluctuation in −5 dB might be because the network has not seen a −5 db SNR during the training.

OBJECTIVE EVALUATION

Methods

The envelope-based correlation measure (36), an objective metric to evaluate speech intelligibility by CI recipients, was adopted

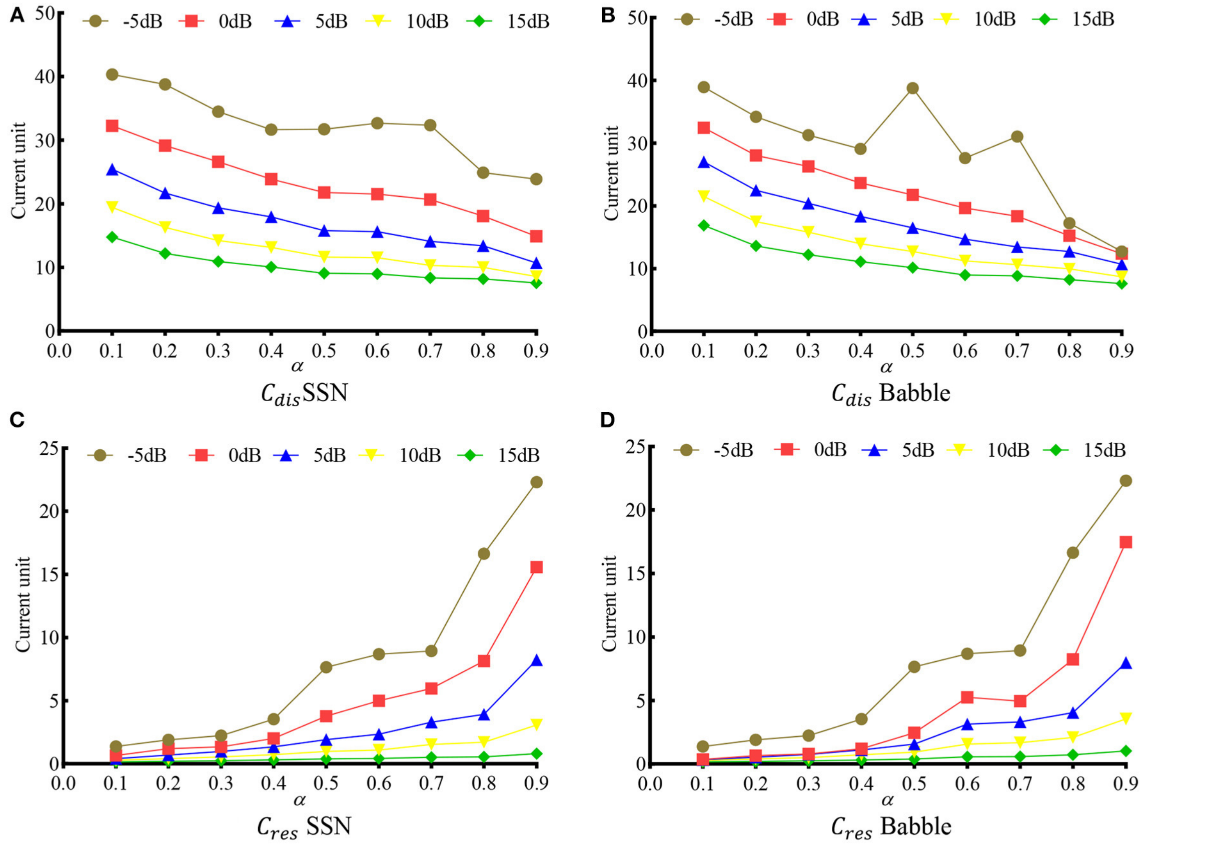


FIGURE 6 | Speech distortion (C_{dis}) and noise residue (C_{res}) as a function of α in various noise conditions. **(A)** C_{dis} in SSN, **(B)** C_{dis} in Babble, **(C)** C_{res} in SSN, and **(D)** C_{res} in Babble.

to measure the performance of different SE systems. In this study, the CCi-Mobile platform was adopted for extracting channel envelopes with the Advanced Combination Encoder (ACE) strategy (37). Given two versions of a speech, e.g., a target one and its distorted one, ECM computes the correlation of their extracted channel envelopes, which will modulate the pulsatile carriers and stimulate the electrodes. In the signal processing strategy of CI, the recipient-dependent MAP parameters are used in computing the channel envelopes. Therefore, ECM computed from such subject-dependent envelopes well represents the speech intelligibility of the corresponding CI recipient (36). The score of ECM is between 0 and 1. The higher ECM, the better intelligibility.

In this experiment, noisy test speech signals were first denoised by different SE front ends and then processed by the ACE strategy in the CCi-Mobile CI research development platform. In addition, the sample MAP file provided in the CCi-Mobile demo system was used in generating the envelopes. ECM was computed on each pair of extracted envelopes, i.e., reference one and distorted one.

Results

Table 1 shows the mean ECM scores for SSN-corrupted noisy speech and their denoised versions with different SE front ends, i.e., WF, MSE-MASK, and WL-MASK with $\alpha = 0.1, 0.2, \cdots, 0.9$. For WL-MASK, the highest score among the nine α is highlighted in red fonts. As illustrated in **Table 1**, all SE front ends achieved a certain ECM gain over noisy speech, except for those highlighted in blue fonts, i.e., WF at high SNRs (10 and 15 dB) and WL-MASK with $\alpha = 0.1$ at SNR of 15 dB. Both DL-based SEs outperformed WF in all SNRs. The performance of the WL-MASK front end varied at different α. Nevertheless, there always exists some α, although the values vary at different SNRs, with which the WL-MASK front end achieved better performance than MSE-MASK.

Table 2 shows the mean ECM scores for Babble-corrupted noisy speech and their denoised versions with different SE front ends. Most SE front ends achieved a certain ECM gain over noisy speech, except for those highlighted in blue fonts. Both DL-based SEs outperformed WF in all SNRs. The performance of the WL-MASK front end varied at different α. Unlike in SSN, in SNR of 5 and 10 dB, the WL-MASK with optimal α showed comparable performance to MSE-MASK.

Tables 1, **2** tell that, although the optimal α varies across SNRs, it generally increases as the SNR increases. Note that α is the weight for loss induced by speech distortion. Therefore, a larger α forces the network to output less distorted speech; in contrast, a smaller α forces the network to suppress more

TABLE 1 | The mean ECM score results for different systems under SSN.

SNR	Noisy	WF	MSE-MASK	WL-MASK								
				$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$	$\alpha = 0.5$	$\alpha = 0.6$	$\alpha = 0.7$	$\alpha = 0.8$	$\alpha = 0.9$
−5 dB	0.143	0.190	0.235	**0.313**	0.256	0.254	0.264	0.282	0.243	0.258	0.234	0.226
0 dB	0.238	0.296	0.456	0.430	0.451	0.485	**0.493**	0.485	0.492	0.489	0.458	0.365
5 dB	0.365	0.428	0.596	0.502	0.556	0.584	0.599	0.586	**0.607**	0.604	0.588	0.503
10 dB	0.512	**0.507**	0.702	0.585	0.652	0.676	0.691	0.683	**0.707**	0.702	0.697	0.643
15 dB	0.712	**0.560**	0.801	**0.673**	0.738	0.764	0.781	0.777	0.800	0.800	**0.805**	0.782

The red bold indicates the best score under each SNRs. The blue bold indicates that worse than noisy condition.

TABLE 2 | The mean ECM score results for different systems under Babble.

SNR	Noisy	WF	MSE-MASK	WL-MASK								
				$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$	$\alpha = 0.5$	$\alpha = 0.6$	$\alpha = 0.7$	$\alpha = 0.8$	$\alpha = 0.9$
−5 dB	0.201	0.228	0.291	0.276	0.293	**0.315**	0.310	0.313	0.292	0.283	0.274	0.253
0 dB	0.295	0.322	0.447	0.413	0.437	**0.464**	0.462	0.463	0.450	0.438	0.424	0.374
5 dB	0.427	0.440	0.592	0.508	0.557	0.583	**0.590**	**0.590**	0.589	0.586	0.576	0.518
10 dB	0.602	**0.528**	0.719	**0.600**	0.664	0.691	0.708	0.704	0.717	**0.719**	0.716	0.676
15 dB	0.789	**0.577**	0.825	**0.695**	**0.760**	**0.787**	0.810	0.801	0.824	0.829	**0.833**	0.809

The red bold indicates the best score under each SNRs. The blue bold indicates that worse than noisy condition.

noise. At low SNRs, noise is the dominant component in noisy speech. Hence, the network must put more attention on noise suppression to improve speech intelligibility. On the other hand, speech dominates the noisy signal at high SNRs, and a larger α is preferred to avoid significant speech distortion. Note that, even for SNR of 0 dB, where speech and noise have the same energy, ECM evaluation shows that noise suppression biased α (0.4 for SSN, 0.3 for Babble) achieved better results. It is reasonable since it is well-known that, unlike NH people, CI recipients are much more sensitive to noise than distortion.

SUBJECTIVE EVALUATION: VOCODER SIMULATION WITH NH SUBJECTS

Methods

Speech reception threshold in noise (38), an SNR level at which the listener could correctly recognize 50% of words in a sentence, was adopted to investigate how the speech distortion and noise residue trading-off would affect the intelligibility of the enhanced speech.

Ten college students, all are normal hearing (pure-tone thresholds not >25 dB HL) and native Mandarin speakers, were recruited with a reward for the test. Each subject underwent 12 SRT measure blocks, each for one of the 12 SE front ends, i.e., Noisy (no SE), WF, MSE-MASK, and WL-MASK with nine α. The 12 test lists from MHINT-M were used, each for a block. Before the formal test, the subject had taken a practice session with the two practice lists in MHINT-M. Due to the limit of speech materials, each subject was tested with one noise type, either SSN or Babble.

The noisy speech signals were first processed by the SE front ends. Then, the vocoded speeches were generated from a Gaussian-enveloped-tone vocoder (39), which directly mapped the electric stimulus of a CI to the acoustic signal. Meng et al. (39) and Kong et al. (40) have demonstrated that this vocoder better predicts the potential CI performance than classical continuous-carrier vocoders. The CCi-Mobile platform was used to generate electric stimuli (electrodogram) of the ACE strategy, where the n-of-m strategy was set to 8 of 22, which is the same as that of Meng et al. (39). The vocoded speech was presented diotically to the NH subjects *via* a Sennheiser HD650 headphone in a soundproof booth.

In each block, SRT was measured with an adaptive procedure using a 20-utterances test list. SNR was adaptively modified in a one-down, one-up way (41). An utterance was "intelligible" when more than half of its syllables were repeated correctly by the subject. The SNR was initialized at 12 dB and changed by 4 dB before the second reversal and by 2 dB afterward. Each sentence could be replayed up to three times upon the request of the subjects. The SRT was computed as the average of the intermediate SNRs of the last six reversals to reduce the measurement deviation.

Results

Figure 7 shows the SRTs for the 12 SE front ends, the left panel for SSN, and the right panel for Babble. For WL-MASK, the mean and standard deviation of SRT for each α are depicted, while for Noisy, WF, and MSE-MASK, mean SRTs are given as constant lines for comparison. As shown, the DL-based SEs outperform the unprocessed noisy speech and that with WF in both noises. Furthermore, by properly trading off the errors induced by speech distortion and residual noise, WL-MASK SE attained lower SRTs than MSE-MASK. In SNN, the best SRT of −2 dB was obtained at $\alpha = 0.3$ and, in Babble, SRT of 0 dB was obtained at $\alpha = 0.4$. The results coincided with the ECM ones in

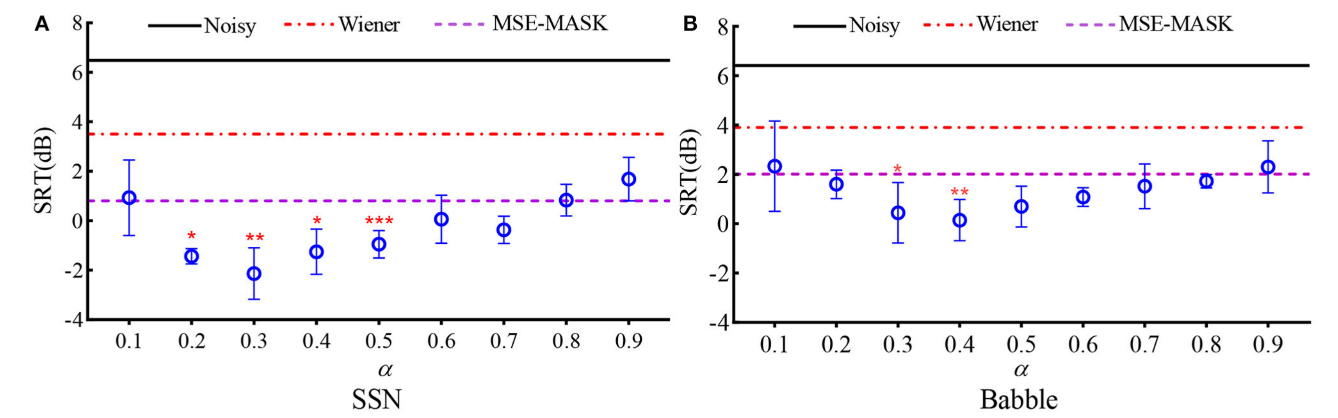


FIGURE 7 | SRTs for different SE front ends with the vocoded speech, **(A)** SSN and **(B)** Babble. The circles represent the means, and the error bars represent the standard deviations for each WL-MASK. The asterisk represents the result of pairwise comparison of MSE-MASK and WL-MASK, * indicates $p < 0.0332$, ** indicates $p < 0.0021$, and *** indicates $p < 0.0001$.

Tables 1, **2**, where the best ECMs were obtained at $\alpha = 0.4$ (SSN) and 0.3 (Babble) at SNR of 0 dB. One-way repeated measure analysis of variance (RM-ANOVA) was applied to analyze the influences of different factors. The results show that there is a significant difference among different DL-SE methods {[$F_{(9, 36)} = 1.262$, $p < 0.001$]}. Dunnett's test was used to pairwise compare MSE-MASK and different WL-MASKs. In SSN, WL-MASK with α from 0.2 to 0.5 obtained significant improvement over MSE-MASK ($p < 0.0332$ at $\alpha = 0.2$ and 0.4; $p < 0.0021$ at $\alpha = 0.3$; $p < 0.0001$ at $\alpha = 0.5$). In Babble, WL-MASK with $\alpha = 0.3$ and 0.4 significantly outperformed MSE-MASK ($p < 0.0332$ at $\alpha = 0.3$; $p < 0.0021$ at $\alpha = 0.4$).

SUBJECTIVE EVALUATION WITH CI RECIPIENTS

Participants

Nineteen CI recipients were recruited for the evaluation. They were all single-side implanted with the CI24M series of Cochlear corporation and could normally communicate in Mandarin in quiet environments. **Table 3** lists the individual information of the subjects. Subjects with processors having noise reduction built-in, i.e., Nucleus 6, Nucleus 7, and Kanso, turned off the noise reduction function during the evaluation. Before the evaluation, each subject read the informed consent and agreed with it. All the subjects were paid after the listening test. This subjective evaluation has been approved by the Medical Ethics Committee of Shenzhen University.

SRT Test

Seven CI recipients, i.e., C1–C7 as listed in **Table 3**, participated in the SRT test. Each subject underwent 10 blocks, each measuring the SRT with one of the 10 experimental conditions: five front ends (Noisy, WF, MSE-MASK, and WL-MASK with α of 0.3 and 0.4), each for two noises (SSN and Babble). Here, only two αs for WL-MASK were tested because those, as illustrated by **Figure 3**, SRTs were mostly around 0 dB, where the optimal α was 0.3 or 0.4. Each block used a 20-utterances list in MHINT-M as the original speech materials. The processed signals were presented to subjects *via* two Genelec 8030A loudspeakers at a comfortable level (about 65 dB SPL) in a soundproof booth. The two loudspeakers were placed front-right or front-left to the participant such that the participants with either left or right side implanted could be equally presented. Before the formal test, each subject had finished two practice blocks. The SRT search process was the same as the vocoder simulation experiments.

Figure 8 shows the SRTs for different SE front ends measured on the seven CI subjects. The left and middle panels give the individual results in SSN and Babble, respectively, and the right panel gives the statistical results on all the subjects. For the individual results, most subjects achieved better SRTs with enhanced speech than noisy speech, except that WF and Noisy performed comparably for C2, C3, and C4. WL-MASK with $\alpha = 0.3$ and 0.4 performed better than MSE-MASK for most subjects except C2. Two-way repeated measure ANOVA (RM-ANOVA) was applied to analyze the influences of different factors. Results showed that there were significant interactions between noise types and SE methods [$F_{(1.56, 9.361)} = 7.152$, $p = 0.02$], and there were significant differences within SE methods [$F_{(1.825, 10.97)} = 19.69$, $p < 0.001$] and within noise types [$F_{(1, 6)} = 43.6$, $p < 0.001$]. Tukey's multiple comparisons test was used for pairwise comparison. In SSN, WL-MASK with $\alpha = 0.3$ and 0.4 significantly improved speech intelligibility against Noisy (both $p < 0.0021$), and significantly outperformed WF (at least $p < 0.0332$), and the superiority of WL-MASK over MSE-MASK was significant ($p < 0.0332$) with $\alpha = 0.3$ but not significant with $\alpha = 0.4$. In Babble, WL-MASK with $\alpha = 0.3$ showed significant superiority over Noisy ($p < 0.0332$) and MSE-MASK ($p < 0.0021$), WL-MASK with $\alpha = 0.4$ showed significant superiority over Noisy ($p < 0.0332$), but its superiority over WF and MSE-MASK was not significant.

Speech Recognition (SR) Test

The ECM results show that the optimal α increases as the SNR increases. To verify this phenomenon, we conducted a

TABLE 3 | The individual subject information for speech recognition evaluation.

Participant	Age at testing (years old)	Etiology	CI experience (years)	Implanted side	Processor
C1	23	Drug induced	20	R	Freedom
C2	23	Drug induced	17	R	Freedom
C3	23	Drug induced	17	R	Sprint
C4	22	Acute meningitis	3	L	Freedom
C5	37	Otitis media	10	R	Esprit 3G
C6	21	Otitis media	12	R	Nucleus 7
C7	26	Drug induced	20	R	Nucleus 5
C8	17	Unknown	14	R	Nucleus 5
C9	12	Unknown	10	R	Kanso
C10	13	LAVS	7	R	Nucleus 5
C11	16	Unknown	12	L	Nucleus 6
C12	15	Drug induced	14	R	Esprit 3G
C13	11	Unknown	8	R	Freedom
C14	28	Drug induced	18	L	Kanso
C15	17	Unknown	16	L	Nucleus 6
C16	12	LAVS	7	R	Nucleus 6
C17	12	Unknown	4	R	Nucleus 5
C18	16	Drug induced	14	R	Nucleus 7
C19	17	Unknown	12	R	Nucleus 5

CI, cochlear implant; F, female; L, left; M; male, and R, right.

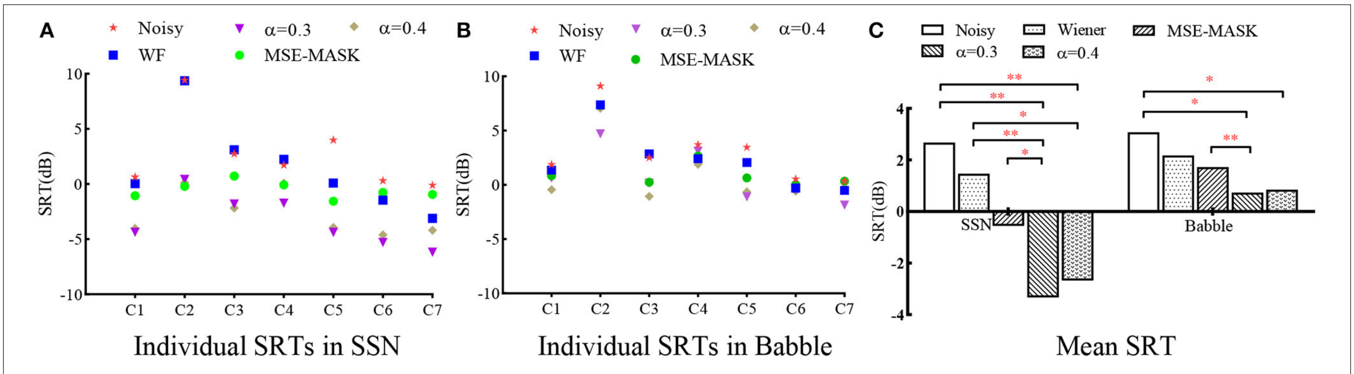


FIGURE 8 | CI users' SRT results for different systems. **(A)** Individual results in SSN, **(B)** individual results in Babble, and **(C)** statistical analysis of the different methods. The asterisk in **(C)** represent significance analysis results. * indicates $p < 0.0332$, ** indicates $p < 0.0021$, *** indicates $p < 0.0001$.

speech recognition test to investigate how the trading-off weight α can maximize the SE gain for CI recipients in different noise conditions.

Twelve CI recipients, i.e., C8–C19 as listed in **Table 3**, participated in the SR test. Due to the limitation on speech materials, only three different αs, i.e., 0.3, 0.5, and 0.7, were tested. Each subject conducted 12 SR blocks, each with one of the 12 test conditions: three SNRs (0 dB, 5 dB, and 10 dB) * four systems (Noisy, WL-MASK with the three αs)* one noise type (either SSN or Babble). Each block randomly took one of the 12 lists in the MHINT-M database as the speech materials. The utterance order was also random in each list. The processed speeches were presented to the subjects in the same way as in the SRT test. Before the formal measurement, each subject had finished two practice blocks in 10 dB noisy condition. In each block, the sentences were presented in random order. Each sentence can be replayed up to three times upon the request of the subjects. The mean word recognition rate (WRR) of the 20 utterances was calculated as the final result in each trial.

Figure 9 shows the mean and standard deviation of WRR (over all the subjects) at different SNRs. The same methods as in section SRT Test were used for significance analysis. RM-ANOVA test showed that, in SSN, there was no significant [$F_{(1.267, 6.333)} = 5.375$, $p = 0.05$] interaction between SNRs and SE methods, there were significant differences within SE methods [$F_{(2.026, 10.13)} = 19.65$, $p < 0.001$] and within SNRs [$F_{(1.404, 7.003)} = 61.46$, $p < 0.001$]; in Babble, there was no significant [$F_{(1.941, 9.704)} = 3.421$, $p = 0.08$] interaction between SNRs and SE methods, and there were significant differences within SE methods [$F_{(1.613, 8.064)} = 11.97$, $p = 0.005$] and within SNRs

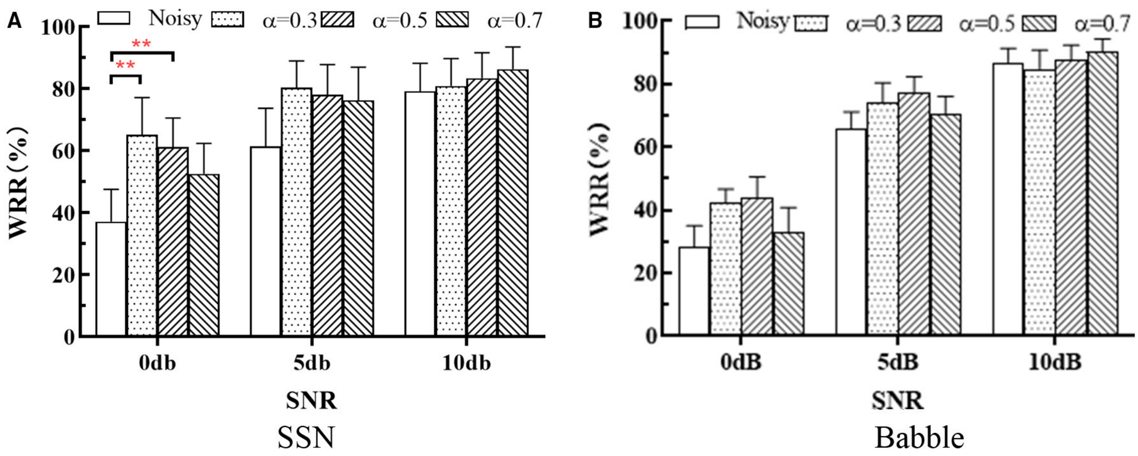


FIGURE 9 | WRR of different systems at different SNRs in **(A)** SSN and **(B)** Babble. The bars show the mean WRR of all the subjects, and error bars indicate the standard deviations. The asterisk indicates a significant level, * indicates $p < 0.0332$, ** indicates $p < 0.0021$, and *** indicates $p < 0.0001$.

[$F_{(1.616,\ 5.803)} = 80.38$, $p < 0.001$]. Tukey's multiple comparisons tests showed that WL-MASK has no significant performance improvement over Noisy in all noise conditions except for 0 dB SSN, where WL-MASK with $\alpha = 0.3$ and 0.5 achieved significant improvement ($p = 0.007$ and 0.004).

DISCUSSIONS

The residual noise and speech distortion in the enhanced signal generally determine the effect of SE. It is well-known that, unlike NH people, CI recipients are more sensitive to noise than to speech distortion in their daily speech perception. The distinct noise- and distortion-perception properties of CI have been investigated and adopted to design the enhancement algorithms (10, 29, 42).

In this study, we developed a deep learning-based SE to systematically investigate how the noise residue and speech distortion could affect the intelligibility of the enhanced speech for patients with CI, and how such noise and distortion sensitivities could be adopted for SE system design. An LSTM-based time-frequency masking system was developed as an SE front end to CI, different loss functions were used to train the system such that different levels of residual noise and speech distortion in the output speech signal could be retained for the investigation. Several objective and subjective experiments were conducted to evaluate the performance of the SE system at different loss conditions.

The objective evaluation with the ECM metric (**Tables 1**, **2**) showed that the MSE loss aiming at minimizing the overall difference between the target and enhanced speech signals usually had suboptimal results in both SSN and Babble noises. By training the LSTM with weighting loss, the SE performance varied at different weighting parameters α. In general, smaller α, which tends to remove more noise components (and induce more speech distortion), was preferred for noisy speech with lower SNRs in both SSN and Babble. Whatever the SNR, there exist some specific α with which the WL-MASK system outperformed the MSE-MASK system. The superiority was more evident in SNN than in Babble. In particular, for SNR of 0 dB, where speech and noise had the same energy, α of 0.4 (in SSN) and 0.3 (in Babble) achieved better results.

Vocoder simulation evaluation with SRT in noise by NH people (**Figure 7**) gave consistent results to that by ECM. Compared with MSE-MASK, WL-MASK had an SRT benefit of about 2.8 dB in SSN (with $\alpha = 0.3$) and about 1.7 dB in Babble (with $\alpha = 0.4$). Compared to Noisy, WL-MASK had an SRT benefit of about 8.6 dB in SSN and about 6.3 dB in Babble.

SRT in noise test with seven CI users showed similar results. Compared with MSE-MASK, WL-MASK had an SRT benefit of about 3.3 dB in SSN and about 1.2 dB in Babble. Compared with Noisy, WL-MASK had an SRT benefit of about 6 dB in SSN and about 2.2 dB in Babble. The SRT gains obtained by WL-MASK are compatible in both NH and CI tests, except for the case with WL-MASK over Noisy in Babble, where the SRT gains drop from 6.3 dB by NHs to 2.2 dB by CIs. In this study, we did find that Babble noise is a relatively more challenging condition for the NN-based SE systems.

Speech recognition tests with 12 CI recipients showed that the proposed WL-MASK had no significant improvement over Noisy in all noise conditions, except for the low SNR (0 dB) SSN case, although the mean word recognition rates did demonstrate that the preferred α was SNR dependent. The lack of significance in high SNRs might be due to the ceiling effect.

The same noise recording used for training and testing could be a limitation of this work in considering the generation of the NN-based SE for real-world applications. Nevertheless, 4- or 10-min recordings are relatively long enough to cover possible variations of a specific noise type. Therefore, this study could reflect the NN-based SEs performance for CI where NNs are well trained with enough noise data, i.e., all noise had been seen by the NNs after training. It is true that real-world settings could be more challenging, which requires more sophisticated NNs and a much larger amount of training noise as well.

Compared to traditional SE methods, the computation load of a DL-based system should be considered in real-world implementations. Furthermore, the WRR test implies that noise conditions-dependent SE systems need to be pre-trained, and real-time noise estimation is required to maximize the benefit from the noise-dependent SE. These drawbacks might restrict the implementation of the proposed WL-MASK SE in clinical CI systems. Nevertheless, this research indicates that it would be promising to further explore the hearing properties of patients with CI and utilize such properties for designing new signal-processing strategies.

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.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Medical Ethics Committee of Shenzhen University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

YK, NZ, and QM contributed to the design and writing of this paper. YK and NZ planned the experimental program, collected data, and wrote the manuscript. All authors agree to submit this version of the article.

FUNDING

This work was jointly supported by Guangdong Key Area R&D Project (No. 2018B030338001), National Natural Science Foundation of China (61771320), Guangdong Basic and Applied Basic Research Foundation (2020A1514325010386), and Science and Technology Program of Guangzhou (202102020944).

ACKNOWLEDGMENTS

The authors are grateful to all the subjects, including people with normal hearing and CI recipients. They thank them for their energetic cooperation in subjective evaluation. At the same time, thanks Huali Zhou for her technical guidance.

REFERENCES

1. Winn MB, Nelson PB. *Cochlear Implants. In Oxford Research Encyclopedia of Linguistics* (2021).
2. Deep NL, Dowling EM, Jethanamest D, Carlson ML. Cochlear implantation: an overview. *JNLS B.* (2019) 80:169–77. doi: 10.1055/s-0038-1669411
3. Ren C, Yang J, Zha D, Lin Y, Liu H, Kong Y, et al. Spoken word recognition in noise in Mandarin-speaking pediatric cochlear implant users. *Int J Pediatr Otorhinolaryngol.* (2018) 113:124–30. doi: 10.1016/j.ijporl.2018.07.039
4. Hast A, Schlücker L, Digeser F, Liebscher T, Hoppe U. Speech perception of elderly cochlear implant users under different noise conditions. *Otol Neurotol.* (2015) 36:1638–43. doi: 10.1097/MAO.0000000000000883
5. Choi CT, Lee YH. A review of stimulating strategies for cochlear implants. In: Umat C, editor. *Cochlear Implant Research Updates*. InTech Open Access Publisher.
6. Fischer T, Schmid C, Kompis M, Mantokoudis G, Caversaccio M, andWimmer W. Effects of temporal fine structure preservation on spatial hearing in bilateral cochlear implant users. *J Acoust Soc Am.* (2021) 150:673–86. doi: 10.1121/10.0005732
7. Meng Q, Zheng N, Li X. Mandarin speech-in-noise and tone recognition using vocoder simulations of the temporal limits encoder for cochlear implants. *J Acoust Soc Am.* (2016) 139:301–10. doi: 10.1121/1.4939707
8. Zhou H, Wang N, Zheng N, Yu G, Meng Q. A new approach for noise suppression in cochlear implants: a single-channel noise reduction algorithm1. *Front Neurosci.* (2020) 14:301. doi: 10.3389/fnins.2020.00301
9. Wang NYH, Wang HLS, Wang TW, Fu SW, Lu X, Wang HM, et al. Improving the intelligibility of speech for simulated electric and acoustic stimulation using fully convolutional neural networks. *IEEE Trans Neural Syst Rehabilitation Eng.* (2020) 29:184–95. doi: 10.1109/TNSRE.2020.3042655
10. Wang D, and Hansen JH. Speech enhancement for cochlear implant recipients. *J Acoust Soc Am.* (2018) 143:2244–54. doi: 10.1121/1.5031112
11. Huber R, Bisitz T, Gerkmann T, Kiessling J, Meister H, Kollmeier B. Comparison of single-microphone noise reduction schemes: can hearing impaired listeners tell the difference? *International Journal of Audiology*. (2017) 57:55–61. doi: 10.1080/14992027.2017.1279758
12. Yang LP, Fu QJ. Spectral subtraction-based speech enhancement for cochlear implant patients in background noise. *J Acoust Soc Am.* (2005) 117:1001–4. doi: 10.1121/1.1852873
13. Loizou PC, Lobo A, Hu Y. Subspace algorithms for noise reduction in cochlear implants. *J Acoust Soc Am.* (2005) 118:2791–3. doi: 10.1121/1.2065847
14. Guevara N, Bozorg-Grayeli A, Bebear JP, Ardoint M, Saa,ï S, Gnansia D, et al. The voice track multiband single-channel modified Wiener-filter noise reduction system for cochlear implants: patients' outcomes and subjective appraisal. *Int J Audiol.* (2016) 55:431–8. doi: 10.3109/14992027.2016.1172267
15. Koning R, Madhu N, andWouters J. Ideal time-frequency masking algorithms lead to different speech intelligibility and quality in normal-hearing and cochlear implant listeners. *IEEE Trans Biomed Eng.* (2014) 62:331–41. doi: 10.1109/TBME.2014.2351854
16. Xu Y, Du J, Dai LR, Lee CH. A regression approach to speech enhancement based on deep neural networks. *IEEE/ACM Trans Audio, Speech, Language Process.* (2014) 23:7–19. doi: 10.1109/TASLP.2014.2364452
17. Wang Y, Narayanan A, Wang D. On training targets for supervised speech separation. *IEEE/ACM Trans Audio, Speech, Language Process.* (2014) 22:1849–58. doi: 10.1109/TASLP.2014.2352935
18. Lai YH, Chen F, Wang SS, Lu X, Tsao Y, Lee CH. A deep denoising autoencoder approach to improving the intelligibility of vocoded speech in cochlear implant simulation. *IEEE Trans Biomed Eng.* (2016) 64:1568–78. doi: 10.1109/TBME.2016.2613960
19. Lai YH, Tsao Y, Lu X, Chen F, Su YT, Chen KC, et al. Deep learning-based noise reduction approach to improve speech intelligibility for cochlear implant recipients. *Ear Hear.* (2018) 39:795–809. doi: 10.1097/AUD.0000000000000537
20. Goehring T, Keshavarzi M, Carlyon RP, Moore BC. Using recurrent neural networks to improve the perception of speech in non-stationary noise by people with cochlear implants. *J Acoust Soc Am.* (2019) 146:705–18. doi: 10.1121/1.5119226
21. Hu Y, and Loizou PC. Environment-specific noise suppression for improved speech intelligibility by cochlear implant users. *J Acoust Soc Am.* (2010) 127:3689–95. doi: 10.1121/1.3365256

22. Mamun N, Khorram S, Hansen JH. Convolutional neural network-based speech enhancement for cochlear implant recipients. *Proc Interspeech.* (2019) 2019:4265–9. doi: 10.21437/Interspeech.2019-1850
23. Bolner F, Goehring T, Monaghan J, Van Dijk B, Wouters J, andBleeck S. (2016. March). Speech enhancement based on neural networks applied to cochlear implant coding strategies. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Shanghai: IEEE. p. 6520-24.doi: 10.1109/ICASSP.2016.7472933
24. Goehring T, Bolner F, Monaghan JJ, Van Dijk B, Zarowski A, and Bleeck S. Speech enhancement based on neural networks improves speech intelligibility in noise for cochlear implant users. *Hear Res.* (2017) 344:183–94. doi: 10.1016/j.heares.2016.11.012
25. Zheng N, Shi Y, Kang Y, Meng Q. (2021). A Noise-Robust signal processing strategy for cochlear implants using neural networks. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech Signal Processing (ICASSP).* IEEE. p. 8343–7. doi: 10.1109/ICASSP39728.2021.9413452
26. Xu Z, Elshamy S, and Fingscheidt T. Using Separate Losses for Speech Noise in Mask-Based Speech Enhancement. In *2020 IEEE International Conference on Acoustics, Speech Signal Processing (ICASSP).* p. 7519–23. IEEE. doi: 10.1109/ICASSP40776.2020.9052968
27. Akbarzadeh S, Lee S, Chen F, Tuan-Tan C. The effect of speech and noise levels on the quality perceived by cochlear implant and normal hearing listeners. *Speech Commun*. (2021) 132:106–13. doi: 10.1016/j.specom.2021.06.001
28. Kressner AA, May T, Dau T. Effect of noise reduction gain errors on simulated cochlear implant speech intelligibility. *Trends Hear.* (2019) 23:2331216519825930. doi: 10.1177/2331216519825930
29. Qazi OU, Van Dijk B, Moonen M, and Wouters J. Understanding the effect of noise on electrical stimulation sequences in cochlear implants and its impact on speech intelligibility. *Hear Res.* (2013) 299:79–87. doi: 10.1016/j.heares.2013.01.018
30. Loizou PC, Kim G. Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions. *IEEE Trans Audio, Speech, Language Process.* (2010) 19:47–56. doi: 10.1109/TASL.2010.2045180
31. Wang D, and Zhang X. THCHS-30: A free Chinese speech corpus. *arXiv preprint arXiv:*1512.01882 (2015).
32. Wong LL, Soli SD, Liu S, Han N, Huang MW. Development of the Mandarin hearing in noise test (MHINT). *Ear Hear.* (2007) 28:70S−4S. doi: 10.1097/AUD.0b013e31803154d0
33. Varga A, andSteeneken HJ. Assessment for automatic speech recognition: II. NOISEX-92: a database and an experiment to study the effect of additive noise on speech recognition systems. *Speech Commun*. (1993) 12:247–51. doi: 10.1016/0167-6393(93) 90095-3
34. Adami A, Burget L, Dupont S, Garudadri H, Grezl F, Hermansky H, et al. Qualcomm-ICSI-OGI features for ASR. In: *Seventh International Conference on Spoken Language Processing.* (2002).
35. Shekar RC, Hansen JH. An evaluation framework for research platforms to advance cochlear implant/hearing aid technology: a case study with CCi-MOBILE. *J Acoust Soc Am.* (2021) 149:229–45. doi: 10.1121/10.0002989
36. Yousefian N, andLoizou PC. Predicting the speech reception threshold of cochlear implant listeners using an envelope-correlation based measure. *J Acoust Soc Am.* (2012) 132:3399–405. doi: 10.1121/1.4754539
37. Vandali AE, Whitford LA, Plant KL, Clark GM. Speech perception as a function of electrical stimulation rate: using the nucleus 24 cochlear implant system. *Ear Hear.* (2000) 21:608–24. doi: 10.1097/00003446-200012000-00008
38. Plomp R, and Mimpen AM. Speech-reception threshold for sentences as a function of age and noise level. *J Acoust Soc Am.* (1979) 66:1333–42. doi: 10.1121/1.383554
39. Meng Q, Yu G, Wan Y, Kong F, Wang X, Zheng N. Effects of Vocoder Processing on Speech Perception in Reverberant Classrooms. In 2018 Asia-Pacific Signal Information Processing Association Annual Summit Conference (APSIPA ASC). Hawaii: IEEE (2018). p. 761–5. doi: 10.23919/APSIPA.2018.8659662
40. Kong F, Wang X, Teng X, Zheng N, Yu G, Meng Q. Reverberant speech recognition with actual cochlear implants: verifying a pulsatile vocoder simulation method. In: *Proceeding of the 23rd International Congress on Acoustics (ICA)*. Aachen: Universitätsbibliothek der RWTH Aachen.
41. Levitt H. Transformed up-down methods in psychoacoustics. *J Acoust Soc Am.* (1971) 49:467–77. doi: 10.1121/1.1912375
42. Mauger SJ, Dawson PW, Hersbach AA. Perceptually optimized gain function for cochlear implant signal-to-noise ratio based noise reduction. *J Acoust Soc Am.* (2012) 131:327–36. doi: 10.1121/1.36 65990

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.

frontiers in Digital Health

ORIGINAL RESEARCH
published: 12 November 2021
doi: 10.3389/fdgth.2021.759723

Involving Children and Teenagers With Bilateral Cochlear Implants in the Design of the BEARS (Both EARS) Virtual Reality Training Suite Improves Personalization

Deborah Vickers[1], Marina Salorio-Corbetto[1], Sandra Driver[2], Christine Rocca[2], Yuli Levtov[3], Kevin Sum[4], Bhavisha Parmar[1], Giorgos Dritsakis[1], Jordi Albanell Flores[4], Dan Jiang[2], Merle Mahon[5], Frances Early[6], Nejra Van Zalk[7] and Lorenzo Picinali[4]*

[1] Sound Laboratory, Cambridge Hearing Group, Clinical Neurosciences, University of Cambridge, Cambridge, United Kingdom, [2] St Thomas' Hearing Implant Centre, Guys and St Thomas' NHS Foundation Trust, London, United Kingdom, [3] Reactify Music, London, United Kingdom, [4] Audio Experience Design, Dyson School of Design Engineering, Imperial College London, London, United Kingdom, [5] Psychology and Language Sciences, Faculty of Brain Sciences, University College London, London, United Kingdom, [6] Department of Respiratory Medicine, Cambridge University Hospital NHS Foundation Trust, Cambridge, United Kingdom, [7] Design Psychology Lab, Dyson School of Design Engineering, Imperial College London, London, United Kingdom

OPEN ACCESS

Edited by:
*Fan-Gang Zeng,
University of California, Irvine,
United States*

Reviewed by:
*Soha Garadat,
The University of Jordan, Jordan
Shekh Md Mahmudul Islam,
University of Dhaka, Bangladesh*

***Correspondence:**
*Deborah Vickers
dav1000@cam.ac.uk*

Specialty section:
*This article was submitted to
Connected Health,
a section of the journal
Frontiers in Digital Health*

Received: *16 August 2021*
Accepted: *06 October 2021*
Published: *12 November 2021*

Citation:
Vickers D, Salorio-Corbetto M, Driver S, Rocca C, Levtov Y, Sum K, Parmar B, Dritsakis G, Albanell Flores J, Jiang D, Mahon M, Early F, Van Zalk N and Picinali L (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

Older children and teenagers with bilateral cochlear implants often have poor spatial hearing because they cannot fuse sounds from the two ears. This deficit jeopardizes speech and language development, education, and social well-being. The lack of protocols for fitting bilateral cochlear implants and resources for spatial-hearing training contribute to these difficulties. Spatial hearing develops with bilateral experience. A large body of research demonstrates that sound localisation can improve with training, underpinned by plasticity-driven changes in the auditory pathways. Generalizing training to non-trained auditory skills is best achieved by using a multi-modal (audio-visual) implementation and multi-domain training tasks (localisation, speech-in-noise, and spatial music). The goal of this work was to develop a package of virtual-reality games (BEARS, Both EARS) to train spatial hearing in young people (8–16 years) with bilateral cochlear implants using an action-research protocol. The action research protocol used formalized cycles for participants to trial aspects of the BEARS suite, reflect on their experiences, and in turn inform changes in the game implementations. This participatory design used the stakeholder participants as co-creators. The cycles for each of the three domains (localisation, spatial speech-in-noise, and spatial music) were customized to focus on the elements that the stakeholder participants considered important. The participants agreed that the final games were appropriate and ready to be used by patients. The main areas of modification were: the variety of immersive scenarios to cover age range and interests, the number of levels of complexity to ensure small improvements were measurable, feedback, and reward schemes to ensure positive reinforcement, and an additional implementation on an iPad for those who had difficulties with the headsets due to age or balance issues. The effectiveness of the BEARS training suite will be evaluated in a large-scale clinical trial to determine if using the games lead to improvements in speech-in-noise, quality of life, perceived benefit, and cost utility. Such interventions allow patients to take control of their own management reducing the reliance on outpatient-based rehabilitation. For young people, a virtual-reality implementation is more engaging than traditional rehabilitation methods, and the participatory design used here has ensured that the BEARS games are relevant.

Keywords: spatial hearing, bilateral, cochlear implant, virtual reality, training, action research, participatory design, children

INTRODUCTION

Advances in mobile technologies have resulted in the development of flexible platforms providing personalized interventions that enable patients to take control of their own health care.

In recent years, the importance of involving patients in the development of clinical interventions has become apparent to maximize engagement, to improve usability and potential success, and more importantly to ensure that the intervention is ultimately relevant for the targeted patient group (1).

Hakobyan et al. (2) highlighted the importance of incorporating patient groups in the design of mobile technology-based interventions to ensure that they meet the needs of the specific population. Participatory design recognizes and involves the key stakeholders in the design and development of the intervention. Without such input, historically, information technologies typically only achieve 40% of population engagement with the intervention, as reported by Hakobyan et al. (2). In spite of this, most interventions still do not involve patients in the development phase. From 18 articles describing the development of mobile technologies reviewed by Hakobyan et al. (2), only four incorporated quality participatory design in the process.

Here we have used participatory design for the development of a virtual reality training suite for improving spatial hearing for 8–16 year-olds with bilateral cochlear implants (CI). This training suite is called BEARS (Both EARS). The development of the BEARS training suite is driven by the fact that normal-hearing listeners use subtle differences in timing and level of sounds reaching each ear to provide directional cues (3–6) that help to separate speech from noise and the ability to attend to a particular speaker (7–9). Although language development, sound localization, speech-in-noise perception and listening effort are better for people with bilateral cochlear implants compared to those with a unilateral implant, these skills remain far below those of normally-hearing children (10–18). Neural plasticity exists for spatial hearing improvements through training (19). Improvements are driven by two processes: (1) cue remapping (the use of new spatial cues to construct a new localisation map, most likely the use of monaural frequency cues in the unprocessed ear), and (2) cue reweighting (the reliance on any unaltered cues while ignoring the altered ones) (20–23). Evidence about these processes is found in several reports. For instance, listeners can adapt to changes in spectral cues, which are critical for judging elevation, as well as solving front-back confusions, even when these are altered (24–26). There has been some discussion about what type of training is most appropriate for feasible delivery and maximization of any benefits. Computer-based training has great potential as it can be delivered anywhere without requiring a face-to-face appointment, and is engaging for most people, especially for children and teenagers. It has been shown that computer-based training can improve speech-in-noise perception for people with cochlear implants (27–30). Green et al. (31) observed an average 2 dB improvement in speech reception thresholds for sentence recognition in babble after 12 h of computer-based training. The nature of the training stimuli has also been explored by previous research. Cai et al. (32) found that audio-visual training is more effective than auditory-only training, and Steadman et al. (23) outlined the importance of auditory-based interaction during the training. A systematic review by Rayes et al. (33) that looked at the effectiveness of training in children with CIs found that the most effective intervention involves the use of multiple modalities or a combination of bottom-up and top-down training tasks. Finally, Whitton et al. (34) explored the transfer and generalization of the acquired training, outlining that audio-motor perceptual training can enhance speech in noise intelligibility by up to 25%.

Based on the evidence summarized above, we designed and carried out the research reported here. The objective was to collaboratively design and develop a training intervention for young people with bilateral cochlear implants, aiming at improving their listening skills, and specifically focusing on the spatial sound cues provided by their cochlear implants.

METHODS

An action research study design was employed (35, 36). Within action research, development and change are achieved through the simultaneous process of taking action whilst conducting research, all informed by user involvement and governed by critical reflection. All stakeholders and researchers are equal members of the research team. Within this study, three phases were employed in the process of development of the BEARS training suite, to ensure that it is appropriate for the intended population.

The stakeholders involved in the process were: young people and young adults using bilateral CIs, family and friends, teachers, engineers and developers, speech and language therapists, music

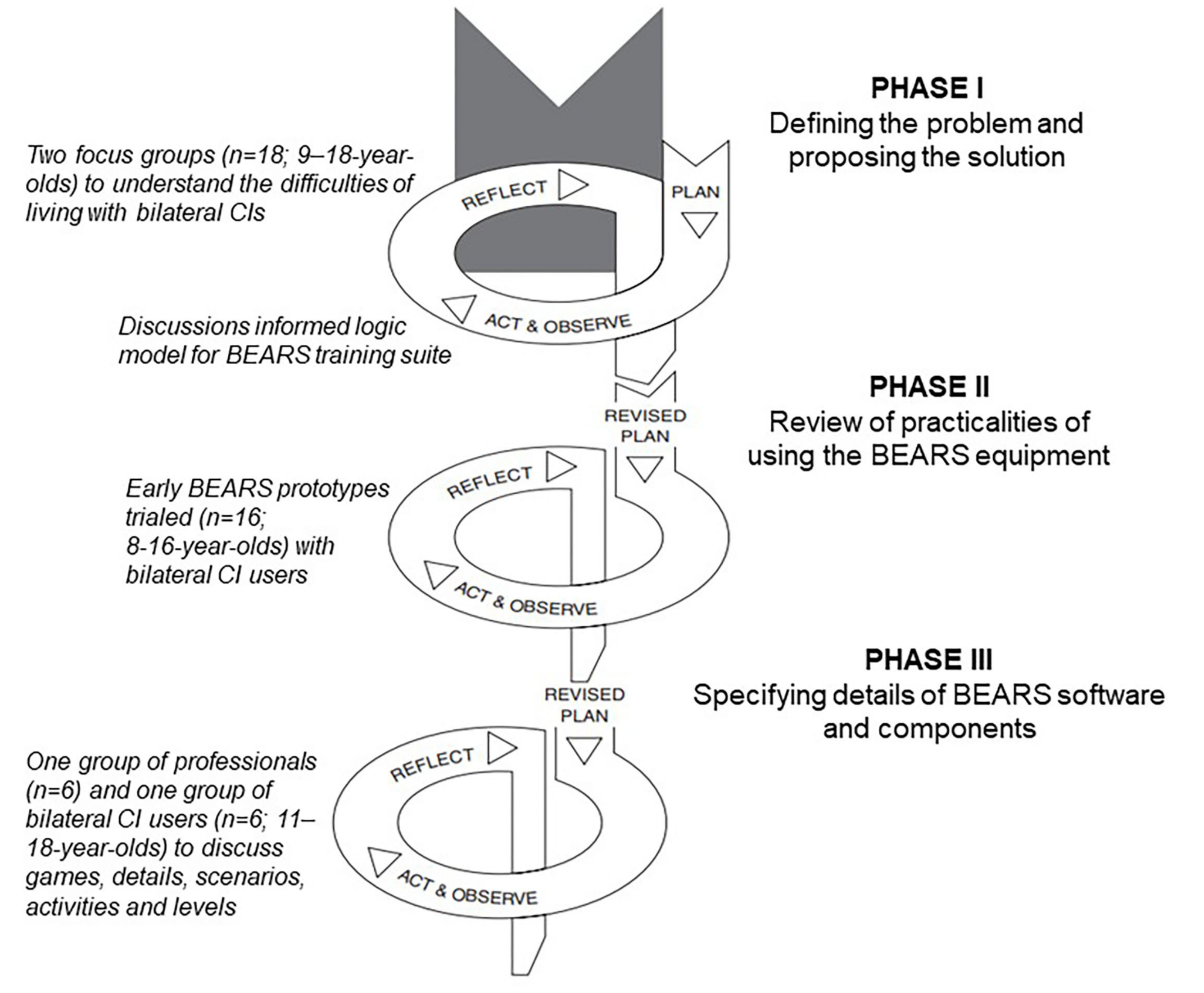


FIGURE 1 | The cycles for the three participatory design phases.

therapists, and audiologists. The process involved multiple focus groups for feedback, reflection, and critical appraisal, each of which was run by an independent facilitator. In advance of each focus group, the goals and topic guides were developed amongst the research team and summary notes were produced for each meeting. Where the meetings were broken down into discussion groups, a note taker and facilitator were assigned to each group. The feedback was reviewed after the focus groups and a plan, consisting of a set of actions, was created and prioritized for the next stages of action and implementation. See **Figure 1** for an outline of the three phases.

Phase I—defining the problem and proposing the solution. Facilitated in-person discussions were conducted with two focus groups ($n = 18$) of CI users aged between 9 and 18 years of age. One group consisted of 10 participants ($n = 6$ male, $n = 4$ female) at a school for deaf and hard-of-hearing (DHH) children, who volunteered to help in response to an advert. The other eight ($n = 3$ male, $n = 5$ female) were recruited from an advert circulated by a charity and attended a meeting held in London.

The goal of these meetings was to understand the difficulties of living with bilateral CIs. The issues were discussed by the two groups and prioritized in terms of overcoming the difficulties discussed to understand what the acceptable interventions were. Based on the discussions, a logic model was created by the research team to underpin the planned multi-modal BEARS virtual reality auditory training intervention.

Phase II—review of the practicalities of using the BEARS equipment. The first implementations of BEARS were trialed by 16 children aged 8–16 years. This group was made up of from older children and teenagers from mainstream and special schools ($n = 6$ male, $n = 5$ female) with an additional 5 younger participants ($n = 2$ male, $n = 3$ female) to help determine if the BEARS training suite would be appropriate for a wider age range of CI users. The goals of the cycles that made up phase II were to understand the practical limitations of BEARS with respect to how frequently training should be conducted, whether the head-mounted display worked well for all listeners, and what sort of age adaptations were required. This phase involved two cycles of action and reflection. In the first cycle, participants were given the BEARS headsets with one game to provide feedback on ease of use. In the second cycle, we held an in-person event with multiple stations for participants to visit to give feedback on different aspects that had been developed and to discuss these with their peers.

Phase III—definition of the details of the BEARS software tools and components. This phase was conducted with two groups. The first were adult professionals working across CI Centres and in the local educational support services. This included teachers of the deaf, speech and language therapists, and audiologists ($n = 6$). They reviewed the tools from a clinical and educational perspective. The second group were bilateral CI users (aged 11–18 years; $n = 3$ male, $n = 3$ female).

Due to the COVID pandemic, the groups in phase III were conducted virtually using Microsoft Teams. For the teachers and clinicians, the entire group discussed the different software tools together. For the bilateral CI group, pre-allocated breakout rooms were set up to discuss the different elements of the applications. Each group was made up of two bilateral CI users who were matched to work well together (based on teacher opinion), one facilitator and one notetaker.

RESULTS

Phase I—Defining the Problem and Proposing the Solution

The CI focus groups were asked to discuss freely about their listening difficulties that they faced in everyday life, what they thought might cause or have an effect on these, and any activities (interventions) that they had found helpful. The notes on the difficulties were fed back to the groups and they grouped points together and set priorities for the project.

The combined prioritized statements from the two groups were:

1) Everyday listening requires "extra effort" which makes communication "tiring" and ultimately "challenging." These problems were particularly reported with respect to noisy environments.
2) It can be difficult to "combine" sounds from two CIs because the two ears often do not sound the same. Sequential implantees (>1 year between two implants) reported that second CI could seem "annoying," "distracting," and "lop-sided," which in some cases resulted in non-use.
3) Listening training can be helpful but current rehabilitation techniques are not always engaging, and relevant, and computer-based approaches may be more motivating.

Based on the feedback, the research team reviewed the information, discussed ideas for addressing the issues and decided to develop the BEARS training suite to address the issues. The research team created goals that the BEARS training suite needed to meet. These goals were:

1) To include age-appropriate and engaging listening games.
2) To use multiple training tasks (speech in noise, localisation, and music) to optimize effectiveness
3) To use visual cues to support engagement
4) To implement in virtual reality to enhance the gaming aspect
5) To use gaming head mounted display headsets for flexibility and usability
6) To develop the game soundscapes using the established 3D Tune-In toolbox (37).

The research team developed the following logic model to underpin the BEARS intervention. The logic model was developed following the UKs Medical Research Council and National Institute of Health and Care Excellence (38) advice that the development of new interventions should outline the mechanism of change for an intervention. A logic model outlines the theory of how an intervention will lead to the desired outcomes:

People with bilateral CIs struggle to understand speech in noise, making communication tiring and challenging, having a negative effect on social integration and well-being. The problem underpinning this is that bilateral CI users are not effectively using the cues from both ears to maximise spatial hearing. The use of the BEARS training suite with audio-visual information and multiple listening modalities (speech in noise, localisation, and music) should improve spatial hearing, speech-in-noise perception and ease of listening. The change mechanisms for these effects will be plasticity-driven processes enhancing learning and maximising spatial listening skills performance. Factors supporting these training-induced effects are audio-visual integration, multimodal stimulation, and cognitive engagement that drives generalisation to other auditory stimuli. These changes will lead to better communication and social engagement skills (with reduced fear of embarrassment) which in turn will improve quality of life through healthier social and emotional-regulation development. As communication becomes easier, self-confidence continues to build up, improving development in multiple areas of life such as building relationships and education.

Motivation, engagement, rewards, time commitment, and developmental processes will act as modulators. Behaviour changes will also increase uptake and usability.

See **Figure 2** for schematic of logic model.

The target population is older children and teenagers (aged 11–16 years) with at least 12 months of bilateral CI experience rather than newly implanted bilateral users. This therefore precludes influencing post-activation plasticity-driven processes for developing spatial listening skills. Plasticity is assumed to be ongoing even after 12 months of usage, as research indicates that the associations between learning and cognitive control performance only emerges with age and appears most prominent for late adolescents (39). Studying bilateral CI users 12 months after implantation allows for the effects of habituation, as they would have grown accustomed to wearing and interacting with their implants, thereby minimizing the confounding effects

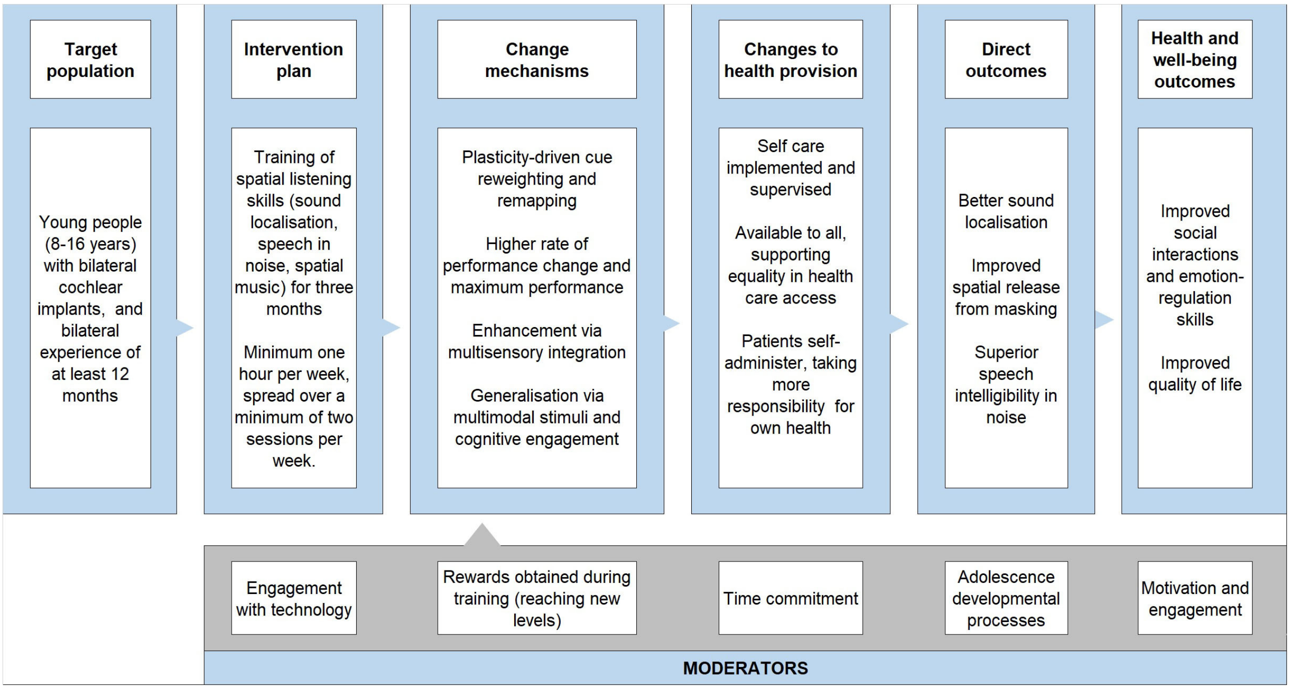


FIGURE 2 | The logic model to explain the BEARS training suite rationale.

of novelty and/or any issues arising from initial use. Also, the numbers of older children and teenagers who receive bilateral cochlear implants are not large. The patient groups who reported difficulties had typically received their implants as young children, so this is the population who are focussed on initially but the work can be applied to other groups.

Phase II—Practicalities of Using BEARS

The feedback from the first cycle of phase II indicated for 67% (10 out of 15) of the CI users that the use of the head-mounted display with headphones was practical. The remaining 33% of the CI users found the systems either too bulky (three people) or had balance issues because of eyes being covered with the virtual reality goggles (two people). The comments about bulkiness came from young people aged 8–10 years who were younger and smaller than the majority of the participants.

All participants reported that they found the games enjoyable and the instructions were straightforward. They requested that the differences between levels were made smaller. The initial game that was trialed was aimed at improving localisation ability, and its effectiveness had already been validated in previous work (23). In the game, the goal was to identify the position of an audio-emitting alien ball and shoot at it. In the lower levels there were audio and visual cues but in the higher levels there were just audio cues and they found the move to sound alone too sudden. They also requested more rewards and the ability to play the game online with friends. Parents liked that the young people were able to play the games independently.

The research team reviewed the feedback and implemented an option for an iPad-based game to avoid wearing bulky equipment or covering eyes. A greater number of levels were introduced so that the visual cues were faded out more gradually and positive feedback was added in.

It was not possible to make the game interactive with other online players or to allow an online scoresheet comparison because the BEARS training suite will be evaluated in a clinical trial and this sort of engagement with others might contaminate the trial. However, the wish for a BEARS users' community has been recorded for future implementations.

In the second cycle of phase II, the participants attended a group event and visited the aforementioned stations to provide ideas for how to expand and implement new games that would be engaging for training speech-in-noise perception and music. For speech-in-noise training, the groups suggested having café scenarios where the user has to listen out for different food/drink orders or key words given by customers presented at different locations. The complexity of the game would change based on the range of locations or background noise. For music training, they recommended a game where one can make music by drawing in different musical instruments and adjust the sound of the instruments. They also liked the idea of identifying and discriminating different songs. The final idea that they wanted to incorporate was to identify when a specific instrument was present and the complexity would be built up by adding in other musical instruments and moving the location which could also be identified.

The group had ideas for making the localisation training game more appropriate for the younger group by catching butterflies or popping bubbles.

The research team were able to incorporate all of these suggestions into initial game prototypes for the three training tasks. These took the form of:

1. A localisation training game, involving identifying the position of an audio-emitting alien ball and shooting it. The ball is visible only in the early levels, and gradually disappears as soon as the player advances.
2. A spatial speech-in-noise recognition training game, involving listening to and identifying customer food/drink orders and serving the correct item.
3. A music training game, involving completing a number of music-oriented challenges to progress through an escape room.

The games were then reviewed in phase III. It was decided that having realistic lip-syncing for the speech-in-noise games was not practical and that the characterisation would be more appropriate as cartoon characters than realistic/natural appearance.

Phase III—Definition of the Details of the BEARS Software Tools and Components

For phase III, video demonstrations of the games were prepared to show to the participants in the online groups. All participants had the opportunity to review the videos in advance (see **Figure 3**) for images of the 3 different games themes. For the first group, the clinicians gave feedback on how to maximize speech and language development with the games and gave ideas for vocabulary to use in café scenes to provide minimal contrasts for example "peas" vs. "cheese." They also suggested changing the carrier phrases to ensure that the contrastive word did not always fall at the end. They also recommended a shop as an alternative to a café.

The young participants reviewed the appearance of the avatars, provided their opinions, and voted on the preferred style to use. They also recommended adding in a wildlife park scenario to increase engagement and interest for a wider age range of listeners. For the music games, they recommended additional artists that would be more appropriate to include.

Note that in the next iteration, phase III will have an additional face-to-face cycle to verify the final version of the games and the implementation using the latest hardware (the initial equipment review was conducted three years ago). We will also verify that the 8–10 year-olds are happy with the modifications that were implemented for their age group. At that stage the BEARS training suite will be finalized for use in a clinical trial.

DISCUSSION

We have outlined the formalized participatory design approach that was used to develop the BEARS training suite, based on multiple action research cycles.

In spite of their potential to maximize patients' adoption of new technologies for delivering training, formalized participatory designs have not been extensively reported in the field of hearing research. There are only a few articles reporting research with young people (older children and teenagers). For instance, Hallewell et al. (40), already mentioned, and Hanssen and Dahl (41), who used participatory design in the development of an interactive sound environment simulator to facilitate communication between audiologists and patients. The authors promoted the value of participatory design to maximize effectiveness of complex interventions that affect both patients and practitioners. Ferguson et al. (42) worked closely with clinicians and hearing-aid users to develop the content and delivery approach for a series of video tutorials that support first-time hearing-aid users. Frost et al. (43) developed an auditory-cognitive training application which was intended to delay the onset of dementia. Their stakeholders were clinicians from audiology and cognitive disorders specialties. All of these reports highlight the importance and value of involving patients and clinicians together to maximize the effectiveness of new hearing healthcare interventions.

BEARS is considered to be a complex intervention because of the multiple training elements within the package and the possibility for tailoring the intervention for individual needs, for instance, those arising from factors such as age.

As a complex intervention it is recommended by the MRC-NIHR (38), guidance that the development phase should have a clearly stated outcome that the intervention should achieve. The outcomes were determined and prioritized by young bilateral CI users to be an improvement in speech-in-noise perception such that not only lead to improved accuracy but also that the level of listening effort is reduced. Based on theoretical and clinical knowledge, a logic model was developed to define

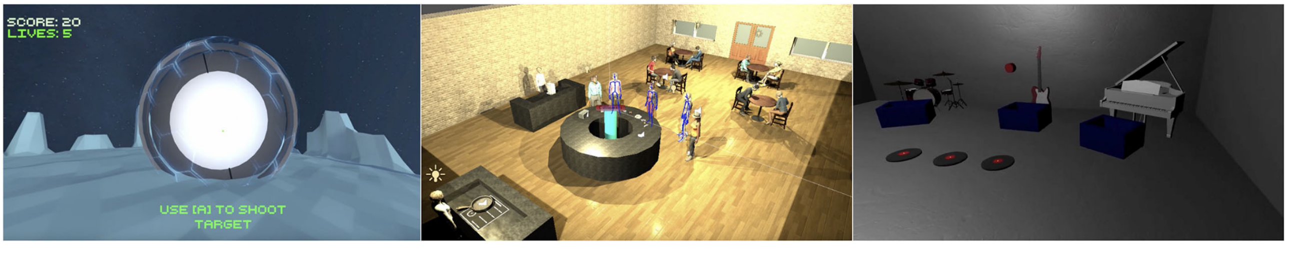


FIGURE 3 | (Left) Localisation game. (Middle) Speech-in-noise game. (Right) Music training game.

the likely mechanism for change for the use of the BEARS training suite.

Many stakeholders were involved in this development phase and the approach for running the different focus groups had to be adapted to be appropriate for the participants themselves. We separated out the professionals from the young bilateral CI users. In addition, we separated out the young bilateral CI users into those at primary school (8–11) and those at secondary school (11–16). The purpose of the separation of the different groups was to ensure that all participants felt comfortable to engage and contribute to the discussions. Each of the primary-school-aged children were accompanied by a caregiver which also changed the dynamics of the focus group. It is important that the facilitation of such groups takes into account the age of the participants and group dynamics (44), and that all participants are given the opportunity to contribute.

One factor that should be borne in mind is that the participants in our focus groups may not represent all backgrounds because they were self-selecting. It is possible that their issues may not be representative of the entire population. However, as difficulties with communication in noisy environments, listening effort and mismatch between sounds in the two ears are well reported in the literature they form reasonable goals for the BEARS training suite. It is possible that the focus groups were made up of young people with a particular interest in virtual reality training games. The future clinical trial to evaluate the effectiveness of the BEARS training suite will enroll young people from a wide range of backgrounds and interests to fully understand the value of the intervention.

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/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Cambridge, Psychology Panel. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

DV, LP, DJ, and MM: conceptualization and methodology. DV: draft article writing. LP, MS-C, and MM: review and editing. MS-C, SD, CR, GD, and DJ: supported focus groups. SD, CR, and DJ: ran recruitment campaigns for both patients and clinicians. GD: project managed the workshops and participant groups. LP, YL, KS, and JA: engineering team who implemented changes and demonstrated concepts at workshops. BP: facilitated workshops and created prioritized action plans. MM, FE, NV, and DV: built the logic model through discussion with clinicians and exploration of the literature. All members of the team were involved in intellectual discussions following each phase and cycle of the project.

FUNDING

This project was funded by NIHR PGfAR (201608). DV and MS-C salaries were funded by a Medical Research Council Senior Fellowship in Hearing (MR/2002537/1).

ACKNOWLEDGMENTS

We acknowledge support from Amanda Stranks Patient, Public, Involvement Lead from the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014).

REFERENCES

1. Hallewell M, Salanitri D, D'Cruz M, Cobb S, Picinali L, Frost E, et al. Dartanan: prototype evaluations of a serious game to engage children in the calibration of their hearing aid functionalities. *J Rehabil Assist Technol Eng.* (2021) 8:20556683211021527. doi: 10.1177/20556683211021527
2. Hakobyan L, Lumsden, J, O'Sullivan D. Participatory design: How to engage older adults in participatory design activities. *Inter J Mobile Human Comp Inter.* (2015) 7:78–92.
3. Blauert J. *Spatial Hearing: The Psychophysics of Human Sound Localization.* Cambridge, Mass: MIT Press (1997). doi: 10.7551/mitpress/6391.001.0001
4. Kobler S, Rosenhall U. Horizontal localization and speech intelligibility with bilateral and unilateral hearing aid amplification. *Int J Audiol.* (2002) 41:395–400. doi: 10.3109/14992020209090416
5. McAlpine D, Jiang D, Palmer AR. A neural code for low-frequency sound localization in mammals. *Nat Neurosci.* (2001) 4:396–401. doi: 10.1038/86049
6. Noble W, Gatehouse S. Effects of bilateral versus unilateral hearing aid fitting on abilities measured by the speech, spatial, and qualities of hearing scale (SSQ). *Int J Audiol.* (2006) 45:172–81. doi: 10.1080/14992020500376933
7. Bronkhorst AW. The cocktail-party problem revisited: early processing and selection of multi-talker speech. *Attent Percept Psychophys.* (2015) 77:1465–87. doi: 10.3758/s13414-015-0882-9
8. Nabelek AK, Pickett JM. Monaural and binaural speech perception through hearing aids under noise and reverberation with normal and hearing-impaired listeners. *J Speech Hear Res.* (1974) 17:724–39. doi: 10.1044/jshr.1704.724
9. McArdle RA, Killion M, Mennite MA, Chisolm TH. Are two ears not better than one? *J Am Acad Audiol.* (2012) 23:171–81. doi: 10.3766/jaaa,.23.3.4
10. Sarant J, Harris D, Bennet L, Bant S. Bilateral versus unilateral cochlear implants in children: a study of spoken language outcomes. *Ear Hear.* (2014) 35:396–409. doi: 10.1097/AUD.0000000000000022
11. Boons T, Brokx JP, Frijns JH, Peeraer L, Philips B, Vermeulen A, et al. Effect of pediatric bilateral cochlear implantation on language development. *Arch Pediatr Adolesc Med.* (2012) 166:28–34. doi: 10.1001/archpediatrics.2011.748
12. Sparreboom M, Langereis MC, Snik AF, Mylanus EA. Long-term outcomes on spatial hearing, speech recognition and receptive vocabulary after sequential bilateral cochlear implantation in children. *Res Dev Disabil.* (2015) 36C:328–37. doi: 10.1016/j.ridd.2014.10.030
13. Lovett RE, Vickers DA, Summerfield AQ. Bilateral cochlear implantation for hearing-impaired children: criterion of candidacy derived from an observational study. *Ear Hear.* (2015) 36:14–23. doi: 10.1097/AUD.0000000000000087
14. Litovsky RY. Review of recent work on spatial hearing skills in children with bilateral cochlear implants. *Cochlear Implants Int.* (2011) 12(Supp. 1):S30–4. doi: 10.1179/146701011X13001035752372

15. Grieco-Calub TM, Litovsky RY. Sound localization skills in children who use bilateral cochlear implants and in children with normal acoustic hearing. *Ear Hear.* (2010) 31:645. doi: 10.1097/AUD.0b013e3181e50a1d
16. Zheng Y, Godar SP, Litovsky RY. Development of sound localization strategies in children with bilateral cochlear implants. *PLoS ONE.* (2015) 10:e0135790. doi: 10.1371/journal.pone.0135790
17. Hughes KC, Galvin KL. Measuring listening effort expended by adolescents and young adults with unilateral or bilateral cochlear implants or normal hearing. *Cochlear Implants Int.* (2013) 14:121–9. doi: 10.1179/1754762812Y.0000000009
18. Lammers MJ, van der Heijden GJ, Pourier VE, Grolman W. Bilateral cochlear implantation in children: a systematic review and best-evidence synthesis. *Laryngoscope.* (2014) 124:1694–9. doi: 10.1002/lary.24582
19. Firszt JB, Reeder RM, Dwyer NY, Burton H, Holden LK. Training results in individuals with unilateral severe to profound hearing loss. *Hear Res.* (2015) 319:48–55. doi: 10.1016/j.heares.2014.11.005
20. Keating P, Rosenior-Patten O, Dahmen JC, Bell O, King A. Behavioral training promotes multiple adaptive processes following acute hearing loss. *Elife.* (2016) 5:e12264. doi: 10.7554/eLife.12264
21. Keating P, Johannes C, Dahmen A, King J. Context-specific reweighting of auditory spatial cues following altered experience during development. *Curr Biol.* (2013) 23:1291–9. doi: 10.1016/j.cub.2013.05.045
22. Stitt P, Picinali L, Katz B. Auditory accommodation to poorly matched non-individual spectral localization cues through active learning. *Sci Rep.* (2019) 9:1–14. doi: 10.1038/s41598-018-37873-0
23. Steadman M, Kim C, Lestang JH, Goodman D, Picinali L. Short-term effects of sound localization training in virtual reality. *Sci Rep.* (2019) 9:1–17. doi: 10.1038/s41598-019-54811-w
24. Carlile S. The plastic ear and perceptual relearning in auditory spatial perception. *Front Neurosci.* (2014) 8:237. doi: 10.3389/fnins.2014.00237
25. Trapeau R, Aubrais V, Schönwiesner M. Fast and persistent adaptation to new spectral cues for sound localization suggests a many-to-one mapping mechanism. *J Acoust Soc Am.* (2016) 140:879–90. doi: 10.1121/1.4960568
26. Watson CJ, Carlile S, Kelly H, Balachandar K. Generalization of auditory accommodation to altered spectral cues. *Sci Rep.* (2017) 7:11588. doi: 10.1038/s41598-017-11981-9
27. Ingvalson EM, Lee B, Fiebig P, Wong PC. The effects of short-term computerized speech-in-noise training on postlingually deafened adult cochlear implant recipients. *J Speech Lang Hear Res.* (2013) 56:81–8. doi: 10.1044/1092-4388(2012/11-0291)
28. Oba SI, Fu QJ, Galvin JJ III. Digit training in noise can improve cochlear implant users' speech understanding in noise. *Ear Hear.* (2011) 32:573–81. doi: 10.1097/AUD.0b013e31820fc821
29. Zhang T, Dorman MF, Fu QJ, Spahr AJ. Auditory training in patients with unilateral cochlear implant and contralateral acoustic stimulation. *Ear Hear.* (2012) 33:e70–9. doi: 10.1097/AUD.0b013e318259e5dd
30. Casserly ED, Barney EC. Auditory training with multiple talkers and passage-based semantic cohesion. *J Speech Lang Hear Res.* (2017) 60:159–71. doi: 10.1044/2016_JSLHR-H-15-0357
31. Green T, Faulkner A, Rosen S. Computer-based connected-text training of speech-in-noise perception for cochlear implant users. *Trends Hear.* (2019) 23:2331216519843878. doi: 10.1177/2331216519843878
32. Cai Y, Chen G, Zhong X, Yu G, Mo H, Jiang J, et al. Influence of audiovisual training on horizontal sound localization and its related erp response. *Front Hum Neurosci.* (2018) 12:423. doi: 10.3389/fnhum.2018.00423
33. Rayes H, Al-Malky G, Vickers D. Systematic review of auditory training in pediatric cochlear implant recipients. *J Speech Lang Hear Res.* (2019) 21:1574–93. doi: 10.1044/2019_JSLHR-H-18-0252
34. Whitton JP, Hancock KE, Shannon JM, Polley DB. Audiomotor perceptual training enhances speech intelligibility in background noise. *Curr Biol.* (2017) 27:3237–47. doi: 10.1016/j.cub.2017.09.014
35. Lewin K. Action research and minority problems. *J Soc Issues.* (1946) 2:34–46. doi: 10.1111/j.1540-4560.1946.tb02295.x
36. Reason P, Heron J. Co-operative inquiry. In: *Rethinking in Psychology.* Harre R, Smith J, Van Langenhove L, editors. London: Sage (1995). p. 122–42. doi: 10.4135/9781446221792.n9
37. Cuevas-Rodríguez M, Picinali L, González-Toledo D, Garre C, de la Rubia-Cuestas E, Molina-Tanco L, et al. 3D Tune-in toolkit: an open-source library for real-time binaural spatialisation. *PLoS ONE.* (2019) 14:e0211899. doi: 10.1371/journal.pone.0211899
38. MRC-NIHR (36). *Developing and Evaluating Complex Interventions.* Available online at: http://www.mrc.ac.uk/complexinterventionsguidance (accessed July 12, 2021).
39. Insel C, Charifson M, Somerville LH. Neurodevelopmental shifts in learned value transfer on cognitive control during adolescence. *Dev Cogn Neurosci.* (2019) 40:100730. doi: 10.1016/j.dcn.2019.100730
40. Hallewell M, Patel H, Salanitri D, Picinali L, Cobb S, Velzen J, et al. Play and tune: user feedback in the development of a serious game for optimizing hearing aid orientation. *Ergon Des.* (2021) 29:14–24. doi: 10.1177/1064804619899558
41. Hanssen GK, Dahl Y. A participatory design approach to develop an interactive sound environment simulator. *Am J Audiol.* (2016) 25:268–71. doi: 10.1044/2016_AJA-16-0005
42. Ferguson M, Leighton P, Brandreth M, Wharrad H. Development of a multimedia educational programme for first-time hearing aid users: a participatory design. *Int J Audiol.* (2018) 57:600–9. doi: 10.1080/14992027.2018.1457803
43. Frost E, Porat T, Malhotra P, Picinali L. (2020). A novel auditory-cognitive training app for delaying or preventing the onset of dementia: participatory design with stakeholders. *JMIR Hum Fact.* (1980) 7:e19880. doi: 10.2196/19880
44. Adler K, Salanterä S, Zumstein-Shaha M. Focus group interviews in child, youth, and parent research: an integrative literature review. *Int J Qual Res.* (2019) 18:1–15. doi: 10.1177/1609406919887274

Author Disclaimer: The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

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.

frontiers in Digital Health

ORIGINAL RESEARCH
published: 25 November 2021
doi: 10.3389/fdgth.2021.740323

Alternative Pathways for Hearing Care May Address Disparities in Access

Amber Willink[1,2], Lama Assi[2], Carrie Nieman[2,3], Catherine McMahon[4], Frank R. Lin[2,3] and Nicholas S. Reed[2,5]*

[1] Menzies Centre for Health Policy, University of Sydney, Sydney, NSW, Australia, [2] Cochlear Center for Hearing and Public Health, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, United States, [3] Department of Otolaryngology, Johns Hopkins School of Medicine, Baltimore, MD, United States, [4] Department of Linguistics, Macquarie University, Sydney, NSW, Australia, [5] Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, United States

OPEN ACCESS

Edited by:
Fan-Gang Zeng, University of California, Irvine, United States

Reviewed by:
Laura Gaeta, California State University, Sacramento, United States
Haotian Liu, Sichuan University, China
Qinglin Meng, South China University of Technology, China

***Correspondence:**
Amber Willink
amber.willink@sydney.edu.au

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: 12 July 2021
Accepted: 19 October 2021
Published: 25 November 2021

Citation:
Willink A, Assi L, Nieman C, McMahon C, Lin FR and Reed NS (2021) Alternative Pathways for Hearing Care May Address Disparities in Access. Front. Digit. Health 3:740323. doi: 10.3389/fdgth.2021.740323

Background/Objectives: Low-uptake of hearing aids among older adults has long dogged the hearing care system in the U.S. and other countries. The introduction of over-the-counter hearing aids is set to disrupt the predominantly high-cost, specialty clinic-based delivery model of hearing care with the hope of increasing accessibility and affordability of hearing care. However, the current model of hearing care delivery may not be reaching everyone with hearing loss who have yet to use hearing aids. In this study, we examine the group of people who do not use hearing aids and describe their characteristics and health care utilization patterns. We also consider what other healthcare pathways may be utilized to increase access to hearing treatment.

Design: Cross-sectional, the 2017 Medicare Current Beneficiary Survey.

Setting: Non-institutionalized adults enrolled in Medicare, the U.S. public health insurance program for older adults (65 years and older) and those with qualifying medical conditions and disabilities.

Participants: A nationally representative sample of 7,361 Medicare beneficiaries with self-reported trouble hearing and/or hearing aid use.

Measurements: Survey-weighted proportions described the population characteristics and health care utilization of those with hearing loss by hearing aid use, and the characteristics of those with untreated hearing loss by health care service type utilized.

Results: Women, racial/ethnic minorities, and low-income Medicare beneficiaries with self-reported hearing trouble were less likely to report using hearing aids than their peers. Among those who do not use hearing aids, the most commonly used health care services were obtaining prescription drugs (64%) and seeing a medical provider (50%). Only 20% did not access either service in the past year. These individuals were more likely to be young and to have higher educational attainment and income.

Conclusion: Alternative models of care delivered through pharmacies and general medical practices may facilitate access to currently underserved populations as they are particularly high touch-points for Medicare beneficiaries with untreated hearing trouble. As care needs will vary across a spectrum of hearing loss, alternative models of hearing care should look to complement not substitute for existing access pathways to hearing care.

Keywords: hearing impairment, hearing aid, direct to consumer, health services utilization, older adults

IMPACT STATEMENT

We certify that this work is novel and provides important contributions to the literature by highlighting the disparities in access to hearing care and alternative pathways of providing hearing care that could address existing disparities among Medicare beneficiaries.

KEY POINTS

- Women, racial/ethnic minorities, and low-income Medicare beneficiaries are less likely to use hearing aids.
- They more commonly visit pharmacies and general medical practices for health care.
  Why does this matter?
  Less that one in five Medicare beneficiaries with hearing loss use hearing aids. Delivering hearing care through alternative pathways may improve hearing care access.

INTRODUCTION

In 2020, 44.11 million adults in the U.S. were estimated to have hearing loss; with the aging of the population, the number is expected to increase to 73.50 millions in the next 40 years (1). Hearing loss, the enact of encoding peripheral environmental auditory information for central decoding in the brain, can have a great toll on well-being and communication (2), and may lead to poor quality of life (3), and disability (4). It has also been associated with negative health consequences such as increased risk of falls (5), and cognitive decline (6), and has been recognized as a modifiable risk factor for dementia (7).

Hearing aids have been shown to improve hearing-related quality of life, listening ability, communication, and social and emotional function (8, 9). Early observational studies suggest that hearing aids may improve cognitive functions by preventing auditory deprivation which can result in insufficient cognitive stimulation (10, 11). Yet, despite the association of hearing loss with negative health consequences, and potential benefits of hearing aid use, <15% of adults with hearing loss aged 50 and over in the U.S. report using hearing aids (12).

Barriers to hearing aid use include high cost and lack of, or inadequate, insurance coverage, perceived stigma by others, significant resources required to navigate current processes, including transportation, mobility, and know-how as well as a lack of clear recommendations or guidance by primary care providers (13, 14).

Current hearing care delivery in the U.S. follows a medical model of specialty clinic-based care, primarily through audiologists and otolaryngologists, which may be costly and time-consuming; even those who have the time and money needed may find it frustrating (15). The current model is grounded in dispensing hearing aids through licensed individuals which requires multiple visits to a hearing aid dispenser or audiologist for identification of hearing needs, customization of the product, and continual maintenance and fine-tuning. It is also common for the services of the professional to be bundled into the sale of the hearing aid as a markup. When this hearing delivery model was developed in the 1970's, hearing aids were extremely complicated and potentially produced dangerous noise levels. This required in-person fitting and tuning by trained professionals. Advances in technology have increasingly allowed for integration and easy adjustment via smartphone, opening the possibilities for alternative models of care.

Alternative approaches to the specialty clinic-based hearing care model tackle some of the shortcomings of the typical clinic-based model, and include community-delivered hearing care, mobile health applications, tele-audiology, pharmacies, retail clinics such as Costco and Walgreens, and involve primary care providers in hearing care (16). Community-delivered hearing care models can include community health workers, peer educators, community health aides, among other trained paraprofessionals who can provide education on hearing loss, basic aural rehabilitation, as well as fitting and orientation to OTC devices (17–19). Some large retail clinics in the United States, such as Costco and Sam's Club, have integrated hearing aid centers into their stores. This model generally recreates best-practice hearing aid delivery models used in private clinics but increases accessibility by putting the clinic where customers already are shopping and increases affordability by leveraging buying power from the large corporations.

Recognizing the importance of treating hearing loss and the presence of barriers to hearing care access at the national level, the President's Council of Advisors on Science and Technology and the National Academies of Sciences, Engineering, and Medicine released reports with recommendations to improve the accessibility of hearing care for older adults (16, 20), and in 2017, the Over-the-Counter (OTC) Hearing Aid Act law was passed with the aim of making hearing aids more accessible and affordable to those with mild or moderate hearing loss (21).

The OTC Hearing Aid Act required the Food and Drug Administration (FDA) to develop regulations by August 2020 for the sale of over-the-counter hearing aids to treat mild to moderate hearing loss. The FDA has missed this statutory deadline but are scheduled to release these recommendations by the end of the year (22).

By allowing the sale of OTC hearing aids that would be regulated to ensure the safety and efficacy of these devices, people with self-perceived hearing loss will be able to purchase hearing aids without assessment or counseling from a hearing care professional. However, ensuring proper access to hearing care services is essential to promote optimal hearing loss management and maximal benefit from hearing aids for those who need them. In a randomized-controlled trial, people who self-selected their own pre-programmed hearing aids via an OTC service-delivery model, compared to those who received hearing aids via an audiologist-based service-delivery model, were less likely to be satisfied with their hearing aid or purchase one after the study (23).

From a public health perspective, it is unclear to what extent alternative models currently available reach people with hearing loss who have yet to use hearing aids. People with untreated hearing loss have different health care utilization patterns and costs compared to those without hearing loss, (24) and compared to those with hearing loss who use hearing aids (25). Those with untreated hearing loss are more likely to visit emergency departments (24), report unmet health care needs (26), and not have a usual source of care (14). Understanding how people with untreated hearing loss access the general health care system and what characteristics set them apart from those who already use hearing aids is fundamental to understanding how we can reach those Medicare beneficiaries with hearing loss who are not served by the current model of care.

In this study, we identify who the current model of hearing care is serving and how those individuals differ from Medicare beneficiaries with untreated hearing trouble. We then explore the other health care service patterns of the population with untreated hearing trouble and describe the populations accessing the most common health care services by sociodemographic characteristics and health status, to demonstrate the population that could be potentially reached via alternative delivery models of hearing care. We focus in on medical providers and pharmacies as those services have the highest utilization across the Medicare population.

METHODS

Study Sample

We used the 2017 Medicare Current Beneficiary Survey and Cost Supplement file, a nationally representative survey of Medicare beneficiaries linked to administrative claims data. Medicare is the publicly-funded health insurance program in the United States for adults aged 65 years and older and those under 65 years who qualify based on medical condition or disability. For this study, 7,361 Medicare beneficiaries who self-reported a little or a lot of trouble hearing or reported using hearing aids were included in the analytic sample.

Measures

The analysis was separated into two sub-populations: untreated functional hearing impairment (those who self-reported a little or a lot of trouble hearing but no hearing aid use), and those who did report using a hearing aid.

The primary outcome of interest was healthcare utilization by service type, including inpatient, outpatient (e.g., hospital outpatient department or clinic visits), medical provider (e.g., physician, primary care, or allied health), prescription drugs, home health, and skilled nursing facilities services. These utilization variables were derived from both survey report and administrative claims data and through an adjudication process developed by the MCBS administrators (27). As the objective of this analysis is to assess in-person utilization of services as possible avenues for hearing care treatment, we have refined the measure of medical provider to be limited to those most likely seen in a primary care setting, including medical doctors (excluding specialists), nurse practitioners, and physicians' assistants. This measure was refined as delivering hearing care through a primary care service, rather than through specialists, is more practical and within the scope of general practice. We have also refined the prescription drug category to identify those who visit a pharmacy to receive their prescriptions. If a Medicare beneficiary reported often receiving their prescriptions in the mail or *via* the internet, they were counted as not having a prescription drug touch point as we are trying to identify in-person visits that might lend themselves to receiving other healthcare services.

Medicare beneficiaries who reported untreated functional hearing loss were described according to the health care service types utilized. Population characteristics used to describe the groups included sociodemographic characteristics, health and functional status, and access to online information. Sociodemographic characteristics included age, gender, race/ethnicity, educational attainment, income relative to federal poverty level, living arrangement, urban/rural status, and supplemental insurance coverage. Supplemental insurance coverage includes Medicaid for low-income adults, Medicare Advantage (the private arm of Medicare), and Medigap the supplemental plans that correspond with the public program. Medicare Advantage plans may include coverage of hearing aids. Health-related variables included severity of hearing trouble, functional vision impairment, number of chronic conditions, cognitive impairment, number of activities of daily living limitations (ADLS), and having a helper to carry out ADLs. Access to online information included having a personal computer at home and using the Internet to get information.

Statistical Analysis

This study is a descriptive analysis of socio-demographic characteristics among Medicare beneficiaries with untreated hearing loss across different groups based on health care utilization. All analyses used the survey weights provided by the Centers for Medicare and Medicaid to account for the complex survey design of the MCBS including the over-sampling of black and Hispanic populations, and survey non-response. Weighted proportions were used to describe the characteristics and health care service utilization of Medicare beneficiaries with functional hearing loss by hearing aid use, and the characteristics of people with untreated functional hearing loss by health care service type utilized. Group comparisons were made using Pearson's chi-squared test of independence. StataSE 14 was used to conduct all analyses.

RESULTS

Descriptive Characteristics

Overall, 7,361 Medicare beneficiaries reported hearing trouble or hearing aid use, representing a weighted sample of 28,195,657 Medicare beneficiaries. Among them, 87% reported not using hearing aids (**Table 1**). A greater proportion of persons who did not use hearing aids compared to those who did were young, women, and non-White. Those who did not use hearing aids were also more likely to have lower educational attainment and lower income. A greater proportion of those not using hearing aids had two or more limitations in activities of daily living compared to those using hearing aids.

Health Care Utilization by Service Type

The health care services most accessed by Medicare beneficiaries with hearing trouble in 2017 were obtaining prescription drugs, seeing a medical provider, and utilizing outpatient services (**Figure 1**). Among those with untreated functional hearing loss, 64% reported obtaining prescription drugs, 50% saw a provider, and 46% accessed outpatient services. Less commonly, they accessed inpatient services (12%), home health (6%), and skilled nursing facilities (3%). The distribution of health care utilization by service type was similar among those who used hearing aids.

Figure 2 shows the extent of the utilization overlap among the two most common services types. Among the Medicare beneficiaries with untreated functional hearing loss, 34% saw a medical provider and obtained prescription drugs from a pharmacy in the past year, 16% saw a medical provider only, and 30% only obtained prescription drugs from a pharmacy. Twenty percent did not access any of these two health care service types.

Characteristics of Medicare Beneficiaries With Untreated Functional Hearing Loss by Health Care Services Utilized

The characteristics of Medicare beneficiaries with untreated functional hearing loss are presented by service type utilized (**Table 2**). Fifty percent of Medicare beneficiaries with untreated functional hearing loss saw a medical provider in the past year. Compared to those who did not see a medical provider, those who did had a greater proportion of individuals who were older, women, living alone, and with supplemental insurance coverage, especially Medicare Advantage. They also had a greater proportion with a high number of chronic conditions, functional vision impairment, and cognitive impairment, but fewer reporting limitations in activities of daily living or to have a helper. A larger proportion of those who did see a provider, compared to those who did not, had a personal computer at home, but a similar proportion in both groups used the Internet to access information.

Sixty-four percent of Medicare beneficiaries with untreated functional hearing loss obtained prescription drugs in the past year. Compared to those who did not, a greater proportion of those who obtained prescription drugs were non-White, a woman, and younger than 65 years old and older than 75 years. A larger proportion of individuals had low educational attainment and income, had supplemental insurance coverage, especially Medicaid, Medicare Advantage, or Medigap, were living alone or with children and other family members, rather

TABLE 1 | Characteristics of Medicare beneficiaries with functional hearing trouble by hearing aid use.

	Untreated functional hearing trouble	Hearing trouble with hearing aid
Unweighted sample (*n*)	5,139	2,222
Weighted population (*N*)	20,803,386	7,392,271
Population distribution (%)	87%	13%
Age	Column percentages	
<65	14%	4%
65-74	48%	37%
75-84	26%	33%
85+	11%	26%
Women	51%	40%
Race/ethnicity		
White	86%	92%
Black	9%	4%
Hispanic	2%	1%
Asian	2%	1%
Other	2%	2%
Education		
Less than HS	16%	12%
HS Graduate	52%	50%
Completed College	32%	38%
Income relative to FPL		
<100%	14%	7%
100-149%	14%	10%
150-199%	11%	10%
200-399%	28%	32%
400%+	32%	41%
Number of chronic conditions		
0	7%	5%
1-2	36%	35%
3-5	44%	47%
6+	13%	12%
Number of ADLs		
0 ADLs	65%	70%
1 ADL	15%	14%
2+ ADLs	20%	16%

ADLs, activities of daily living; FPL, federal poverty level; HS, high school. Source: Authors' analysis of the Medicare Current Beneficiary Survey 2017. All characteristics have statistically significant distributions between those with hearing aids and those without at $p < 0.05$, column percentages may not add up to 100% due to rounding.

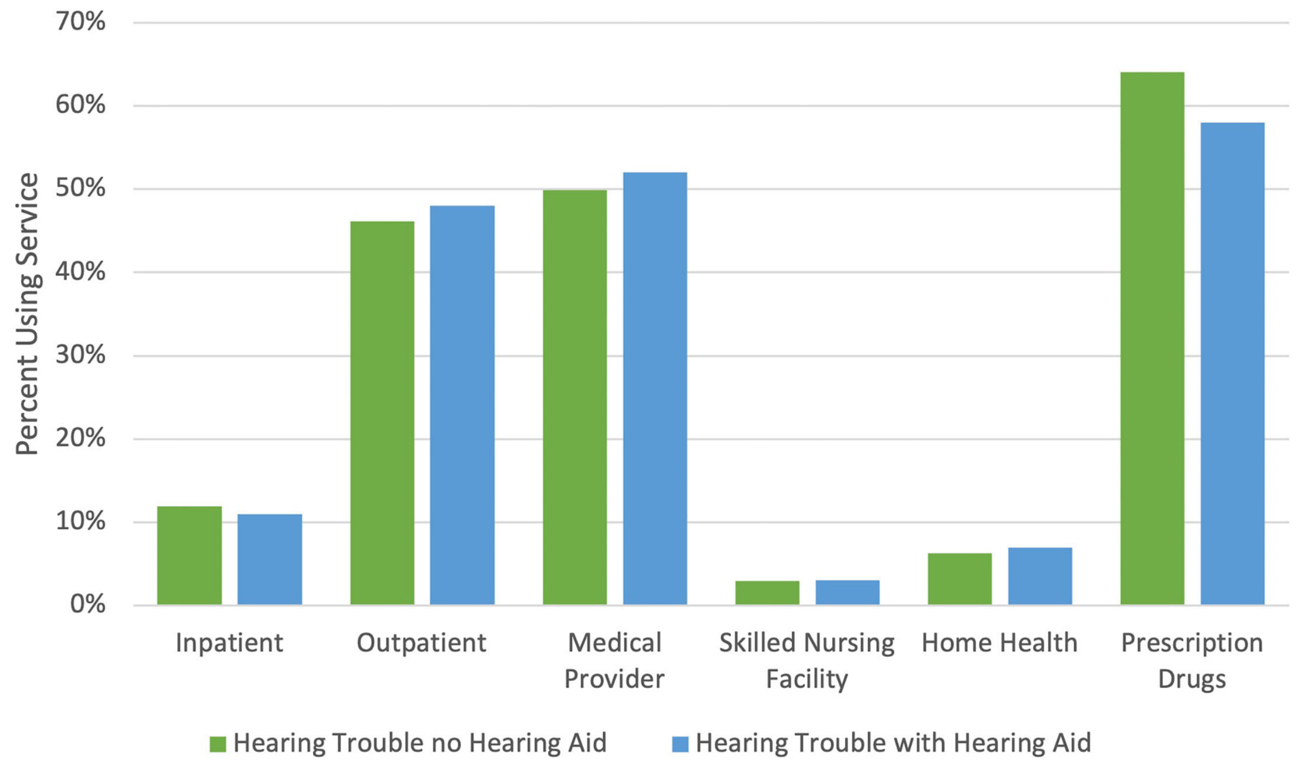


FIGURE 1 | Percent of Medicare beneficiaries with functional hearing trouble who report any utilization in 2017 by service type and hearing aid use. Source: Authors' analysis of the Medicare Current Beneficiary Survey 2017. Medical provider includes medical doctor, nurse practitioner, physician assistant, and nurses. Prescription drugs do not include beneficiaries who report often receiving their prescription drugs in the mail or through the internet.

than living with a spouse. A larger proportion of people who obtained prescription drugs in the past year relative to those who did not obtain prescription drugs had chronic conditions, cognitive impairment, a lot of trouble (rather than a little trouble) hearing, trouble seeing, limitations in activities of daily living, and had a helper. Compared to those who did not obtain prescription drugs, fewer Medicare beneficiaries who did obtain prescription drugs reported having a computer or using the internet for information.

Among Medicare beneficiaries with untreated functional hearing loss, 20% did not see a medical provider or obtain prescription medicines in the past year. Compared to those who utilized any of these services, those who did not were younger (<65 years old), White, men and had higher educational attainment and income. Among those who did not obtain prescription drugs or see a medical provider in the past year, 69% did not have supplemental insurance coverage, compared to 89% of those who utilized at least one of the services, and 57% lived with a spouse, compared to 49%. A smaller proportion of those who did not access any of the services compared to those who did had chronic conditions (87 vs. 95%), cognitive impairment (6 vs. 11%), limitations in activities of daily living, a lot of trouble hearing (rather than a little trouble hearing), trouble with vision, and had a helper. Of those who did not obtain prescription drugs or see a medical provider in the past year, 76% had a personal computer at home and 70% used the Internet to get information, compared to 67 and 61%, respectively, among those who accessed any of the services.

DISCUSSION

This study reinforces the low uptake of hearing aids among Medicare beneficiaries who report having hearing trouble and the differences in socio-demographic characteristics and health services utilization between those who use hearing aids and those who do not. Almost nine in ten (87%) beneficiaries who identify hearing problems are not serviced by the current model of hearing care, which involves high-cost devices and a predominantly specialty clinic-based approach to hearing care. Those with untreated functional hearing loss are more often women and racial/ethnic minorities. They have lower incomes, and more functional limitations than those who use hearing aids. This corroborates the financial and physical barriers to accessing hearing aids as found in previous analyses (13, 28).

Many studies highlight the importance of hearing to one's health and well-being. From a health system perspective, untreated hearing loss is associated with higher health care costs and poor health outcomes (24). The current model of care is not serving the majority of those in need. In fact, among minority groups access to hearing treatment has decreased over the last decade (29). Alternative pathways need to be considered.

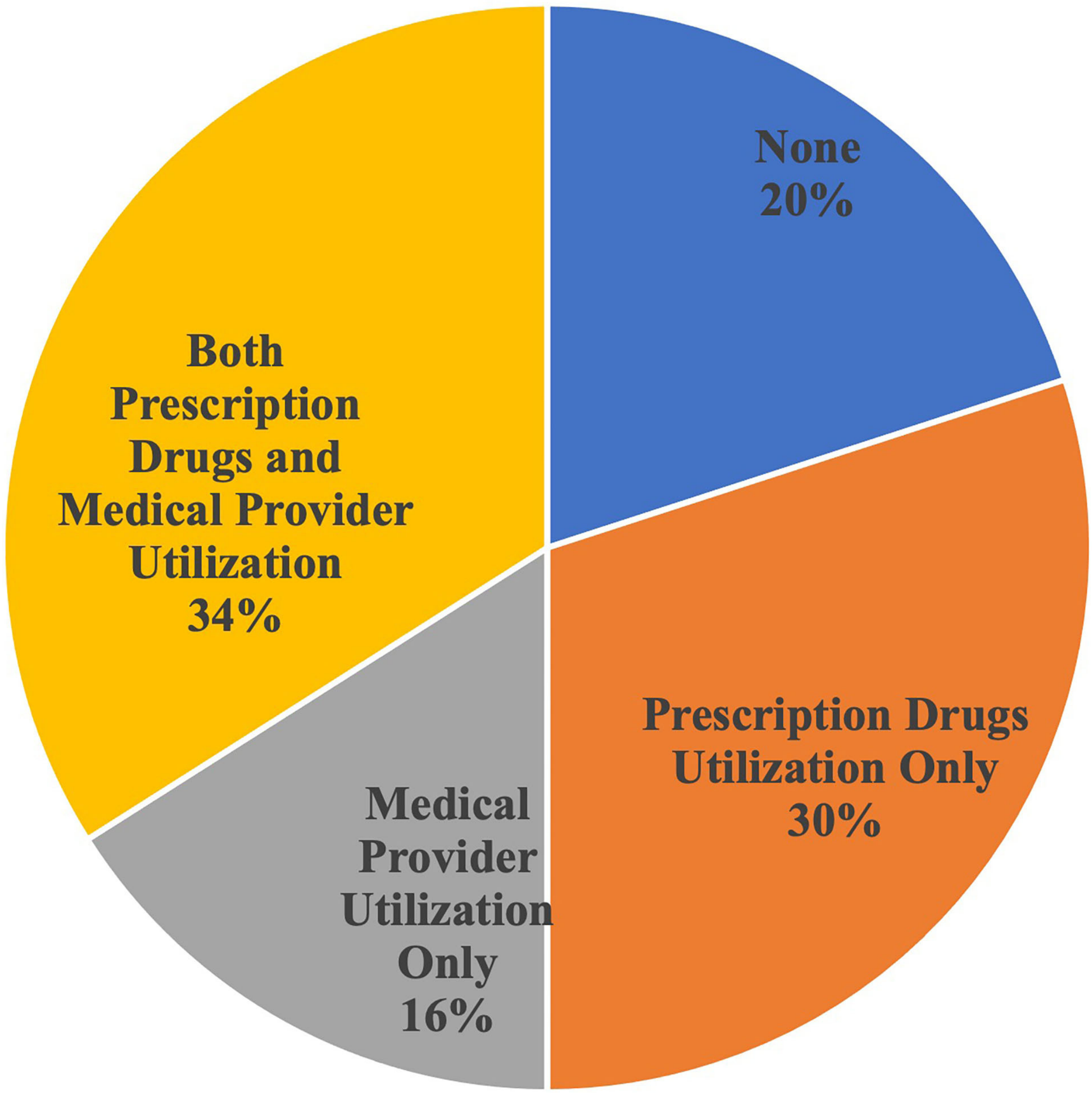


FIGURE 2 | Intersection of medical provider and prescription drug utilization by Medicare beneficiaries with untreated functional hearing trouble. Source: Authors' analysis of the Medicare Current Beneficiary Survey 2017. Medical provider includes medical doctor, nurse practitioner, physician assistant, and nurses. Prescription drugs do not include beneficiaries who report often receiving their prescription drugs in the mail or through the internet.

Some alternative delivery models are currently being piloted in the community [e.g., HEARS (17), Oyendo Bien (19)], through retail clinics [e.g., Walgreens, COSTCO (30)], and within adult day clinics [e.g., PACE clinics (31)] and tele-health (32–34) [e.g., Veterans Affairs (35)] (see **Supplementary Table 1** for more detail).

While these alternative delivery models are being trialed in specific locations, this analysis provides a nationally representative picture of health care access patterns among Medicare beneficiaries with self-reported hearing trouble to provide insights into which alternative models of hearing care may be the most accessible to these individuals. Sixty-four percent of beneficiaries with untreated functional hearing trouble visited a pharmacy and 50% visited a medical provider. Our analysis suggests that a hearing care program run through a pharmacy would reach a population with greater financial and physical barriers, than a program run through a medical provider clinical only. Those who visited the pharmacy were more likely to be lower income, in a minority racial or ethnic group, with greater comorbidities, and functional limitations, than those who did not go to the pharmacy. Pharmacies have taken on a greater role in delivering health care to communities over time across other aspects of health including administering vaccinations, medication reconciliation, and patient education (36).

Interestingly, 20% of the untreated functional hearing trouble group did not visit a pharmacy or medical provider in the

TABLE 2 | Characteristics of Medicare beneficiaries with untreated functional hearing trouble by service utilization, 2017.

	Any medical provider or prescriptive drug use		Medical provider visit		Prescriptive drugs	
	None	**Either MP or PD Use**	**No**	**Yes**	**No**	**Yes**
Unweighted sample	873	4,266	2,423	2,680	1,636	3,467
Weighted population	4,214,409	16,587,213	10,432,013	10,369,609	7,480,263	13,321,359
Population distribution (%)	20%	80%	50%	50%	36%	64%
Age						
<65	12%	15%***	15.38%	13%***	11%	16%***
65-74	58%	45%	49%	48%	55%	44%
75-84	22%	27%	26%	27%	24%	28%
85+	8%	12%	10%	13%	10%	12%
Women	40%	54%***	49%	53%***	42%	56%***
Race/ethnicity						
White	88%	83%*	84%	85%	89%	82%***
Black	7%	9%	8%	8%	6%	10%
Other	5%	8%	8%	7%	5%	9%
Education						
Less than high school	12%	17%***	17%	15%	10%	19%***
High school graduate	53%	52%	53%	52%	52%	53%
Completed college	35%	31%	31%	33%	38%	28%
Income relative to FPL						
<100%	9%	15%***	14%	13%	7%	18%***
100-149%	11%	15%	15%	14%	10%	17%
150-199%	10%	12%	11%	11%	10%	12%
200-399%	28%	28%	27%	29%	29%	28%
400%+	43%	30%	33%	32%	45%	25%
Living arrangement						
Alone	23%	28%***	26%	28%***	24%	28%***
Spouse	57%	49%	50%	51%	58%	46%
Children/family	11%	15%	14%	14%	10%	17%
Other	9%	8%	10%	7%	8%	9%
Rural	25%	23%	23%	23%	24%	23%
Cognitive impairment	6%	11%***	8%	12%***	8%	10%***
Supplemental insurance coverage						
Medicare only	31%	11%***	18%	13%***	28%	8%***
Medicaid	5%	21%	18%	17%	4%	25%
Employer	36%	17%	21%	20%	37%	12%
MA	14%	28%	23%	28%	15%	31%
Medigap	14%	23%	19%	22%	16%	24%
Has trouble hearing						
A little trouble	91%	88%*	89%	88%	91%	88%***
A lot of trouble	9%	12%	11%	12%	9%	12%
Has trouble with vision	37%	47%***	44%	47%*	40%	48%***
Number of chronic conditions						
0	13%	5%	9%	5%***	11%	5%***
1-2	44%	34%	40%	31%	40%	33%
3-5	35%	47%	40%	49%	40%	47%
6+	8%	14%	11%	15%	9%	15%
Number of ADLs						
0 ADLs	71%	64%***	64%	67%*	71%	62%***
1 ADL	14%	15%	15%	14%	15%	15%
2+ ADLs	15%	21%	21%	19%	14%	23%
Has a helper	28%	34%***	34%	32%*	27%	36%***
Has a personal computer at home	76%	67%***	67%	70%*	78%	64%***
Ever use the Internet to get info	70%	61%***	62%	63%	70%	58%***

*ADLs, activities of daily living; FPL, federal poverty level; HS, high school; MP, medical provider; PD, prescription drugs. Source: Authors' analysis of the Medicare Current Beneficiary Survey 2017. P-value <0.001*** <0.01** <0.05*.*

previous 12 months, suggesting that these approaches would not capture all beneficiaries with self-reported hearing trouble. These non-users tended to be aged 65-74, men, white, with higher income and living with a spouse. They were more likely to report a little rather than a lot of hearing trouble. They were also more likely to have a personal computer at home and use the internet, suggesting they might be good candidates for an online model of hearing care, such as through mobile health applications and tele-audiology. The coronavirus pandemic, which has significantly disrupted access to clinic-based health care, has increased the need for alternative models of hearing care, particularly those that incorporate an online or tele-audiology component. As evidenced in the UK during the pandemic, there remain both provider- and patient-side barriers to delivering tele-audiology services (37).

The purpose of this analysis is to better inform the planning for increasing access to hearing care at a time when the existing model is already undergoing substantial change. The introduction of the Over-the-Counter Hearing Aid Act (2017) which will regulate the sale of hearing aids to treat mild- to moderate-hearing loss over the counter and is expected to spawn a broader array of more affordable, quality-controlled devices available direct to the consumer. If these low-cost devices are placed in pharmacies, or doctors' offices, it may result in greater uptake of hearing care among older adults who have not previously engaged in the existing model of care.

With many alternative delivery models focused on increasing access to affordable devices without hearing care services come questions of quality of care. While early analysis of self-fitting devices compared to audiologist supported fitting suggests that comparable outcomes can be achieved (23), the literature suggests that hearing outcomes are optimized when receiving supportive hearing services in addition to the device (23, 38, 39). Further, greater perceived self-efficacy of managing hearing aids is associated with a more successful outcome in using and benefiting from a device (40). OTC hearing aids are designed to be self-fitting however, it is too early to tell whether device adherence and hearing outcomes under an OTC model will be maintained at least to the extent observed in hearing care service approaches. Certainly, device-focused approaches do not support or promote the development of coping strategies for psychosocial impacts of hearing loss (41), partly resulting from the chronic nature of hearing loss and the limitations of hearing aids to fully compensate for the impairment (42).

Ultimately, there is unlikely to be a one-size-fits-all approach to hearing care. Even across two common delivery models such as medical provider and prescription drug there was 34% overlap among Medicare beneficiaries with untreated functional hearing trouble. Alternative delivery methods should consider how to complement the existing model rather than substitute for it. Defining a pathway for hearing care in the U.S. that covers the spectrum from prevention of hearing loss to treatment, and addresses the physical, financial, and emotional barriers to seeking care is a crucial goal of the system. This will require engagement from older adults and current hearing care providers, as well as other stakeholders both inside and outside the health care system.

Many alternative or complementary models of care will require changes to existing workforce and reimbursement structures, including forward thinking on scope-of-practice legislation, investments in training and certification of paraprofessionals, such as community health workers, and the reimbursement of education and counseling separate from the cost of the device (18, 43).

The limitations of the study reflect the challenge of refining the measures for hearing loss and utilization based on the survey and administrative data available. Firstly, hearing loss is measured by self-reported trouble hearing, not by a professional examination. The population captured in this analysis is therefore those who recognize they have some degree of hearing trouble and does not reflect the entire population who could potentially benefit from some degree of hearing care. Previous studies have shown that individuals often underestimate their hearing loss and that underestimation can vary by sociodemographic characteristics (44). For the purposes of this analysis, however, we are interested in how to better reach those who recognize they have hearing difficulties, but are not using hearing aids.

Secondly, the categories of utilization provided by the MCBS contain broader categories of medical provider and prescription drugs than primary care and pharmacy visits, respectively. Our methods detail the ways that we attempted to refine these variables; however, it is possible that our estimates of potential reach of alternative delivery models may over-estimate those attending primary care clinics or pharmacies. The prescription drug measure does not account for visits to the pharmacy for over-the-counter purchases. It is therefore also possible that this pathway is under-estimated. Finally, we do not assume that attendance at these alternative pathways equates to receiving hearing care treatment.

This analysis highlights the ongoing inequities in the existing hearing care system which may be attributable to the financial and physical barriers associated with high-cost devices and clinic-based hearing care. Disruption to the clinic-based delivery model brought on by the introduction of FDA-regulated, direct-to-consumer hearing aids, has resulted in piloting of models in the community, pharmacies, and medical systems. This analysis suggests that these new delivery models may create greater equity in hearing care access by reaching currently unserved populations.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available from the Centers for Medicare and Medicaid Services (CMS) but restrictions apply to the availability of these data, which were used under license for the current study, and are not publicly available. Data can be requested from cms.gov.

AUTHOR CONTRIBUTIONS

AW, LA, and NR designed the study and drafted the manuscript. AW conducted the data analysis. AW, LA, CN, CM, FL, and NR interpreted the data. CN, CM, and FL critically revised the manuscript for important intellectual content. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Commonwealth Fund (grant number 20192345). This manuscript was also supported in part by funding from the Cochlear Center for Hearing and Public Health at the Johns Hopkins Bloomberg School of Public Health. CN was funded by NIA/NIH grant K23AG059900. NR was funded by NIA/NIH grant 1K23AG065443. Funders and sponsors had no role in the design, methods, analysis, or preparation of the paper.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fdgth.2021.740323/full#supplementary-material

REFERENCES

1. Goman AM, Reed NS, Lin FR. Addressing estimated hearing loss in adults in 2060. *JAMA Otolaryngol Head Neck Surg.* (2017) 143:733-4. doi: 10.1001/jamaoto.2016.4642
2. Gopinath B, Hickson L, Schneider J,McMahon CM, Burlutsky G, Leeder SR, et al. Hearing-impaired adults are at increased risk of experiencing emotional distress and social engagement restrictions five years later. *Age Ageing.* (2012) 41:618-23. doi: 10.1093/ageing/afs058
3. Polku H, Mikkola TM, Rantakokko M, Portegijs E, Törmäkangas T, Rantanen T, et al. Hearing and quality of life among community-dwelling older adults. *J Gerontol B Psychol Sci Soc Sci.* (2018) 73:543-52. doi: 10.1093/geronb/gbw045
4. Lin TC, Yen M, Liao YC. Hearing loss is a risk factor of disability in older adults: a systematic review. *Arch Gerontol Geriatr.* (2019) 85:103907. doi: 10.1016/j.archger.2019.103907
5. Viljanen A, Kaprio J, Pyykko I, Sorri M, Pajala S, Kauppinen M, et al. Hearing as a predictor of falls and postural balance in older female twins. *J Gerontol A Biol Sci Med Sci.* (2009) 64:312-7. doi: 10.1093/gerona/gln015
6. Deal JA, Betz J, Yaffe K, Harris T, Purchase-Heizner E, Satterfield S, et al. Hearing impairment and incident dementia and cognitive decline in older adults: the health ABC study. *J Gerontol A Biol Sci Med Sci.* (2017) 72:703-9. doi: 10.1093/gerona/glw069
7. Livingston G, Sommerlad A, Orgeta V, Costafreda SG, Huntley J, Ames D, et al. Dementia prevention, intervention, and care. *Lancet.* (2017) 390:2673-734. doi: 10.1016/S0140-6736(17)31363-6
8. Mulrow CD, Aguilar C, Endicott JE, Tuley MR, Velez R, Charlip WS, et al. Quality-of-life changes and hearing impairment. A randomized trial. *Ann Intern Med.* (1990) 113:188-94. doi: 10.7326/0003-4819-113-3-188
9. Ferguson MA, Kitterick PT, Chong LY, Edmondson-Jones M, Barker F, Hoare DJ. Hearing aids for mild to moderate hearing loss in adults. *Cochrane Database Syst Rev.* (2017) 9:CD012023. doi: 10.1002/14651858.CD012023.pub2
10. Maharani A, Dawes P, Nazroo J, Tampubolon G, Pendleton N, group SE-CW. Longitudinal relationship between hearing aid use and cognitive function in older Americans. *J Am Geriatr Soc.* (2018) 66:1130-6. doi: 10.1111/jgs.15363
11. Mahmoudi E, Basu T, Langa K, Mckee MM, Zazove P, Alexander N, et al. Can hearing aids delay time to diagnosis of dementia, depression, or falls in older adults? *J Am Geriatr Soc.* (2019) 67:2362-9. doi: 10.1111/jgs.16109
12. Chien W, Lin FR. Prevalence of hearing aid use among older adults in the United States. *Arch Intern Med.* (2012) 172:292-3. doi: 10.1001/archinternmed.2011.1408
13. McKee MM, Choi H, Wilson S, DeJonckheere MJ, Zazove P, Levy H. Determinants of hearing aid use among older Americans with hearing loss. *Gerontologist.* (2019) 59:1171-81. doi: 10.1093/geront/gny051
14. Assi L, Reed NS, Nieman CL, Willink A. Factors associated with hearing aid use among medicare beneficiaries. *Innov Aging.* (2021) 5:igab021. doi: 10.1093/geroni/igab021
15. Pacala JT, Yueh B. Hearing deficits in the older patient: "I didn't notice anything". *JAMA.* (2012) 307:1185-94. doi: 10.1001/jama.2012.305
16. National Academies of Sciences Engineering and Medicine. *Hearing Health Care for Adults: Priorities for Improving Access and Affordability.* Washington, DC: The National Academies Press (2016).
17. Nieman CL, Marrone N, Mamo SK, Betz J, Choi JS, Contrera KJ, et al. The Baltimore HEARS pilot study: an affordable, accessible, community-delivered hearing care intervention. *Gerontologist.* (2017) 57:1173-86. doi: 10.1093/geront/gnw153
18. Suen JJ, Bhatnagar K, Emmett SD, Marrone N, Kleindienst Robler S, Swanepoel DW, et al. Hearing care across the life course provided in the community. *Bull World Health Organ.* (2019) 97:681-90. doi: 10.2471/BLT.18.227371
19. Marrone N, Ingram M, Somoza M, Jacob DS, Sanchez A, Adamovich S, et al. Interventional audiology to address hearing health care disparities: oyendo bien pilot Study. *Semin Hear.* (2017) 38:198-211. doi: 10.1055/s-0037-1601575
20. The President's Council of Advisors on Science and Technology. *Aging America & Hearing Loss: Imperative of Improved Hearing Technologies.* Washington, DC: The President's Council of Advisors on Science and Technology (2015).
21. U.S. Food and Drug Administration. *FDA Reauthorization Act of 2017 (FDARA).* Available online at: https://www.fda.gov/regulatory-information/selected-amendments-fdc-act/fda-reauthorization-act-2017-fdara (accessed May 5, 2021).
22. Ostermiller I. *Medical Devices; Ear, Nose and Throat Devices; Establishing Over-the-Counter Hearing Aids and Aligning Other Regulations.* In: Services DoHaH, ed. Vol. 0910-AI21. MD: Silver Spring, Office of Management and Budget (2021).
23. Humes LE, Rogers SE, Quigley TM, Main AK, Kinney DL, Herring C. The effects of service-delivery model and purchase price on hearing-aid outcomes in older adults: a randomized double-blind placebo-controlled clinical trial. *Am J Audiol.* (2017) 26:53-79. doi: 10.1044/2017_AJA-16-0111
24. Reed NS, Altan A, Deal JA, Yeh C, Kravetz AD, Wallhagen M, et al. Trends in health care costs and utilization associated with untreated hearing loss over 10 years. *JAMA Otolaryngol Head Neck Surg.* (2019) 145:27-34. doi: 10.1001/jamaoto.2018.2875
25. Mahmoudi E, Zazove P, Meade M, McKee MM. Association between hearing aid use and health care use and cost among older adults with hearing loss. *JAMA Otolaryngol Head Neck Surg.* (2018) 144:498-505. doi: 10.1001/jamaoto.2018.0273
26. Reed NS, Assi L, Horiuchi W, Hoover-Fang JE, Lin FR, Ferrante LE, et al. Medicare beneficiaries with self-reported functional hearing difficulty have unmet health care needs. *Health Aff.* (2021) 40:786-94. doi: 10.1377/hlthaff.2020.02371
27. Office of Enterprise Data and Analytics (OEDA). *Medicare Current Beneficiary Survey 2017: Methodology Report.* In: Centers for Medicare and Medicaid, ed. Bethesda, MD: Office of Enterprise Data and Analytics (OEDA) (2019).
28. Nieman CL, Marrone N, Szanton SL, Thorpe RJ, Jr., Lin FR. Racial/ethnic and socioeconomic disparities in hearing health care among older Americans. *J Aging Health.* (2016) 28:68-94. doi: 10.1177/08982643155 85505
29. Reed NS, Stolnicki W, Gami A, Myers C, Kohn C, Willink A. Association of self-reported trouble hearing and patient-provider communication with hospitalizations among medicare beneficiaries. *Semin Hear.* (2021) 42:26-36. doi: 10.1055/s-0041-1725998
30. Kasewurm G. An independent practice's guide to battling big box retail and commoditization in hearing healthcare. *Hearing Rev.* (2014) 21:18-21.

31. Mamo SK, Nirmalasari O, Nieman CL, McNabney MK, Simpson A, Oh ES, et al. Hearing care intervention for persons with dementia: a pilot study. *Am J Geriatr Psychiatry.* (2017) 25:91-101. doi: 10.1016/j.jagp.2016.08.019
32. Swanepoel DW, Clark JL, Koekemoer D, Hall JW, Krumm M, Ferrari DV, et al. Telehealth in audiology: the need and potential to reach underserved communities. *Int J Audiol.* (2010) 49:195-202. doi: 10.3109/14992020903470783
33. Swanepoel W, De Sousa KC, Smits C, Moore DR. Mobile applications to detect hearing impairment: opportunities and challenges. *Bull World Health Organ.* (2019) 97:717-8. doi: 10.2471/BLT.18.227728
34. van Wyk T, Mahomed-Asmail F, Swanepoel W. Supporting hearing health in vulnerable populations through community care workers using mHealth technologies. *Int J Audiol.* (2019) 58:790-7. doi: 10.1080/14992027.2019.1649478
35. Jacobs PG, Saunders GH. New opportunities and challenges for teleaudiology within Department of Veterans Affairs. *J Rehabil Res Dev.* (2014) 51:vii-xii. doi: 10.1682/JRRD.2014.04.0093
36. Goode JV, Owen J, Page A, Gatewood S. Community-based pharmacy practice innovation and the role of the community-based pharmacist practitioner in the United States. *Pharmacy.* (2019) 7:106. doi: 10.3390/pharmacy7030106
37. Parmar B, Beukes E, Rajasingam S. The impact of COVID-19 on provision of UK audiology services and on attitudes towards delivery of telehealth services. *Int J Audiol.* (2021) 2021:1-11. doi: 10.1080/14992027.2021.1921292
38. Willink A, Reed NS, Lin FR. Access to hearing care services among older medicare beneficiaries using hearing aids. *Health Aff.* (2019) 38:124-31. doi: 10.1377/hlthaff.2018.05217
39. Willink A, Reed NS, Lin FR. Cost-benefit analysis of hearing care services: what is it worth to medicare? *J Am Geriatr Soc.* (2019) 67:784-9. doi: 10.1111/jgs.15757
40. Hickson L, Meyer C, Lovelock K, Lampert M, Khan A. Factors associated with success with hearing aids in older adults. *Int J Audiol.* (2014) 53(Suppl 1):S18-27. doi: 10.3109/14992027.2013.860488
41. Gomez RG, Madey SF. Coping-with-hearing-loss model for older adults. *J Gerontol B Psychol Sci Soc Sci.* (2001) 56:P223-5. doi: 10.1093/geronb/56.4.P223
42. Lesica NA. Why do hearing aids fail to restore normal auditory perception? *Trends Neurosci.* (2018) 41:174-85. doi: 10.1016/j.tins.2018.01.008
43. Yong M, Willink A, McMahon C, Mcpherson B, Nieman Cl, Reed NS, et al. Access to adults' hearing aids: policies and technologies used in eight countries. *Bull World Health Organ.* (2019) 97:699-710. doi: 10.2471/BLT.18.228676
44. Goman AM, Reed NS, Lin FR, Willink A. Variations in prevalence and number of older adults with self-reported hearing trouble by audiometric hearing loss and sociodemographic characteristics. *JAMA Otolaryngol Head Neck Surg.* (2019) 146:201-3. doi: 10.1001/jamaoto.2019.3584

Conflict of Interest: AW reports receiving a speaker honorarium from the American Speech-Language-Hearing Association and consulting honorarium from BioMedical Insights. CN serves on the non-profit board of directors for Access HEARS and the Hearing Loss Association of America. NR is a scientific advisory board member of Shoebox, Inc and Good Machine Studio but does not receive financial compensation. FL reports being a consultant to Frequency Therapeutics, receiving speaker honoraria from Caption Call, and being the director of a public health research center funded in part by a philanthropic gift from Cochlear Ltd to the Johns Hopkins Bloomberg School of Public Health. CM is on the scientific advisory board for Good Machine Inc but does not receive financial compensation.

The remaining author declares 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.

ORIGINAL RESEARCH
published: 04 January 2022
doi: 10.3389/fpsyg.2021.640300

Speech Perception in Noise Is Associated With Different Cognitive Abilities in Chinese-Speaking Older Adults With and Without Hearing Aids

Yuan Chen[1†], Lena L. N. Wong[2†], Shaina Shing Chan[2] and Joannie Yu[3]*

[1] Department of Special Education and Counseling, The Education University of Hong Kong, Tai Po, Hong Kong SAR, China, [2] Clinical Hearing Sciences (CHearS) Laboratory, Faculty of Education, The University of Hong Kong, Pokfulam, Hong Kong SAR, China, [3] Audiology Centre, Union Hospital, Tai Wai, Hong Kong SAR, China

OPEN ACCESS

Edited by:
Qinglin Meng,
South China University of Technology,
China

Reviewed by:
Helen Nuttall,
Lancaster University, United Kingdom
Christian Füllgrabe,
Loughborough University,
United Kingdom

***Correspondence:**
Yuan Chen
cheny@eduhk.hk

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology

Received: 11 December 2020
Accepted: 03 December 2021
Published: 04 January 2022

Citation:
Chen Y, Wong LLN, Chan SS and Yu J (2022) Speech Perception in Noise Is Associated With Different Cognitive Abilities in Chinese-Speaking Older Adults With and Without Hearing Aids. Front. Psychol. 12:640300. doi: 10.3389/fpsyg.2021.640300

Chinese-speaking older adults usually do not perceive a hearing problem until audiometric thresholds exceed 45 dB HL, and the audiometric thresholds of the average hearing-aid (HA) user often exceed 60 dB HL. The purpose of this study was to examine the relationships between cognitive and hearing functions (measured as audiometric or speech reception thresholds) in older Chinese adults with HAs and with untreated hearing loss (HL). Participants were 49 Chinese older adults who used HAs and had moderate to severe HL (HA group), and 46 older Chinese who had mild to moderately severe HL but did not use HAs (untreated; or UT group). Multiple linear regression analysis was employed to evaluate how well age, education level, audiometric thresholds, and speech perception in noise were related to performance on general cognitive function, working memory, executive function, attention, and verbal learning tests. Results showed that speech perception in noise alone accounted for 13–25% of the variance in general cognitive function, working memory, and executive function in the UT group, and 9–21% of the variance in general cognitive function and verbal learning in the HA group (i.e., medium effect sizes). Audiometric thresholds did not explain any proportion of the variance in cognitive functioning in the HA or UT group. Thus, speech perception in noise accounts for more variance in cognitive performance than audiometric thresholds, and is significantly associated with different cognitive functions in older Chinese adults with HAs and with untreated HL.

Keywords: cognitive function, hearing loss, speech perception, older adults, Chinese

---

Abbreviations: EF, executive function; GCF, general cognitive function; HL, hearing loss; HA, hearing aid; SRT, speech recognition threshold; CHINT, Cantonese Hearing in Noise Test; NF, noise front; NBE, noise better ear; NEW, noise worse ear; SNR, signal-to-noise ratio; SA, sustained attention; MoCA, Montreal Cognitive Assessment; CSB, CogState Battery; MCI, mild cognitive impairment; OBK, One-Back Test; GML, Groton Maze Learning Test; IDN, Identification Test; ISL, International Shopping List Task; SD, standard deviation; PTA, pure tone average; UT, untreated; VL, verbal learning; WM, working memory.

INTRODUCTION

Hearing Loss and Hearing Aid Uptake in Chinese-Speaking Older Adults

Hearing loss (HL) is reported in nearly two-thirds of adults aged 60 years and older (Gong et al., 2018). However, Chinese older adults often do not report hearing problems until the HL exceeds 45 dB hearing level, and typically do not acquire a hearing aid (HA) until an average loss of about 65 dB hearing level, even though normal hearing sensitivity is defined as pure-tone audiometric thresholds not exceeding 25 dB HL (Doyle and Wong, 1996). Similarly, studies by the Institute of Human Communicative Research (2005) and Wong et al. (2014) found that the majority of participants with untreated HL exhibit mild to moderately severe audiometric HL, and those with HAs exhibit moderate to severe HL. The HA uptake rate is lower among older Chinese speakers with HL (less than 10%) than the rate of approximately 25% reported in the United States and United Kingdom (Chien and Lin, 2012; Zhao et al., 2015; Bisgaard and Ruf, 2017).

The Relationship Between Hearing Loss and Cognitive Function

Epidemiological studies suggest that HL in older adults is independently associated with an increased incidence of cognitive decline, after accounting for factors such as age, gender, income, education, general physical health, and HA use (Lin et al., 2011, 2013; Gurgel et al., 2014; Harrison Bush et al., 2015). Specifically, older adults with HL experience a 30–40% faster cognitive decline than that in the general older population (Lin et al., 2013). Cross-sectional and longitudinal studies have shown an association between HL and cognitive decline in general, as well as in specific domains such as verbal learning (VL), sustained attention (SA), executive function (EF), and working memory (WM) in older adults (Wingfield et al., 2005; Pichora-Fuller and Singh, 2006; Arlinger et al., 2009; Lin et al., 2011, 2013). In a meta-analysis, Taljaard et al. (2016) found a medium effect in individuals with treated HL, and a small effect in those with untreated HL, when evaluating the relationship between HL and general cognitive function (GCF). However, some studies, such as the one by Harrison Bush et al. (2015), reported very small effect sizes or were underpowered in relating HL to cognitive function, while others did not find any such relationship (Gallacher, 2005).

As better education contributes to better cognitive functioning in older adults (Livingston et al., 2017), cognitive decline is of particular concern for Chinese speakers aged 60 years and above, as their median education level is 5.89 years (National Bureau of Statistics in China, 2011). This is much lower than that of HA users in studies on the relationship between hearing and cognitive functions conducted in Western societies (Lin et al., 2011; Dawes et al., 2015; Maharani et al., 2018).

Measurement of Hearing Function

The weak association between HL and cognitive functioning may be attributed to the use of pure-tone audiometric thresholds as an indicator of hearing function (i.e., the ability to perceive sounds) (Plack, 2014). Audiometric hearing thresholds primarily reflect peripheral hearing impairment and may not reflect auditory cortical processing (Tun et al., 2012). In a large-scale study involving a 60-h battery of various tests of cognitive and auditory functioning, Humes et al. (2013) found that decreased auditory sensory functioning, which was a composite of non-speech psychophysical and speech perception measures, was associated with reduced cognitive functioning, and the authors concluded that relying on audiometric thresholds alone may underestimate the relationship between hearing and cognitive functions. Therefore, being able to recognize speech in noise, which involves not only peripheral hearing but also higher auditory cortical processing, and reflects daily listening ability (Humes et al., 2013), may be a better indicator of a decline in auditory sensory functioning.

Different degrees of HL (i.e., ranging from mild to profound HL) were mixed and whether HA was fitted were not specified in previous studies, rendering comparisons in findings across studies difficult (Harrison Bush et al., 2015; Taljaard et al., 2016). More severe HL is associated with poorer suprathreshold auditory processing skills, which could lead to greater temporal information distortion in speech (Füllgrabe and Moore, 2018). Similarly, HA processing that alters certain acoustic information to enhance hearing could introduce spectral and temporal distortions (Stone et al., 2009a; Wong et al., 2018). Such distortions could disrupt automatic lexical retrieval, leading to explicit, effortful processing mechanisms that rely on cognitive processing system (Rönnberg et al., 2019). Thus, when analyzing the relationship between HL and cognitive functioning, individuals with different degrees of HL should be considered separately. Similarly, individuals using HAs should not be treated in the same way as those who do not.

Therefore, the first aim of the current study was to examine whether pure-tone audiometric and speech reception thresholds were significantly related to cognitive function in two groups: older adult HA users with moderate to severe HL, and older adults with untreated mild to moderately severe HL. These two groups represent typical older HA users and non-HA users in Hong Kong and mainland China (Wong et al., 2014), for whom the relationship between hearing and cognitive functioning has not been studied. Findings from the literature may not apply to Chinese speakers, as typical HA users in Western societies exhibit different characteristics (i.e., higher education level and a higher proportion of HA uptake, especially in older adults with mild HL) (Doyle and Wong, 1996).

Furthermore, the causal mechanisms underlying the link between auditory sensory and cognitive decline is still unclear. Wayne and Johnsrude (2015) reviewed several potential mechanisms. Among them, three hypotheses have emerged as strong contenders: (1) the information-degradation hypothesis, which suggests that auditory sensory decline leads to impoverished (but reversible) cognitive function; (2) the sensory deprivation hypothesis, which indicates that auditory sensory decline causes more permanent cognitive decline; and (3) the common cause hypothesis, which suggests that a third variable contributes to declines in auditory sensory and cognitive functions. While it is difficult to separately test these hypotheses, it is important to know which cognitive skill is significantly related to hearing function. Therefore, the second aim of the current study was to examine which cognitive skill was related to audiometric thresholds and speech perception in the two participant groups.

There are three differences between the current and previous studies examining the relationship between audiometric thresholds, speech perception and cognitive function (e.g., Akeroyd, 2008; Humes et al., 2013). First, as mentioned above, cognitive functions of HA users and non-users were examined separately. Second, instead of using speech perception as a dependent variable (e.g., Humes, 2007; Lunner and Sundewall-Thorén, 2007), the current study used cognitive function test scores as dependent variables, with speech perception and audiometric thresholds as independent variables, while controlling for the effects of age, peripheral HL, and education level. Accordingly, the results indicated which cognitive function(s) were more likely associated with auditory sensory decline in the UT and HA groups. Third, although examining speech perception and cognitive function in a pre-post design (i.e., HA users before and after fitting) may be better at controlling confounders, the untreated comparison group (i.e., HA users before fitting) would not represent the largest untreated HL population in China, as Chinese older adults typically do not acquire HAs until an average loss of about 65 dB is reached (Doyle and Wong, 1996). In contrast, the two independent samples included in the present study represent the largest relevant populations in China: the untreated HL group, representing those with mild to moderately severe audiometric HL, and the HA group, representing those with moderate to severe HL.

MATERIALS AND METHODS

Participants

A convenience sample of 50 current HA users with moderate to severe HL (i.e., the HA group), and 50 untreated participants (non-HA users) with mild to moderately severe HL (i.e., the UT group) were recruited from the Audiological Center at the Prince of Wales Hospital and Alice Ho Miu Ling Nethersole Hospital, in Hong Kong. Participants in the HA group must have worn HAs for at least 2 years. This restriction was employed because it normally takes 2–5 years for acclimatization to HAs (Ng and Rönnberg, 2020). All participants were required to be older than 60 years and speak Cantonese. All had normal or corrected-to-normal vision as screened using the Snellen Chart. Patients were excluded if they had a history, as documented in medical records, of neurodegenerative disorders, brain tumors, significant head trauma, epilepsy, significant psychiatric disorders (such as major depression or schizophrenia), substance abuse, or alcoholism.

Materials

Hearing thresholds were obtained in a standard audiometric booth at the Audiology Center of the Prince of Wales Hospital using a GSI 61 Audiometer (Grason-Stadler, Eden Prairie, MN, United States). Aided and unaided soundfield hearing thresholds at 0.5, 1, 2, and 4 kHz were measured binaurally using warble tones, in the same room, using the GSI61 Audiometer, and were used to determine the severity of HL with and without HAs. Sounds were presented via a loudspeaker (Cerwin-Vega) located 1 m in front of the participant. In addition, because soundfield hearing thresholds cannot determine the better/worse ear for SRT testing, ear-specific air-conduction thresholds at 0.25, 0.5, 1, 2, 4, and 8 kHz were obtained using TDH-49 headphones.

Speech reception thresholds (SRTs), defined as the presentation level at which 50% of sentences were repeated entirely correctly by the participant (Wong and Soli, 2005), were obtained using the Cantonese version of the Hearing in Noise Test (CHINT) (Wong and Soli, 2005) in four test conditions: quiet (SF), with noise originating from the front (noise front: NF), on the side of the better ear (noise better ear: NBE), and on the side of the worse ear (noise worse ear: NWE). Each test condition included 20 sentences. Speech was always presented at 0° azimuth using a loudspeaker, and the intensity level of the sentences was adjusted adaptively, depending on the correctness of the response. Speech-spectrum shaped noise was used as a masker and fixed at 65 dB A in the noise conditions. SRTs in the quiet condition were measured in dB A, with lower SRT suggesting better speech perception in quiet. SRTs in noise were measured as signal-to-noise ratios (SNR), with lower SRTs suggesting better ability to extract speech from noise. The starting level was 65 and +5 dB SNR for quiet and noise conditions, respectively. The step size was 4 dB for the first four sentences and 2 dB for the 5th to 20th sentences. A noise composite score was calculated to represent the overall performance in noise, and was based on the performance when noise was from the front and from the side using the following formula: $[NF + \frac{1}{2}(NBE + NWE)]/2$.

The Cantonese version of the Montreal Cognitive Assessment (MoCA; Wong et al., 2009), and four tests selected from the CogState Battery (CSB),[1] were used to assess cognitive functioning.

The MoCA, covering eight domains of cognitive function: attention and concentration, EF, memory, language, visuospatial skills, conceptual thinking, calculations, and orientation, was designed to assess GCF and detect mild cognitive impairment (MCI). The test takes 10 mins to administer and the total score ranges from 0 to 30, with higher scores indicating better cognitive functioning. The Cantonese Chinese version was retrieved from the official website[2]. It has been adapted into Chinese with good reliability and validity (Zhong et al., 2013). It has been validated in Chinese older adults in Hong Kong, and was determined to be brief and feasible for administration in clinical settings (Yeung et al., 2014). Although the MoCA is often used to detect MCI, no exclusion of participants was made based on the MoCA score. This was because individuals with HL tend to have a higher risk of cognitive impairment than those without HL (Fritze et al., 2016). The relationship between HL and cognitive function may be obscured if participants with MCI were excluded.

[1]http://www.Cogstate.com
[2]http://www.mocatest.org/

The four selected CSB subtests comprised the One-Back Test (OBT), Groton Maze Learning Test (GML), Identification Test (IDN), and International Shopping List Task (ISL), which were used to examine WM, EF, SA, and VL, respectively. The CSB test battery has been shown to be sensitive in detecting MCI and Alzheimer's disease in older adults (de Jager et al., 2009).

The OBT was used to evaluate WM. A series of playing cards were presented one-by-one in the middle of a computer screen. Participants were requested to identify whether the playing card presented on the screen was the same as the previous one by pressing the "yes" button (the right button on the mouse) or the "no" button (the left button). Participants were encouraged to work as quickly as they could and be as accurate as possible. Accuracy of performance is reported as the arcsine transformation of the square root of the proportion of correct responses. Higher scores represent better performance. This test was relatively easy compared to other WM tests (e.g., the Reading Span Test). It was the only non-verbal WM test that has been validated in Cantonese speakers during the data collection, and Chen et al. (2020) demonstrated a lack of ceiling effects in a group of Chinese-speaking older adults with HAs (Chen et al., 2020).

The GML was used to evaluate EF. Participants were asked to learn the same hidden pathway in a maze on five consecutive trials. This learning process requires goal-directed problem-solving skills including set shifting, WM, and inhibitory control, which are important elements of EF (Carlson et al., 2013). The total number of errors made in attempting to learn the same hidden pathway is reported, and higher scores indicate poorer EF.

The IDN was used to evaluate SA. A playing card was presented face down in the center of the screen. Participants were asked to press the "yes" button if the card was red when flipped over, and press the "no" button if the card was black when flipped over. Participants were asked to complete the trials as fast and accurately as possible. The speed of performance is reported as the mean of the log10 transformed reaction times for correct responses. Lower scores represent better performance.

The ISL was developed to assess VL in populations with diversity in language and cultural backgrounds. A total of 12 food items were presented on a screen facing the test administrator (a trained research assistant) who read the words one by one to the participants. Participants were instructed to repeat each word one by one, along with the administrator, to ensure that all words were intelligible to them. Subsequently, the participants were asked to recall as many of the items as they could. The same list and procedure were repeated two more times. The total number of items from the list that were correctly recalled were summed to compute the score; thus, higher scores indicate better performance.

Procedures

This was a cross-sectional and observational study with two distinct groups of participants. Ethical clearance was obtained from the University of Hong Kong and Chinese University of Hong Kong. After an explanation of the nature of the study and procedures, written consent was obtained from participants. Demographic information was collected from all participants, followed by hearing assessments. HA verification and fine-tuning were conducted by audiologists to ensure the best fit before testing. During the actual testing, the MoCA, OBT, GML, IDN, and ISL were administered according to the instructions provided in the manuals. The order of SRT testing in quiet and noise conditions was randomized across participants. SRTs were obtained with sentence lists randomly selected by the CHINT program. Participants were encouraged to make guesses even when unsure. A pocket-talker was fitted to participants in the UT group and the volume was set to a comfortable listening level to ensure good reception of the test instructions, and test stimuli of the ISL. Participants in the HA group set HAs to their usual settings during the cognitive assessment. Verbal instructions were repeated and participants' understanding of the instructions was checked by asking them to verbally recall the instructions prior to the administration of each test. Each test (except the MoCA) started with a practice session.

The entire testing session took approximately 2 h. To avoid fatigue, a break was given to participants after 1 h of assessment or upon request. A transportation allowance of HKD 200 (equivalent to USD 25) was provided to each participant.

Data Analysis

Independent samples *t*-tests or Mann-Whitney *U* tests were used to examine whether there were significant differences in audibility and SRTs between two groups. Shapiro Wilk tests were used to examine whether data were normally distributed. Multiple linear regression analyses using a forward method were employed to evaluate how well age, education level, soundfield hearing thresholds (unaided soundfield hearing thresholds for the UT groups and aided soundfield hearing thresholds for the HA group), and SRTs in noise were associated with performance in each cognitive domain (i.e., GCF, EF, SA, WM, and VL). Education level was coded as 0 (uneducated), 1 (primary school), 2 (secondary school), and 3 (tertiary school or above). Pearson product-moment correlation coefficients were calculated to examine the correlation among variables before conducting multiple linear regression analysis. Data analyses were conducted using IBM SPSS Statistic 21.0 for Windows (IBM Corp., Armonk, NY, United States).

RESULTS

Demographics and Hearing Assessment

Due to scheduling conflicts, 49 participants in the UT group and 46 participants in the HA group completed all tests. All the following analysis were based on these participants. Most HA participants (94%) were unilaterally fitted with a HA, while the remaining were bilaterally fitted (**Table 1**). The mean duration of HA use was 7.28 years (SD = 4.77; range: 2–20 years). Participants were wearing different brands of HAs including Resound (n = 15), Beltone (n = 19), Widex (n = 4), Phonak (n = 4), Oticon (n = 1), and Siemens (n = 3). All participants were using their HAs at least 3–4 days/week, with 33 using their HA every day. For 11/46, 14/46, and 20/46 participants, HA use was less than 4, 4–8 h, and more than 8 h per day, respectively.

Although an independent samples t-test [t (93) = −7.91, p < 0.001] showed the HA group exhibited a significantly worse unaided soundfield hearing threshold (mean = 64.89, SD = 10.73) compared to that in the UT group (mean = 47.50, SD = 10.69), there was no significant difference [independent t-test, t (93) = 0.23, p = 0.82] between the aided soundfield hearing threshold in the HA group (mean = 47.96, SD = 8.30) and the unaided soundfield hearing threshold in the UT group (mean = 47.50, SD = 10.69). These results suggest that although the HA group possessed more severe unaided HL than did the UT group, aided hearing thresholds in the HA group were comparable to the unaided hearing thresholds in the UT group (i.e., audibility) when measured in the soundfield. However, the UT group was significantly better at perceiving sentences in noise (mean = 2.69, SD = 4.52) than the HA group (mean = 5.87, SD = 4.10) [independent t-test, t (93) = −3.58, p = 0.001; see **Table 1**]. Furthermore, using a cutoff score of 22 for the Cantonese version of MoCA, as recommended by Yeung et al. (2014), 27% of participants in the UT group, and 26% of participants in the HA group, were regarded as exhibiting cognitive impairment.

TABLE 1 | Demographic characteristics, speech recognition thresholds, and cognitive functions of the untreated (UT) and hearing aid (HA) groups.

	UT group	HA group
Mean age in years (SD) [range]	71.41 (6.33) [61–87]	68.74 (5.05) [60–83]
Gender (%male/%female)	51%/49%	41%/59%
Education level		
Uneducated (0 years)	14%	0%
Primary (1–6 years)	43%	63%
Secondary (7–13 years)	41%	37%
Tertiary (>13 years)	2%	0%
Hearing level		
Unaided soundfield hearing thresholds at 0.5, 1, 2, and 4 kHz in dB HL (SD) [range]	47.50 (10.69) [27.50–67.5]	64.89 (10.73) [41.25–88.75]
Aided bilateral soundfield hearing thresholds at 0.5, 1, 2, and 4 kHz in dB HL (SD) [range]	N/A	47.96 (8.30) [28.75–65.00]
SRT		
Quiet in dBA (SD)	56.16 (10.90) (unaided)	61.49 (9.40) (aided)
Noise from front in dB SNR (SD)	3.29 (4.34) (unaided)	5.67 (4.02) (aided)
Noise from better ear side in dB SNR (SD)	3.13 (4.99) (unaided)	7.23 (4.98) (aided)
Noise from worse ear side in dB SNR (SD)	0.89 (5.29) (unaided)	4.66 (5.11) (aided)
*Noise composite in dB SNR (SD)	2.69 (4.52) (unaided)	5.87 (4.11) (aided)
Cognitive assessment		
MoCA, general cognitive function (SD)	23.86 (4.32)	23.85 (4.18)
OBT, working memory (SD)	1.13 (0.21)	1.09 (0.24)
GML, executive function (SD)	118.94 (77.12)	115.89 (57.61)
IDN, attention (SD)	2.81 (0.08)	2.81 (0.08)
ISL total, verbal learning (SD)	18.47 (3.85)	20.22 (4.53)
ISL trial 1 (SD)	4.22 (1.48)	4.93 (1.57)
ISL trial 2 (SD)	6.47 (1.76)	7.15 (1.73)
ISL trial 3 (SD)	7.78 (1.77)	8.13 (2.13)

SD, standard deviation; N/A, not applicable; MoCA, Montreal Cognitive Assessment; OBT, One-Back Test; GML, Groton Maze Learning Test; IDN, Identification Test; ISL, International Shopping List Task; UT, untreated; HA, hearing aid; SNR, signal-to-noise ratio.

**A noise composite score was calculated to represent the overall performance in noise, and was based on the performance when noise was from the front and from the side using the following formula: [NF + $^1/_2$ (NBE + NWE)]/2. NF, noise from front; NBE, noise from the better ear side; NEW, noise from worse ear side.*

Factors Associated With Cognitive Functions

The Pearson product-moment correlation coefficients indicated that age was significantly correlated with all cognitive functions, with the exception of the GCF, in the UT group, and GCF, SA, and EF in the HA group. The soundfield hearing threshold was significantly correlated with GCF, WM, and EF in the UT group. Sentence perception in noise was significantly correlated with GCF, WM, EF in the UT group, and GCF, and VL in the HA group. Education level was correlated with all cognitive functions, with the exception of VL, in the UT group, and significantly correlated with EF in the HA group (**Tables 2**, **3**).

Untreated Group

Multiple linear regression analyses showed that SRTs in noise were significantly related to WM and EF, accounting for 25% of the variance of these cognitive functions (**Table 4**). In addition, when education level was included in these analyses, a further 8 and 14% of the variance in WM and EF could be further explained, respectively. Together, SRTs in noise and education level accounted for 30–39% of the variance in GCF, WM, and EF.

However, SRTs in noise were not significantly related to SA or VL, while 20% of the variance in SA was accounted for by education level alone. Finally, age was the only variable related to VL, accounting for 15% of the variance.

Hearing Aid Group

Speech reception thresholds in noise were significantly associated with GCF and VL, accounting for 9 and 21% of the variance in GCF and VL, respectively (**Table 5**). Education level was the only variable associated with EF, accounting for 11% of the variance. Age significantly contribute to the WM. No factor was significantly associated with SA in the HA group.

DISCUSSION

Audiometric Thresholds Versus Speech Perception

Several previous studies reported that cognitive function weakly related to unaided audiometric thresholds which ranged from normal to severe. For example, Harrison Bush et al. (2015) reported that unaided audiometric thresholds of the better ear, measured under headphones, accounted for 0.9, 0.6–1, 0.5–1.7, and 0.4–2.2% of the variance in GCF, speed of processing, EF, and memory, respectively, in 894 older adults from the Staying

Keen in Later Life study cohort. Baltes and Lindenberger (1997) reported that unaided audiometric thresholds measures under headphones accounted for 1.1% of the variance of a cognitive composite score. Additionally, Valentijn et al. (2005) found that the effect sizes of the relationship between unaided audiometric thresholds for the better ear measured under headphones and cognitive functions (e.g., verbal memory, attention, speed of information processing, cognitive flexibility) were small ($R^2 \leq 0.01$). However, Anstey et al. (2001) did not find a relationship between unaided bilateral audiometric thresholds measured under headphones and cognitive function. Similarly, in the current study, unaided soundfield audiometric thresholds in the UT group and aided soundfield audiometric thresholds in the HA group did not account for any variance in the evaluated cognitive functions.

In the present study, speech perception in noise alone accounted for 13–25% of the variance in GCF, WM, and EF in the UT group and 9–15% of the variance in GCF and VL in the HA group. Thus, the association between speech perception in noise and cognitive function found in the present study was much stronger than that in other studies reporting a significant relationship between audiometric thresholds and cognitive function, as discussed above. As detection of pure tones in audiometric testing depends on cochlear transduction and neuronal afferents to brainstem nuclei, and speech perception in noise involves higher auditory cortical processing taxing cognitive resources (Lin et al., 2013), it is not surprising that SRTs exhibited a stronger relationship with cognitive performance than did audiometric thresholds, especially for the ISL test, which requires an understanding of the test stimuli.

Relationships Between Speech Perception in Noise and Cognition Functions

The finding that WM and EF were associated with speech perception in noise could be interpreted under the framework of the Ease of Language Understanding (ELU) Model (Rönnberg et al., 2013). According to this model, WM and EF could come into play when there is mismatch between perceptual input (e.g., phonology, prosody, syntax, and semantics) and the phonological representation stored in long-term memory. This mismatch is more severe in listeners with HL as background noise and HL could both introduce this mismatch (Rönnberg et al., 2019). This may explain why speech perception in noise was significantly associated with WM and

TABLE 2 | Correlations among variables in the UT group.

	Age	Education level	Unaided soundfield hearing thresholds	Sentence perception in noise	General cognitive function	sustained Attention	Working memory	Executive function
Education level	**−0.36***							
Unaided soundfield hearing thresholds	**0.54****	−0.28						
Sentence perception in noise	**0.47****	−0.12	**0.64****					
General cognitive function	−0.22	**−0.41****	**−0.29***	**−0.41****				
Sustained attention	**0.35***	**−0.45****	0.01	0.18	**0.42****			
Working memory	**0.42****	**0.45****	**−0.40****	**−0.51****	**0.52****	−0.26		
Executive function	**0.41****	**−0.35***	**0.30***	**0.50****	**−0.56****	**0.40****	**−0.56****	
Verbal learning	**−0.39****	0.23	−0.23	−0.16	**−0.29***	−0.25	0.20	**-0.31***

*Correlation is significant at the 0.05 level (2-tailed).
**Correlation is significant at the 0.01 level (2-tailed).
Correlation coefficients marked with bold indicate statistically significant relationships.

TABLE 3 | Correlations among variables in the HA group.

	Age	Education level	Aided soundfield hearing thresholds	Sentence perception in noise	General cognitive function	sustained Attention	Working memory	Executive function
Education level	−0.00							
Aided soundfield hearing thresholds	−0.04	−0.14						
Sentence perception in noise	0.26	−0.11	**0.41****					
General cognitive function	−0.27	0.18	−0.14	**−0.30***				
Sustained attention	0.12	−0.06	−0.01	0.21	−0.17			
Working memory	**-0.32***	0.20	0.06	−0.27	**0.42****	−0.21		.
Executive function	0.17	**−0.33***	0.14	0.06	**−0.37***	0.00	**−0.42****	
Verbal learning	**−0.31***	0.22	−0.05	**−0.46****	**0.31***	**−0.34****	0.19	-0.18

*Correlation is significant at the 0.05 level (2-tailed).
**Correlation is significant at the 0.01 level (2-tailed).
Correlation coefficients marked with bold indicate statistically significant relationships.

TABLE 4 | Results from multiple linear regression analyses in the UT group.

Model	B	SE B	Beta	R^2
General cognitive function (MoCA)				
Step 1				0.17
Constant***	18.34	1.88		
Education level**	2.39	0.78	−0.41	
Step 2				0.30
Constant***	19.86	1.82		
Education level**	2.14	0.73	−0.37	
SRTs in noise**	−0.35	0.12	0.36	
Working memory				
Step 1				0.25
Constant***	1.19	0.03		
SRTs in noise***	−0.02	0.01	−0.51	
Step 2				0.39
Constant***	0.93	0.08		
SRTs in noise***	−0.02	0.01	−0.46	
Education level**	0.11	0.03	0.39	
Executive function				
Step 1				0.25
Constant***	96.06	11.26		
SRTs in noise***	8.52	2.16	0.50	
Step 2				0.33
Constant***	167.40	31.64		
SRTs in noise**	7.92	2.07	0.46	
Education level*	−30.24	12.62	−0.29	
Sustained attention				
Step 1				0.20
Constant***	2.92	0.03		
Education level**	−0.05	0.01	−0.45	
Verbal learning				
Step 1				0.15
Constant***	35.39	5.85		
Age**	−0.24	0.08	−0.39	

Independent variables included age, education level, soundfield hearing thresholds, and SRTs in noise. Education level was coded as 0 = uneducated, 1 = primary school, 2 = secondary school, and 3 = tertiary school or above. UT, untreated; MoCA, Montreal Cognitive Assessment; SRT, speech reception threshold; B, raw coefficients; SE B, standard error of b; ΔR^2, the change of R^2 for each subsequent step.

$*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

TABLE 5 | Results from multiple linear regression analyses in the HA group.

Model	B	SE B	Beta	R^2
General cognitive function (MoCA)				
Step 1				0.09
Constant***	25.63	1.04		
SRTs in noise*	−0.31	0.15	-0.30	
Working memory				
Step 1				0.10
Constant***	2.19	0.46		
Age*	−0.02	0.01	-0.34	
Executive function				
Step 1				0.11
Constant***	207.68	40.65		
Education level*	−38.74	16.81	-0.33	
Sustained attention				
	No variables were entered			
Verbal learning				
Step 1				0.21
Constant***	23.17	1.06		
SRTs in noise**	−0.50	0.15	-0.46	

Independent variables included age, education level, aided soundfield hearing thresholds, and SRTs in noise. Education level was coded as 0 = uneducated, 1 = primary school, 2 = secondary school, and 3 = tertiary school or above.

HA, hearing aid; MoCA, Montreal Cognitive Assessment; SRT, speech reception threshold; B, raw coefficients; SE B, standard error of b; ΔR^2, the change of R^2 for each subsequent step.

$*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

EF in the UT group. However, this relationship has not been found in all listeners. For example, Füllgrabe and Rosen (2016) found that in young normal-hearing listeners, WM capacity only explained 2% of the variance in speech-in-noise identification scores. Similarly, we did not find this relationship in the HA group.

HA signal processing could also cause mismatch between perceptual input and the phonological representation, and previous studies have reported WM significantly correlated with aided speech-in-noise perception in older listeners with HL (Lunner, 2003; Foo et al., 2007; Rudner et al., 2008, 2009, 2011). However, after consistent exposure to information despite being distorted via HA, newly established and recalibrated phonological representations would gradually supplement the existing long-term memory representations in the lexicon. With the establishment of these new long-term memory representations, the role of WM/EF in speech perception would decrease (Rönnberg et al., 2013). This prediction, based on the ELU model, has been verified by Ng and Rönnberg (2020), who found that the relationship between WM and speech perception in noise decreased as HA experience increased. In the current study, HA users had 2–10 years (mean = 7.25 years) of HA use experience. Some of the participants might have already developed new long-term memory representations, decoupling the relationship between WM and speech perception in noise.

One might also argue that the same logic (i.e., acclimatization) should be applied to the UT group. That is, sentence perception in noise should not significantly associate with WM and EF in the UT group due to acclimatization. The average SNR in daily environments is approximately 5 dB (Smeds et al., 2015), which approximates the mean aided SRT in the HA group (i.e., 5.87 dB), and is much higher than the mean SRT in the UT group (2.83 dB) (**Table 1**). This suggests that the HA group had more opportunities and experience in practicing speech perception at an SNR close to the SRT in daily situations. The reliance on WM/EF in speech perception may thus decrease. This speculation could be verified by including another UT group with SRTs matched to those in the HA group in a future study. The lack of a relationship between WM/EF and sentence perception in noise in both groups would support this speculation.

Although aided audiometric hearing thresholds in the HA group were comparable to the unaided hearing thresholds in the UT group, the UT group was better at perceiving sentences in noise than the HA group. This may be attributed to greater suprathreshold auditory processing deficits associated with more severe HL (as observed in the HA group in the present study) (Füllgrabe and Moore, 2018). This deficit could also have affected performance in cognitive measures (Füllgrabe, 2020), although we checked to ensure participants had no problem hearing the test instructions and test stimuli. Such deficits may, thus, explain the significant relationship between speech perception in noise and VL in the HA group. Meister et al. (2013) has also reported a significant relationship between speech perception in noise and VL in older adults, and speculated that the ability to process fine structural cues may mediate the relationship. Assessment of suprathreshold processing deficits in this group in a future study could verify this speculation.

In the present study, age did not contribute to GCF, WM, EF, and SA in the UT group, and GCF, EF, SA, and VL in the HA group, when the effects of education level were controlled. These findings do not necessarily contrast with those from Western societies, for which a relationship between age and cognitive function has been reported (Naveh-Benjamin, 2000; Fisk and Sharp, 2004; Treitz et al., 2007). As mentioned above, the education level of the older Chinese adults in the current study was much lower than that of the participants in these other studies. Education level, which also was significantly related to age in the UT group, overshadow the effects of age on WM, EF, SA in the regression analyses (**Tables 2**, **4**). As for the HA group, HA may affect the developmental trajectories of the cognitive functions and thus, altering the relationship between cognitive functions and age. This speculation could be examined using a longitudinal study with randomization of participants to the HA and UT group.

There are several limitations to the present study. First, potential confounders, such as social participation, perceived hearing difficulties, personality traits, attitudes, income, and other health aspects, and their impact on cognitive function might not have been equivalent in the two groups. Thus, the cognitive skills in the HA and UT groups were not directly compared to examine whether HA use enables older adults to retain cognitive function to a similar level as those with milder, unaided HL. To better control bias due to these confounders, a longitudinal study with a larger sample size and randomization of participants to HA and UT groups could be carried out.

Second, although the OBT scores in the HA group were slightly skewed (skewness = 0.38), those in the UT group were moderately skewed (skewness = 0.70), suggesting that the OBT was relatively easy for this group, and accordingly, the relationship between WM and speech perception in noise might have been obscured.

Third, the current study included a wide range of HA brands and models. HA features (e.g., compression parameters, noise reduction, and directional microphone) were not included in the analysis. As some HA features (e.g., compression parameters) have been found to be significantly related to speech perception in noise (Stone et al., 2009b; Chen et al., 2021), it is possible that these HA features might have mediated the relationship between cognitive function and sentence perception in noise in the HA group.

Fourth, there was no significant difference in audibility (aided soundfield hearing thresholds in the HA group and unaided soundfield hearing thresholds in the UT group) between the HA and UT groups. Whether these results could generalize to other populations, in which audibility between UT group and HA groups differ, requires further study.

Finally, although a relationship between speech perception and cognitive function was established in the current study, little is known about the causality or underlying mechanisms of this relationship. The current study has demonstrated which cognitive functions are more likely associated with HL in UT and HA populations, and this information could be used to understand the potential mechanisms underlying auditory sensory and cognitive functions in future studies.

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 Human Research Ethics Committee at the University of Hong Kong (EA1706014) and the Joint Chinese University of Hong Kong—New Territories East Cluster Clinical Research Ethics Committee (CRE-2013.481). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SSC and JY collected the data. LW, YC, JY, and SSC contributed to the study design. LW, YC, and SSC interpreted, wrote, and reviewed this publication. All authors approved the final manuscript.

FUNDING

This research project was supported by the Faculty Research Fund, administered by the Faculty of Education, the University of Hong Kong, and Internal Research Grant (RG50/18-19R) funded by the Committee on Research and Development, the Education University of Hong Kong. The funders did not have a role in the design of the study, in the collection, analysis, and interpretation of data, or in the writing of the manuscript.

ACKNOWLEDGMENTS

The authors are grateful for the participation of their patients from the two hospitals, as well as for the help from various research assistants in data collection.

REFERENCES

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *Int. J. Audiol.* 47(Suppl. 2), S53–S71. doi: 10.1080/14992020802301142

Anstey, K. J., Luszcz, M. A., and Sanchez, L. (2001). Two-year decline in vision but not hearing is associated with memory decline in very old adults in a population-based sample. *Gerontology* 47, 289–293. doi: 10.1159/000052814

Arlinger, S., Lunner, T., Lyxell, B., and Pichora-Fuller, M. K. (2009). The emergence of cognitive hearing science. *Scand. J. Psychol.* 50, 371–384. doi: 10.1111/j.1467-9450.2009.00753.x

Baltes, P. B., and Lindenberger, U. (1997). Psychology and aging emergence of a powerful connection between sensory and cognitive functions across the adult life span: A new window to the study of cognitive aging? *Psychol. Aging* 12:12. doi: 10.1037//0882-7974.12.1.12

Bisgaard, N., and Ruf, S. (2017). Findings From EuroTrak Surveys From 2009 to 2015: Hearing Loss Prevalence, Hearing Aid Adoption, and Benefits of Hearing Aid Use. *Am. J. Audiol.* 26, 451–461. doi: 10.1044/2017_aja-16-0135

Carlson, S. M., Zelazo, P. D., and Faja, S. (2013). "Executive function," in *The Oxford handbook of developmental psychology (Vol. 1): Body and mind*, ed. P. D. Zelazo (Oxford: Oxford University Press), 706–743.

Chen, Y., Wong, L. L., Kuehnel, V., Qian, J., Voss, S. C., and Shangqiguo, W. (2021). Can Dual Compression Offer Better Mandarin Speech Intelligibility and Sound Quality Than Fast-Acting Compression? *Trends Hear.* 25:2331216521997610. doi: 10.1097/aud.0000000000000774

Chen, Y., Wong, L. L. N., Qian, J., Kuehnel, V., Christina Voss, S., and Chen, F. (2020). The Role of Lexical Tone Information in the Recognition of Mandarin Sentences in Listeners With Hearing Aids. *Ear. Hear.* 41, 532–538. doi: 10.1097/AUD.0000000000000774

Chien, W., and Lin, F. R. (2012). Prevalence of hearing aid use among older adults in the United States. *Arch. Intern. Med.* 172, 292–293. doi: 10.1001/archinternmed.2011.1408

Dawes, P., Cruickshanks, K. J., Fischer, M. E., Klein, B. E. K., Klein, R., and Nondahl, D. M. (2015). Hearing-aid use and long-term health outcomes: Hearing handicap, mental health, social engagement, cognitive function, physical health, and mortality. *Int. J. Audiol.* 54, 838–844. doi: 10.3109/14992027.2015.1059503

de Jager, C. A., Schrijnemaekers, A. C., Honey, T. E., and Budge, M. M. (2009). Detection of MCI in the clinic: evaluation of the sensitivity and specificity of a computerised test battery, the Hopkins Verbal Learning Test and the MMSE. *Age Ageing* 38, 455–460. doi: 10.1093/ageing/afp068

Doyle, J., and Wong, L. L. (1996). Mismatch between aspects of hearing impairment and hearing disability/handicap in adult/elderly Cantonese speakers: some hypotheses concerning cultural and linguistic influences. *J. Am. Acad. Audiol.* 7, 442–446.

Fisk, J. E., and Sharp, C. A. (2004). Age-related impairment in executive functioning: Updating, inhibition, shifting, and access. *J. Clin. Exp. Neuropsy.* 26, 874–890. doi: 10.1080/13803390490510680

Foo, C., Rudner, M., Rönnberg, J., and Lunner, T. (2007). Recognition of speech in noise with new hearing instrument compression release settings requires explicit cognitive storage and processing capacity. *J. Am. Acad. Audiol.* 18:618. doi: 10.3766/jaaa.18.7.8

Fritze, T., Teipel, S., Óvári, A., Kilimann, I., Witt, G., and Doblhammer, G. (2016). Hearing impairment affects dementia incidence. an analysis based on longitudinal health claims data in germany. *PLoS One* 11:e0156876. doi: 10.1371/journal.pone.0156876

Füllgrabe, C. (2020). On the possible overestimation of cognitive decline: the impact of age-related hearing loss on cognitive-test performance. *Front. Neurosci.* 14:454. doi: 10.3389/fnins.2020.00454

Füllgrabe, C., and Moore, B. C. (2018). The association between the processing of binaural temporal-fine-structure information and audiometric threshold and age: A meta-analysis. *Trends Hear.* 22:2331216518797259. doi: 10.1177/2331216518797259

Füllgrabe, C., and Rosen, S. (2016). On the (un)importance of working memory in speech-in-noise processing for listeners with normal hearing thresholds. *Front. Psychol.* 7:1268. doi: 10.3389/fpsyg.2016.01268

Gallacher, J. (2005). Hearing, cognitive impairment and aging: a critical review. *Rev. Clin. Gerontol.* 14, 199–209. doi: 10.1017/S095925980500153X

Gong, R., Hu, X., Gong, C., Long, M., Han, R., Zhou, L., et al. (2018). Hearing loss prevalence and risk factors among older adults in China. *Int. J. Audiol.* 57:354. doi: 10.1080/14992027.2017.1423404

Gurgel, R. K., Ward, P. D., Schwartz, S., Norton, M. C., Foster, N. L., and Tschanz, J. T. (2014). Relationship of hearing loss and dementia: a prospective, population-based study. *Otol. Neurotol.* 35, 775–781. doi: 10.1097/mao.0000000000000313

Harrison Bush, A. L., Lister, J. J., Lin, F. R., Betz, J., and Edwards, J. D. (2015). Peripheral Hearing and Cognition: Evidence From the Staying Keen in Later Life (SKILL) Study. *Ear. Hear.* 36, 395–407. doi: 10.1097/aud.0000000000000142

Humes, L. E. (2007). The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults. *J. Am. Acad. Audiol.* 18, 590–603. doi: 10.3766/jaaa.18.7.6

Humes, L. E., Kidd, G. R., and Lentz, J. J. (2013). Auditory and cognitive factors underlying individual differences in aided speech-understanding among older adults. *Front. Syst. Neurosci.* 7:55. doi: 10.3389/fnsys.2013.00055

Institute of Human Communicative Research (2005). *Hearing problems of older adults in Hong Kong. Synopsis*. Hong Kong: The Chinese University of Hong Kong.

Lin, F. R., Ferrucci, L., Metter, E. J., An, Y., Zonderman, A. B., and Resnick, S. M. (2011). Hearing loss and cognition in the Baltimore Longitudinal Study of Aging. *Neuropsychology* 25, 763–770. doi: 10.1037/a0024238

Lin, F. R., Yaffe, K., Xia, J., Xue, Q. L., Harris, T. B., and Purchase-Helzner, E. (2013). Hearing Loss and Cognitive Decline Among Older Adults. *JAMA Internal. Med.* 173:4. doi: 10.1001/jamainternmed.2013.1868

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

Lunner, T. (2003). Cognitive function in relation to hearing aid use. *Internat. J. Audiol.* 42, S49–S58. doi: 10.3109/14992020309074624

Lunner, T., and Sundewall-Thorén, E. (2007). Interactions between cognition, compression, and listening conditions:Effects on speech-in-noiseperformance in a two-channel hearing aid. *J.Am Acad Audiol* 18, 604617. doi: 10.3766/jaaa.18.7.7

Maharani, A., Dawes, P., Nazroo, J., Tampubolon, G., and Pendleton, N. (2018). Longitudinal Relationship Between Hearing Aid Use and Cognitive Function in Older Americans. *J. Am. Geriatr. Soc.* 66, 1130–1136. doi: 10.1111/jgs.15363

Meister, H., Schreitmüller, S., Grugel, L., Beutner, D., Walger, M., and Meister, I. (2013). Examining speech perception in noise and cognitive functions in the elderly. *Am. J. Audiol.* 22, 310–312. doi: 10.1044/1059-0889(2012/12-0067)

National Bureau of Statistics in China (2011). *Sixth National Population Census of the People's Republic of China*. Beijing: National Bureau of Statistics in China.

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: tests of an associative deficit hypothesis. *J. Exp. Psychol.* 26, 1170–1187.

Ng, E. H. N., and Rönnberg, J. (2020). Hearing aid experience and background noise affect the robust relationship between working memory and speech recognition in noise. *Int. J. Audiol.* 59, 208–218.

Pichora-Fuller, M. K., and Singh, G. (2006). Effects of age on auditory and cognitive processing: implications for hearing aid fitting and audiologic rehabilitation. *Trends Amplif.* 10, 29–59. doi: 10.1177/108471380601000103

Plack, C. J. (2014). *The Sense of Hearing*. Hove: Psychology Press, 978–1848725157.

Rönnberg, J., Holmer, E., and Rudner, M. (2019). Cognitive hearing science and ease of language understanding. *Internat. J. Audiol.* 58, 247–261. doi: 10.1080/14992027.2018.1551631

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., et al. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Front. Syst. Neurosci.* 7:31.

Rudner, M., Foo, C., Rönnberg, J., and Lunner, T. (2009). Cognition and aided speech recognition in noise: Specific role for cognitive factors following nine-week experience with adjusted compression settings in hearing aids. *Scand. J. Psychol.* 50, 405–418. doi: 10.1111/j.1467-9450.2009.00745.x

Rudner, M., Foo, C., Sundewall-Thorén, E., Lunner, T., and Rönnberg, J. (2008). Phonological mismatch and explicit cognitive processing in a sample of 102 hearing-aid users. *Internat. J. Audiol.* 47, S91–S98. doi: 10.1080/14992020802304393

Rudner, M., Rönnberg, J., and Lunner, T. (2011). Working memory supports listening in noise for persons with hearing impairment. *J. Am. Acad. Audiol.* 22, 156–167.

Smeds, K., Wolters, F., and Rung, M. (2015). Estimation of Signal-to-Noise Ratios in Realistic Sound Scenarios. *J. Am. Acad. Audiol.* 26:183. doi: 10.3766/jaaa.26.2.7

Stone, M. A., Füllgrabe, C., and Moore, B. C. (2009a). High-rate envelope information in many channels provides resistance to reduction of speech intelligibility produced by multi-channel fast-acting compression. *J. Acoust. Soc. Am.* 126, 2155–2158.

Stone, M. A., Moore, B. C., Füllgrabe, C., and Hinton, A. C. (2009b). Multichannel fast-acting dynamic range compression hinders performance by young, normal-hearing listeners in a two-talker separation task. *J. Audio Eng. Soc.* 57, 532–546.

Taljaard, D. S., Olaithe, M., Brennan-Jones, C. G., Eikelboom, R. H., and Bucks, R. S. (2016). The relationship between hearing impairment and cognitive function: a meta-analysis in adults. *Clin. Otolaryngol.* 41, 718–729. doi: 10.1111/coa.12607

Treitz, F. H., Heyder, K., and Daum, I. (2007). Differential course of executive control changes during normal aging. *Aging Neuropsychol. Cogn.* 14, 370–393.

Tun, P. A., Williams, V. A., Small, B. J., and Hafter, E. R. (2012). The effects of aging on auditory processing and cognition. *Am. J. Audiol.* 21, 344–350. doi: 10.1044/1059-0889(2012/12-0030)

Valentijn, S. A. M., Van Boxtel, Martin, P. J., Van Hooren, Susan, A. H., Bosma, H., et al. (2005). Change in sensory functioning predicts change in cognitive functioning: Results from a 6-year follow-up in the maastricht aging study. *J. Am. Geriat. Soc. (JAGS)* 53, 374–380. doi: 10.1111/j.1532-5415.2005.53152.x

Wayne, R. V., and Johnsrude, I. S. (2015). A review of causal mechanisms underlying the link between age-related hearing loss and cognitive decline. *Ageing Res. Rev.* 23, 154–166. doi: 10.1016/j.arr.2015.06.002

Wingfield, A., Tun, P. A., and McCoy, S. L. (2005). Hearing Loss in Older Adulthood: What It Is and How It Interacts With Cognitive Performance. *Curr. Direct. Psychol. Sci.* 14, 144–148. doi: 10.1111/j.0963-7214.2005.00356.x

Wong, A., Xiong, Y. Y., Kwan, P. W., Chan, A. Y., Lam, W. W., Wang, K., et al. (2009). The validity, reliability and clinical utility of the Hong Kong Montreal Cognitive Assessment (HK-MoCA) in patients with cerebral small vessel disease. *Dement Geriatr. Cogn. Disord* 28, 81–87. doi: 10.1159/000232589

Wong, L. L., and Soli, S. D. (2005). Development of the Cantonese Hearing In Noise Test (CHINT). *Ear. Hear.* 26, 276–289.

Wong, L. L. N., Chen, Y., Wang, Q., and Kuehnel, V. (2018). Efficacy of a Hearing Aid Noise Reduction Function. *Trends Hear.* 22:2331216518782839. doi: 10.1177/2331216518782839

Wong, L. L. N., Yu, J. K. Y., Chan, S. S., and Tong, M. C. F. (2014). Screening of Cognitive Function and Hearing Impairment in Older Adults: A Preliminary Study. *BioMed. Res.* Intern. 2014:7. doi: 10.1155/2014/867852

Yeung, P. Y., Wong, L. L., Chan, C. C., Leung, J. L., and Yung, C. Y. (2014). A validation study of the Hong Kong version of Montreal Cognitive Assessment (HK-MoCA) in Chinese older adults in Hong Kong. *Hong Kong medical journal = Xianggang yi xue za zhi* 20, 504–510. doi: 10.12809/hkmj144219

Zhao, F., Manchaiah, V., St Claire, L., Danermark, B., Jones, L., Brandreth, M., et al. (2015). Exploring the influence of culture on hearing help-seeking and hearing-aid uptake. *Int. J. Audiol.* 54, 435–443. doi: 10.3109/14992027.2015.1005848

Zhong, N., Jiang, H., Wu, J., Chen, H., Lin, S., Zhao, Y., et al. (2013). Reliability and Validity of the CogState Battery Chinese Language Version in Schizophrenia.(Research Article)(Report). *PLoS One* 8:e74258. doi: 10.1371/journal.pone.0074258

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.

frontiers in Digital Health

REVIEW
published: 10 January 2022
doi: 10.3389/fdgth.2021.788103

Check for updates

Tele-Audiology: Current State and Future Directions

Kristen L. D'Onofrio[1*†] *and Fan-Gang Zeng*[2†]

[1] *Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN, United States,*
[2] *Department of Otolaryngology - Head and Neck Surgery, Center for Hearing Research, University of California, Irvine, Irvine, CA, United States*

The importance of tele-audiology has been heightened by the current COVID-19 pandemic. The present article reviews the current state of tele-audiology practice while presenting its limitations and opportunities. Specifically, this review addresses: (1) barriers to hearing healthcare, (2) tele-audiology services, and (3) tele-audiology key issues, challenges, and future directions. Accumulating evidence suggests that tele-audiology is a viable service delivery model, as remote hearing screening, diagnostic testing, intervention, and rehabilitation can each be completed reliably and effectively. The benefits of tele-audiology include improved access to care, increased follow-up rates, and reduced travel time and costs. Still, significant logistical and technical challenges remain from ensuring a secure and robust internet connection to controlling ambient noise and meeting all state and federal licensure and reimbursement regulations. Future research and development, especially advancements in artificial intelligence, will continue to increase tele-audiology acceptance, expand remote care, and ultimately improve patient satisfaction.

Keywords: tele-audiology, telemedicine, telehealth, hearing aid, cochlear implant, hearing loss, smartphone, tablet

OPEN ACCESS

Edited by:
Timo Jämsä,
University of Oulu, Finland

Reviewed by:
Pasin Israsena,
National Electronics and Computer Technology Center, Thailand
Rosie Dobson,
The University of Auckland, New Zealand

***Correspondence:**
Kristen L. D'Onofrio
kristen.l.donofrio.1@vumc.org

†ORCID:
Kristen L. D'Onofrio
orcid.org/0000-0002-2312-655X
Fan-Gang Zeng
orcid.org/0000-0002-4325-2780

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: *01 October 2021*
Accepted: *13 December 2021*
Published: *10 January 2022*

Citation:
D'Onofrio KL and Zeng F-G (2022) Tele-Audiology: Current State and Future Directions. Front. Digit. Health 3:788103. doi: 10.3389/fdgth.2021.788103

INTRODUCTION

Telemedicine is defined as "the delivery of healthcare services and information via high-tech telecommunications technologies" (1). This delivery model has been used in various forms since the late 1950s as a means of providing remote services to underserved populations who would not otherwise have access to care (2). Telemedicine has been adapted to the field of audiology, known as *tele-audiology*, to provide remote hearing screenings, diagnostic testing, intervention, and/or rehabilitation services (e.g., hearing aid adjustment, cochlear implant programming) (3). Recently the COVID-19 pandemic has caused widescale disruption to healthcare services worldwide and has consequently accelerated the need for a remote hearing healthcare service model (4, 5). The present review examines the history and current state of tele-audiology care, while also discussing its current limitations and potential future directions.

In 2021, the World Health Organization (WHO) estimated that 1 in 5 or about 1.5 billion people have some degree of hearing loss. Of those, approximately one-third need hearing healthcare intervention, and 80% live in low- and middle-income nations. Despite the global prevalence of hearing loss, only a fraction of those who could benefit from hearing healthcare services actually receive care (6).

Of hearing aid candidates, Grundfast and Liu (7) reported that the rate of hearing aid adoption is 33%. More recent literature suggests that number is closer to 17% and may be significantly less in low- and middle-income areas, such as Africa at ∼10% (8). With respect to cochlear implantation, market penetration in developed countries has been estimated at only 20% (9)—and even lower by more recent data [4.4% (10); <10% (11)]. In developing countries, that number is <1% (9). Indeed, it is estimated that under 15% of cochlear implant (CI) candidates have been implanted worldwide (12, 13).

Failure to receive appropriate intervention comes at an enormous price. The global economic impact of unaddressed hearing loss is \$980 billion, annually (6). The impact of inadequate hearing care can also be devasting to an individual's well-being and development. If children experience delays in hearing loss detection and intervention, they may suffer deleterious effects on language development, literacy, social and mental well-being, and overall academic progress (14, 15). Similarly, adults with untreated hearing loss may experience social isolation, reduced quality of life, and poorer academic and/or job performance. To that end, hearing loss has contributed to higher rates of both unemployment and under-employment (16). Moreover, hearing loss is the number one modifiable factor that can reduce or even prevent Alzheimer's disease and dementia (17).

Thus, it begs the question—what most contributes to the lack of care, and how can this be improved? While reasons vary from community-to-community and from individual-to-individual, common factors include cost of services (18–20), a shortage of clinical providers/specialized services (21–26), inconvenience of travel (24, 27), communication barriers (28, 29), and racial/ethnic disparities (30–35). The COVID-19 pandemic, with its associated lockdowns and social distancing requirements, led to mass disruption in the delivery of healthcare services worldwide. For example, Alqudah et al. (36) found that daily hearing aid usage dropped dramatically during the pandemic—a finding potentially caused by limited access to device programming, maintenance and/or repair services, a shortage of hearing aid batteries, and/or the cessation of speech and language therapy sessions. Thus, providers were forced to accelerate implementation of tele-audiology as a means of continuing care during the pandemic. While traditional, in-person care remains the "gold standard," tele-audiology may mitigate several of the aforementioned challenges, expand the reach of audiological care to underserved populations, and help narrow the gap between those who could benefit from services and those who ultimately receive care.

METHODS

A review of the literature was completed between September 2020—November 2021. The search was primarily conducted by the first author (KD) with additional contribution from the second author (FGZ) using two electronic databases, PubMed and Google Scholar. Eligible peer-reviewed articles were searched for using combinations of keywords including telemedicine, tele-audiology, telehealth, audiology, hearing, hearing aid, cochlear implant, otoscopy, audiometry, DPOAE, ABR, smartphone, tablet. For inclusion, articles were required to address the following: (1) barriers to hearing healthcare, (2) tele-audiology services, and/or (3) tele-audiology key issues, challenges, and future directions. Articles were excluded if they were not peer-reviewed, if they were not written in English, and if they were not directly germane to the topics and purpose of this review. Based on our specified criteria, our literature review identified 70 total articles, including 19 articles on barriers to hearing healthcare, 45 articles on tele-audiology services (4 home-based otoscopy studies, 12 tele-audiology hearing screening studies, including 8 audiometric screening and 4 distortion product otoacoustic emissions (DPOAE) testing studies, 16 tele-audiology diagnostic studies, including 11 audiometry studies and 5 ABR studies, and 13 tele-audiology intervention and rehabilitation studies, including 4 on hearing aids and 9 on cochlear implants), and 6 articles on challenges and future directions.

BARRIERS TO HEARING HEALTHCARE

Cost

Hearing technology, including its associated components (e.g., batteries), can be financially burdensome to patients (19). It is recommended that the cost of hearing technology not exceed 3% of the gross national income (GNI) per capita (37). Therefore, with a GNI per capita in the United States of \$65,910.00 in 2019, the cost of hearing technology should not exceed \$1977.30. With a GNI per capita in Ethiopia of \$890.00 in 2020, the cost of hearing technology should not exceed \$26.70 (38). However, the cost of CI technology in developing countries can be as much as 30 times one's annual income (20). Thus, despite an exponential increase in CI market size over the past 2–3 decades, this growth has not resulted in a commensurate reduction in unit price. The resulting high unit prices severely limit accessibility and adoption rates of hearing technology (20).

In addition to being a financial hardship for patients, hearing healthcare providers can also be burdened by the high purchase price of testing/programming equipment (e.g., audiometer, hearing aid fitting software, cochlear implant programming software), as well as routine equipment maintenance. In areas that lack repair and/or calibration services, necessary maintenance can be cost prohibitive as it may require sending equipment overseas (18). Thus, cost may not only prevent some patients from seeking out hearing healthcare and/or hearing technology, it may also prevent providers from offering specialized services.

Providers/Specialized Services

The number of clinical professionals and/or specialized services is disproportionately low in developing countries. Compared with one audiologist per ∼25,000 individuals in the United States (39), the proportion of audiologists in developing countries ranges from ∼1 per 500,000 to 1 per 6,250,000 individuals (25).

The shortage of providers and/or specialized services also exists in many rural and remote communities in developed countries. Powell et al. (26) cited a lack of audiology providers as a barrier for adults with hearing loss in rural parts of Kentucky, USA. Similarly, Barr et al. (21) reported that children with hearing loss in rural parts of Canada and the United States may experience barriers to specialized services (e.g., hearing aids, cochlear implants, therapy services) when compared to those in urban communities. Children in rural parts of Kentucky were less likely to access specialized healthcare services and were more likely to have a delayed diagnosis compared to those in urban areas (22, 23). With a delay in diagnosis, a delay in intervention is inevitable. To that end, Bush et al. (24) reported that children in rural communities were fitted with hearing aids at a median age of 11 months and cochlear implants at 42 months, while children in urban communities were fitted with hearing aids at a median age of 6 months and cochlear implants at 23 months. Thus, children in rural areas experienced about a 5-month delay for hearing aids, and a nearly 20-month delay for cochlear implantation.

Travel

Inconvenience, time, and cost associated with travel to an on-site clinic can be a significant barrier to care, especially for those in developing countries or rural areas of developed countries. Bush et al. (24) reported average travel distance for patients in rural communities was 96 miles, compared to just 13.5 miles in urban communities. Further, they reported that the distance a patient lived from hearing healthcare services was moderately correlated ($r = 0.5$) with the delay in both hearing aid and cochlear implant intervention.

While travel distance is often shorter for patients in non-rural areas, unique travel challenges also exist for patients in urban communities. For example, patients in urban communities cite overall health issues, mobility challenges, and/or difficulty accessing transportation as barriers to care—all of which can contribute to a lack of or a reduction in traditional, in-person healthcare visits [e.g., (27)].

Communication

Patients or families of patients with hearing loss routinely report a lack of communication or information as a barrier to care, especially regarding available services and financial support. In rural Kentucky, primary care physicians were found to have limited resources and knowledge about hearing loss (28), which can be a barrier to appropriate, timely referrals and intervention (29).

Racial/Ethnic Disparities

Minority groups face additional barriers to care, and importantly, disparities may be present even when minorities have reasonable access to care, have advanced levels of education, and are of high socioeconomic status (31, 40). With respect to hearing healthcare specifically, previous studies have demonstrated significantly higher rates of hearing aid use among White older adults as compared to minority groups (28.6–35.4% of White older adults vs. 10–17.1% of minorities) (30, 32, 34). In pediatric patients, Zhang et al. (35) found a relationship between race/ethnicity and the time delay between a failed newborn hearing screening and initial auditory brainstem response (ABR) testing. The mean interval for White patients was 6.3 months [standard deviation (SD): 5.6 months] compared to 12.3 months (SD: 11.8 months) for racial/ethnic minority patients. The mean difference between White and minority patients was 6.0 months (95% confidence interval: 2.3–9.7 months). Mahendran et al. (33) likewise reported racial disparities in hearing healthcare, specifically that cochlear implant evaluations and implantation rates were disproportionately lower among Black patients. Furthermore, they found that hearing was significantly worse among Black patients at the time of referral for cochlear implantation as compared to their White and Asian counterparts.

TELE-AUDIOLOGY SERVICES

Home-Based Otoscopy

The efficacy of home-based otoscopy is of particular interest, as tele-otoscopy holds significant potential in reducing expenditure and costs associated with travel to a clinic. About 80% of children have at least one episode of acute otitis media (AOM) prior to age 3. In fact, otitis media is the most frequent cause of healthcare visits in the pediatric population. Because AOM diagnosis is heavily dependent upon visualization of the tympanic membrane via otoscopy, home-based otoscopy devices can allow parents/caregivers to complete otoscopy and then transmit videos to a physician for remote diagnosis and treatment recommendations. While the viability of smartphone otoscope use by healthcare professionals has previously been demonstrated [e.g., (41, 42)], its home-based use relies on parents or caregivers to independently perform an otoscopic exam. The ability of a physician to make an accurate diagnosis depends entirely on the quality of tympanic membrane visualization.

Shah et al. (43) evaluated the reliability of the CellScope iPhone device (CellScope, Inc., San Francisco, CA) for at-home use and subsequent remote diagnosis of AOM. Participants included children between the ages of 3 months and 17 years. The results revealed low inter-rater agreement between parent-obtained iPhone video-otoscopy and conventional otoscopy by a physician, and high inter-rater agreement between physician-obtained iPhone video-otoscopy and conventional otoscopy. Thus, tele-otoscopy was able to be completed successfully via use of a smartphone, but only when images/videos were obtained by trained healthcare professionals.

Using the same CellScope iPhone device, Erkkola-Anttinen et al. (44) examined whether parents, following a 60-min training session, could perform home-based otoscopy on children between 6 and 35 months of age. The authors found that, *with instruction*, parents are capable of obtaining adequate video of the tympanic membrane. Importantly, physicians were able to detect or exclude presence of AOM in the majority of smartphone otoscopy videos obtained by trained parents. Further, parents reported that at-home use of smartphone otoscopy was both feasible and easy to perform.

Recently, artificial intelligence algorithms have been developed to improve the sensitivity and specificity of home-based otoscopy, making it a reliable and useful alternative to the current standard care model. Chan et al. (45) utilized a smartphone-based machine learning algorithm to detect middle ear fluid in children between the ages of 18 months and 17 years. The authors report a sensitivity and specificity of 85 and 82%, respectively, which is comparable to that of conventional methods (i.e., tympanometry and pneumatic otoscopy). Importantly, parents demonstrated the ability to use the smartphone-based technology with results comparable to trained clinicians, and similar results were obtained across both Android and iPhone smartphone platforms. These results indicate that a smartphone can be used reliably by both parents and/or trained professionals to detect middle ear fluid. Similarly, Cha et al. (46) used a machine learning model to diagnose a variety of ear diseases from 10,544 images of patient tympanic membranes and external auditory canals. Results offer significant promise, as the authors report accuracy via this method was comparable or even *better than* conventional methods.

Hearing Screening

Audiometric Screening

Lancaster et al. (47) compared conventional hearing screenings with hearing screenings completed remotely using portable audiometers in a group of 3rd graders ($n = 32$). On average, the two test methods produced no significant differences, except for a discrepancy in results for 5 of the 32 children. In such cases, the discrepancies were simply due to the lack of a response at one of the three test frequencies (1, 2, or 4 kHz).

More recently, investigators have examined the reliability of tablet-based, computer-based, and smartphone-based audiometry, but results have been mixed. Khoza-Shangase and Kassner (48) compared test results obtained via conventional audiometry and an iPad method, specifically UHear™, in a group of 86 children. Thresholds were significantly poorer with the iPad method as compared to conventional testing, which the authors attributed to higher ambient noise levels, differences in transducers used, and inadequate calibration. On the other hand, Rourke et al. (49) reported that testing with an iPad audiometer was both reliable and cost effective in 220 children between 5 and 11 years of age. Samelli et al. (50) likewise found results obtained by tablet-based testing to be valid and suggested tablet-based screenings may hold particular promise for use in school settings.

Similarly, Dillon et al. (51) examined use of a computer-based hearing screening program (Sound Scouts) which presents stimuli in the form of a game. Participants included 491 children ages 5–14 years ($n = 394$ with normal hearing, $n = 97$ with known hearing loss), as well as adults with normal hearing ($n = 50$). The screening program tested speech-in-quiet, speech-in-noise, and tones-in-noise. The goal of the study was to investigate whether the program was engaging and held the children's attention, whether it detected hearing loss, and whether it could differentiate between conductive and sensorineural hearing loss types. The authors concluded that sensitivity and specificity was sufficiently high, particularly when all three tests were averaged; however, hearing loss type was only identified correctly in two-thirds of cases. Thus, this program is an appropriate hearing screening tool for children in the 5–14 year old age range, particularly when accompanied by follow-up testing to more accurately determine hearing loss type.

Smartphone-based audiometry applications have also been examined for hearing screening use in recent years. In a study of 6,288 children, Wu et al. (52) examined the validity of a smartphone hearing screening application. They determined that although specificity was high (93%), sensitivity was low (37%), thereby suggesting improvements to sensitivity were necessary before widespread use. However, Swanepoel et al. (53) reported no difference in results between smartphone and conventional hearing screenings so long as the smartphone application could be accurately calibrated with noise monitoring in real-time. To this end, Eksteen et al. (54) demonstrated that hearing screenings using an application on a Smartphone connected to supra-aural Sennheiser HD280 headphones performed by staff with minimal training were affordable and offer a promising large-scale service delivery model.

DPOAE Testing

Several studies have examined whether remote synchronous (real-time) DPOAE testing could be completed reliably. Krumm et al. (55) compared DPOAE measures completed in 30 adults via traditional, in-person methods with measures completed remotely (via interactive video and screensharing software). No significant differences were found between the two test methods.

Ciccia et al. (56) likewise examined remote synchronous hearing screenings (pure-tone, DPOAE, and tympanometry) in children 6 years old and younger ($n = 411$) in the United States. Compared to traditional, in-person testing, the reliability was 100% for remote pure-tone and DPOAE screenings, and around 84% for tympanometry. Monica et al. (57) conducted remote synchronous audiometric and DPOAE screenings in school-aged children ($n = 31$) in India using teachers as facilitators and found results similar to traditional, in-person testing. Ameyaw et al. (58) examined remote synchronous DPOAE screenings in a group of newborn infants ($n = 50$) in Ghana. Screenings were completed in-person and remotely by an audiologist via the internet with real-time audio, video, and text messaging between the facilitator and audiologist. Again, no differences were found between the two screening methods. Although these studies provided support for the effectiveness of remote DPOAE screening, they also noted technical challenges, such as high ambient noise levels and slow internet speed in rural areas. Further, they highlighted the need for additional research in difficult-to-test populations (e.g., patients with disabilities).

Diagnostic Testing

Audiometry

Givens and Elangovan (59) compared air and bone conduction thresholds via traditional, in-person audiometry with thresholds completed via tele-audiometry. The audiologist accessed the audiometer at the remote site via an internet connection on his or her own computer. No significant differences were found between traditional and remote testing (mean threshold difference: < 1.3 dB for air conduction, < 1.2 dB for bone conduction). In more recent years, several additional studies have corroborated these findings, suggesting strong reliability and accuracy between tele-audiometry and traditional audiometry for air conduction (60–62), bone conduction (63, 64), and contralateral masking (60, 61).

While many researchers offer support for the reliable use of tele-audiometry in children (65, 66), others have found that due to increased ambient noise/environmental interference, tele-audiometry resulted in more children failing testing than via conventional audiometry (67). For sites where a conventional audiometer and sound-treated test booth are not available, devices such as the KUDUwave 5000 (GeoAxon, Pretoria, South Africa)—a portable audiometer that "uses insert earphones covered by circumaural earphones fitted with internal and external microphones to monitor ambient noise levels" (68)—have been developed to reduce ambient noise levels.

More recently, smartphone-based automated audiometric testing has been examined, as it presents an affordable, cost-effective testing mechanism. van Tonder et al. (69) compared conventional air-conduction audiometry to air-conduction thresholds determined via hearTest, a smartphone-based application for Android devices. Calibrated supra-aural headphones that monitor noise levels in real-time were utilized, and testing was conducted in a soundproof booth. Of the 95 total participants tested, 94.4% of adult smartphone thresholds were within 10 dB of conventional audiometry thresholds—98% for adolescents. Using the same smartphone application (hearTest), Sandström et al. (70) sought to examine its accuracy and reliability in low-income communities. Testing was conducted in a non-sound treated environment. Of the 63 total participants tested, 80.1% of adult smartphone thresholds were within 10 dB of conventional audiometry thresholds—threshold agreement was lowest at 500 Hz (69.4%) and highest at 2,000 Hz (88.8%). Sensitivity for hearing loss detection was 90.6%, and specificity was 94.2%. The authors suggest results indicate a satisfactory mean difference between the hearTest smartphone application and conventional audiometry, though additional noise monitoring could improve agreement, particularly in the low frequencies. Thus, the hearTest smartphone application provides a low-cost method for obtaining air-conduction thresholds with sufficient accuracy and reliability; however, thresholds are best-obtained when real-time noise monitoring is incorporated, particularly in settings with unfavorable levels of ambient noise.

Online testing and machine learning will likely further improve not only the efficiency of tele-audiometry but more importantly its diagnostic power. Barbour et al. (71) compared an online machine learning audiogram method with the traditional modified Hughson-Westlake method also completed online. Adults between 19 and 79 years of age ($n = 21$) completed air conduction pure-tone audiometry. Similar threshold estimates were obtained using both methods (mean absolute difference: 3.24 ± 5.15 dB). Thus, online machine learning can be utilized with similar reliability and accuracy with the important benefit of a shorter test duration. Additional advantages include its flexibility for expansion to bone conduction, speech perception, and masking. Crowson et al. (72) utilized deep learning in the form of "Auto Audio," a proof-of-concept model to interpret diagnostic audiograms. Audiograms consisting of various hearing loss types (e.g., conductive, sensorineural, mixed) were used to train several neural networks. While challenges still remain (i.e., mixed hearing losses were most likely to be misclassified), the authors report that this technology holds promise and may enable an automatic and efficient audiogram interpretation method. Pitathawatchai et al. (73) compared a machine learning algorithm and a common approach (based on slope calculations) for predicting the full audiograms of children with sensorineural hearing loss for cases in which only 1 or 2 thresholds between 500 and 4,000 Hz were labeled. Results indicated that the machine learning approach was not only reliable, but also predicted the full set of thresholds with greater accuracy than the common approach.

Auditory Brainstem Response (ABR)

Towers et al. (74) compared ABR results obtained between in-person and remote methods in a group of 15 adults. ABRs were completed using either broadband clicks or tone bursts at 500 and 3,000 Hz. There were no significant differences across test sites; specifically, wave latencies across the two methods were within the clinically acceptable range of variability. Similarly, Ramkumar et al. (75) examined the role of real-time diagnostic tele-ABR in a mobile van with satellite connection, which allowed videoconferencing between an off-site audiologist and a trained on-site facilitator. A total of 30 newborns were tested via tele-ABR, and latency results were comparable to conventional in-person ABR measures.

Dharmar et al. (76) conducted remote diagnostic audiological evaluations in infants who did not pass their newborn hearing screening ($n = 22$) in California and examined parent/caregiver satisfaction, as well as the impact these remote services had on improving what had been a high loss to follow-up rate. The procedure included a case history, video otoscopy, tympanometry, acoustic reflexes, DPOAEs, and diagnostic ABR testing. An audiologist conducted the testing remotely via an on-site facilitator who positioned the otoscope and tympanometry probe, and also prepped the skin and placed the electrodes for ABR testing. Thirteen of the 22 children tested were diagnosed with hearing loss. All parents/caregivers rated the importance of remote audiology services as a 7 on a 7-point Likert scale (7 = "extremely important"). Importantly, *all* infants completed the diagnostic testing with zero loss to follow-up, representing a marked improvement compared to a 22% loss to follow-up rate in the region prior to the study. Thus, a remote option can significantly reduce loss to follow-up rates in infants who fail their newborn hearing screenings and experience barriers to traditional, in-person diagnostic test services. Ramkumar et al. (77) likewise reported that offering a remote option improved the loss to follow-up rates in their examination of remote diagnostic ABR testing among children 5 years of age and under who had previously failed a hearing screening. A third study by Hatton et al. (78) examined remote diagnostic ABR testing in infants who did not pass their newborn hearing screenings ($n = 102$) in British Columbia. Remote testing with an on-site trained facilitator determined that 50 of the 102 children had hearing loss, with efficiency and accuracy being comparable to traditional, in-person testing. Most importantly, remote testing resulted in a significant reduction in travel costs ($91,250). There are clear benefits to offering a remote option, though the current lack of standardized procedures and insufficient information technology support across test centers needs to be addressed prior to widespread usage.

Intervention and Rehabilitation

Hearing Aids

Remote Fitting and Verification

Ferrari and Bernardez-Braga (79) compared probe microphone measures completed via traditional, in-person procedures and remotely in a group of 60 adult hearing aid (HA) users. The remote setup consisted of application sharing software, desktop videoconferencing, and an on-site facilitator to place the probe tube. No clinically-significant differences across methods were noted, with differences in real-ear unaided response (REUR), real-ear aided response (REAR), and real-ear insertion gain (REIG) measurements varying by only 0–2.2 dB.

Novak et al. (80) examined HA fittings completed remotely by audiology students and faculty, with nursing students and faculty serving as on-site trained facilitators. The tele-audiology setup included video-conferencing and remote desktop access. As part of this study, 181 patients were fit with hearing aids remotely and had probe-microphone verification completed successfully. Significant improvement in quality of life and communication was reported by the majority of patients.

Pross et al. (81) examined the effectiveness of remote HA services at the Veterans Health Administration. In total, 42,697 veterans were fit with HAs and completed outcome measures, 1,009 of whom did so via tele-audiology and 41,688 via traditional, in-person services. An on-site facilitator was present with the patient, and probe-microphone measures and adjustments were completed by an audiologist via video-conferencing. Hearing aid satisfaction was comparable to traditional, in-person fittings, suggesting that remote fittings may be a viable and cost effective alternative.

Self-Fitting Hearing Aids/User Programmable Hearing Aids

The ultimate means of facilitating access to care is "self-fitting" hearing aids, which would allow the patient to control and manage the device independently, without dedicated equipment and without assistance from a hearing specialist. Convery et al. (82) discussed a self-fitting behind-the-ear (BTE) hearing aid with an instant-fit tip (National Acoustics Laboratory version). The device utilizes a tone generator for automated, *in situ* hearing threshold measurements, followed by a prescriptive algorithm for hearing aid programming. However, the success of self-fitting depends not only on accurate threshold measurements [e.g., (83, 84)], but also on accurately identifying medical contraindications (e.g., the presence of a conductive or mixed hearing loss).

An intermediate means is user-programmable hearing aids, which, unlike self-fitting hearing aids, require a previously obtained audiogram, a computer with internet, hearing aid programming software, and/or a hardware interface between the hearing aids and a computer (82). Despite increased risk of inappropriate fittings and a lack of supervision by an audiologist, both self-fitting and user-programmable hearing aids can potentially lower cost, increase accessibility by reducing the need for travel, improve performance and satisfaction by adjusting the hearing aids in the patient's real-world listening environment, and ultimately increase hearing aid adoption rates by giving the patient a sense of personal ownership.

Cochlear Implants

Remote Programming

Several studies have examined the feasibility of remote cochlear implant programming, specifically seeking to answer the question of whether mapping via tele-audiology is equivalent to traditional, in-person programming.

Ramos et al. (85) compared traditional, in-person and remote cochlear implant programming in five adult CI users. Remote programming was completed using an on-site trained facilitator and internet and/or video conferencing to connect with a remote programmer. Twenty-four different sessions were completed (half in-person and half remote) across four intervals, each separated by 3 months, and both methods produced comparable results with respect to thresholds and word recognition scores. Several follow-up studies have replicated the safety and effectiveness of remote CI programming in both adult and pediatric populations (86–91). These follow-up studies also identified limitations and potential means of improving remote CI programming. For example, Hughes et al. (86) found lower speech recognition scores via tele-audiology, which was likely due to higher ambient noise levels in the absence of a sound-treated test booth. Additionally, when the CI was connected to the programming interface, communication with the patient was difficult since the CI microphone was inactive. Hughes et al. recommend using an alternative method of communication, such as video-conferencing, so speech reading or sign language could be possible.

Schepers et al. (90) found no significant differences between the local and remote fittings in terms of Maximum Comfortable Levels (MCL), Threshold Levels (THR), Impedance Field Telemetry (IFT), audiometry, or speech perception results, except for a slightly longer duration for the remote fitting. Like Schepers et al. (90), Luryi et al. (87) found similar results among 20 adult CI users at the Connecticut Veterans Affairs (VA) Healthcare System. Most importantly, both studies found that patients were highly satisfied with remote programming, which is a reliable and cost effective means of providing follow-up care to patients in remote areas or those with limited mobility.

Self-Fitting Cochlear Implants

Meeuws et al. (92) examined the feasibility of an autonomous "self-fitting" method in adult CI users ($n = 6$). Study participants completed a self-fitting session, including audiometry and spectral discrimination testing, 2 weeks after initial traditional, in-person activation. An artificial intelligence software system (FOX) was used for interpretation, analysis, and map recommendation. Specifically, it analyzes the test results and the patient's current map, then calculates predicted outcomes with alternative maps. The alternative map with the best predicted outcome is recommended to the patient. Importantly, this method does not completely preclude the role of an audiologist; a CI audiologist was still required to review the recommended map. Following programming of the new map, participants were then re-tested again after 2 months. A questionnaire was also completed. Four of the six participants were able to complete all tests without any additional assistance from the audiologist. Four were fitted with a new map without physical intervention. All six participants reported feeling comfortable with the autonomous process, but initial audiologist supervision may be required or preferred.

KEY ISSUES

Recent literature has shown that tele-audiology can be completed accurately and reliably; however, additional measures and regulations *specific to the provision of remote care* are necessary. Thus, prior to providing remote services, audiologists must be familiar with the unique licensure, reimbursement, and privacy/security regulations required for tele-audiology services.

Licensure

Current policy with respect to tele-audiology stipulates that the site of service is determined by the patient's physical location. Thus, an audiologist must be licensed in both the location where services are provided from *and* the location the patient is in when services are received. At present, American Speech-Language-Hearing Association (ASHA) is exploring licensing options that would facilitate multi-state service delivery (3). No global licensure agreement currently exists, thus limiting the usage of tele-audiology across countries or regions.

Reimbursement

Tele-audiology payment and coverage is variable across state, federal, and commercial payers (e.g., private health insurance, Medicare, Medicaid). For example, while commercial payers and Medicaid have the discretion to provide coverage for tele-audiology services, Medicare does *not* consider audiologists to be eligible telehealth service providers (3). Thus, it is imperative that audiologists confirm billing and coverage policies prior to providing tele-audiology services.

Privacy and Security

Tele-audiology services must adhere to state, federal, and international regulations with respect to patient privacy and security, particularly that which includes transmission and storage of patient data. Tele-audiology providers must abide by the same regulations applicable to traditional, in-person services. Current federal legislation includes (1) Health Insurance Portability and Accountability Act of 1996 (HIPAA), (2) Health Information Technology for Economic and Clinical Health Act of 2009 (HITECH), and (3) Family Educational Rights and Privacy Act of 1974 (FERPA) (3). It is possible that state requirements may be more rigorous than those at the federal level; thus, it is important for audiologists to familiarize themselves with both federal and state-specific guidance prior to providing tele-audiology care.

Additional important considerations include security of patient rooms, security of electronic documents, security of telecommunications, identification of all individuals present in the rooms at both locations, and documentation of informed consent from the patient. The informed consent document should describe how tele-audiology services may be different from traditional, in-person care, the equipment to be used, the patient's right to switch to traditional, in-person services at any point (if available), modifications (if any) to clinical protocols and procedures, and any potential issues with patient confidentiality. In order to ensure patient confidentiality, the audiologist must be familiar with state and federal regulations regarding electronic storage of patient data, privacy protections (e.g., firewalls, encryption, VPN), configuring software and hardware for use with firewalls, encryption, or VPN, and policies for breach notification (3).

DISCUSSION

Taken together, the current body of literature suggests that tele-audiology provides a viable service delivery model. Remote hearing screenings, diagnostic testing, intervention, and rehabilitation can be completed safely and effectively, in both children and adults. Moreover, the accurate provision of tele-audiology care has been demonstrated both in developed and developing countries. Benefits of such services include but are not limited to improvement in loss to follow-up, reductions in travel time and costs, and improved access to services not otherwise available in one's physical location.

Challenges and Future Directions

Still, tele-audiology is not without its logistical challenges. Several modifications to the traditional testing paradigm may be needed in order to successfully implement remote testing. For example, the presence of ambient noise during testing can reduce the accuracy of results. While use of a soundproof or sound-treated test booth is ideal, this may not be feasible in all circumstances. To improve diagnostic accuracy in the absence of a sound-treated environment, headphones with real-time ambient noise monitoring are recommended, as well as detailed instructions for proper headphone use by the patient (93). Further, different protocols may need to be developed for different age groups and for those with varying degrees of hearing loss. For CI care specifically, offering alternative communication strategies (text messaging, video-conferencing, or sign language) may be particularly useful for this population.

Access to digital technology also presents an ongoing challenge. While tele-audiology could help narrow the healthcare gap evident in many underserved communities, it also runs the risk of exacerbating existing inequalities and amplifying the "digital divide" (94, 95). The current digital divide most negatively impacts racial/ethnic minorities, individuals of low socioeconomic status, those in rural areas, and the elderly. For example, while rural communities are very much in need of telehealth opportunities due to a shortage of physicians that is higher than the national average, fewer than 50% of rural households actually have broadband access (96, 97). Thus, a telehealth option is only practical and realistic for less than half the rural community. Further, in order for tele-audiology care to be reliable, the remote test site's internet connection must be strong enough to support real-time communication and data transfer, particularly if a synchronous service delivery model is utilized. In other words, it's not enough to simply have *access* to digital services, the *quality* of the services must be sufficiently reliable (95). Clearly there are a number of logistical barriers that must be overcome to ensure patients in *all* communities have equitable access to high-quality tele-audiology services.

Digital literacy, on both the part of the provider and the patient, presents another challenge. For example, downloading and setting up a mobile application can be quite daunting for those unfamiliar or uncomfortable with digital technology. During the COVID-19 pandemic, social distancing and other safety provisions forced many patients and healthcare providers to rely on computer programs and other mobile tools that may have been beyond their digital literacy level. In the future, formal assessment of a patient's digital proficiency will allow the provider to meet the patient at their comfort level and tailor online intervention accordingly. Questionnaires validated for assessing digital proficiency include the Mobile Device Proficiency Questionnaire (MDPQ-16) and the Computer Proficiency Questionnaire (CPQ-12). The MDPQ-16 consists of 8 domains (mobile device basics, communication, data and file storage, internet, calendar, entertainment, privacy, and trouble shooting and software management) with 16 questions. The CPQ-12 consists of 6 domains (computer basics, printer, communication, internet, calendar, and entertainment) with 12 questions (98).

Additional challenges include reluctance or even resistance from payers, as well as the increased regulation required, particularly with respect to licensure, reimbursement, and privacy and security. Specifically, remote care must be delivered in accordance with clinical guidelines, payer policies, and state and federal law. Further, audiologists must ensure that tele-audiology clinical care is equivalent in quality to care delivered in-person. While undoubtedly necessary, the additional levels of regulation and requirements may, at least in the short-term, serve as hurdles to the widespread implementation of tele-audiology care.

Future advancements in technology, especially in artificial intelligence, may help facilitate the provision of remote care and may accelerate the adoption of tele-audiology services worldwide at affordable costs (99). To date, artificial intelligence has been implemented in a CI software system (FOX) to allow for autonomous "self-fitting" of the CI device, as well as home-based otoscopy. Future applications of such technology may include commercially-available diagnostic testing (e.g., conventional audiometry) and commercially-available "self-fitting" or even "cognitive-controlled" hearing aids.

CONCLUSION

Recent research shows that tele-audiology services can increase patient accessibility and engagement, improve loss to follow-up rates, and reduce cost and travel time. Furthermore, remote services can be completed in a manner that is safe, valid, reliable, and satisfactory. Still, logistical challenges do remain. For example, careful attention must be given to controlling ambient noise (particularly when testing is completed in the absence of a sound-treated test booth), modifications to current testing procedures may be needed to tailor to the provision of remote care, and regulatory and reimbursement hurdles need to be overcome before tele-audiology may be implemented on a widescale. The current COVID-19 pandemic has caused mass disruption to the delivery of healthcare services and has consequently accelerated the pace of development and acceptance of tele-audiology. Future research, including advancements in artificial intelligence, will continue to improve not only the effectiveness and efficiency of tele-audiology services but also most importantly, patient acceptance and satisfaction.

AUTHOR CONTRIBUTIONS

KD'O completed a review of the existing tele-audiology literature between September 2020–November 2021, wrote the initial draft of this manuscript, and reviewed and edited the final version. F-GZ provided input on topic coverage, wrote portions of the final manuscript, and reviewed and edited the final version. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported in part by NIH NIDCD (3R01DC015587).

ACKNOWLEDGMENTS

The authors would like to extend thanks to Dr. Katelyn Berg and two reviewers for their helpful and constructive comments on an earlier draft of this manuscript.

REFERENCES

1. Wootton R, Patil NG, Scott RE, Ho K. *Telehealth in the Developing World. Royal Society of Medicine*. London: Press/IDRC (2009).
2. Bashshur R, Shannon GW. *History of Telemedicine: Evolution, Context, Transformation* (Vol. 2009). New Rochelle, NY: Mary Ann Liebert (2009).
3. American Speech-Language-Hearing Association (ASHA). *Telepractice*. (2021). Available online at: https://www.asha.org/practice-portal/professional-issues/telepractice/#collapse_1 (accessed September 25, 2021).
4. Mao W, Ogbuoji O, Watkins D, Bharali I, Nsiah-Boateng E, Diab MM, et al. Achieving global mortality reduction targets and universal health coverage: the impact of COVID-19. *PLoS Med*. (2021) 18:e1003675. doi: 10.1371/journal.pmed.1003675
5. Muñoz K, Nagaraj NK, Nichols N. Applied tele-audiology research in clinical practice during the past decade: a scoping review. *Int J Audiol*. (2021) 60(sup1):S4–12. doi: 10.1080/14992027.2020.1817994
6. Word Health Organization. *World Report on Hearing*. (2021). Available online at: https://www.who.int/teams/noncommunicable-diseases/sensory-functions-disability-and-rehabilitation/highlighting-priorities-for-ear-and-hearing-care (accessed November 20, 2021).
7. Grundfast K, Liu S. What otolaryngologists need to know about hearing aids. *JAMA Otolaryngol Head Neck Surg*. (2017) 143:109–10. doi: 10.1001/jamaoto.2016.3416
8. Orji A, Kamenov K, Dirac M, Davis A, Chadha S, Vos T. Global and regional needs, unmet needs and access to hearing aids. *Int J Audiol*. (2020) 59:166–72. doi: 10.1080/14992027.2020.1721577

9. Zeng FG. Cochlear implants: why don't more people use them?. *Hear J.* (2007) 60:48–9. doi: 10.1097/01.HJ.0000286715.37618.c1
10. Perkins E, Dietrich MS, Manzoor N, O'Malley M, Bennett M, Rivas A, et al. Further evidence for the expansion of adult cochlear implant candidacy criteria. *Otol Neurotol.* (2021) 42:815–23. doi: 10.1097/MAO.0000000000003068
11. Ebrahimi-Madiseh A, Eikelboom RH, Bennett RJ, Upson GS, Friedland PL, Swanepoel DW, et al. What influences decision-making for cochlear implantation in adults? Exploring barriers and drivers from a multistakeholder perspective. *Ear Hear.* (2020) 41:1752–63. doi: 10.1097/AUD.0000000000000895
12. De Raeve L, van Hardeveld R. Prevalence of cochlear implants in Europe: what do we know and what can we expect?. *J Hear Sci.* (2013) 3:9–16. doi: 10.17430/889707
13. Fagan JJ, Tarabichi M. Cochlear implants in developing countries: practical and ethical considerations. *Curr Opin Otolaryngol Head Neck Surg.* (2018) 26:188–9. doi: 10.1097/MOO.0000000000000457
14. Moeller MP, Tomblin JB, Yoshinaga-Itano C, Connor CM, Jerger S. Current state of knowledge: language and literacy of children with hearing impairment. *Ear Hear.* (2007) 28:740–53. doi: 10.1097/AUD.0b013e318157f07f
15. Yoshinaga-Itano C, Sedey AL, Coulter DK, Mehl AL. Language of early- and later-identified children with hearing loss. *Pediatrics.* (1998) 102:1161–71. doi: 10.1542/peds.102.5.1161
16. Olusanya BO, Ruben RJ, Parving A. Reducing the burden of communication disorders in the developing world: an opportunity for the millennium development project. *JAMA.* (2006) 296:441–4. doi: 10.1001/jama.296.4.441
17. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet.* (2020) 396:413–46. doi: 10.1016/S0140-6736(20)30367-6
18. McPherson B. Audiology: a developing country context. In: McPherson B, Brouillette R, editors. *Audiology in Developing Countries*. New York, NY: Nova Science (2008). p. 5–20.
19. Vo QT, Pham D, Choi KJ, Nguyen UT, Le L, Shanewise T, et al. Solar-powered hearing aids for children with impaired hearing in Vietnam: a pilot study. *Paediatr Int Child Health.* (2018) 38:40–5. doi: 10.1080/20469047.2016.1276119
20. Zeng FG. Challenges in improving cochlear implant performance and accessibility. *IEEE Trans Biomed Eng.* (2017) 64:1662–4. doi: 10.1109/TBME.2017.2718939
21. Barr M, Dally K, Duncan J. Service accessibility for children with hearing loss in rural areas of the United States and Canada. *Int J Pediatr Otorhinolaryngol.* (2019) 123:15–21. doi: 10.1016/j.ijporl.2019.04.028
22. Bush ML, Osetinsky M, Shinn JB, Gal TJ, Ding X, Fardo DW, et al. Assessment of Appalachian region pediatric hearing healthcare disparities and delays. *Laryngoscope.* (2014) 124:1713–7. doi: 10.1002/lary.24588
23. Bush ML, Hardin B, Rayle C, Lester C, Studts CR, Shinn JB. Rural barriers to early diagnosis and treatment of infant hearing loss in Appalachia. *Otol Neurotol.* (2015) 36:93. doi: 10.1097/MAO.0000000000000636
24. Bush ML, Burton M, Loan A, Shinn JB. Timing discrepancies of early intervention hearing services in urban and rural cochlear implant recipients. *Otol Neurotol.* (2013) 34:1630–5. doi: 10.1097/MAO.0b013e31829e83ad
25. Goulios H, Patuzzi R. Education and practice of audiology internationally: affordable and sustainable education models for developing countries. In: McPherson B, Brouillette R, editors. *Audiology in developing countries*. New York, NY: Nova Science (2008). p. 51–74.
26. Powell W, Jacobs JA, Noble W, Bush ML, Snell-Rood C. Rural adult perspectives on impact of hearing loss and barriers to care. *J Community Health.* (2019) 44:668–74. doi: 10.1007/s10900-019-00656-3
27. Coco L, Champlin CA, Eikelboom RH. Community-based intervention determines tele-audiology site candidacy. *Am J Audiol.* (2016) 25:264–7. doi: 10.1044/2016_AJA-16-0002
28. Bush ML, Alexander D, Noblitt B, Lester C, Shinn JB. Pediatric hearing healthcare in Kentucky's Appalachian primary care setting. *J Commun Health.* (2015) 40:762–8. doi: 10.1007/s10900-015-9997-0
29. Zeng FG. Where is my hearing plan?. *Hear J.* (2019) 72:6. doi: 10.1097/01.HJ.0000616116.69080.a7
30. Bainbridge KE, Ramachandran V. Hearing aid use among older U.S. adults: the national health and nutrition examination survey, 2005- 2006 and 2009-2010. *Ear Hear.* (2014) 35:289–94. doi: 10.1097/01.aud.0000441036.40169.29
31. Duran DG, Pérez-Stable EJ. Novel approaches to advance minority health and health disparities research. *Am J Public Health*. (2019) 109:S8–10. doi: 10.2105/AJPH.2018.304931
32. Lee DJ, Carlson DL, Lee HM, Ray LA, Markides KS. Hearing loss and hearing aid use in Hispanic adults: results from the Hispanic health and nutrition examination survey. *Am J Public Health.* (1991) 81:1471–4. doi: 10.2105/AJPH.81.11.1471
33. Mahendran GN, Rosenbluth T, Featherstone M, Vivas EX, Mattox DE, Hobson CE. Racial disparities in adult cochlear implantation. *Otolaryngol Head Neck Surg.* (2021) 1–7. doi: 10.1177/01945998211027340. [Online ahead of print].
34. Tomita M, Mann WC, Welch TR. Use of assistive devices to address hearing impairment by older persons with disabilities. *Int J Rehabil Res.* (2001) 24:279–90. doi: 10.1097/00004356-200112000-00004
35. Zhang L, Links AR, Boss EF, White A, Walsh J. Identification of potential barriers to timely access to pediatric hearing aids. *JAMA Otolaryngol Head Neck Surg.* (2020) 146:13–9. doi: 10.1001/jamaoto.2019.2877
36. Alqudah S, Zaitoun M, Alqudah O, Alqudah S, Alqudah Z. Challenges facing users of hearing aids during the COVID-19 pandemic. *Int J Audiol.* (2021) 60:747–53. doi: 10.1080/14992027.2021.1872806
37. World Health Organization. *Guidelines for Hearing Aids and Services for Developing Countries*. 2nd ed. Geneva: World Health Organization (2004). Retrieved from: www.who.int/pbd/deafness/en/hearing_aid_guide_en.pdf (accessed November 20, 2021).
38. World Bank International Comparison Program database. Available online at: https://data.worldbank.org/indicator/NY.GNP.PCAP.PP.CD?year_high_desc=true (accessed November 21, 2021).
39. U.S. Bureau of Labor Statistics. *Occupational Employment and Wage Statistics*. (2020). Available online at: https://www.bls.gov/oes/current/oes291181.htm#(1) (accessed September 20, 2021).
40. Gurney J, Fraser L, Ikihele A, Manderson J, Scott N, Robson B. Telehealth as a tool for equity: pros, cons and recommendations. *N Zealand Med J.* (2021) 134:111–5.
41. Rappaport KM, McCracken CC, Beniflah J, Little WK, Fletcher DA, Lam WA, et al. Assessment of a smartphone otoscope device for the diagnosis and management of otitis media. *Clin Pediatr.* (2016) 55:800–10. doi: 10.1177/0009922815593909
42. Richards JR, Gaylor KA, Pilgrim AJ. Comparison of traditional otoscope to iPhone otoscope in the pediatric ED. *Am J Emerg Med.* (2015) 33:1089–92. doi: 10.1016/j.ajem.2015.04.063
43. Shah MU, Sohal M, Valdez TA, Grindle CR. iPhone otoscopes: currently available, but reliable for tele-otoscopy in the hands of parents?. *Int J Pediatr Otorhinolaryngol.* (2018) 106:59–63. doi: 10.1016/j.ijporl.2018.01.003
44. Erkkola-Anttinen N, Irjala H, Laine MK, Tähtinen PA, Löyttyniemi E, Ruohola A. Smartphone otoscopy performed by parents. *Telemed e-Health.* (2019) 25:477–84. doi: 10.1089/tmj.2018.0062
45. Chan J, Raju S, Nandakumar R, Bly R, Gollakota S. Detecting middle ear fluid using smartphones. *Sci Transl Med.* (2019) 11:eaav1102. doi: 10.1126/scitranslmed.aav1102
46. Cha D, Pae C, Seong SB, Choi JY, Park HJ. Automated diagnosis of ear disease using ensemble deep learning with a big otoendoscopy image database. *EBioMedicine.* (2019) 45:606–14. doi: 10.1016/j.ebiom.2019.06.050
47. Lancaster P, Krumm M, Ribera J, Klich R. Remote hearing screenings via telehealth in a rural elementary school. *Am J Audiol.* (2008) 17:114–22. doi: 10.1044/1059-0889(2008/07-0008)
48. Khoza-Shangase K Kassner L. Automated screening audiometry in the digital age: exploring uhear$^{\mathrm{TM}}$ and its use in a resource-stricken developing country. *Int J Technol Assess Health Care.* (2013) 29:42. doi: 10.1017/S0266462312000761
49. Rourke R, Bromwich M, Chan D. iPad audiometry in Canada's North: a portable and cost-effective method for hearing screening. *Otolaryngol Head Neck Surg.* (2014) 151(1_suppl):P104. doi: 10.1177/0194599814541627a237
50. Samelli AG, Rabelo CM, Sanches SGG, Martinho AC, Matas CG. Tablet-Based tele-audiometry: automated hearing screening for schoolchildren. *J Telemed Telecare.* (2020) 26:140–149. doi: 10.1177/1357633X18800856

51. Dillon H, Mee C, Moreno JC, Seymour J. Hearing tests are just child's play: the sound scouts game for children entering school. *Int J Audiol.* (2018) 57:529–37. doi: 10.1080/14992027.2018.1463464
52. Wu W, Lü J, Li Y, Kam ACS, Tong MCF, Huang Z, et al. A new hearing screening system for preschool children. *Int J Pediatr Otorhinolaryngol.* (2014) 78:290–5. doi: 10.1016/j.ijporl.2013.11.026
53. Swanepoel DW, Myburgh HC, Howe DM, Mahomed F, Eikelboom RH. Smartphone hearing screening with integrated quality control and data management. *Int J Audiol.* (2014) 53:841–9. doi: 10.3109/14992027.2014.920965
54. Eksteen S, Launer S, Kuper H, Eikelboom RH, Bastawrous A, Swanepoel DW. Hearing and vision screening for preschool children using mobile technology, South Africa. *Bull World Health Organ.* (2019) 97:672. doi: 10.2471/BLT.18.227876
55. Krumm M, Ribera J, Klich R. Providing basic hearing tests using remote computing technology. *J Telemed Telecare.* (2007) 13:406–10. doi: 10.1258/135763307783064395
56. Ciccia AH, Whitford B, Krumm M, McNeal K. Improving the access of young urban children to speech, language and hearing screening via telehealth. *J Telemed Telecare.* (2011) 17:240–4. doi: 10.1258/jtt.2011.100810
57. Monica SD, Ramkumar V, Krumm M, Raman N, Nagarajan R, Venkatesh L. School entry level tele-hearing screening in a town in South India–Lessons learnt. *Int J Pediatr Otorhinolaryngol.* (2017) 92:130–5. doi: 10.1016/j.ijporl.2016.11.021
58. Ameyaw GA, Ribera J, Anim-Sampong S. Interregional Newborn hearing screening via telehealth in Ghana. *J Am Acad Audiol.* (2019) 30:178–86. doi: 10.3766/jaaa.17059
59. Givens GD, Elangovan S. Internet application to tele-audiology— "Nothin' but net". *Am J Audiol*. (2003) 12:59–65. doi: 10.1044/1059-0889(2003/011)
60. Ho ATP, Hildreth AJ, Lindsey L. Computer-assisted audiometry versus manual audiometry. *Otol Neurotol.* (2009) 30:876–83. doi: 10.1097/MAO.0b013e3181b120d0
61. Margolis RH, Glasberg BR, Creeke S, Moore BC. AMTAS®: automated method for testing auditory sensitivity: validation studies. *Int J Audiol.* (2010) 49:185–94. doi: 10.3109/14992020903092608
62. Swanepoel DW, Mngemane S, Molemong S, Mkwanazi H, Tutshini S. Hearing assessment—reliability, accuracy, and efficiency of automated audiometry. *Telemed e-Health.* (2010) 16:557–63. doi: 10.1089/tmj.2009.0143
63. Margolis RH, Moore BC. AMTAS®: automated method for testing auditory sensitivity: III. sensorineural hearing loss and air-bone gaps. *Int J Audiol.* (2011) 50:440–7. doi: 10.3109/14992027.2011.575085
64. Swanepoel DW, Biagio L. Validity of diagnostic computer-based air and forehead bone conduction audiometry. *J Occup Environ Hyg.* (2011) 8:210–4. doi: 10.1080/15459624.2011.559417
65. Maclennan-Smith F, Swanepoel DW, Hall JW III. Validity of diagnostic pure-tone audiometry without a sound-treated environment in older adults. *Int J Audiol.* (2013) 52:66–73. doi: 10.3109/14992027.2012.736692
66. Margolis RH, Frisina R, Walton JP. AMTAS®: automated method for testing auditory sensitivity: II. air conduction audiograms in children and adults. *Int J Audiol.* (2011) 50:434–9. doi: 10.3109/14992027.2011.553206
67. Botasso M, Sanches SGG, Bento RF, Samelli AG. Teleaudiometry as a screening method in school children. *Clinics.* (2015) 70:283–8. doi: 10.6061/clinics/2015(04)11
68. Edwards AL. Evaluation of the Kuduwave 5000 audiometer for compliance with standards for hearing conservation purposes. *Telemed J E-Health.* (2010) 16:557–63.
69. Van Tonder J, Swanepoel DW, Mahomed-Asmail F, Myburgh H, Eikelboom RH. Automated smartphone threshold audiometry: validity and time efficiency. *J Am Acad Audiol.* (2017) 28:200–8. doi: 10.3766/jaaa.16002
70. Sandström J, Swanepoel D, Laurent C, Umefjord G, Lundberg T. Accuracy and reliability of smartphone self-test audiometry in community clinics in low income settings: a comparative study. *Ann Otol Rhinol Laryngol.* (2020) 129:578–84. doi: 10.1177/0003489420902162
71. Barbour DL, Howard RT, Song XD, Metzger N, Sukesan KA, DiLorenzo JC, et al. Online machine learning audiometry. *Ear Hear.* (2019) 40:918. doi: 10.1097/AUD.0000000000000669
72. Crowson MG, Lee JW, Hamour A, Mahmood R, Babier A, Lin V, et al. AutoAudio: deep learning for automatic audiogram interpretation. *J Med Syst.* (2020) 44:1–7. doi: 10.1007/s10916-020-01627-1
73. Pitathawatchai P, Chaichulee S, Kirtsreesakul V. Robust machine learning method for imputing missing values in audiograms collected in children. *Int J Audiol.* (2021) 1–12. doi: 10.1080/14992027.2021.1884909. [Online ahead of print].
74. Towers AD, Pisa J, Froelich TM, Krumm M. The reliability of click-evoked and frequency-specific auditory brainstem response testing using telehealth technology. In: Palmer CV, Givens GD, editors. *Seminars in Hearing.* Vol. 26. New York, NY: Thieme Medical Publishers, Inc. (2005). p. 26–34. doi: 10.1055/s-2005-863792
75. Ramkumar V, Hall JW, Nagarajan R, Shankarnarayan VC, Kumaravelu S. Tele-ABR using a satellite connection in a mobile van for newborn hearing testing. *J Telemed Telecare.* (2013) 19:233–37. doi: 10.1177/1357633X13494691
76. Dharmar M, Simon A, Sadorra C, Friedland G, Sherwood J, Morrow H, et al. Reducing loss to follow-up with tele-audiology diagnostic evaluations. *Telemed e-Health.* (2016) 22:159–64. doi: 10.1089/tmj.2015.0001
77. Ramkumar V, Nagarajan R, Shankarnarayan VC, Kumaravelu S, Hall JW. Implementation and evaluation of a rural community-based pediatric hearing screening program integrating in-person and tele-diagnostic auditory brainstem response (ABR). *BMC Health Serv Res.* (2019) 19:1. doi: 10.1186/s12913-018-3827-x
78. Hatton JL, Rowlandson J, Beers A, Small S. Telehealth-enabled auditory brainstem response testing for infants living in rural communities: the British Columbia Early Hearing Program experience. *Int J Audiol.* (2019) 58:381–92. doi: 10.1080/14992027.2019.1584681
79. Ferrari DV, Bernardez-Braga GRA. Remote probe microphone measurement to verify hearing aid performance. *J Telemed Telecare*. (2009) 15:122–4. doi: 10.1258/jtt.2009.003005
80. Novak RE, Cantu AG, Zappler A, Coco L, Champlin CA, Novak JC. The future of healthcare delivery: IPE/IPP audiology and nursing student/faculty collaboration to deliver hearing aids to vulnerable adults via telehealth. *J Nurs Interprofess Leadership Qual Safety*. (2016) 1:1–11.
81. Pross SE, Bourne AL, Cheung SW. TeleAudiology in the Veterans health administration. *Otol Neurotol.* (2016) 37:847–50. doi: 10.1097/MAO.0000000000001058
82. Convery E, Keidser G, Dillon H, Hartley L. A self-fitting hearing aid: need and concept. *Trends Amplif.* (2011) 15:157–66. doi: 10.1177/1084713811427707
83. Smith-Olinde L, Nicholson N, Chivers C, Highley P, Williams DK. Test–retest reliability of *in situ* unaided thresholds in adults. *Am J Audiol*. (2006) 15:75–80. doi: 10.1044/1059-0889(2006/009)
84. O'Brien A, Keidser G, Yeend I, Hartley L, Dillon H. Validity and reliability of in-situ air conduction thresholds measured through hearing aids coupled to closed and open instant-fit tips. *Int J Audiol.* (2010) 49:868–76. doi: 10.3109/14992027.2010.503664
85. Ramos A, Rodríguez C, Martinez-Beneyto P, Perez D, Gault A, Falcon JC, et al. Use of telemedicine in the remote programming of cochlear implants. *Acta Otolaryngol.* (2009) 129:533–40. doi: 10.1080/00016480802294369
86. Hughes ML, Goehring JL, Baudhuin JL, Diaz GR, Sanford T, Harpster R, et al. Use of telehealth for research and clinical measures in cochlear implant recipients: a validation study. *J Speech Lang Hear Res.* (2012) 55:1112–27. doi: 10.1044/1092-4388(2011/11-0237)
87. Luryi AL, Tower JI, Preston J, Burkland A, Trueheart CE, Hildrew DM. Cochlear implant mapping through telemedicine—a feasibility study. *Otol Neurotol.* (2020) 41:e330–3. doi: 10.1097/MAO.0000000000002551
88. Rodríguez C, Ramos A, Falcon JC, Martínez-Beneyto P, Gault A, Boyle P. Use of telemedicine in the remote programming of cochlear implants. *Cochlear Implants Int.* (2010) 11(sup1):461. doi: 10.1179/146701010X12671177204624
89. Slager HK, Jensen J, Kozlowski K, Teagle H, Park LR, Biever A, et al. Remote programming of cochlear implants. *Otol Neurotol.* (2019) 40:e260. doi: 10.1097/MAO.0000000000002119
90. Schepers K, Steinhoff HJ, Ebenhoch H, Böck K, Bauer K, Rupprecht L, et al. Remote programming of cochlear implants in users of all ages. *Acta Otolaryngol.* (2019) 139:251–57. doi: 10.1080/00016489.2018.1554264
91. Wasowski A, Skarzynski H, Lorens A, Obrycka A, Walkowiak A, Skarzynski P, et al. The telefitting method used in the national network of teleaudiology:

assessment of quality and cost effectiveness. *J Hear Sci.* (2012) 2:81–5. doi: 10.17430/882767
92. Meeuws M, Pascoal D, Janssens de Varebeke S, De Ceulaer G, Govaerts PJ. Cochlear implant telemedicine: remote fitting based on psychoacoustic self-tests and artificial intelligence. *Cochlear Implants Int.* (2020) 21:260–8. doi: 10.1080/14670100.2020.1757840
93. Chen CH, Lin HYH, Wang MC, Chu YC, Chang CY, Huang CY, et al. Diagnostic accuracy of smartphone-based audiometry for hearing loss detection: meta-analysis. *JMIR mHealth uHealth.* (2021) 9:e28378. doi: 10.2196/28378
94. Blandford A, Wesson J, Amalberti R, AlHazme R, Allwihan R. Opportunities and challenges for telehealth within, and beyond, a pandemic. *Lancet Global Health.* (2020) 8:e1364–5. doi: 10.1016/S2214-109X(20)30362-4
95. Cortelyou-Ward K, Atkins DN, Noblin A, Rotarius T, White P, Carey C. Navigating the digital divide: barriers to telehealth in rural areas. *J Health Care Poor Underserved.* (2020) 31:1546–1556. doi: 10.1353/hpu.2020.0116
96. Federal Communications Commission. *Broadband Deployment Report*. Washington, DC: Federal Communications Commission (2018). Retrieved from: https://www.fcc.gov[Google~Scholar] (accessed November 20, 2021).
97. Jetty A, Moore MA, Coffman M, Petterson S, Bazemore A. Rural family physicians are twice as likely to use telehealth as urban family physicians. *Telemed e-Health.* (2018) 24:268–76. doi: 10.1089/tmj.2017.0161
98. Ratanjee-Vanmali H, Swanepoel DW, Laplante-Lévesque A. Digital proficiency and teleaudiology: key implications in hearing care. *Hear J.* (2020) 73:18–20. doi: 10.1097/01.HJ.0000717160.38022.5c
99. Lesica NA, Mehta N, Manjaly JG, Deng L, Wilson BS, Zeng FG. Harnessing the power of artificial intelligence to transform hearing healthcare and research. *Nat Mach Intell.* (2021) 3:840–9. doi: 10.1038/s42256-021-00394-z

Conflict of Interest: F-GZ owns stock in Axonics, Dianavi, Nurotron, Syntiant, Velox, Voibook, and Xense.

The remaining author declares 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.

frontiers in Neurology

ORIGINAL RESEARCH
published: 11 January 2022
doi: 10.3389/fneur.2021.724800

Check for updates

Visualization of Speech Perception Analysis *via* Phoneme Alignment: A Pilot Study

J. Tilak Ratnanather[1]*, Lydia C. Wang[1], Seung-Ho Bae[1], Erin R. O'Neill[2], Elad Sagi[3] and Daniel J. Tward[1,4]

[1] Center for Imaging Science and Institute for Computational Medicine, Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, United States, [2] Center for Applied and Translational Sensory Sciences, University of Minnesota, Minneapolis, MN, United States, [3] Department of Otolaryngology, New York University School of Medicine, New York, NY, United States, [4] Departments of Computational Medicine and Neurology, University of California, Los Angeles, Los Angeles, CA, United States

Objective: Speech tests assess the ability of people with hearing loss to comprehend speech with a hearing aid or cochlear implant. The tests are usually at the word or sentence level. However, few tests analyze errors at the phoneme level. So, there is a need for an automated program to visualize in real time the accuracy of phonemes in these tests.

Method: The program reads in stimulus-response pairs and obtains their phonemic representations from an open-source digital pronouncing dictionary. The stimulus phonemes are aligned with the response phonemes via a modification of the Levenshtein Minimum Edit Distance algorithm. Alignment is achieved via dynamic programming with modified costs based on phonological features for insertion, deletions and substitutions. The accuracy for each phoneme is based on the F1-score. Accuracy is visualized with respect to place and manner (consonants) or height (vowels). Confusion matrices for the phonemes are used in an information transfer analysis of ten phonological features. A histogram of the information transfer for the features over a frequency-like range is presented as a phonemegram.

Results: The program was applied to two datasets. One consisted of test data at the sentence and word levels. Stimulus-response sentence pairs from six volunteers with different degrees of hearing loss and modes of amplification were analyzed. Four volunteers listened to sentences from a mobile auditory training app while two listened to sentences from a clinical speech test. Stimulus-response word pairs from three lists were also analyzed. The other dataset consisted of published stimulus-response pairs from experiments of 31 participants with cochlear implants listening to 400 Basic English Lexicon sentences via different talkers at four different SNR levels. In all cases, visualization was obtained in real time. Analysis of 12,400 actual and random pairs showed that the program was robust to the nature of the pairs.

Conclusion: It is possible to automate the alignment of phonemes extracted from stimulus-response pairs from speech tests in real time. The alignment then makes it possible to visualize the accuracy of responses via phonological features in two ways. Such visualization of phoneme alignment and accuracy could aid clinicians and scientists.

Keywords: phoneme alignment, speech tests, phoneme accuracy, relative information transfer, F1-score

OPEN ACCESS

Edited by:
Jing Chen,
Peking University, China

Reviewed by:
Christine Rogers,
University of Cape Town, South Africa
Lynne E. Bernstein,
George Washington University, United States

****Correspondence:***
J. Tilak Ratnanather
tilak@cis.jhu.edu

Specialty section:
This article was submitted to Neuro-Otology, a section of the journal Frontiers in Neurology

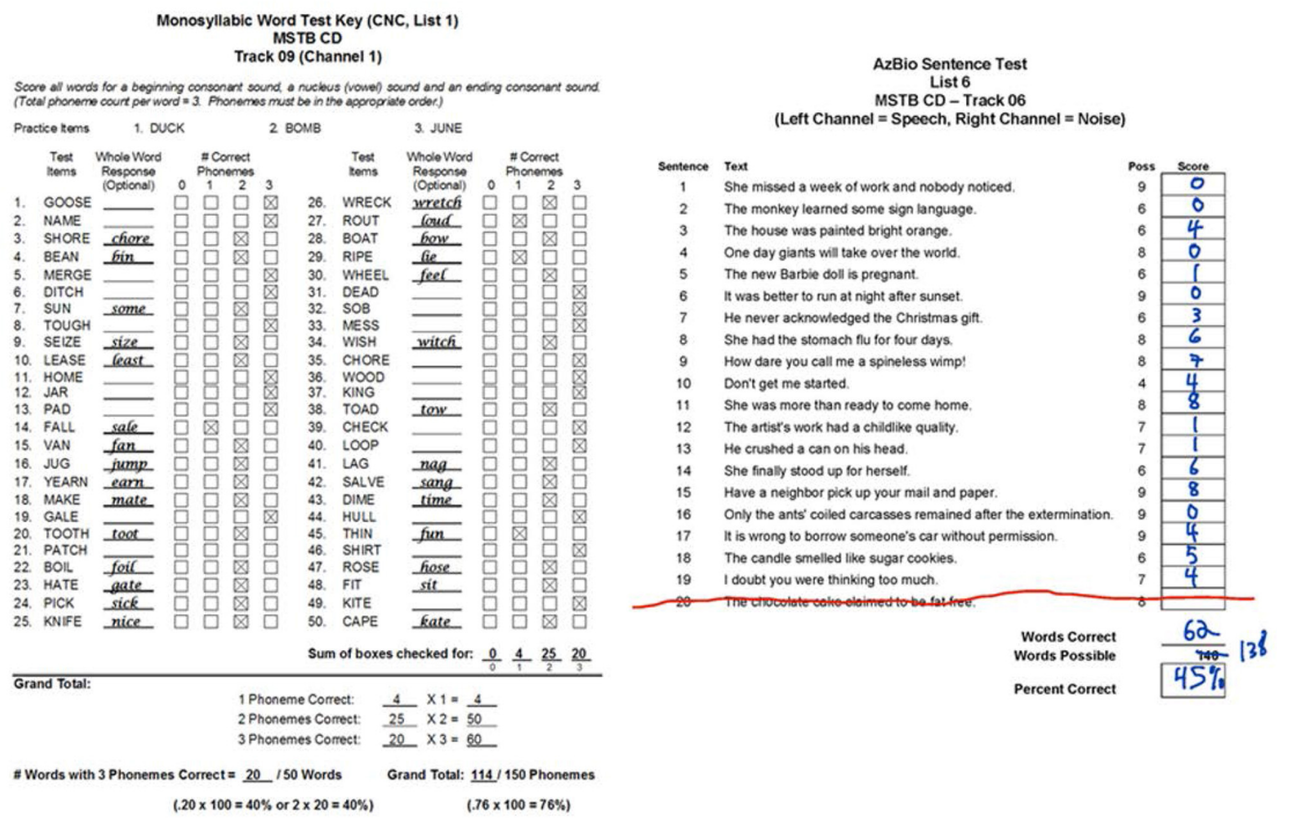
Received: *14 June 2021*
Accepted: *13 December 2021*
Published: *11 January 2022*

Citation:
Ratnanather JT, Wang LC, Bae S-H, O'Neill ER, Sagi E and Tward DJ (2022) Visualization of Speech Perception Analysis via Phoneme Alignment: A Pilot Study. Front. Neurol. 12:724800. doi: 10.3389/fneur.2021.724800

INTRODUCTION

Audiologists and speech pathologists use speech perception tests to analyze speech comprehension in people who are learning to hear with hearing aids and cochlear implants. Specifically, the tests provide an objective measure of how the listener processes spoken words and sentences from the acoustic signal. The words and sentences are composed of sequences of phonemes that are characterized as either consonants or vowels. Further, phonemes are differentiated by how they are produced in the vocal tract, i.e. phonological features (1). For consonants, these features are place, manner, voicing and associated subtypes, and for vowels, these are height, place and associated subtypes. Typically, speech tests are based on lists of words or sentences and are presented in a sound booth in the clinic, sometimes with noise, e.g. PBK-50 (2), AB (3), NU-6 (4), BKB (5), CUNY (6), HINT (7) and AzBio (8). Usually, the clinician records the numbers of correct words and/or sentences, and sometimes the number of correct phonemes, as illustrated by two examples in **Figure 1**. One example is a typical list of 50 words, each with an initial consonant followed by a nucleus (vowel) and then a final consonant. Here, the correct response and number of correct phonemes are recorded. The result is a tally of the number of correct words and phonemes together with incorrect words transcribed. The other example is a typical list of 20 phonetically balanced sentences. Here, the number of correct words is recorded and then summed. It is clear in both cases that the person does not always hear the whole stimulus. There is potentially more useful data to be extracted from these tests, namely the analysis of phonemes with respect to their phonological features. To do so in the clinic would be time consuming. The challenge then is to present information about phonemic comprehension in a manner that can be understood in real time.

At the same time, many people learning to hear with a new hearing aid or a cochlear implant use auditory training apps such as the Speech Banana app which is freely available (10). Progress tracking provides the user a record of correct sentences, correct words, and number of repetitions in the quizzes. Additional information such as accuracy for the phonemes could help



Monosyllabic Word Test Key (CNC, List 1)
MSTB CD
Track 09 (Channel 1)

Score all words for a beginning consonant sound, a nucleus (vowel) sound and an ending consonant sound. (Total phoneme count per word = 3. Phonemes must be in the appropriate order.)

Practice Items 1. DUCK 2. BOMB 3. JUNE

	Test Items	Whole Word Response (Optional)	0	1	2	3		Test Items	Whole Word Response (Optional)	0	1	2	3
1.	GOOSE		☐	☐	☐	☒	26.	WRECK	wretch	☐	☐	☒	☐
2.	NAME		☐	☐	☐	☒	27.	ROUT	loud	☐	☒	☐	☐
3.	SHORE	chore	☐	☐	☒	☐	28.	BOAT	bow	☐	☐	☒	☐
4.	BEAN	bin	☐	☐	☒	☐	29.	RIPE	lie	☐	☒	☐	☐
5.	MERGE		☐	☐	☐	☒	30.	WHEEL	feel	☐	☐	☒	☐
6.	DITCH		☐	☐	☐	☒	31.	DEAD		☐	☐	☐	☒
7.	SUN	some	☐	☐	☒	☐	32.	SOB		☐	☐	☐	☒
8.	TOUGH		☐	☐	☐	☒	33.	MESS		☐	☐	☐	☒
9.	SEIZE	size	☐	☐	☒	☐	34.	WISH	witch	☐	☐	☒	☐
10.	LEASE	least	☐	☐	☒	☐	35.	CHORE		☐	☐	☐	☒
11.	HOME		☐	☐	☐	☒	36.	WOOD		☐	☐	☐	☒
12.	JAR		☐	☐	☐	☒	37.	KING		☐	☐	☐	☒
13.	PAD		☐	☐	☐	☒	38.	TOAD	tow	☐	☐	☒	☐
14.	FALL	sale	☐	☒	☐	☐	39.	CHECK		☐	☐	☐	☒
15.	VAN	fan	☐	☐	☒	☐	40.	LOOP		☐	☐	☐	☒
16.	JUG	jump	☐	☐	☒	☐	41.	LAG	nag	☐	☐	☒	☐
17.	YEARN	earn	☐	☐	☒	☐	42.	SALVE	sang	☐	☐	☒	☐
18.	MAKE	mate	☐	☐	☒	☐	43.	DIME	time	☐	☐	☒	☐
19.	GALE		☐	☐	☐	☒	44.	HULL		☐	☐	☐	☒
20.	TOOTH	toot	☐	☐	☒	☐	45.	THIN	fun	☐	☒	☐	☐
21.	PATCH		☐	☐	☐	☐	46.	SHIRT		☐	☐	☐	☒
22.	BOIL	foil	☐	☐	☒	☐	47.	ROSE	hose	☐	☐	☒	☐
23.	HATE	gate	☐	☐	☒	☐	48.	FIT	sit	☐	☐	☒	☐
24.	PICK	sick	☐	☐	☒	☐	49.	KITE		☐	☐	☐	☒
25.	KNIFE	nice	☐	☐	☒	☐	50.	CAPE	kate	☐	☐	☒	☐

Sum of boxes checked for: 0 (0) 4 (1) 25 (2) 20 (3)

Grand Total:

1 Phoneme Correct: 4 X 1 = 4
2 Phonemes Correct: 25 X 2 = 50
3 Phonemes Correct: 20 X 3 = 60

 # Words with 3 Phonemes Correct = 20 / 50 Words **Grand Total: 114 / 150 Phonemes**

(.20 x 100 = 40% or 2 x 20 = 40%) **(.76 x 100 = 76%)**

AzBio Sentence Test
List 6
MSTB CD – Track 06
(Left Channel = Speech, Right Channel = Noise)

Sentence	Text	Poss	Score
1	She missed a week of work and nobody noticed.	9	0
2	The monkey learned some sign language.	6	0
3	The house was painted bright orange.	6	4
4	One day giants will take over the world.	8	0
5	The new Barbie doll is pregnant.	6	1
6	It was better to run at night after sunset.	9	0
7	He never acknowledged the Christmas gift.	6	3
8	She had the stomach flu for four days.	8	6
9	How dare you call me a spineless wimp!	8	7
10	Don't get me started.	4	4
11	She was more than ready to come home.	8	8
12	The artist's work had a childlike quality.	7	1
13	He crushed a can on his head.	7	1
14	She finally stood up for herself.	6	6
15	Have a neighbor pick up your mail and paper.	9	8
16	Only the ants' coiled carcasses remained after the extermination.	9	0
17	It is wrong to borrow someone's car without permission.	9	4
18	The candle smelled like sugar cookies.	6	5
19	I doubt you were thinking too much.	7	4
~~20~~	~~The chocolate cake claimed to be fat free.~~	~~8~~	

Words Correct 62
Words Possible ~~140~~ 138
Percent Correct 45%

FIGURE 1 | Two examples of typical speech perception tests performed in the clinic. Both lists were obtained from the new Minimum Speech Test Battery (MSTB) for Adult Cochlear Implant Users [Auditory Potential LLC, (9)]. The one on the left (from page 12 of MSTB manual, https://www.auditorypotential.com/MSTB_Nav.html) shows the actual results from 50 monosyllabic consonant-nucleus-consonant (CNC) words. The clinician records the incorrect response and the number of correct phonemes for each stimulus. A tally of the number of correct words and phonemes is presented. The one on the right (from page 6 of the MSTB score sheets, https://www.auditorypotential.com/MSTB_Nav.htm) shows the results from a volunteer (see Methods - Datasets) listening to 19 sentences from the AzBio list #6. The clinician records the number and the total percentage of correct sentences. See **Figure 6** for the corresponding visual representation of these scores.

the user work remotely or in person with the clinician to identify areas of weaknesses. To that end, visualization of phonemic accuracy could be useful as motivation and diagnostic tool for patient and clinician respectively especially in the telemedicine era.

Hence, there is a need for an automated program to compute and visualize the accuracy of phonemes from responses to speech stimuli in real time. Specifically, given a stimulus-response pair of words or sentences, the problem is to develop and implement the automated program in four steps. First, use an online pronunciation dictionary to express the stimulus and response as two ASCII sequences of phonemes. Second, use an alignment technique to align the sequences. Third, calculate and visualize phoneme accuracy with respect to phonological features and associated subtypes. Fourth, make the program available to the computational audiology community.

The first two steps can be accomplished by leveraging two tools commonly used in speech recognition research. For the first, there are several online pronunciation dictionaries: Pronlex, CMUDict, CELEX and UNISYN to name but a few (11). Of these, CMUDict is publicly available and has been widely used in open source automatic speech recognition software such as Kaldi (12). For the second, several sequence alignment algorithms are available from scLite, which is part of an open source library (13, 14). The third step makes use of two commonly used metrics: a F1-score (Sørenson-Dice coefficient) for the phonemes and relative information transfer for the phonological features. The fourth step deploys the program in MATLAB so that it can be converted for open-source usage.

Using a pronunciation dictionary followed by automated sequence alignment for analyzing speech comprehension by people with hearing loss is not new. Previous uses include analyses of lipreading by people with normal hearing and hearing loss (15–18), estimating intelligibility from atypical speech (19–21) and more recently, listening to speech in noise by people with normal hearing (22). Using relative information transfer to analyze speech comprehension based on phonological features of transcribed phonemes is also not new. In addition to Bernstein (15), previous uses include analyses of listening by people with hearing loss (23–31). There was also a study of bimodal hearing with hearing aid and cochlear implant that manually transcribed phonemes with the aid of a digital dictionary (32). The approach here differs from earlier work in that the program is made publicly available by adopting and modifying two open-source algorithms and two commonly used metrics, with the goal of providing a visual representation of results similar to those shown in **Figure 1**.

This paper describes a pilot study of the design and implementation of the automated program. It reports the program's validation and the results of using it in several cases. Finally, it discusses the advantages and disadvantages of the program and provides suggestions for clinical usage.

METHODS

This section describes: (i) the design of the program; (ii) how stimulus-response pairs of words or sentences are formatted as two sequences of phonemes; (iii) how two sequences are aligned; (iv) how the F1-score is used to compute the accuracy of the stimulus phonemes; (v) how relative information transfer is used to assess the accuracy based on phonological features; (vi) how the preceding two metrics can be visualized for a set of stimulus-response pairs; (vii) the different datasets used for testing; and (viii) program validation.

Design

Figure 2 illustrates the overall design for analyzing the response of a person with hearing loss listening to sentences or words in speech tests in real time. The program first takes as input stimulus-response pairs in the form of sentences or words. Both are converted to phonemes using a digital pronunciation dictionary for each word, and the phonemes are entered into the alignment algorithm. Then the accuracy for the stimulus phonemes is computed in two ways via a F1-score for each phoneme and relative information transfer for ten different phonological features.

Input

The program uses the Carnegie Mellon University Pronouncing Dictionary (CMUDict) which is an open-source machine-readable pronunciation dictionary for North American English that contains over 134,000 words and their pronunciations (33). CMUDict has been widely used (11) for speech recognition and synthesis, as its entries map words to their pronunciations as ASCII symbols in the ARPAbet format (34). The ARPAbet format contains 39 phonemes with vowels each carrying a lexical stress marker. Transcriptions are expressed as strings of phonemes. The raw text file for the most stable version of CMUDict (0.7b) was downloaded from http://www.speech.cs.cmu.edu/cgi-bin/cmudict, and saved as a MATLAB map data structure. Also, lexical stress markers were removed as they did not affect the subsequent analysis. Misspelled or incorrectly pronounced words, however, need to be modified by the user. For example, in a YouTube demonstration of a subset of PBK-50, the word "pinch" was misheard as "kints," which is a nonsense word. Since CMUDict is not able to translate "kints" into phonemes, the user is directed to the online dictionary and enters real words such as "mints" and "key" yielding "M IH N T S" and "K IY" respectively so that "K IH N T S," is manually entered as the phonemic representation of "kints". Since the program splits its input sentences into words, it only requires manual input for nonsense words, not the entire sentence containing them.

Alignment

Given a paired strings of phonemes for the stimulus and response, the next step is to align the phonemes. Algorithms for aligning strings arise in other areas including bioinformatics (11, 35). The goal is to minimize the distance between two strings. The minimum edit distance (MED), based on the classic

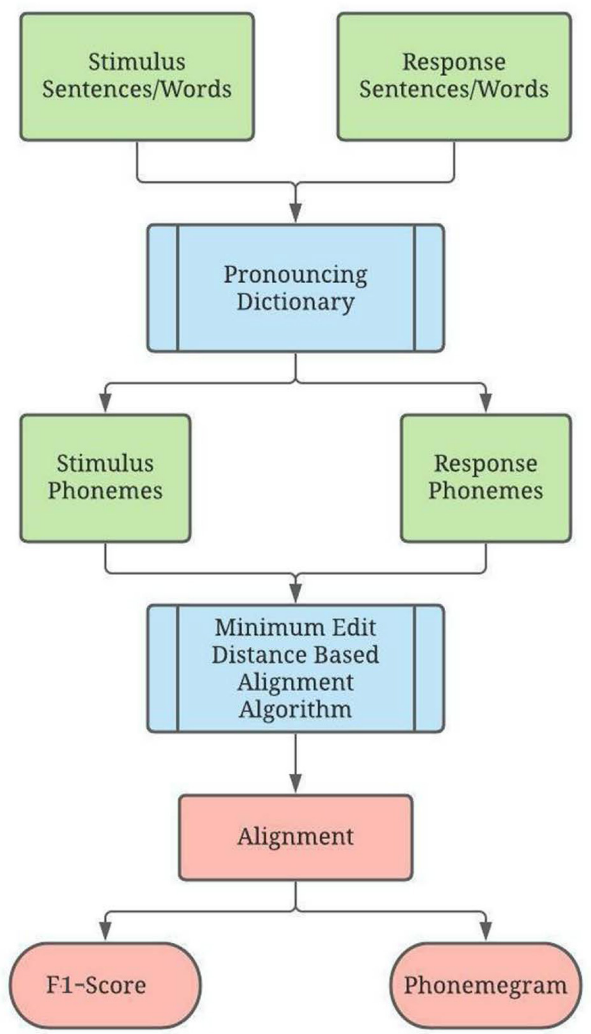


FIGURE 2 | Program Overview. Here, green represents input, blue represents the core algorithm, and red represents output. The program takes a set of stimuli and the set of corresponding responses as parameters. The stimuli and responses are translated from words into phonemes using a digital pronouncing dictionary. The phonemes for each stimulus-response pair are passed into the alignment algorithm, which displays an alignment and phoneme accuracy. Once all stimulus-response pairs have been evaluated, graphics of phonemegram and phoneme accuracy for vowels, voiced and unvoiced consonants are generated.

Levenshtein distance algorithm (36), computes the number of editing operations (insertion, deletion, and substitution) needed to transform one string to the other. Each operation is associated with a numerical cost or weight. Here the costs are modified for the particular case of aligning strings of phonemes. The MED is computed by applying dynamic programming (35) to generate an edit distance matrix which is a table of transitions from one string to the other. A global solution is built by solving and remembering the solutions to simpler subproblems, resulting in an alignment with the minimum associated cost.

The first version of alignment was implemented in MATLAB and used the Levenshtein algorithm from scLite (14) that used simple costs—1 for insertion or deletion, 2 for substitution, and 0 for a match. These favored quick matches, sometimes aligning a response phoneme far from any others—for example, if "bite" was the response to the stimulus of "birds bite," the initial "B" consonant in "bite" was aligned with that in "birds". As this caused issues with longer sentences, the costs were modified to discourage switching from a deletion (a space within the aligned response) to an insertion or substitution (of a response phoneme), and vice versa. As a first step toward avoiding multiple alignments, modification was accomplished by adding a 0.5 cost for deletion if the previous operation was insertion or substitution, and a 0.5 cost for insertion/substitution if the previous operation was deletion. The two exceptions are within the substitution cost. To favor matches, the cost for a match after a deletion or an insertion is an extra 0.2 or 0.1, respectively, instead of 0.5. These costs are summarized in the left half of the first row and the second and third columns of **Table 2**.

Initially, the algorithm was coded to generate one alignment once the edit distance matrix was filled. However, this did not guarantee the best alignment. Multiple alignments led to the same MED if, for example, fewer phonemes than expected were entered, and the algorithm aligned incorrect phonemes in different places. Previously, the algorithm would assign each cell in the edit distance matrix a single operation, even if two or more operations led to the same MED. Consequently, the algorithm would generate a single alignment, arbitrarily based on the order of costs evaluated. By logging all of the operations that led to the same MED in a cell of the edit distance matrix, this single alignment was found to be a result of these simple costs. Many alignments—even over 1000—led to the same MED. The costs were then modified to favor substitutions for phoneme alignments that have similar phonological features (1). **Table 1** maps the following 10 phonological features to the 39 phonemes: nasality, vowel height, manner, voicing, contour, vowel place, vowel length, affrication, sibilance and consonant place. These features and their subtypes (described in the caption for **Table 1**) are used to deem consonant-consonant and vowel-vowel alignments sharing all or most of their attributes as similar, and given a substitution cost deduction to favor substitution of "similar" phonemes. For example, a voiced "F" results in a "V", so the program will prefer the substitution of these two phonemes over any other incorrect substitutions. Even after implementing the similarity cost deductions, the algorithm often generated several alignments, some of which were preferable to others. To further favor alignments that represent probable errors, a slight consonant manner cost deduction was implemented, in order to prefer substitution between two manner subtypes such as stops, fricatives, or glides. For example, if the algorithm must choose between aligning the stop phoneme "K" with the fricative "S" or the stop "P", the algorithm will choose to align the stops together. Details of possible consonant-consonant, vowel-vowel and consonant manner pairs are given in the **Supplementary Data**. Last but not least, vowel-consonant substitution is heavily penalized to prevent alignment

TABLE 1 | Phonological features of consonants and vowels based on Ladefoged and Johnstone (1).

Phonological Features (Vowels)

Phoneme	Vowel height	Contour	Vowel place	Vowel length
AA	0	1	2	0
AE	0	1	1	0
AH	1	1	1	0
AO	1	1	2	0
AW	0	2	1	1
AY	0	0	1	1
EH	1	1	0	0
ER	1	1	0	0
EY	1	1	1	0
IH	1	0	0	1
IY	2	1	0	0
OW	1	0	2	1
OY	1	2	2	1
UH	2	1	2	0
UW	2	1	2	1

Phonological Features (Consonants)

Phoneme	Nasality	Manner	Voicing	Affrication	Sibilance	Place
B	0	0	1	0	0	0
CH	0	4	0	1	1	1
D	0	0	1	0	0	1
DH	0	2	0	1	0	0
F	0	2	0	1	0	0
G	0	0	1	0	0	2
HH	0	2	0	1	0	2
JH	0	4	1	1	1	1
K	0	0	0	0	0	2
L	0	3	1	0	0	1
M	1	1	1	0	0	0
N	1	1	1	0	0	1
NG	1	1	1	0	0	2
P	0	0	0	0	0	0
R	0	3	1	0	0	1
S	0	2	0	1	1	1
SH	0	2	0	1	1	1
T	0	0	0	0	0	1
TH	0	2	1	1	0	0
V	0	2	1	1	0	0
W	0	3	1	0	0	0
Y	0	3	1	0	0	1
Z	0	2	1	1	1	1
ZH	0	2	1	1	1	1

Values of subtypes for vowel height are: 0 = low, 1 = mid, 2 = high; vowel place are 0 = front, 1 = central, 2 = back; contour are 0 = rising, 1 = flat, 2 = falling; vowel length are 0 = short, 1 = long; consonant manner are: 0 = stop, 1 = nasal; 2 = fricative; 3 = glide; 4 = affricate and consonant place are: 0 = front, 1 = center, 2 = back.

of consonants with vowels. With these modified costs, the alignment should then accurately reflect the response. These costs are summarized in **Table 2**.

Figure 3 shows an example of the operations used in MED with the costs from **Table 2** for aligning the response "thin" with the stimulus "fun". **Figure 4** shows the differences between the

TABLE 2 | Costs for each operation (left–insertion and deletion; right–substitution), depending on the previous operation.

Current operation	Insertion		Deletion		Current operation	Substitution		
Previous operation	Ins	Sub/Del	Del	Ins/Sub	Previous operation	Sub	Ins	Del
Cost	1	1.5	1.5	1	Vowel-cons or vice versa	5	5.1	5.5
					Consonant-consonant	1.75	1.85	2.25
					Same manner consonants	1.3	1.4	1.8
					Similar consonants	1.2	1.3	1.7
					Vowel-vowel	0.9	0.8	1.4
					Similar vowels	0.65	0.75	1.15
					Match	0	0.1	0.2

The previous operation is considered in order to prefer alignments with the fewest phoneme-to-space and space-to-phoneme transitions. The left half of the first row shows the addition of a 0.5 cost for deletion if the previous operation was insertion or substitution, and a 0.5 cost for insertion/substitution if the previous operation was deletion. The first column on the right shows the costs for the substitutions. The second and third columns show the two exceptions for the substitution cost—to favor matches, the cost for a match after a deletion or an insertion is an extra 0.2 or 0.1. See ***Supplementary Data*** *for examples of similar consonants, same manner consonants, and similar vowels.*

outputs for aligning the response "We live on the earth" with the stimulus "These are your books" (from participant V1-HA in the test data, see below and **Table 3**) from three different alignments: the diff function first used in UNIX (37) and available in scLite, the original Levenshtein algorithm with simple costs, and the finalized modified algorithm. diff gave no weight to consonants or vowels. The unmodified algorithm yielded two alignments generated with no similarity substitution costs whereas the modified algorithm with similarity substitution yielded just one alignment, because "S" and "TH" are two consonants with similar manner that are assigned a substitution cost deduction.

F1-Score

For each phoneme in each stimulus, the true positive (TP), false positive (FP) and false negative (FN) values were used to compute the F1-Score, or the Sørensen-Dice coefficient, which is defined as the harmonic mean of precision ($TP/(TP + FP)$) and sensitivity ($TP/(TP + FN)$), i.e., $2TP/(2TP + FP + FN)$. Consider the phoneme "K" as an example. A TP occurs when a "K" response is matched with a "K" stimulus; a FN occurs when not recording a "K" stimulus; a FP occurs when recording a non-existent "K". **Figure 5** shows examples of alignments and phoneme F1-scores for four challenging stimulus-response pairs. The first two are examples of the consequences of insertion and deletion [see Table 4 from (38)]. The third example is one of phonemic ambiguity but with different alignments caused by one substitution. The fourth illustrates the use of all three MED operations in the alignment.

Phonemegram

Following ideas by Danhauer and Singh (29–31), Blamey et al. (25) and others (15, 32, 39–41), an alternative way of visualizing speech comprehension performance is to construct a phonemegram. Specifically, a histogram of relative information transfer for the phonological features from **Table 1** over a range from low to high frequency was created as follows. First, confusion matrices for the consonants and vowels were generated. Each matrix consisted of N rows of phonemes in the stimulus set and $N + 1$ columns of phonemes in the response set with the extra column reserved for unclassified phonemes due to empty responses (40, 42). The matrices were regenerated as several smaller ones based on the prescribed phonological features. For example, within the vowel height feature, vowels can be further divided into three separate categories: high, mid, and low. In this way, the relative information transfer can be obtained for different features. Following Miller and Nicely (43) and others, the information transfer for each feature was computed via $IT = \log(n) + H_x + H_y - H_{xy}$ where H_x, H_y, and H_{xy} refer to the row (stimulus), column (response), and element entropy respectively, while n refers to the total number of entries within the feature matrix. The entropies are characterized by:

$$H = \frac{1}{n}\sum s\ \log(s)$$

where s refers to either the individual elements, row sums, or column sums of the feature matrix for computing H_{xy}, H_x, and H_y respectively. Then the relative information transfer is given by:

$$H_{stim} = -\sum_{i=1}^{n}\left(\frac{p_i}{p_{total}}\right)\log\left(\frac{p_i}{p_{total}}\right)$$

where n refers to the number of different sub-categories within the feature, p_i refers to the number of phonemes presented that are within the given sub-category, and p_{total} refers to the total number of phonemes (regardless of subcategory). For example, if out of 16 consonants presented, seven are voiced and nine are unvoiced, then $H_{stim} = -(7/16)\log(7/16) - (9/16)\log(9/16)$.

Output

For each stimulus-response pair, the program displays the best alignment, as well as the unique phonemes in the stimulus and their F1-scores. After all responses are analyzed, the program generates three plots showing the averaged F1-scores (expressed as percentages) for individual phonemes with respect to the classic two dimensional representation of phonological features (1). In these plots, the averaged F1-score is color-coded and assigned at the (x, y) coordinates corresponding to the place (x) and manner or vowel height (y) for each phoneme. A color bar

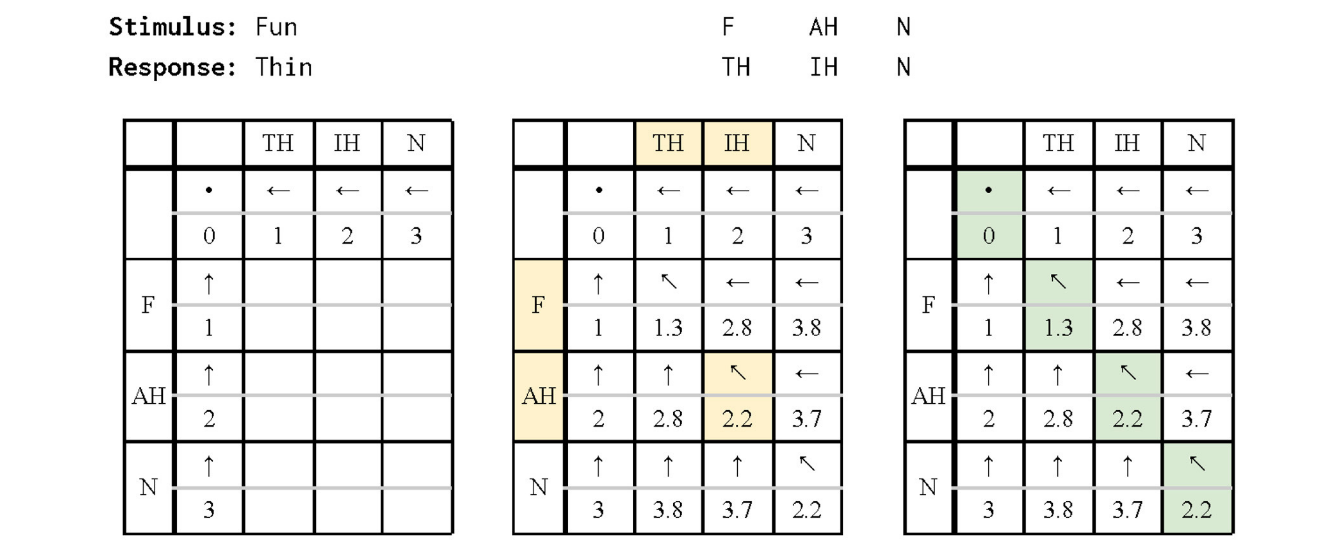


FIGURE 3 | The response phonemes are placed on the top row of the edit distance matrix, while the stimulus phonemes are on the left column. Each square represents the minimum edit distance (MED) for the substrings on each axis, and shows what operation was executed to get to that MED (⟵ is insertion, ↑ is deletion, ↖ is substitution). **Left**: These squares (comparing all substrings of response or stimulus sentence to an empty string) are filled in first, to provide base cases for the rest of the matrix. The MED between an empty string and any string of length *n* is equal to *n*. **Middle**: The highlighted square finds the MED between the response of "TH IH" and the stimulus of "F AH." It does this by building on the squares of the matrix that have already been filled. Insertion entails aligning the IH with a space (cost 1.5) and adding onto the optimal alignment of "TH" and "F AH" (cost 2.8), for a total cost of 4.3; deletion aligns a space with the AH (1.5) and adds onto the alignment of "TH IH" and "F" (2.8), for a total cost of 4.3; substitution aligns the IH with the AH (0.9) and adds to the alignment of "TH" and "F" (1.3), for a total cost of 2.2. The substitution cost is the lowest, so the matrix records the cost of 2.2 and the substitution operation. **Right**: Once the entire matrix has been filled, the algorithm finds how it generated the MED by tracing back the recorded operations. In this case, the MED of "TH IH N" and "F AE N" is 2.2, and the alignment consists of three substitutions.

shows the range from 0 to 100 for the F1-score. A fourth plot shows the phonemegram with the relative information transfer for each feature computed as a percentage. Histogram bars are color-coded corresponding to the frequency ranges associated for the features: black was assigned to low frequency for nasality, vowel height, manner and voicing; dark blue assigned to medium frequency for the vowels–contour, vowel place, vowel length; light blue to medium frequency for the consonants–affrication; and white to high frequency for the consonants–sibilance and place.

Datasets

Two datasets were used. One dataset consisted of test data at sentence and word levels. For the sentences, six volunteers with hearing loss recorded their responses to stimuli. In 2017, four people with various degrees of hearing loss tested the alpha version of the Speech Banana iPad app for auditory training (10); testing was approved by JHU Homewood Institutional Research Board Protocol HIRB00001670. Specifically, the volunteers provided their responses to different sets of 30 sentences recorded in Clear Speech (44) by male and female American English speakers, extracted as WAV audio files from the app which is based on an auditory training book (45). At the same time, two clinical audiologists who also use cochlear implants donated their responses to 19 sentences from AzBio lists #1 and #6 (8) with stimuli presented at 60 dB SPL with 12-talker babble at 50 dB SPL. For the words, stimulus-response pairs were obtained from three sources: a) MSTB [page 12 in (9)], b) List 1 of PBK-50 (2, 46) with the "kints" response to "pinch" observed in a YouTube video clip (https://www.youtube.com/watch?v=GPRwA9BG-m4), and c) erroneous responses to AB word lists (3) by several adults with hearing loss [Table 1-1 in (47)] including the "she's" response to "cheese". The other dataset consisted of stimulus-response pairs of 31 participants (age range: 22–79 years), each listening to 16 lists of 25 Basic English Lexicon (BEL) sentences (48) at four different SNRs (0, 5, 10, quiet) obtained from speech perception experiments (49, 50); these lists are akin to and more extensive than the BKB-SIN lists (51). For this dataset, protocols (8804M00507) were approved by the Institutional Review Board of the University of Minnesota, and all participants provided written informed consent prior to participating.

Validation

The large dataset of 12,400 actual stimulus-response pairs from 31 participants listening to 400 sentences is used to validate the program. A set of 12,400 random pairs is created by randomizing the responses such that none of the actual pairs are replicated. Following similar approach (15, 17), three computations are performed. First is a frequency histogram of sentences with the number of correct phonemes in the response (indicated by the number of TPs in the calculation of the F1-scores). Second is the entropy or uncertainty for each of the 39 phonemes obtained from the two confusion matrices for the consonants and the

Stimulus: We live on the earth **Response:** These are your books

A - Using Diff

W	IY	L	AY	V	AA	N	DH	AH	**ER**	**TH**	**_**	**_**	**_**
DH	IY	Z	_	_	AA	R	Y	AO	**R**	**B**	**UH**	**K**	**S**

B - Primitive algorithm (multiple alignments)

W	IY	L	AY	V	AA	N	DH	AH	**_**	**_**	**ER**	**TH**	**_**
DH	IY	Z	_	_	AA	R	Y	AO	**R**	**B**	**UH**	**K**	**S**

W	IY	L	AY	V	AA	N	DH	AH	_	_	ER	**_**	**TH**
DH	IY	Z	_	_	AA	R	Y	AO	R	B	UH	**K**	**S**

C - Substitution with similarity cost deductions

W	IY	L	AY	V	AA	N	DH	AH	_	_	ER	**_**	**TH**
DH	IY	Z	_	_	AA	R	Y	AO	R	B	UH	**K**	**S**

FIGURE 4 | Comparison of three different alignment algorithms for a stimulus-response pair taken from test data V1-HA (see **Table 3**). **(A)** The alignment generated by the UNIX diff function. The function gives no weight to consonants or vowels, and has no issues with aligning consonants with vowels and vice versa, as shown by the bolded area. **(B)** Multiple alignments generated by primitive algorithm, with no similarity substitution costs. Although the most of the response matches the stimulus, the algorithm generated two alignments with the same MED. **(C)** With the similarity substitution cost implemented, the algorithm generates only one alignment, because S and TH are produced in a similar manner, and therefore have a substitution cost deduction.

vowels used for the phonemegram. Similar to above, the entropy is calculated as $-\sum_{k=1}^{40} p_k \, log_2 \, p_k$ where k sums over all the response phonemes as well as unclassified ones due to empty responses (40, 42). Third is the information transfer for the same ten phonological features used in the phonemegram.

RESULTS

The results from the two datasets are shown in **Table 3, Figures 6–11**, and **Supplementary Figures 1–4**. **Figure 6** provides the desired visual representation of results in **Figure 1** from the two examples from the CNC word list (from the MSTB manual) and AzBio List #6 (by one of the two clinical audiologists with a cochlear implant). **Figure 7** shows results for one person with profound congenital bilateral hearing loss (V1), aided bimodally with a cochlear implant and a hearing aid (**top**), unilaterally with just the cochlear implant (**middle**), and unilaterally with just the hearing aid (**bottom**). **Figure 8** shows results for one person with severe hearing loss (V2) without using an in the canal hearing aid (**top**) and one person with partial but progressive hearing loss (V3), aided with bilateral hearing aids since childhood (**bottom**). **Figure 8** should be compared with **Supplementary Figure 1** showing near perfect results from V2 aided with the in the canal hearing aid (**top**), one person with severe progressive hearing loss (V4, **middle**) who has been using bilateral hearing aids for a few years and the other clinical audiologist (V5, **bottom**). **Figure 9** shows the results from the two other word lists. **Table 3** reports the number of total and correct sentences, words and phonemes for the test and validation datasets, with the last column indicating that the program is able to give comprehensive results in real time; note that the one case of manual intervention, such as entering the phonemes for nonsense responses, resulted in a slightly longer run time. Limiting the analysis to only incorrect stimulus-response pairs did not drastically alter the visualization of phoneme accuracy. Of the 361 stimulus-response pairs used for **Figures 6–9**, there were just two instances of double alignments. **Figure 10** visualizes the pooled results of the responses from 31 participants with cochlear implants listening to lists of BEL sentences as spoken by different talkers at different

TABLE 3 | Number of sentence or word stimuli and their responses with the program run time for the examples shown in **Figures 6–11** and **Supplementary Figures 1–4**.

Participant	Figure	Stimuli dataset	# Stimulus sentences	# Correct response sentences	# Stimulus words	# Correct response words	# Stimulus phonemes	# Response phonemes	Time (secs)
V1 - CI+HA	7	SB	30	13	165	122	490	483	6.5
V1 - CI	7	SB	30	12	162	101	484	446	4.3
V1 - HA	7	SB	30	0	160	36	488	317	4.0
V2 - CIC HA	S1	SB	30	28	164	159	470	473	3.5
V2 - No HA	8	SB	30	11	153	66	458	211	3.3
V3 – HA	8	SB	30	9	160	108	488	378	3.5
V4 – HA	S1	SB	30	27	164	158	470	473	3.3
V5 – CI	S1	AzBio#1	19	11	146	128	527	515	3.3
V6 – CI	6	AzBio#6	19	3	138	64	492	396	3.2
Anonymous	6	CNC			50	23	150	148	3.0
Anonymous	9	PBK			25	7	69	82	7.0
Several	9	AB			38	0	114	118	2.8
Actual ($N = 31$)	10,11, S2-S4	BEL	12,400	4,310	74,245	43,514	281,480	212,584	364.5
Random ($N = 31$)	11	BEL	12,400	3	74,245	7,424	281,480	212,584	358.8

CI, cochlear implant; HA, Hearing Aid; CIC, Completely in Canal; SB, Speech Banana; AB, Boothroyd.

SNR levels; the runtimes for the individual participants shown in **Supplementary Figures S2–S4** ranged from 9.3 to 22.7 secs. **Figure 11** shows the validation results by comparing 12,400 actual and random stimulus-response pairs in three different ways. There were just 45 instances of double alignments from the actual pairs. MATLAB scripts including the stimulus-response pairs used to generate these figures (except the validation data) are available from https://github.com/SpeechBanana/SpeechPerceptionTest-PhonemeAnalysis.

DISCUSSION

In this pilot study, an automated program for visualizing phoneme accuracy in speech perception tests has been developed and implemented. Two key features are the use of a digital speech pronouncing dictionary for automated derivation of the phonemes from stimuli and responses, and the modification of the Levenshtein minimum edit distance via dynamic programming for automated alignment of phonemes. Traditionally, speech pronouncing dictionaries have been used in speech recognition research for purposes such as aligning phonemes in speech-to-text translation (38). Here, the open source CMUDict is used for aligning phonemes in text-to-text comparison. The program is able to parse results (**Figure 1**) from standard speech tests at the phoneme level (**Figure 6**) in a robust, efficient, flexible and fast manner.

Several observations can be made. First, there is a benefit from amplification which, however was not an aim of this work. Second, while the averaged F1-scores are informative overall, the phonemegram analysis of the sentences appear to provide less information than that for the word tests which, could be attributed to significant top-down or contextual processing when presented with sentences. Third, there is potentially more information provided by the analysis of phonemes than just the number of correct sentences, words or even phonemes. Here accuracy is viewed in two different ways. The first shows the consequences of inserting, deleting and substituting phonemes and the second shows the perception of the phonological features. Such information about phonemes could help guide auditory training either in the clinic or at home.

A few things can be observed from the validation experiments. First, the likelihood of having five or more exactly matched phonemes for a randomized pair is low (∼44%) compared with that for an actual pair (∼83%). Second, actual responses can be distinguished from the randomly assigned ones. Third, there is higher uncertainty in response phonemes from random pairs (with a difference of about 1–1.5 bits across all phonemes). Fourth, very little information for the features can be discerned from the randomized pairs. Fifth, the tail of the distributions for actual pairs is higher due to better speech comprehension with cochlear implants even across different SNR levels while the tail for random pairs is influenced by a combination of duplicated pairs and mismatches of just a few words. These observations suggest that the program is robust to the nature of the stimulus-response pairs.

Although automated alignment of phonemes have been used for evaluation of speech recognition systems (52, 53), this study is not the first reported use of automated alignment of phonemes to study speech comprehension by people with hearing loss. The earlier work of Bernstein and colleagues (15–18) mainly focused on lipreading i.e., comprehension via audiovisual stimuli for people with normal hearing and hearing loss and only recently has this focus moved to listening to speech in noise by people with normal hearing (22). Alignment of phonemes via dynamic programming was also used by Ghio and colleagues (19–21) to develop intelligibility metrics for atypical speech. Therefore,

Stimulus: ascending **Response:** and sending

AH	_	_	S	EH	N	D	IH	NG
AH	N	D	S	EH	N	D	IH	NG

Unique Phonemes and their F-Scores

AH	D	EH	IH	N	NG	S
100	66.6	100	100	66.6	100	100

Stimulus: crude leaf **Response:** crudely

K	R	UW	D	L	IY	F
K	R	UW	D	L	IY	_

Unique Phonemes and their F-Scores

D	F	IY	K	L	R	UW
100	0	100	100	100	100	100

Stimulus: an app **Response:** a nap

AH	N	AE	P
AE	N	AE	P

Unique Phonemes and their F-Scores

AE	N	P
66.6	100	100

Stimulus: Father is in the car **Response:** Everything in the car

F	AA	DH	_	ER	R	IH	Z	IH	N	DH	AH	K
_	EH	V	R	IY	TH	IH	NG	IH	N	DH	AH	K

AA	R
AA	R

Unique Phonemes and their F-Scores

AA	AH	DH	ER	F	IH	K	N	R	Z
66.6	100	66.6	0	0	100	100	100	66.6	0

FIGURE 5 | Four examples of alignments and phoneme percent accuracy. The first example shows insertion of the phonemes N and D. The second example shows deletion of the phoneme F. The third example shows substitution of the AE phoneme (æ) for the AH phoneme (ə). The fourth example has all three minimum edit distance operations within its alignment.

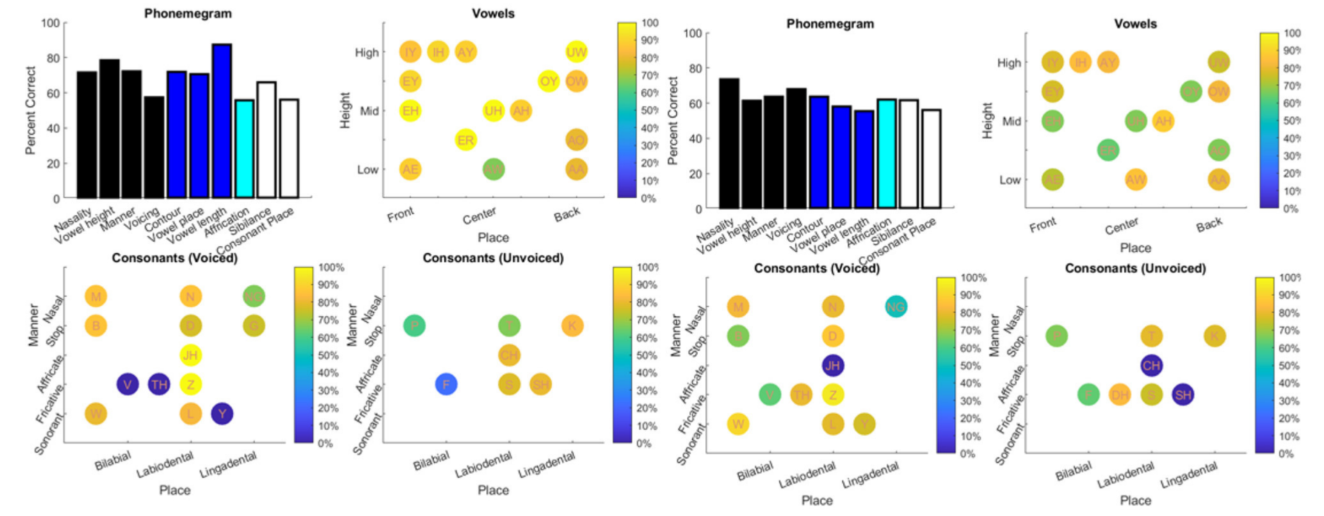


FIGURE 6 | Visual representation of the two typical examples of scoring in the clinic shown in **Figure 1**. The results from the CNC word list and AzBio List #6 are shown on the left and right respectively. See **Table 3** for details.

it is helpful to discuss similarities and differences between these approaches in four areas: pronunciation dictionary, costs, alignment and metrics.

CMUDict is open-source and has more than 134,000 words, which is an order of magnitude larger than 35,000 words in PhLex used by Seitz, Bernstein, Auer Jr, and MacEachern (54). Words not in CMUDict were manually parsed and entered in the CMUDict website to yield the phonemic string while a rule-based transcription system was used by Bernstein et al. (22). Ghio et al. (20) used a French based pronunciation dictionary (55).

The costs in **Table 2** are essentially *ad hoc*, having evolved from the open-source scLite software used for the Levenshtein algorithm. It is worth noting that Bernstein (15) and Bernstein et al. (17) initially used *ad hoc* costs which were fixed with perceptually based costs. Costs for vowel-vowel, consonant-consonant and same consonant manner alignments were modified based on having similar phonological features. In fact, a similar approach has been adopted by Ghio and colleagues (19–21) and Kondrak (56, 57) who used Hamming distance matrices for vowels and consonants based on deviations from features for the costs used by dynamic programming for analyzing atypical speech and different languages, respectively; Ruiz and Federico (38) used a similar approach with constraints for vowels and consonants in analyzing speech translation. The **Supplementary Data** shows that the phoneme pairs deemed to be similar can actually be derived by thresholding the distance matrix for the vowels and the stratified distance matrices for the consonants. It should then be possible to make formal use of these distance matrices. While voicing was not explicitly used in setting costs, it was actually used to determine the costs for consonant-consonant substitution pairs. As described in the **Supplementary Data**, the sibilant consonants were grouped and then non-sibilant consonants were stratified based on first manner, then place and voicing. In contrast, Bernstein et al. (22) perceptually computes costs based on the Euclidean distance between two phonemes derived from multidimensional scaling of confusion matrices for consonants and vowels from people with normal hearing. Since further work should compare the feature- and perceptual- based approaches, this work should be considered as a pilot study.

The use of modified costs in MED operations to align the phonemes in **Figure 2** should be contrasted with that in **Figure 1** in Bernstein (15). Usually, MED operations yield multiple alignments (**Figure 4**); see also Bernstein et al. (17) and **Figure 2** in Bernstein (15). In this work as well as the recent work by Bernstein et al. (22) and Ghio et al. (20), single alignments are achieved in virtually all cases which may be attributed to the use of costs derived from the distance matrices. About 0.6% of the stimulus-response pairs in both test ($n = 2$) and validation ($n = 45$) datasets yielded multiple—actually double—alignments. In the rare case of double alignments, the user is given the manual option of choosing the best one; by default, the program selects the first of the two alignments. It is remarkable that only double alignments occurred; in fact, more than two alignments occurred when a lower cost of two instead of five for consonant-vowel substitution was used. Inspection of the 45 stimulus-response pairs from the validation dataset that yielded double alignments suggests that these arise depending on the type of the response. The response may be nearly complete such that the alignment cannot decide between two similar phonemes, or a purely random guess, or a combination of correct and random words. This is actually borne by instances of double alignments from 2.5% ($n = 305$) of the randomized stimulus-response pairs from the validation dataset. Avoidance or significant reduction of multiple alignments using feature-based costs were also observed in a comparative study of Dutch dialects (58). As this work is a pilot study, further work should explore differences accrued

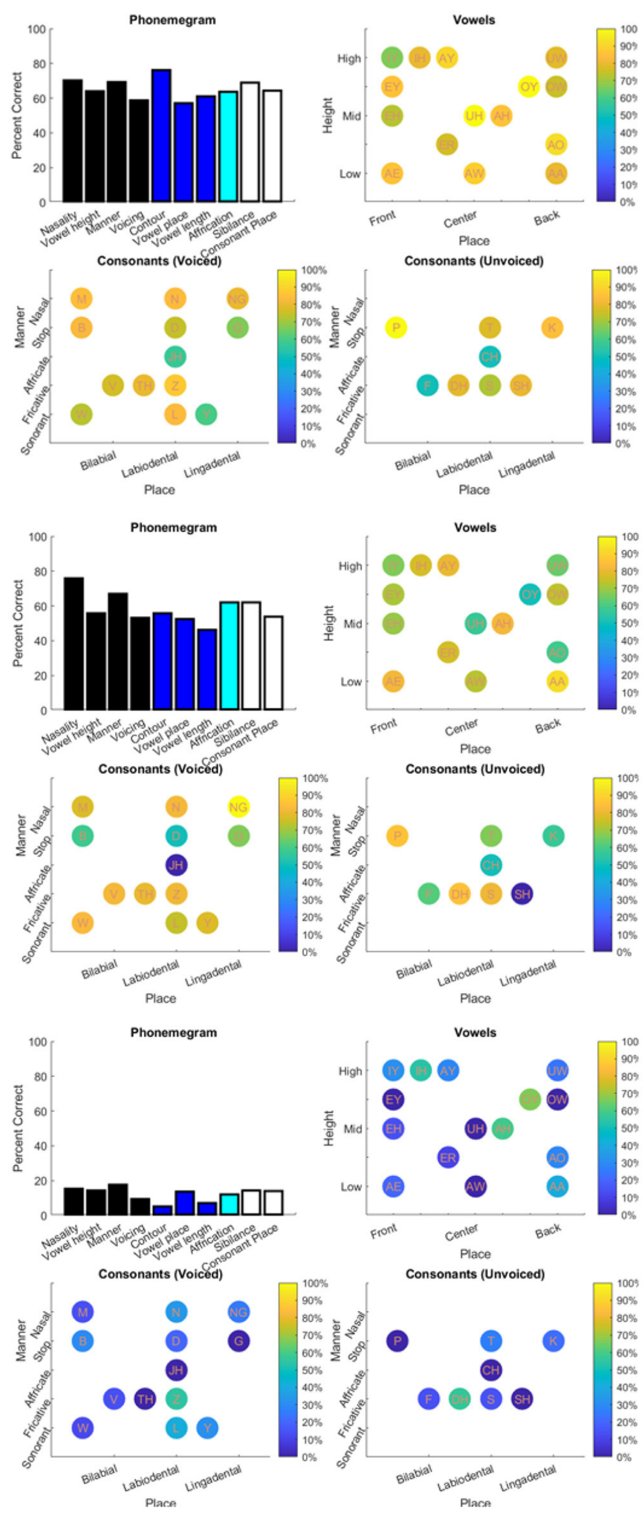


FIGURE 7 | Results for V1 with both cochlear implant (CI) and hearing aid (HA) (**top**), CI only (**middle**) and HA only (**bottom**) in response to a set of 30 sentences extracted from Speech Banana auditory training app. See **Table 3** for details.

from feature-based *ad hoc* and perceptual-based costs. These differences might be reflected by comparing the alignments for 12 stimulus-response pairs listed in **Table 1** of Bernstein et al. (22) with those produced by the program in the **Supplementary Data**. There may be problems with sparse responses such as misaligning one response phoneme in a correct word with the stimulus phoneme in a different (as in a preceding) word, ironically without loss of accuracy so future work should incorporate costs for boundary detection (38). These problems are likely not to occur with word lists or nearly complete sentences which may be more helpful in pinpointing areas of weaknesses for auditory training. Others have used MED for aligning phonetic transcriptions of words based on phonological features (59) and fuzzy string matching with a novel metric for sentences (60), both of which are available as open source. Future work should also explore using costs from confusion matrices from people with normal hearing listening to sentences as opposed to words.

In this work, two sets of commonly used metrics are used. One is the F1-score which is a function of true positives, true negatives and false positives for each phoneme and visualized with respect to manner, place and voicing for consonants and height and place for vowels. The other is the relative informationtransfer or entropy for each of the 10 phonological features used to construct the phonemegram. In contrast, the recent work of Bernstein et al. (22) proposed mining three metrics to analyze listening by people with normal hearing to speech in noise. These were (i) phoneme substitution dissimilarity, which measures the perceptual distance between separate stimulus phonemes and all incorrect phonemes in the response, (ii) number of words correct, and (iii) number of insertions. The former is obtained from dividing the sum of the phoneme-to-phoneme costs for incorrect substitutions by the number of substitutions. The latter is obtained by the count of the number of phonemes that could not be aligned as substitutions. In contrast, the program did not save these types of data except for the number of true positives needed for the validation study (**Figure 11**, top left). It is argued that due to different manipulations of intrinsic data the two different set of metrics are probably related in some way or other. Furthermore, in analyzing people with speech disorders, Ghio et al. (20) used the distance between the expected and actual sequence. As this is a pilot study, future work would be necessary to uncover and explore these relationships particularly in a comparison i.e., statistical study.

Care must be taken to interpret the accuracy for phonological features. Take, for example, analysis of several people with hearing loss in the bottom panel of **Figure 9**. The near-perfect scores for vowel height, contour, and vowel place may seem inaccurate but the phonological analysis of the vowels show an inability to identify IY and IH. Since these vowels are grouped for the vowel height, contour, and vowel place features, accuracy for identifying phonemes with these features remains at 100%. In other words, even though there may have been confusion between IY with IH, since both are identical with respect to their categorization within the vowel height, contour, and vowel place feature groups, the responses showed the ability

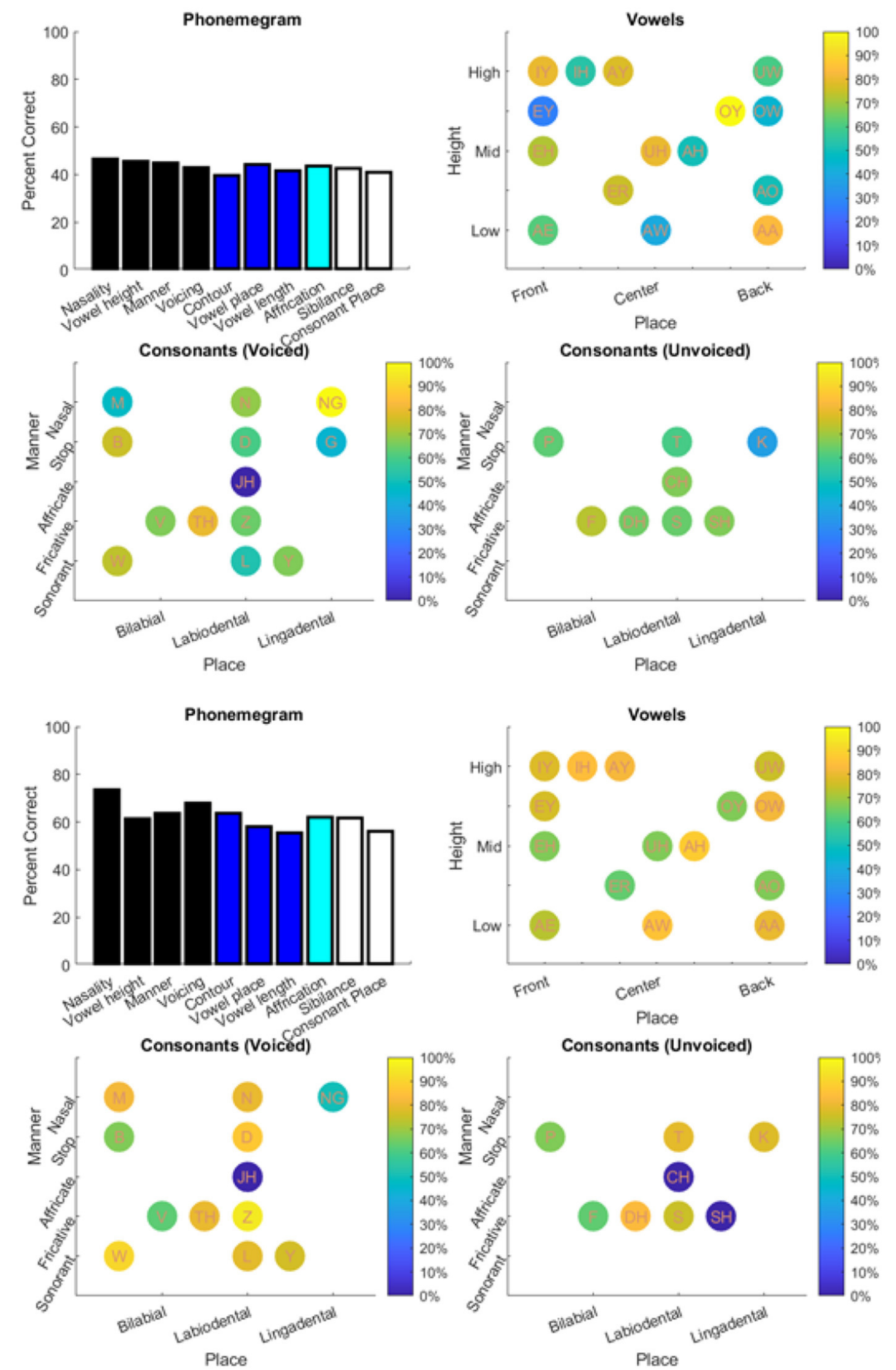


FIGURE 8 | Results for V2 without in the canal HA (**top**) and V3 (**bottom**) with HA responding to different sets of 30 sentences extracted from Speech Banana auditory training app. See **Table 3** for details.

to detect those features at a high rate. Similarly, for the PBK-50 test (**Figure 9**, **top**), since the non-nasal consonants are still categorized as the same the nasality feature is recorded perfectly.

The availability of datasets from recently published experiments provided an opportunity to assess the potential use of phoneme alignment in these experiments. For example, O'Neill et al. (49) recorded the BEL sentences using four different talkers, as well as developed and recorded 20 lists of nonsense sentences derived from the BEL corpus. These stimuli were used in speech perception

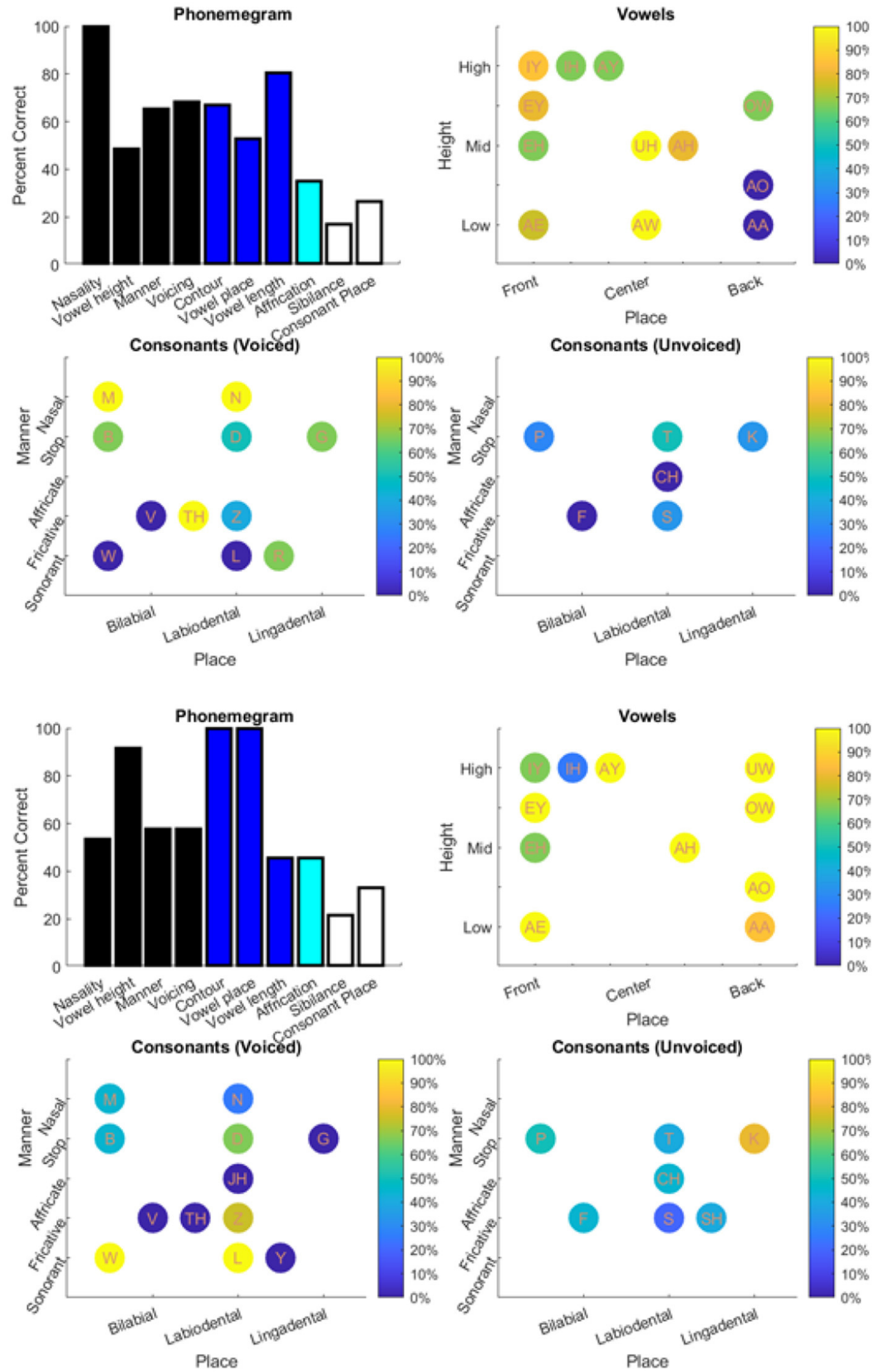


FIGURE 9 | Results for output from two different word tests: PBK-50 (**top**) and AB (**bottom**). See **Table 3** for details.

experiments involving people with normal hearing and hearing loss (49, 50). The visualization of phoneme accuracy from **Supplementary Figures S2–S4** for one experiment (50) provides potentially more information than the reported percentage of correctly identified keywords in sentences.

By construction, the phonemegram offers a different perspective of speech comprehension based on phonological

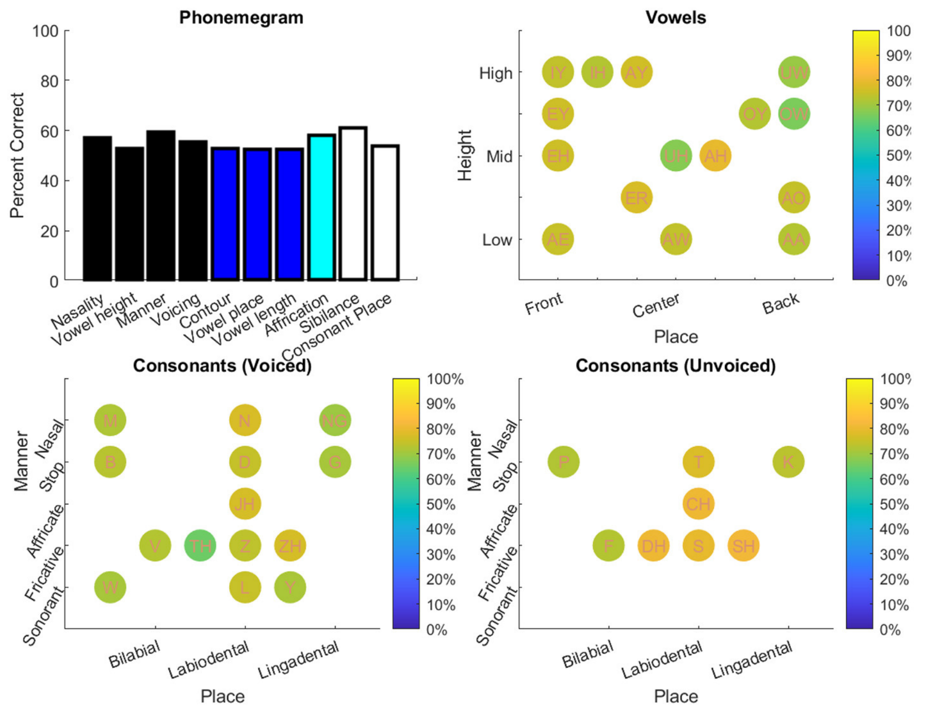


FIGURE 10 | Analysis of the stimulus-response pairs pooled from the O'Neill et al. (50) study of 31 participants with cochlear implants listening to 16 lists of 25 sentences as spoken by four speakers at four different SNR levels. Analysis from the individual participants is shown in **Supplementary Figures 2–4**. See **Table 3** for details.

features of the phonemes, specifically information transfer of features with respect to a frequency range, as used in the Infogram for hearing aid fitting in tele-audiology (23–25). Information transfer analysis has also been adopted (32, 39) who compared low frequency phonemes (diphthongs, semivowels, and nasals) to high frequency phonemes (sibilants, fricatives, bursts, and plosives). The frequency aspect of the phonemegram may be complemented by the averaged F1-scores for the vowels based on the inverse relationship between place from back to front (manner from high to low) and the 1st (2nd) formant (1). Further, the phonemegram can compensate for the absence of variance for the F1-scores since it records just the information transfer for a feature. It is important to provide enough repetitions for each phoneme, otherwise the transmitted information estimate becomes highly erratic and overestimates the stimulus information on average (43, 61). Accumulating responses over time is one way to overcome bias and error which might be useful in mobile apps for auditory training (10). In this case, it may be necessary to use bootstrapping to generate confidence intervals (62). Furthermore, since non-symmetric confusion matrices have been considered in analysis of speech perception by people with hearing loss (40, 42), it is reasonable to consider non-classified phonemes accruing from empty responses. Further work could consider a more appropriate alternative visualization by generating 3D plots of F1-scores for each phoneme with respect to the first three formants.

A challenge for testing the program was obtaining examples of stimulus-response pairs from people with hearing loss. Fortunately, the program was developed at the same time as the development of the Speech Banana mobile app for auditory training which allowed for testers to provide valuable data. In this era of digital hearing health, there is a great need for raw data such as stimulus-response pairs from scientific studies to be made available publicly in the same way as human neuroimaging data are now being made available for the scientific community (63, 64). The use of the data from recently published speech perception experiments is a step in that direction.

The program has several other advantages. First, though currently implemented in Matlab, the program can be implemented in Python, Javascript or even R. Second, it could be self-administered or used in telepractice by people with hearing loss, who are learning to hear with a new hearing aid or cochlear implant. Results are saved over time for feedback with the speech language pathologist or audiologist. Third, the program could be integrated with inputs from NU-6, CUNY, Az-Bio, HINT or BKB for real-time quantification in the clinic; further, the program could be integrated with more challenging tests such as Az-TIMIT (65), STARR (66) and PRESTO (67). Fourth, as implied by the Infogram, the phonemegram may offer audiologists a frame of reference for the ability of the person with hearing loss to perceive speech at different frequencies. Fifth, the visualization of phonemic

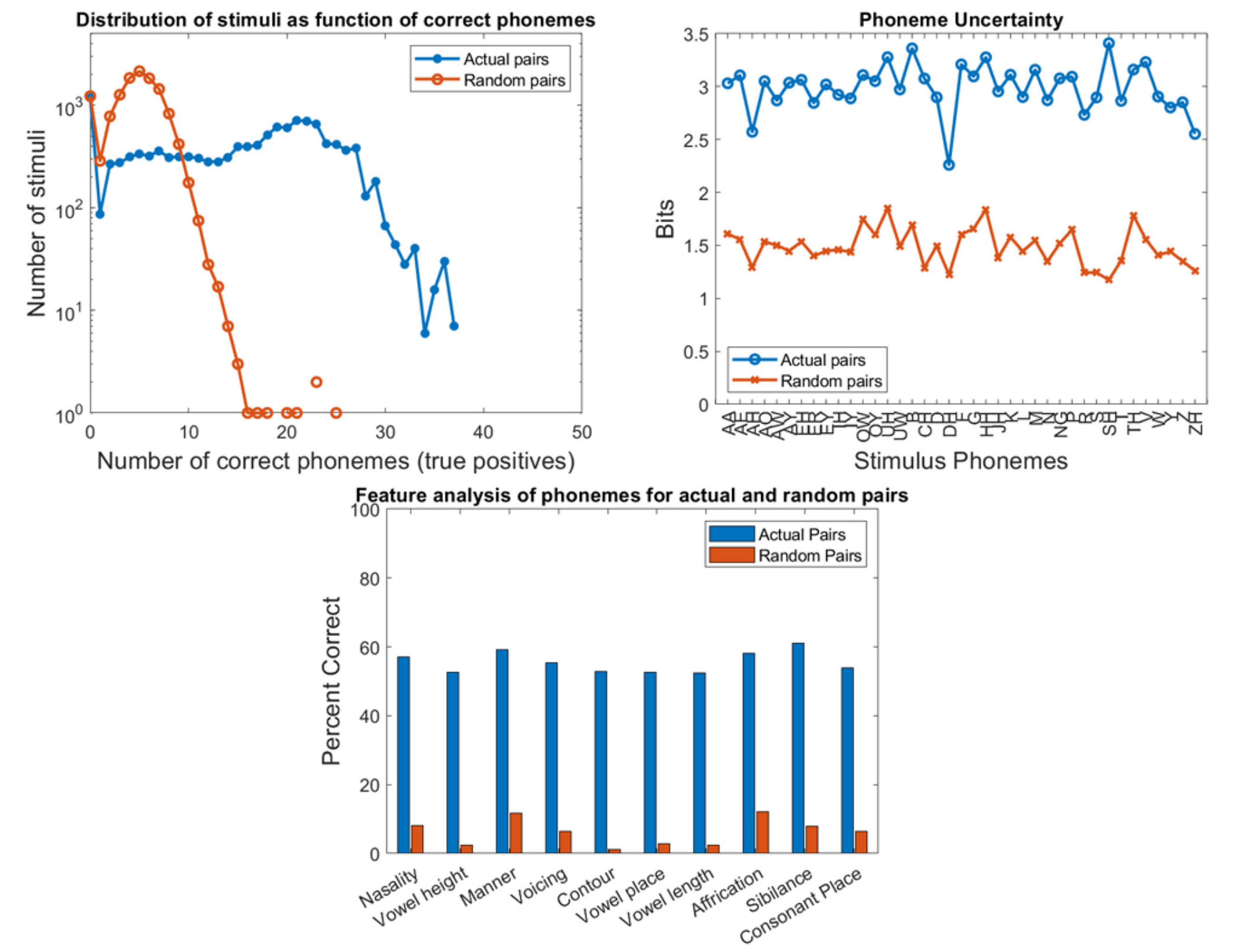


FIGURE 11 | Results of program validation by comparing actual and random 12,400 stimulus-response pairs from the O'Neill et al. (50) study of 31 participants with cochlear implants listening to 16 lists of 25 sentences as spoken by four speakers at four different SNR levels. Top left compares the frequency histograms of stimuli with the number of correct phonemes in the response. Top right compares the entropy or uncertainty for the phonemes. Bottom compares the relative information transfer for the ten phonological features used in the phonemegram.

accuracy may offer speech language pathologists a perspective of how the person with hearing loss processes different phonemes, in order to develop a targeted auditory training program. Sixth, the program can be used for educational purposes. For instance, it was used in the past few years for assessing responses by biomedical engineering undergraduates at Johns Hopkins University doing the Speech Perception module of the Neuroengineering Lab in which they listened to sentences in simulations of different types of hearing loss, number of channels in a cochlear implant, and frequency offsets in a cochlear implant.

There are also several disadvantages. Many people with hearing loss use top-down processing such as using contextual information to fill in words misheard in sentences (50, 68–71), so accuracy of responses to sentences may be overestimated. In fact, this may explain the small differences between the information transfer values for the features with the sentences in the test cases. As alluded above, word lists may be more practical in the clinic (**Figures 6**, **9**). Differences in stresses and emotion may influence perception (72) and therefore, might require using lexical stress information, if available, from the pronouncing dictionary. Finally, the program may not be suitable for people with very poor speech comprehension as they are more than likely to make random or very sparse guesses that may then confound phoneme alignment. For example, the last stimulus-response pair in **Figure 5** yielded alignment that was erratic with respect to the first part of the stimulus due to the volunteer having greater difficulty hearing with just the hearing aid instead of bimodal hearing. Incidentally, this is a good example of a person with hearing loss finding it difficult to process the early part of a stimulus compared with the rest of the stimulus (71).

Future work includes feasibility for clinical usage, user-friendly implementation for mobile auditory training apps such as Speech Banana (10), and exploring alternative approaches such as multidimensional scaling for features (29, 30) as opposed to prescribed ones, other features (69) and other metrics (6, 16, 73).

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 first dataset consisted of responses to sentences by volunteers testing the Speech Banana app. For this dataset, protocol (HIRB00001670) was reviewed and approved by JHU Homewood Institutional Research Board. The other dataset consisted of stimulus-response pairs of 31 participants. For this dataset, protocol (8804M00507) was approved by the Institutional Review Board of the University of Minnesota, and in both cases, all participants provided written informed consent prior to participating.

AUTHOR CONTRIBUTIONS

Concept was developed by JR and DT. Algorithmic development was by LW, S-HB, and ES. Testing and manuscript was written by JR, LW, S-HB, and EO'N. All authors contributed to the article and approved the submitted version.

FUNDING

EO'N was supported by NIH grant R01 DC012262 awarded to Professor Andrew Oxenham of the University of Minnesota.

ACKNOWLEDGMENTS

We thank Professor Hynek Hermansky of the Center for Language and Speech Processing at Johns Hopkins University for suggesting CMUdict and scLite, students and colleagues with hearing loss who tested the program, Kaitlin Stouffer for her critical comments, Nole Lin, Hong Seo Lim, and Zachary Heiman for early work on the program. Finally, we thank the reviewers for their comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fneur.2021.724800/full#supplementary-material

REFERENCES

1. Ladefoged P, Johnstone K. *A Course in phonetics (Seventh edition*. ed.). Stamford, CT: Cengage Learning. (2015).
2. Haskins HL. *A Phonetically Balanced Test of Speech Discrimination for Children*. (Master's thesis). Northwestern University. (1949).
3. Boothroyd A. Statistical theory of the speech discrimination score. *J Acoust Soc Am.* (1968) 43:362–7. doi: 10.1121/1.1910787
4. Tillman TW, Carhart R. An expanded test for speech discrimination utilizing CNC monosyllabic words. Northwestern University Auditory Test No 6 SAM-TR-66-55. *Tech Rep SAM-TR*. (1966) 1–12. doi: 10.21236/AD0639638
5. Bench J, Kowal A, Bamford J. The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *Br J Audiol.* (1979) 13:108–12. doi: 10.3109/03005367909078884
6. Boothroyd A, Nittrouer S. Mathematical treatment of context effects in phoneme and word recognition. *J Acoust Soc Am.* (1988) 84:101–14. doi: 10.1121/1.396976
7. Nilsson M, Soli SD, Sullivan JA. Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J Acoust Soc Am.* (1994) 95:1085–99. doi: 10.1121/1.408469
8. Spahr AJ, Dorman MF, Litvak LM, Van Wie S, Gifford RH, Loizou PC, et al. Development and validation of the AzBio sentence lists. *Ear Hear.* (2012) 33:112–7. doi: 10.1097/AUD.0b013e31822c2549
9. Auditory Potential LLC. *Minimum Speech Test Battery (MSTB) For Adult Cochlear Implant Users*. (2011). Available online at: http://www.auditorypotential.com/MSTBfiles/MSTBManual2011-06-20%20.pdf
10. Ratnanather JT, Bhattacharya R, Heston MB, Song J, Fernandez LR, Lim HS, et al. An mHealth App (Speech Banana) for Auditory Training: App Design and Development Study. *JMIR Mhealth Uhealth.* (2021) 9:e20890. doi: 10.2196/20890
11. Jurafsky D, Martin JH. *Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition (2nd ed.)*. Upper Saddle River, N.J.: Pearson Prentice Hall. (2009).
12. Povey D, Ghoshal A, Boulianne G, Burget L, Glembek O, Goel N, et al. The Kaldi speech recognition toolkit. *Paper Presented at the IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. New York, NY: IEEE Signal Processing Society (2011).
13. Fiscus J. (2018). SCTK, the NIST Scoring Toolkit (Version: 2.4.11). Available online at: https://github.com/usnistgov/SCTK
14. Speech Recognition Scoring Toolkit (SCTK). (2017). *NIST Spoken Language Technology Evaluation and Utility*. Available online at: http://www.nist.gov/speech/tools
15. Bernstein LE. Sequence comparison techniques can be used to study speech perception. *Paper Presented at the Symposium on Speech Communication Metrics and Human Performance*. Wright-Patterson Air Force Base: USAF (1995).
16. Bernstein LE. Response Errors in Females' and Males' Sentence Lipreading Necessitate Structurally Different Models for Predicting Lipreading Accuracy. *Lang Learn.* (2018) 68:127–58. doi: 10.1111/lang.12281
17. Bernstein LE, Demorest ME, Eberhardt SP. A computational approach to analyzing sentential speech perception: phoneme-to-phoneme stimulus-response alignment. *J Acoust Soc Am.* (1994) 95:3617–22. doi: 10.1121/1.409930
18. Bernstein LE, Tucker PE, Demorest ME. Speech perception without hearing. *Percept Psychophys.* (2000) 62:233–52. doi: 10.3758/BF035205546
19. Ghio A. *Achile: un dispositif de décodage acoustico-phonétique et d'identification lexicale indépendant du locuteur à partir de modules mixtes*. New York, NY: Université d'Aix Marseille. (1997).
20. Ghio A, Lalain M, Giusti L, Fredouille C, Woisard V. How to compare automatically two phonological strings: application to intelligibility measurement in the case of atypical speech. *Paper Presented at the LREC Language Ressource and Evaluation Conference*. Marseille: LREC (2020).
21. Ghio A, Rossi M. Reconnaissance analytique par règles dans SYMULDEPHO, un SYstème MULti-locuteurs de DEcodage acoustico-PHOnétique. *Travaux*

Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence (TIPA). (1995) 16:77–92.
22. Bernstein LE, Eberhardt SP, Auer ETJr. Errors on a speech-in-babble sentence recognition test reveal individual differences in acoustic phonetic perception and babble misallocations. *Ear Hear.* (2021) 42:673–90. doi: 10.1097/AUD.0000000000001020
23. Blamey PJ. *An alternative to the audiogram for hearing aid fitting*. Paper presented at the IHCON, Lake Tahoe, CA. (2012).
24. Blamey PJ. The Expected Benefit of Hearing Aids in Quiet as a Function of Hearing Thresholds. In E. Saunders (Ed.), *Tele-Audiology and the Optimization of Hearing Healthcare Delivery*. Hershey, PA, USA: IGI Global. (2019) p. 63–85. doi: 10.4018/978-1-5225-8191-8.ch004
25. Blamey PJ, Blamey JK, Saunders E. Effectiveness of a teleaudiology approach to hearing aid fitting. *J Telemed Telecare.* (2015) 21:474–8. doi: 10.1177/1357633X15611568
26. Blamey PJ, Blamey JK, Taft D, Saunders E. Predicting speech information from the audiogram and vice versa. *Paper presented at the World Congress of Audiology, Brisbane.* (2014).
27. Blamey PJ, Blamey JK, Taft D, Saunders E. *Using acoustic phonetics to reduce the financial and social burdens of hearing loss for individuals*. Paper presented at the IHCON. Lake Tahoe, CA. (2014).
28. Blamey PJ, Saunders E. Predicting Speech Perception from the Audiogram and Vice Versa. *Canadian Audiologist, 2(1).* (2015). Available online at: https://www.canadianaudiologist.ca/issue/volume-2-issue-1-2015/predicting-speech-perception-from-the-audiogram-and-vice-versa/
29. Danhauer JL, Singh S. A multidimensional scaling analysis of phonemic responses from hard of hearing and deaf subjects of three languages. *Lang Speech.* (1975) 18:42–64. doi: 10.1177/002383097501800105
30. Danhauer JL, Singh S. *Multidimensional speech perception by the hearing impaired : a treatise on distinctive features*. Baltimore: University Park Press. (1975).
31. Danhauer JL, Singh S. A study of "feature-gram" profiles for three different hearing impaired language groups. *Scand Audiol.* (1975) 4:67–71. doi: 10.3109/01050397509043068
32. Mok M, Grayden D, Dowell RC, Lawrence D. Speech perception for adults who use hearing aids in conjunction with cochlear implants in opposite ears. *J Speech Lang Hear Res.* (2006) 49:338–51. doi: 10.1044/1092-4388(2006/027)
33. Weide RL. The CMU pronouncing dictionary. (1998). Available online at: http://www.speech.cs.cmu.edu/cgi-bin/cmudict
34. Shoup J. *Phonological aspects of speech processing*. Trends in speech recognition (WA Lea, ed.). Prentice-Hall. (1980).
35. Kruskal JB. An Overview of Sequence Comparison - Time Warps, String Edits, and Macromolecules. *SIAM Review.* (1983) 25:201–37. doi: 10.1137/1025045
36. Levenshtein VI. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady.* (1966) 10:707–10.
37. Hunt JW, MacIlroy MD. *An algorithm for differential file* comparison. Bell Laboratories. (1976).
38. Ruiz N, Federico M. Phonetically-oriented word error alignment for speech recognition error analysis in speech translation. *Paper presented at the 2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*. Cham: IEEE (2015). doi: 10.1109/ASRU.2015.7404808
39. Mok M, Galvin KL, Dowell RC, McKay CM. Speech perception benefit for children with a cochlear implant and a hearing aid in opposite ears and children with bilateral cochlear implants. *Audiol Neurootol.* (2010) 15:44–56. doi: 10.1159/000219487
40. Rodvik AK, Tvete O, Torkildsen JVK, Wie OB, Skaug I, Silvola JT. Consonant and vowel confusions in well-performing children and adolescents with cochlear implants, measured by a nonsense syllable repetition test. *Front Psychol.* (2019) 10:1813. doi: 10.3389/fpsyg.2019.01813
41. van Wieringen A, Wouters J. Natural vowel and consonant recognition by Laura cochlear implantees. *Ear Hear.* (1999) 20:89–103. doi: 10.1097/00003446-199904000-00001
42. Danhauer JL, Lucks LE. The confusion matrix: A new model. *Human Communication Canada.* (1987) 11:7–11.
43. Miller GA, Nicely PE. An analysis of perceptual confusions among some english consonants. *J Acoust Soc.* (1955) 27:338–52. doi: 10.1121/1.1907526
44. Picheny MA, Durlach NI, Braida LD. Speaking clearly for the hard of hearing. II: Acoustic characteristics of clear and conversational speech. *J Speech Hear Res.* (1986) 29:434–46. doi: 10.1044/jshr.2904.434
45. Whitehurst MW, Monsees EK. *Auditory Training for the Deaf*. Washington, D.C.: The Volta Bureau. (1952).
46. Anderson K, Arnoldi K. *Building Skills for Success in the Fast-Paced Classroom*. Hillsboro, OR: Butte Publications. (2011).
47. Erber NP. *Communication Therapy*. Australia: Clavis Publishing. (1988).
48. Calandruccio L, Smiljanic R. New sentence recognition materials developed using a basic non-native English lexicon. *J Speech Lang Hear Res.* (2012) 55:1342–55. doi: 10.1044/1092-4388(2012/11-0260)
49. O'Neill ER, Parke MN, Kreft HA, Oxenham AJ. Development and validation of sentences without semantic context to complement the basic english lexicon sentences. *J Speech Lang Hear Res.* (2020) 63:3847–54. doi: 10.1044/2020_JSLHR-20-00174
50. O'Neill ER, Parke MN, Kreft HA, Oxenham AJ. Role of semantic context and talker variability in speech perception of cochlear-implant users and normal-hearing listeners. *J Acoust Soc Am.* (2021) 149:1224. doi: 10.1121/10.0003532
51. Niquette P, Arcaroli J, Revit L, Parkinson A, Staller S, Skinner M, et al. *Development of the BKB-SIN Test.* Scottsdale, AZ: Paper presented at the Annual meeting of the American Auditory Society. (2003).
52. Fisher WM, Fiscus JG. Better alignment procedures for speech recognition evaluation. *Paper presented at the 1993 IEEE International Conference on Acoustics, Speech, Signal Processing.* New York, NY: IEEE (1993). doi: 10.1109/ICASSP.1993.319229
53. Picone J, Goudie-Marshall K, Doddington G, Fisher W. Automatic text alignment for speech system evaluation. *IEEE Trans Acoust.* (1986) 34:780–4. doi: 10.1109/TASSP.1986.1164912
54. Seitz PF, Bernstein LE, Auer Jr ET, MacEachern M. *PhLex (Phonologically Transformable Lexicon): A 35,000-word computer readable pronouncing American English lexicon on structural principles, with accompanying phonological transformations, word frequencies.* In. Los Angeles: House Ear Institute. (1998).
55. New B, Pallier C, Ferrand L, Matos R. Une base de données lexicales du français contemporain sur internet: LEXIQUE^{TM}//A lexical database for contemporary french: LEXIQUE^{TM}. *L'année psychologique.* (2001) 101:447–62. doi: 10.3406/psy.2001.1341
56. Kondrak G. Alignment of phonetic sequences. Department of Computer Science. *University of Toronto, Tech. Rep. CSR*G-402. (1999).
57. Kondrak G. Phonetic alignment and similarity. *Comput Hum.* (2003) 37:273–91. doi: 10.1023/A:1025071200644
58. Nerbonne J, Heeringa W. Measuring dialect differences. *Language and Space: Theories and Methods.* Berlin: Mouton De Gruyter. (2010) p. 550–566.
59. Bailey DJ, Speights Atkins M, Mishra I, Li S, Luan Y, Seals C. An automated tool for comparing phonetic transcriptions. *Clin Linguist Phon.* (2021) 1–20. doi: 10.1080/02699206.2021.1896783
60. Bosker HR. Using fuzzy string matching for automated assessment of listener transcripts in speech intelligibility studies. *Behav Res Methods.* (2021) 53:1945–53. doi: 10.3758/s13428-021-01542-4
61. Sagi E, Svirsky MA. Information transfer analysis: a first look at estimation bias. *J Acoust Soc Am.* (2008) 123:2848–57. doi: 10.1121/1.2897914
62. Azadpour M, McKay CM, Smith RL. Estimating confidence intervals for information transfer analysis of confusion matrices. *J Acoust Soc Am.* (2014) 13:EL140–146. doi: 10.1121/1.4865840
63. Vogelstein JT, Perlman E, Falk B, Baden A, Gray Roncal W, Chandrashekhar V, et al. A community-developed open-source computational ecosystem for big neuro data. *Nat Methods.* (2018) 15:846–7. doi: 10.1038/s41592-018-0181-1
64. White T, Blok E, Calhoun VD. Data sharing and privacy issues in neuroimaging research: Opportunities, obstacles, challenges, and monsters under the bed. *Hum Brain Mapp*. 43:278–291 (2020). doi: 10.1002/hbm.25120
65. King SE, Firszt JB, Reeder RM, Holden LK, Strube M. Evaluation of TIMIT sentence list equivalency with adult cochlear implant recipients. *J Am Acad Audiol.* (2012) 23:313–31. doi: 10.3766/jaaa.23.5.3
66. Boyle PJ, Nunn TB, O'Connor AF, Moore BC. STARR: a speech test for evaluation of the effectiveness of auditory

prostheses under realistic conditions. *Ear Hear.* (2013) 34:203–12. doi: 10.1097/AUD.0b013e31826a8e82
67. Gilbert JL, Tamati TN, Pisoni DB. Development, reliability, and validity of PRESTO: a new high-variability sentence recognition test. *J Am Acad Audiol.* (2013) 24:26–36. doi: 10.3766/jaaa.24.1.4
68. Dingemanse JG, Goedegebure A. The important role of contextual information in speech perception in cochlear implant users and its consequences in speech tests. *Trends Hear.* (2019) 23:2331216519838672. doi: 10.1177/2331216519838672
69. McGettigan C, Rosen S, Scott SK. Lexico-semantic and acoustic-phonetic processes in the perception of noise-vocoded speech: implications for cochlear implantation. *Front Syst Neurosci.* (2014) 8:18. doi: 10.3389/fnsys.2014.00018
70. O'Neill ER, Kreft HA, Oxenham AJ. Cognitive factors contribute to speech perception in cochlear-implant users and age-matched normal-hearing listeners under vocoded conditions. *J Acoust Soc Am.* (2019) 146:195. doi: 10.1121/1.5116009
71. Winn MB, Teece KH. Listening Effort Is Not the Same as Speech Intelligibility Score. *Trends in Hearing.* (2021) 25:23312165211027688. doi: 10.1177/23312165211027688
72. Jiam NT, Caldwell M, Deroche ML, Chatterjee M, Limb CJ. Voice emotion perception and production in cochlear implant users. *Hear Res.* (2017) 352:30–9. doi: 10.1016/j.heares.2017.01.006
73. Scheidiger C, Allen JB, Dau T. Assessing the efficacy of hearing-aid amplification using a phoneme test. *J Acoust Soc Am.* (2017) 141:1739. doi: 10.1121/1.4976066

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.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 28 January 2022
doi: 10.3389/fnins.2021.725127

A Compact Two-Loudspeaker Virtual Sound Reproduction System for Clinical Testing of Spatial Hearing With Hearing-Assistive Devices

Eric C. Hamdan[1,2]* *and Mark D. Fletcher*[1,2]

[1] University of Southampton Auditory Implant Service, University of Southampton, Southampton, United Kingdom, [2] Institute of Sound and Vibration Research, University of Southampton, Southampton, United Kingdom

Exciting developments in hearing aid and cochlear implant technology for linking signal processing across the ears have improved spatial hearing outcomes. This has resulted in an increased emphasis on clinical assessment of the spatial hearing abilities of hearing-assistive device users. Effective assessment of spatial hearing currently requires a large and costly loudspeaker array system, housed in a heavily acoustically treated testing room. This imposes economic and logistical constraints that limit proliferation of array systems, particularly in developing nations. Despite their size and cost, the ability of current clinical array systems to reproduce realistic spatial sound fields is limited, which substantially reduces the range of realistic acoustic scenes that can be used for diagnostic testing. We propose an alternative low-cost, compact virtual acoustics system with just two loudspeakers. This system uses crosstalk cancelation to reproduce pressure signals at the device microphones that match those for real-world sound sources. Furthermore, in contrast to clinical array systems, the system can adapt to different room acoustics, removing the requirement for a heavily acoustically treated testing environment. We conducted a proof-of-concept study in two stages: in the first, we evaluated the physical performance of the system for a stationary listener in anechoic conditions and in a small audiological testing booth with moderate acoustic treatment. To do this, a head and torso simulator was fitted with specially adapted hearing-assistive devices that allowed direct access to the microphone signals. These microphone signals were compared for real and virtual sound sources at numerous source locations. In the second stage, we quantified the system's robustness to head rotations with and without the system adapting for head position. In the stationary case, the system was found to be highly effective at reproducing signals, such as speech, at all tested source locations. When head rotation was added, it performed well for rotations of up to 2°, even without adapting. However, performance improved markedly for larger rotations when the system adapted. These findings suggest that a compact, low-cost virtual acoustics system can give wider access to advanced and ecologically valid audiological testing, which could substantially improve clinical assessment of hearing-assistive device users.

Keywords: hearing impairment, speech in noise (SIN), sound localization, binaural, clinical audiology, transaural, bilateral, sound field control

OPEN ACCESS

Edited by:
Jing Chen,
Peking University, China

Reviewed by:
Justin Aronoff,
University of Illinois Urbana-Champaign, United States
Qinglin Meng,
South China University of Technology, China

****Correspondence:***
Eric C. Hamdan
eh4n15@southamptonalumni.ac.uk

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Neuroscience

Received: *14 June 2021*
Accepted: *22 December 2021*
Published: *28 January 2022*

Citation:
Hamdan EC and Fletcher MD (2022) A Compact Two-Loudspeaker Virtual Sound Reproduction System for Clinical Testing of Spatial Hearing With Hearing-Assistive Devices. Front. Neurosci. 15:725127. doi: 10.3389/fnins.2021.725127

INTRODUCTION

Bilateral cochlear implant and hearing aid technology has the potential to restore binaural hearing to hearing-impaired listeners. Binaural hearing is critical for locating and separating sounds, such as speech in noisy environments (Litovsky et al., 2004; Brown and Balkany, 2007; Lovett et al., 2010). However, the signal processing used in hearing-assistive devices (HADs) often distorts interaural level and time differences between the ears (Pastore et al., 2021), which are the primary spatial hearing cues. As a result, many HAD users have limited spatial hearing capabilities (Dorman et al., 2016; Fletcher et al., 2020a; Pastore et al., 2021). While there is a growing interest in approaches for improving spatial hearing in hearing-impaired listeners (e.g., Moore et al., 2016; Williges et al., 2018; Fletcher and Zgheib, 2020; Fletcher et al., 2020a,b; Gajecki and Nogueira, 2021), clinical testing of spatial hearing ability remains limited.

There are currently several sound field reproduction methods for assessing spatial hearing abilities of HAD users and the directional processing capabilities of HADs. The most common method is to play back sounds using a spatially distributed array of loudspeakers (Seeber et al., 2004, 2010; Lovett et al., 2010; Kitterick et al., 2011). The loudspeakers are typically arranged in a circle or semicircle around the listener, as in the Crescent of Sound system that is used clinically across the United Kingdom (Kitterick et al., 2011). Because these systems use simple direct-speaker playback or amplitude panning, the sound that reaches the ears can be colored by the acoustics of the room in which the system is housed. The room should therefore be heavily acoustically treated to ensure that system performance is equivalent across clinics. However, this is rarely achieved and systems such as the Crescent of Sound do not have an operating standard for the acoustic treatment of the room they are used in. Furthermore, because the method only allows reproduction of sound sources from a limited set of locations, these systems are unable to accurately reproduce complex auditory scenes that are typically encountered in the real world. This limits the ecological validity of the tests that can be performed. In addition, systems with many loudspeakers, such as the Crescent of Sound, are expensive and need to be housed in a large room. This severely limits proliferation, particularly in low- and middle-income countries.

An alternative to current clinical array systems are virtual acoustics (VA) systems. These seek to simulate the perception of real-world spatial sounds and include a variety of approaches (Lokki and Savioja, 2008). Previously proposed VA systems have used techniques such as higher-order ambisonics and vector base amplitude panning (VBAP) in combination with large loudspeaker arrays, e.g., more than 20 loudspeakers (Minnaar et al., 2013; Grimm et al., 2015, 2016; Cubick and Dau, 2016; Oreinos and Buchholz, 2016). Loudspeaker arrays of this size are impractically large and expensive, and seen as not suitable for clinical use. However, higher-order ambisonics constrained to the horizontal plane would only require $(2N + 1)$ loudspeakers, where N is the order (Zotter and Frank, 2019). Still, the practicality of ambisonics systems could be limited as they require more loudspeakers in exchange for higher accuracy offered by higher orders of reproduction. More recently, Meng et al. (2021) investigated a smaller two-loudspeaker VBAP system for facilitating a minimum audible angle test. However, the virtual source positioning and reproduction accuracy are limited, as VBAP restricts the position of the virtual source to within the span of the loudspeakers. Furthermore, horizontal plane VBAP is not designed to faithfully reproduce sources above and below the listener. This means that the system is substantially limited in its ability to produce realistic acoustic scenes.

Aside from these VA methods, there are binaural methods for evaluating bilateral HADs. Headphones placed over the devices are a common approach to delivering binaural audio. However, headphones can be obtrusive and binaural signals delivered through headphones are most often derived from binaural recordings (i.e., microphones placed at the opening of the ear canals) or binaural synthesis [i.e., simulated using measured head-related transfer functions (HRTFs)]. Neither of these approaches match the pressure signals that would arrive at the HAD microphones in the real world because they are not derived from HAD-related transfer functions. Another possible issue is inconsistent coupling of the headphone loudspeaker with the HAD microphones across headphone fittings, which could further compromise the integrity of the reproduction.

Pausch et al. (2018) used custom-made research hearing aids that allow the microphones to be bypassed and hearing-aid-related binaural signals to be delivered directly to the devices. The research hearing aids were used in tandem with a crosstalk cancelation (CTC) system for reproducing HRTF-based binaural signals at the ear drums. This meant that stimulation was provided for residual hearing as well as through the HAD. However, they did not demonstrate that their CTC system could reproduce accurate target physical pressure signals, making it difficult to evaluate the success of their system. While they reported the channel separation (see section "Metrics") achieved by their system, this metric alone is insufficient for determining the physical accuracy of the reproduction and ruling out audible artifacts that could diminish perceptual outcomes.

Another approach that used direct input to the HADs was proposed by Chan et al. (2008). In this approach, the binaural signals were calibrated and synthesized using transfer functions measured between a loudspeaker and the HADs when mounted on a dummy head. Measurements were made using either the onboard HAD microphones or separate microphones placed near to the HAD microphones. It is possible that, in the future, device manufacturers or a dedicated service could provide clinicians with transfer functions for each of their devices. However, because microphones are bypassed with the direct input approach, this would mean that defective microphones or changes in microphone response over its lifetime would not be accounted for. Alternatively, sound field measurements could be repeated for each device in clinic. However, the measurements would then be susceptible to local room acoustics, meaning they could act as a significant source of variance between clinical measurements. The protracted calibration process would require additional clinician training and may be unsuitable for clinical appointments, where time is typically limited. Furthermore, this would require the use of potentially expensive additional equipment (e.g., a head and torso simulator).

In the current study, we investigated a VA system that uses two loudspeakers. The system utilizes a type of sound field control based on inverse filters, more commonly known as CTC (e.g., Xie, 2013), designed using HAD transfer functions measured *in situ*. We propose that, when given access to the HAD microphone signals in the clinic environment, signal processing steps can be taken that allow rapid transfer function measurement and inverse filter design. These inverse filters enable pressure signal reproduction at the device microphones. Such a system could precisely control the sound field at the device and allow the reproduction of complex real-world auditory scenes, while remaining unobtrusive. Inverse filters could also allow a room agnostic approach, where a measurement standard can be retained across clinical settings. Furthermore, because the system only requires two loudspeakers, it would be inexpensive and have a small physical footprint. The compactness, low cost, room adaptability, and capacity to accommodate tests with high ecological validity would give this system major advantages over current clinical systems.

The first aim of the study was to assess the ability of the system to reproduce physically accurate sound fields at the HAD microphones for a stationary listener. We evaluated the sound fields by measuring and analyzing the sound pressure at the HAD microphones. These measurements were taken in both an ideal (anechoic) and a representative clinical environment. The reproduced sound pressure was analyzed using several metrics that directly measure the system's performance, both in the frequency and time domains. From these metrics, we establish a baseline standard (that does not currently exist) for what sound field control can be physically achieved at the HAD microphones. We set a target channel separation of at least 20 dB between the reproduced left and right binaural signals measured at the HAD microphones. This amount of channel separation has been shown to be the minimum amount needed to give equivalent perception of the original and reproduced binaural signal in normal-hearing listeners (e.g., Parodi and Rubak, 2011). We demonstrated that our system is capable of achieving this target channel separation in both listening environments and that the time domain error was low. In this first part of the experiment, we demonstrated that, for a stationary listener, the system can reproduce target HAD microphone signals to a high degree of accuracy.

The second aim of the study was to assess the impact of minor involuntary head movements ("postural sway"), that are likely to occur in clinic even if participants are instructed to sit still (e.g., Hirahara et al., 2010; Denk et al., 2017). Using the aforementioned metrics, we compared the system performance with and without compensating for head movements to establish whether head movement compensation is required. We show that the system is robust to the small head rotations (of 2° or less) that would be expected in clinical assessments, and that good performance can be achieved for head rotations up to 10° if the system adapts for changes in head position.

MATERIALS AND METHODS

Test Apparatus

The loudspeakers used were Genelec 8020Cs (Genelec Oy, Iisalmi, Finland). A KEMAR head and torso simulator (G.R.A.S. Sound and Vibration, Holte, Denmark) was fitted with custom HADs behind each ear to simulate a stationary seated listener. The KEMAR was placed on a stand to allow easy placement and rotation. The HADs we used were modified Oticon Medical (Smørum, Denmark) Saphyr CI processors (TX9 model; shown in **Figure 1**), with an onboard sampling rate of 16 kHz. The modification allowed direct access of the microphone signals via analog cable outputs so that the sound field could be controlled at the microphones directly. All onboard signal processing was bypassed. For recording the HAD microphone signals, the analog outputs of the microphones were sent to an RME UC (RME Audio, Haimhausen, Germany) for digital conversion. The digital signals were sent from the RME UC via USB to a computer running the measurement and reproduction software. A custom Matlab (version R2020b, The MathWorks, Inc., Natick, MA, United States) script was written to record the microphone signals to the computer during measurements. For loudspeaker reproduction, the same RME UC was used to output audio signals to the loudspeaker array. All audio playback was done with either Matlab or Max/MSP (Version 8.1.2, Cycling '74, Walnut, CA, United States). All signals were recorded and reproduced at a sampling rate of 48 kHz with a bit depth of 24 bits.

In order to measure head rotation, a HTC VIVE tracker (Version 1, HTC Corporation, Xindian, New Taipei, Taiwan) was fitted to the top of the KEMAR head with a plastic cap in between measurements. The head tracker was removed before taking a new measurement.

Testing Environments

Two testing environments were used: the Institute of Sound and Vibration Research (University of Southampton, United Kingdom) anechoic chamber (shown in **Figure 2A**) and a clinical audiological testing booth (shown in **Figure 2B**) located in the Hearing and Balance Centre (University of Southampton, United Kingdom). The anechoic chamber was

FIGURE 1 | Modified behind-the-ear HADs.

chosen to represent an ideal testing environment with heavy acoustic treatment. The clinical booth was chosen to represent a typical clinical testing environment. The test booth was 2.5 m by 2.1 m, with a ceiling height of 2.05 m and had a background noise level conforming to the recommendation of British Society of Audiology (2017).

Sound Field Reproduction

The reproduction system consisted of a six-channel loudspeaker array arranged in a semicircle, with the KEMAR positioned at the center of the semicircle facing the center of the array (see **Figure 2**). Each loudspeaker was placed 1.5 m in the anechoic chamber and 1 m in the booth (due to space constraints) from the center of the KEMAR head and set at ear height. All loudspeakers were used to produce reference signals, to compare against virtual sources. Loudspeakers L1 and R1 (labeled in light blue in **Figure 2**) were chosen for the VA system to retain a compact array.

Pressure Matching Method

The method of sound field control underlying the VA system was the Tikhonov regularized pressure matching method (e.g., Kirkeby et al., 1998; Olivieri et al., 2015). When applied to just two loudspeakers, the pressure matching method has been more commonly known as CTC (see Xie, 2013 for an overview of CTC technology). The overall sound field control problem is described in the frequency domain as:

$$\boldsymbol{G}(\omega)\boldsymbol{q}(\omega) \stackrel{!}{=} \boldsymbol{d}(\omega),$$

where $\stackrel{!}{=}$ means 'ideally equal to'; $\boldsymbol{d}(\omega) = \left[\, d_1(\omega) \;\; d_2(\omega) \,\right]^T \in \mathbb{C}^2$ is the vector of target pressure signals that we wish to reproduce at two HAD microphones; the so-called "plant matrix":

$$\boldsymbol{G}(\omega) = \begin{bmatrix} \boldsymbol{g}_1^T(\omega) \\ \boldsymbol{g}_2^T(\omega) \end{bmatrix} \in \mathbb{C}^{2\times 2},$$

is composed of electroacoustic transfer functions between the two microphones and two loudspeakers; $\boldsymbol{g}_m(\omega) = \left[\, g_{m1}(\omega) \;\; g_{m2}(\omega) \,\right]^T \in \mathbb{C}^2$ is the mth vector of electroacoustic transfer functions between the mth microphone and each loudspeaker, where $m = 1, 2$; and $\boldsymbol{q}(\omega) = \left[\, q_1(\omega) \;\; q_2(\omega) \,\right]^T \in \mathbb{C}^2$ is the vector of *unknown* loudspeaker signals that we wish to calculate. Note that ω is radian frequency. When we apply a calculated set of loudspeaker signals, the physical result is:

$$\boldsymbol{G}(\omega)\boldsymbol{q}_0(\omega) = \boldsymbol{p}(\omega),$$

where $\boldsymbol{p}(\omega) = \left[\, p_1(\omega) \; p_2(\omega) \,\right]^T \in \mathbb{C}^2$ is the vector of reproduced pressure signals at the microphones, having applied a specific set of loudspeaker signals $\boldsymbol{q}_0(\omega)$. The desired result of the system, can be expressed as:

$$\boldsymbol{p}(\omega) = \boldsymbol{d}(\omega)e^{-j\omega\tau},$$

where τ is a delay in seconds. This means that the reproduced pressure signals are an exact, delayed copy of the target pressure signals. To obtain an inverse filter solution, the Tikhonov regularized inverse solution was calculated using the well-known equation:

$$\boldsymbol{q}_0(\omega) = \boldsymbol{G}^H(\omega)\left(\boldsymbol{G}(\omega)\boldsymbol{G}^H(\omega) + \beta\boldsymbol{I}\right)^{-1}\boldsymbol{d}(\omega),$$

where β is the real-valued regularization parameter (here frequency-independent); $\boldsymbol{I}$ is a 2×2 identity matrix; and $(\cdot)^{-1}$ denotes the matrix inverse. Note that in our experiments, the number of microphones was always equal to or less than the number of loudspeakers. In practice, due to the need to regularize (non-zero β) and truncate the inverse filters, an exact solution is generally impossible. However, using this approach allowed filter stability to be obtained.

Inverse Filter Design

The inverse filter design was divided into two primary stages: first, the *in situ* hearing-assistive device transfer function (HADRTF) measurements and, second, the inverse filter calculation. For the HADRTF measurements, we used an exponential sine sweep as proposed by Farina (2000). An exponential sine sweep from 1 Hz to 24 kHz was played from each loudspeaker and simultaneously recorded at the device microphones. The hearing-assistive device impulse responses (HADIRs) were extracted from each measurement. The frequency domain equivalents of these HADIRs were the HADRTFs used for the inverse filter design.

After the HADIRs were obtained, the process outlined in **Figure 3** was used to generate the inverse filters. The first step was to normalize the HADIRs to take full advantage of the available digital dynamic range. The second step was to temporally window the HADIRs to remove later reflections and noise that can cause instability in the inverse filters. We used a modified Tukey window consisting of concatenated raised cosine sections with a flat rectangular window section. This modification allowed for tuneable fade in and out positions and window length. It was found that a balance of temporal windowing and regularization was needed to produce stable filters and accurate performance, in both spaces. The windowed HADIRs were then converted to the frequency domain to obtain the final HADRTFs. For each frequency bin, the HADRTFs were used to populate the plant matrix. Next, the inverse filters were constructed from the Tikhonov regularized pseudoinverse of the plant matrix using a regularization value of $\beta = 0.0005$ for the anechoic chamber and $\beta = 0.001$ for the clinical booth. The frequency domain inverse filters were converted to time domain filters. The resulting time domain inverse filters were low-pass filtered with a linear phase FIR filter (99 taps and cutoff frequency at 8000 Hz) to reduce instabilities due to a roll-off in the HADIR magnitude responses around the Nyquist frequency. Following this, the time domain inverse filters were normalized to a peak amplitude of 1 to reduce the need for further amplification at the loudspeaker stage. Lastly, the final time domain inverse filters were shifted to ensure causality and sufficient decay before and after the main peak of each filter. This helped to ensure that time domain artifacts, such as smearing and echoes, were avoided. Full details of the computation applied is provided in the **Supplementary Material**.

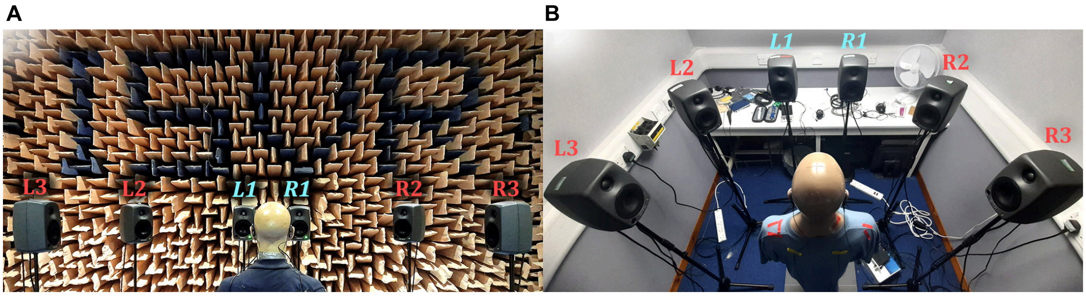


FIGURE 2 | Photographs of the reproduction setups: **(A)** Anechoic chamber; **(B)** Clinical booth. L1 and R1 (labeled in light blue) were used for the VA system.

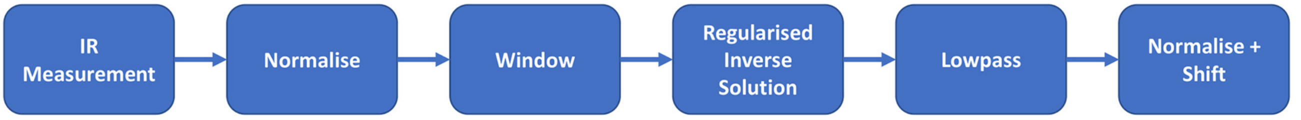


FIGURE 3 | Block diagram showing the inverse filter design process.

Sound Field Evaluation

The VA system's physical performance was evaluated based on its ability to reproduce a target binaural signal in each testing environment. This included assessment of the VA system's ability to reproduce sources from spatial positions well away from the immediate vicinity of the VA loudspeakers. Here, we highlight the performance when the target signal originated from loudspeaker L3 (see **Figure 2**), which was positioned 90° to the left of the KEMAR. We assessed the VA system's room adaptation potential by attempting to reproduce an anechoic signal in the clinical booth. Additionally, we evaluated the system's robustness to minor head rotations with and without adaptation of the inverse filters for the head movement. We limited the study to head rotations of no more than 10° to the left and right, at single degree increments (positive angles were to the right and negative angles were to the left of center). HADIR measurements were taken for each head rotation angle, in each room, and corresponding inverse filters were calculated from these measurements using the same parameters as the static filters (detailed in section "Inverse Filter Design"). We measured the reproduced performance when the head was rotated using inverse filters designed only for the center position (0°), thus evaluating the system's robustness without compensation for head rotation. Additionally, for each head rotation position, performance measurements were made with the inverse filters constructed from the HADIRs at that given head rotation angle (i.e., with the head rotation accounted for).

To create target microphone recordings in the anechoic chamber, a recorded female speech sample (Fletcher et al., 2020a; available at DOI: 10.5258/SOTON/D1206) was played from each loudspeaker successively and recorded by the front HAD microphones, so that six stereo HAD recordings of the speech played from the six loudspeaker positions were obtained. These recordings were then filtered with the same low-pass filter applied to the inverse filters to ensure that any energy beyond the HAD Nyquist frequency (8 kHz) was negligible. The target binaural signals were convolved with the inverse filters in real-time and the resulting loudspeaker signals were played back over the VA loudspeakers. The resulting HAD microphone signals were recorded and compared to the target HAD recordings using the time domain error metrics detailed in the following section. Before comparison, the reproduced recordings were time aligned with the target (removal of the constant modeling and lowpass delays) and all recordings were cropped to 1.6 s (76,800 samples at the recording sampling rate of 48 kHz) to remove unnecessary silence.

Metrics

For the sound field control to work successfully, channel separation must be maintained between the microphones. Additionally, a signal desired at either of the microphones should be as uncolored as possible by the reproduction method itself. Therefore, channel separation alone is insufficient for accurate reproduction. To estimate these qualities, we calculated the left and right microphone signals, $p_1(\omega)$, $p_2(\omega)$, respectively, after applying the inverse filters. The target signals were unit impulses to the left and right input channels (one at a time) while sending zeros to the other channel, i.e.,

$$\boldsymbol{d}(\omega) = \begin{bmatrix}1 & 0\end{bmatrix}^T,$$

$$\boldsymbol{d}(\omega) = \begin{bmatrix}0 & 1\end{bmatrix}^T.$$

For these quantities, the measured HADRTFs between loudspeakers L1 and R1 and the HAD microphones were convolved with the loudspeaker signals (with inverse filters applied to the input) for the forward calculation. We examined the magnitude and phase responses of $p_1(\omega)$, $p_2(\omega)$ to give an indication of any unwanted artifacts imposed by the inverse filters. Additionally, the resulting frequency-dependent channel separation in each of these cases was measured by:

$$CS_1(\omega) = \left|\frac{p_1(\omega)}{p_2(\omega)}\right|,$$
$$CS_2(\omega) = \frac{1}{CS_1},$$

respectively.

The ability to reproduce spatial sounds was assessed by attempting to reproduce a target HADRTF due to each loudspeaker. Additionally, the time domain waveforms of the target microphone signals and reproduced microphone signals were compared using the absolute error:

$$AE_1[n] = \left|d_1[n] - p_1[n]\right|,$$
$$AE_2[n] = \left|d_2[n] - p_2[n]\right|,$$

where $p_1[n]$, $p_2[n]$ and $d_1[n]$, $d_2[n]$ are the discrete time domain reproduced and target signals at microphones 1 and 2 (front left and right), respectively, and n is the time sample index. For visual presentation, the absolute errors were presented with smoothing applied from a Savitzky–Golay filter with window length 1001 and polynomial order 1. Additionally, the mean absolute error (MAE) of each channel was also evaluated and calculated as:

$$MAE_1 = \frac{\sum_{n=0}^{N-1} AE_1[n]}{N}$$
$$MAE_2 = \frac{\sum_{n=0}^{N-1} AE_2[n]}{N},$$

where N is the total number of samples in each time domain recording (here $N = 76800$).

Note that dB quantities in this work were calculated as $20\log_{10} x$, where x is an amplitude quantity in linear scale.

RESULTS

Stationary Measurements

For both the anechoic chamber and clinical booth, inverse filters were created for loudspeakers L1 and R1 (see **Figure 2**) according to the procedure detailed in section "Inverse Filter Design." The head and torso simulator was kept in a forward-facing position centered between the two loudspeakers for all measurements. The achieved channel separation and ability to reproduce a target HADRTF and time domain waveform was evaluated according to the procedure detailed in section "Metrics." The VA system's room adaptation potential was assessed by reproducing anechoic recordings in the booth setting.

Anechoic Chamber

Figure 4 shows the reproduced magnitude responses $|p_1(\omega)|$, $|p_2(\omega)|$ and the corresponding unwrapped phase responses $\angle p_1(\omega)$, $\angle p_1(\omega)$, as functions of frequency in Hz for frequencies 50–8000 Hz, as a result of a target impulse to the left and right binaural input channels, respectively. For each channel, the impulse signal was reproduced with an almost flat magnitude response centered around 0 dB, with fluctuations less than 0.5 dB throughout the passband. There was a roll-off around 55 Hz due to regularization, however, these low frequencies are unimportant in most practical use cases. This result confirms that the target signal magnitude responses can be well reproduced. In each case, the opposite channel, i.e., the side where zero pressure was desired, was substantially attenuated. **Figure 5** shows that the achieved channel separations $CS_1(\omega)$, $CS_2(\omega)$ were never less than 40 dB from around 100–6500 Hz (except for the slightly less amount of 39 dB around 4700 Hz), and no less than 20 dB to the limit of the effective passband (7800 Hz) where the applied lowpass filter had already taken substantial effect. The phase responses were essentially linear in the passband.

Next, the target binaural signal was set to the HADRTFs measured from each loudspeaker. **Figure 6** shows the reproduced magnitude and phase response at the left and right HAD microphones when the target was the HADRTF due to loudspeaker L3. There was excellent agreement between the target and reproduced magnitude responses in the passband, with only minor fluctuations at the lowest frequencies. The reproduced phase responses (with the constant modeling delay removed) were also in excellent agreement, although a small constant shift was seen in each channel, thus the original phase relationships between microphones were retained. Excellent agreement between reproduced and target responses was also observed for the remaining loudspeakers (including when the HADRTF was from either of the VA loudspeakers). However, it was important to verify the performance in the time domain and to assess the results in subjective listening tests. The front HAD microphone recordings of the female speech sample played from each loudspeaker were compared to the VA system's reproduction of that same recording using loudspeakers L1 and R1. **Figure 7** compares a portion of the recorded time domain waveforms when speech originated from loudspeaker L3 (the furthest from the VA speakers). With the constant delay removed, excellent alignment of the time domain waveforms is observed. Overall, as expected from the HADRTF reproduction, we found that the agreement in the time domain reproduction was excellent for all six loudspeakers. **Figure 8A** shows the corresponding AEs (unsmoothed) and **Figure 8B** shows the smoothed AEs and MAEs, all in dB. Smoothed versions of the AEs are denoted by $\widehat{AE}_1[n]$, $\widehat{AE}_2[n]$. This result showed excellent agreement in the time domain between the real and virtualized recordings at both HAD microphones, with MAEs of −82 dB and −84 dB. An MAE of no more than −77 dB was achieved for all loudspeaker positions, and the average MAEs amongst the loudspeaker positions was −81 dB at each microphone. Informal subjective headphone listening tests with five expert listeners found that the reproduced recording was indistinguishable from the target for each loudspeaker position and that the reproduction was free of any time domain artifacts such as echoes or smearing. Binaural audio files of the measured results are available in the **Supplementary Material**.

Clinical Booth

To attenuate certain adverse room reflections (present only in the clinical booth HADIRs), the temporal window parameters were

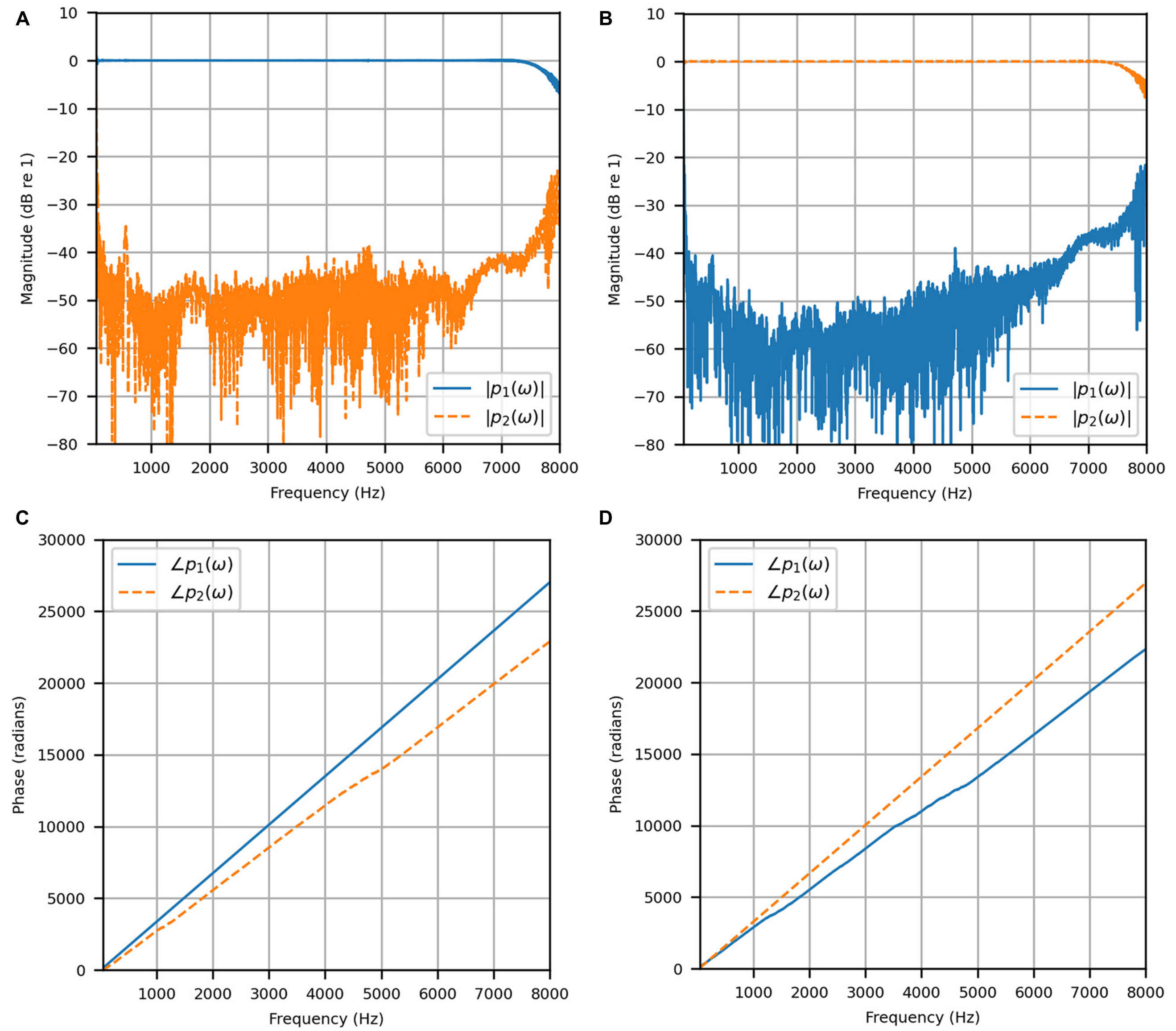


FIGURE 4 | Anechoic chamber: **(A)** Reproduced magnitude responses for a target impulse to the left microphone (blue solid); **(B)** Reproduced magnitude responses for a target impulse to the right microphone (red dashed); **(C)** Reproduced phase responses for a target impulse to the left microphone (blue solid); **(D)** Reproduced phase responses for a target impulse to the right microphone (red dashed).

adjusted for the clinical booth to attenuate later reflections (not present in the anechoic chamber). **Figure 9** shows the reproduced magnitude responses $|p_1(\omega)|$, $|p_2(\omega)|$ and unwrapped phase responses $\angle p_1(\omega)$, $\angle p_1(\omega)$ as functions of frequency in Hz for frequencies 50–8000 Hz, again as a result of a target impulse to the left and right input channels. Like the anechoic reproduction, the target impulse signal was reproduced with an almost flat response centered about 0 dB. There were slightly more fluctuations; however, these remained within 1 dB for frequencies above 100 Hz. The opposite channel had been significantly attenuated, although to a lesser extent than seen in the anechoic chamber. **Figure 10** shows channel separations as functions of frequency in Hz. Still, there was no less than 25 dB of separation at most frequencies between 100 and 7800 Hz (the effective passband). These results suggested that accurate sound field control at the HAD microphones is possible in the audiological testing booth.

To reinforce the accuracy of the booth reproduction, and to evaluate the room adaptability potential of the VA system, the target binaural signal was set as the HADRTF measured from the loudspeakers *from within the anechoic chamber*. Thus, the objective was to reproduce a response with essentially no reverberation within a room with reverberation. **Figure 11** shows the reproduced magnitude and phase responses at each microphone when the target HADRTF was due to loudspeaker L3. There was excellent agreement in the magnitude responses within the passband, again with only negligible fluctuations below

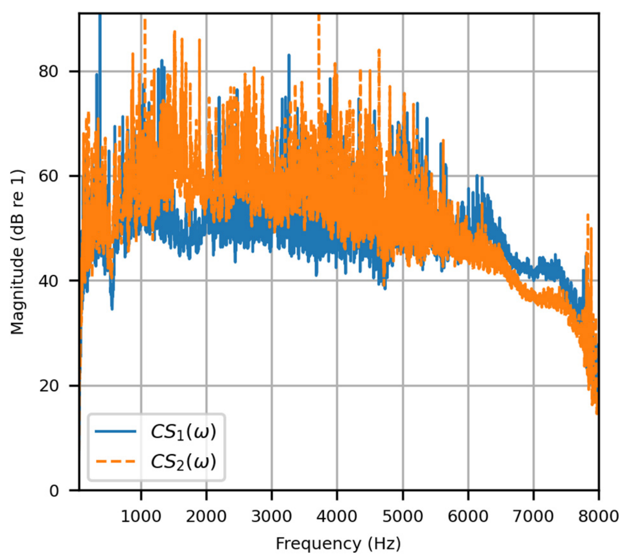


FIGURE 5 | Anechoic chamber: Channel separations for a target impulse to the left (blue solid) and right (red dashed) microphones.

100 Hz (4 dB at 60 Hz), as predicted by the channel separation analysis. As with the anechoic chamber, the phase responses were reproduced accurately, albeit with a constant shift that retained the inter-microphone phase relationships. Excellent agreement was also observed in the frequency responses for the remaining loudspeaker positions. This result showed that the VA system is capable of room adaption *and* accurate reproduction in a real-world clinical space.

As before, the time domain performance was investigated by comparing the front HAD microphone recordings of the female speech sample played from each loudspeaker, although this time from within the anechoic chamber, to a recording of the VA system's reproduction of those same target microphone signals using loudspeakers L1 and R1 within the booth. **Figure 12** compares a portion of the reproduced versus target (played from loudspeaker L3) time domain waveforms. Again, after removal of the constant delay, excellent agreement between the target and reproduced waveforms was achieved (and was observed for the other loudspeaker positions). Again, as expected from the HADRTF reproduction in the booth, we found that the agreement in the time domain reproduction was excellent for all six loudspeakers. Following the format of **Figures 8**, **13A** shows the unsmoothed AEs and **Figure 13B** shows the smoothed AEs and MAEs calculated from the time domain signals. This result reinforced the excellent agreement in the time domain between the real and virtualized recordings at both HAD microphones in the clinical booth environment, albeit to a slightly lesser extent than in the anechoic chamber, with MAEs of −77 dB and −80 dB. An MAE of no more than −75 dB was achieved for all loudspeaker positions, and the average MAEs amongst the loudspeaker positions were −77 dB at each microphone (4 dB less than in the anechoic chamber). Again, in informal subjective listening tests with expert listeners, the target and virtualized recordings could not be differentiated for each loudspeaker position (for audio demonstrations, see the **Supplementary Material**).

Head Rotation Measurements

For each head rotation angle in both uncompensated and compensated modes of operation, we evaluated the achieved channel separation as a function of frequency, and the AEs and the MAEs (see section "Metrics"). Here, we report only the channel separation and the MAEs for brevity (data for the AEs is provided in the supporting data). For the MAEs (time domain performance), the target audio was the female speech sample (as measured from each loudspeaker in the anechoic chamber). The performance of the system when the target loudspeaker was L3 is highlighted.

Anechoic Chamber

Figures 14A,B show the achieved channel separations measured in the anechoic chamber with and without compensation for head rotation, respectively, as 2D functions of head rotation angle in degrees and frequencies 50–8000 Hz. **Figure 14A** shows that substantial channel separation was achieved in the passband (greater than 60 dB for some frequencies and at least 20 dB for most frequencies) for head rotation angles of up to 2° to the left or right when the filters were not updated to account for rotation. Beyond 2°, the channel separation lessened as the head rotation angle increased, as expected. However, it still exceeded our target performance of 20 dB or more for many frequencies, despite the lack of filter adaptation. **Figure 14B** shows that the channel separation was substantially higher (nearly 60 dB or more for many frequencies) and more consistent with increase in head rotation angle when the inverse filters were updated with changes in angle.

Figures 14C,D show the uncompensated and compensated MAEs measured in the anechoic chamber, respectively, as functions of head rotation angle in degrees. **Figure 14C** shows that the MAEs were lowest (−82 dB and −84 dB) for the centered head position (0°), as expected. The MAEs increased with head rotation angle in either direction, reaching a maximum of −54 dB with the largest head rotation angle (10°). In contrast, **Figure 14D** indicates that when the filters were updated with rotation, the MAEs stayed consistently below −80 dB for most head rotation angles (except for two outliers at −3° and 2°, which were around −80 dB and −78 dB, respectively, for MAE_1). These results show that the system is robust to minor head rotation, but that performance is much better and more consistent (particularly in the time domain) when the filters are updated with head rotation angle.

Clinical Booth

The same evaluation was done in the clinical booth as in the anechoic chamber. As for the stationary measurements, the target signal was measured in the anechoic chamber. **Figures 15A,B** show the uncompensated and compensated channel separations measured in the booth, respectively, as functions of head rotation angle in degrees and for frequencies 50–8000 Hz. **Figure 15A** shows that, as in the anechoic chamber, significant channel

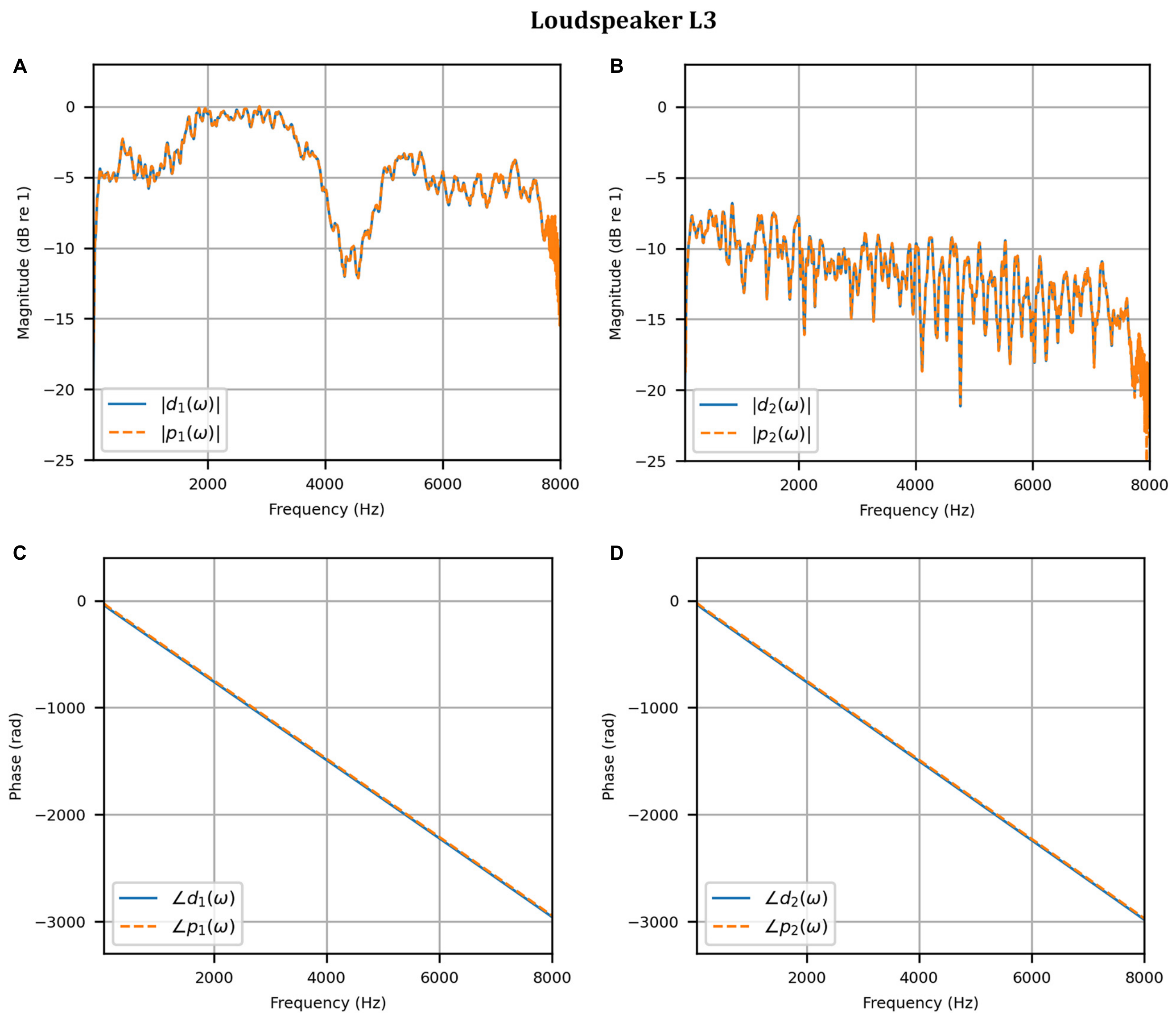


FIGURE 6 | Anechoic chamber: Target (blue solid) versus reproduced (red dashed) frequency responses for the HADRTF due to loudspeaker L3: **(A)** Magnitude responses at the left microphone; **(B)** Magnitude responses at the right microphone; **(C)** Phase responses at the left microphone; **(D)** Phase responses at the right microphone.

separation was achieved in the passband when the head was not rotated (60 dB or more for some frequencies and as little as 20 dB for most frequencies except for some in the range of 3000–3500 Hz). When the head rotation was 1° to the left, the achieved channel separation was 20 dB or more from around 160 to 2700 Hz. Above 2700 Hz, the right channel separation fluctuated around 20 dB while the left channel separation generally stayed above 20 dB (except for some frequencies in the range of 3000–3600 Hz). For 1° rotation to the right, 20 dB or more of channel separation was generally achieved between 160 and 6000 Hz for both channels. Beyond 1°, in either direction, the channel separation lessened and tended to become worse with increasing angle (being as little as 5 dB at some frequencies between 160 and 2700 Hz at 10° of rotation). For all angles, the channel separation tended to be substantially lower for frequencies above 2700 Hz. However, separation remained substantial (20 dB or more) for some frequency bands. **Figure 15B** shows that, like for the anechoic chamber, when the inverse filters were updated to compensate for rotation, the channel separation was significantly higher (30 dB or more for many frequencies) and remained more consistent as head rotation angle increased. However, channel separation was reduced in the range of 3000–4000 Hz for all angles.

Figures 15C,D show the uncompensated and compensated MAEs measured in the clinical booth, respectively, as functions of head rotation angle in degrees. **Figure 15C** shows that the MAEs were lowest (around −75 dB) for 0°, as expected. However, MAEs increased with head rotation angle in either

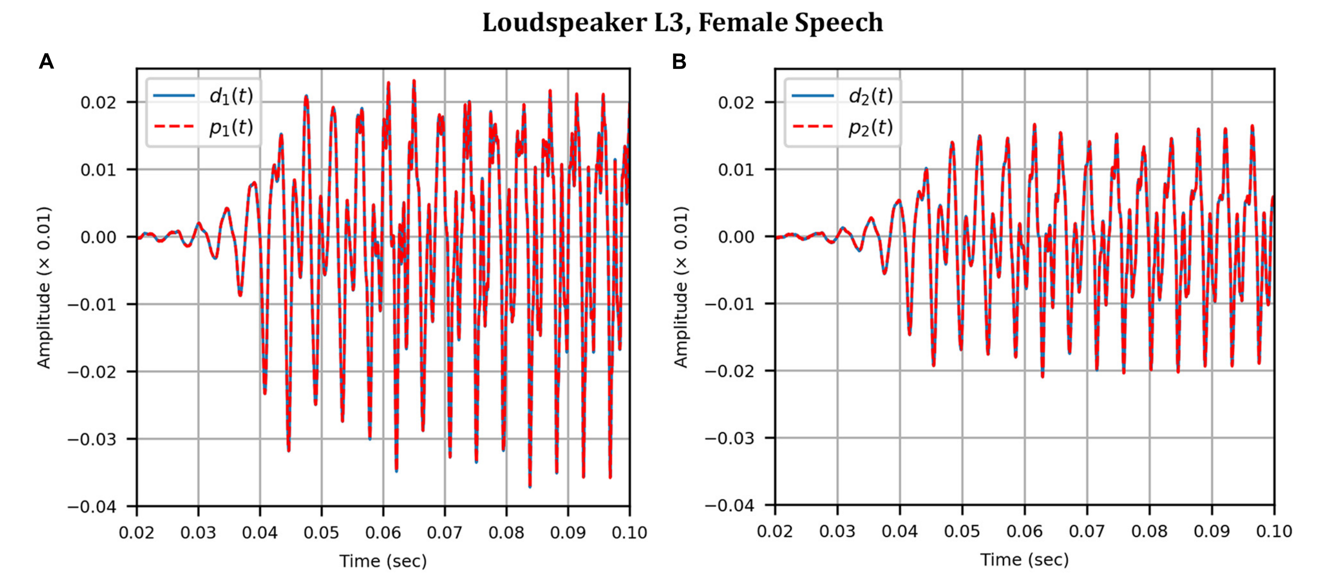


FIGURE 7 | Anechoic chamber: Target (blue solid) versus reproduced (red dashed) time domain waveforms of a female speech sample played from loudspeaker L3: **(A)** Left microphone; **(B)** Right microphone.

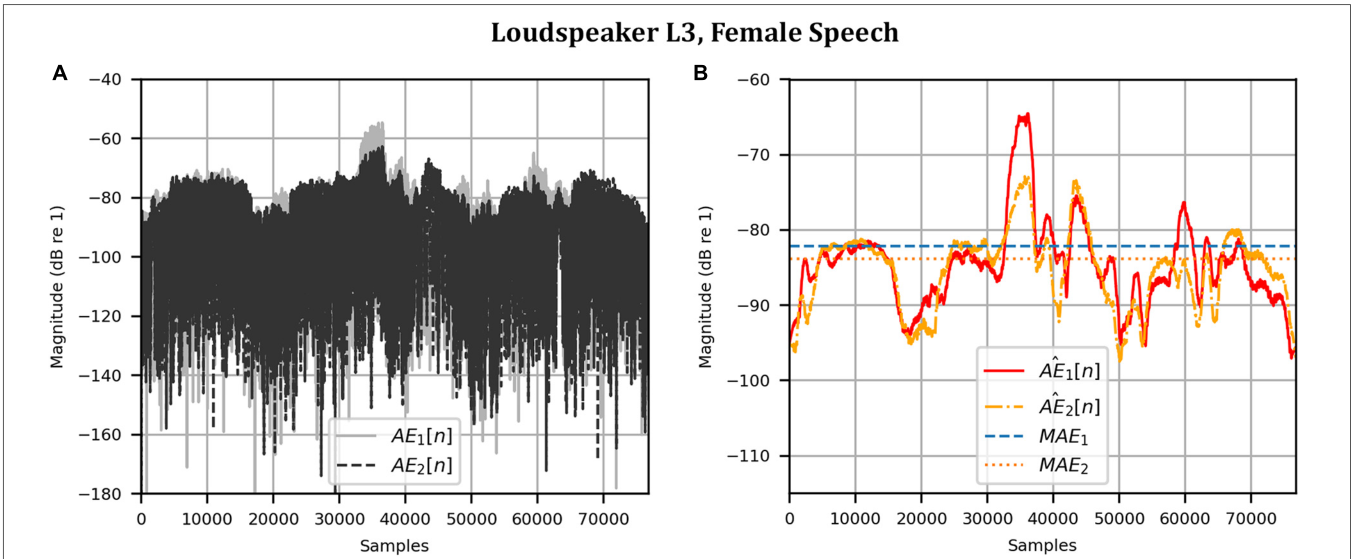


FIGURE 8 | Anechoic chamber: Error metrics quantifying the difference between the target and reproduced time domain waveforms in **Figure 7**: **(A)** Unsmoothed absolute errors; **(B)** Smoothed absolute errors and mean absolute errors.

direction, reaching a maximum on each side with the largest head rotation angle. Note that the MAE for 0° was slightly less than reported in the previous section on stationary measurements (see sections "Stationary Measurements" and "Clinical Booth"). This is likely due to slight alterations in the room acoustics caused by differences in placement of the clinical equipment used in the test booth (which is part of an active clinic). **Figure 15D** indicates that, like for the anechoic chamber, when the filters were updated with rotation, the MAE stayed consistently below −70 dB for most head rotation angles. A slight asymmetry was observed where the error was higher for positive angles, which was likely due to asymmetric room reflections.

DISCUSSION

We investigated the ability of a two-loudspeaker VA system to control HAD microphone pressure signals using CTC both in an anechoic chamber and a representative audiological testing booth. Our system has a small physical footprint and low cost compared to previous approaches, which required large

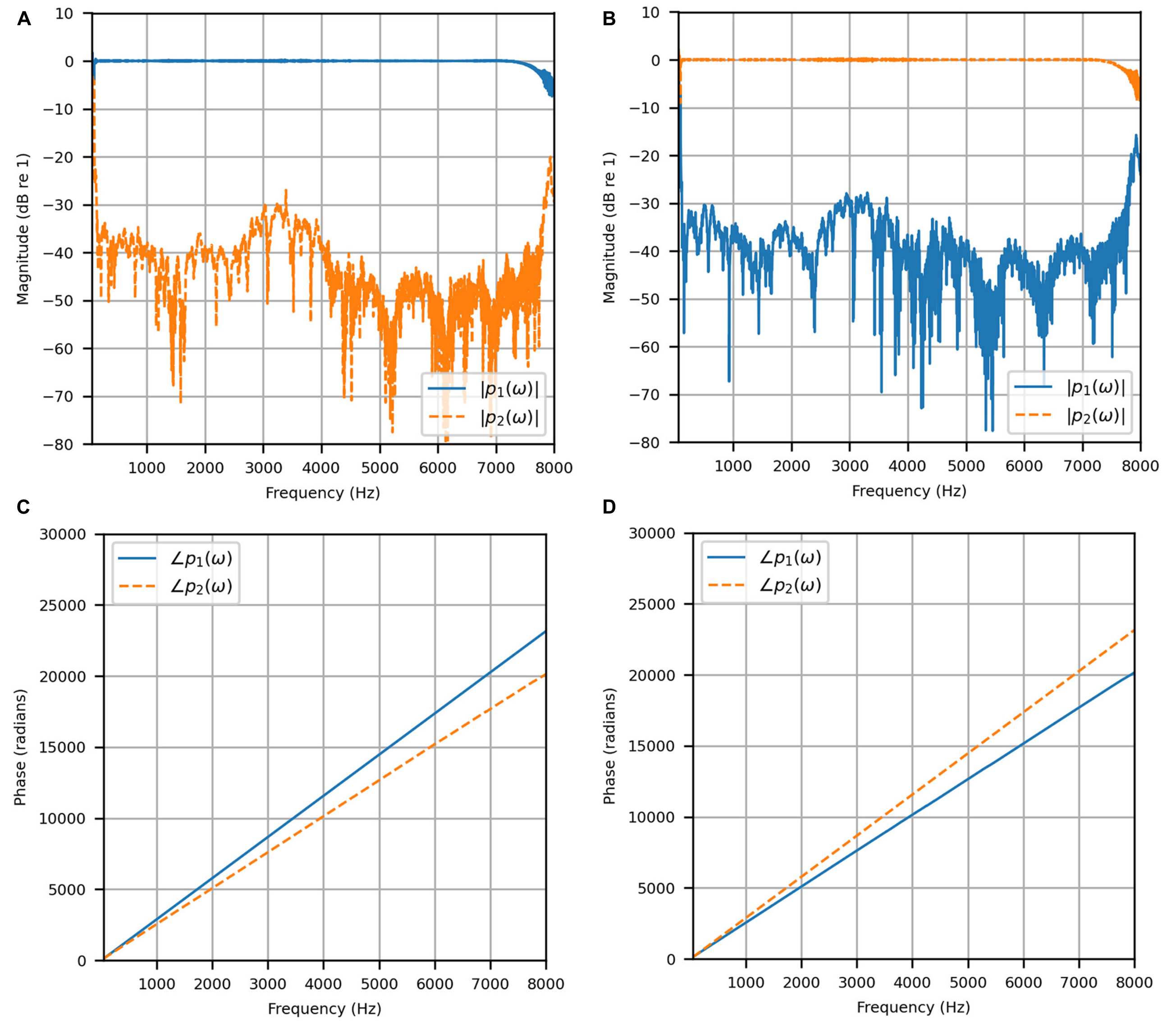


FIGURE 9 | Clinical booth: **(A)** Reproduced magnitude responses for a target impulse to the left microphone (blue solid); **(B)** Reproduced magnitude responses for a target impulse to the right microphone (red dashed); **(C)** Reproduced phase responses for a target impulse to the left microphone (blue solid); **(D)** Reproduced phase responses for a target impulse to the right microphone (red dashed).

loudspeaker arrays (e.g., Kitterick et al., 2011; Grimm et al., 2016). We evaluated the proposed VA system by attempting to reproduce target signals at the front microphones of behind-the-ear HADs worn by a KEMAR head and torso simulator. Using our proposed inverse filter design process, we showed a high reproduction accuracy under anechoic conditions, both in the time and frequency domain, when the head and torso simulator was kept still. To demonstrate the performance in a real-world environment, we repeated this stationary evaluation process in an audiological testing booth with only moderate acoustic treatment. We showed that the VA system performance within the booth was comparably accurate to the anechoic-based reproduction for the same anechoic target signals. In both spaces, we showed that the achievable channel separation in the passband matched, and for some frequencies exceeded, our target performance of 20 dB, which is the reported minimum needed to accurately reproduce the perception of the intended binaural signals for normal-hearing listeners. These findings demonstrate that the VA system can overcome the room acoustics within the testing booth. These results establish a baseline physical performance standard against which alternative systems and approaches can be assessed.

Head-related transfer functions can change markedly even with quite small head movements (Yu et al., 2018). Because of this, in clinical settings, participants are typically instructed to not move their heads. Nonetheless minor head movement (postural sway) is expected (e.g., Hirahara et al., 2010; Denk et al., 2017). Therefore, we investigated the effect

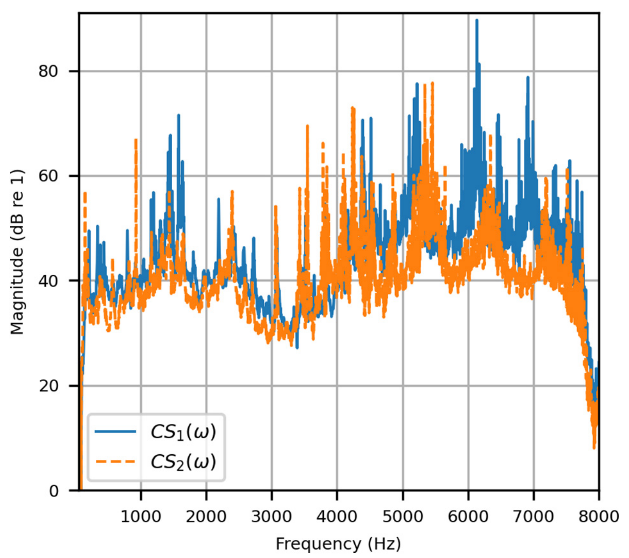


FIGURE 10 | Clinical booth: Channel separations for a target impulse to the left (blue solid) and right (red dashed) microphones.

of head rotation on the system's performance when using inverse filters designed assuming a stationary forward-facing listener (i.e., uncompensated performance). We compared the uncompensated performance to the performance when inverse filters were updated to account for head rotation (i.e., compensated performance). We found that the system was robust to minor head rotations of around 1° or 2° without compensation. However, for larger head rotations (3° or more) notable deviations from the target performance were observed. However, a psychophysical evaluation has not yet been conducted so it is unclear how much these deviations affect the perceived binaural signals. Hirahara et al. (2010) showed that participants who were instructed to stay still during HRTF measurements exhibited postural sway of no more than 2° for 5 minutes of continuous testing, which increased to between 3° and 5° (depending on the subject) after 20 minutes of continuous testing. Denk et al. (2017) showed that postural sway could be reduced to about 0.5° when the listener was given visual feedback on their current head position and allowed to realign before each measurement. A similar visual feedback system could be implemented during VA system measurements to reduce head movement during testing.

We showed that, when head rotation was compensated for, the performance was comparable to the baseline forward-facing and stationary performance across the full range of angles tested. This suggests that a dynamic system that adapts to listener position could markedly improve performance. In our measurements, the inverse filter design parameters, such as regularization and windowing applied to the HADIRs, were not varied with head rotation angle, and thus there is scope to further optimize these parameters. This may particularly benefit performance in the clinical booth, where room reflections influenced physical outcomes. It should be noted that the approach for compensating head movements that we presented requires either prior knowledge of the HADIRs for a given head rotation (e.g., Gálvez et al., 2019) or real-time measurement of HADIRs and updating of inverse filters (e.g., Kabzinski and Jax, 2019). An adaptive approach could be used with direct access to the HAD microphone signals and could allow for more listener movement during measurement, assuming the rate of head movement is not rapid (Kabzinski and Jax, 2019). Alternatively, a head tracking system could be used to reject the small number of trials where excessive head rotation occurs (e.g., Denk et al., 2017). In addition to compensating for head movement, future work should establish how effective these adaptive techniques are for different HADs, as well as for different head and pinna shapes and sizes.

The measured performance of the VA system suggests it has strong potential for clinical use. The high reproduction accuracy shows that the system can reproduce complex spatial auditory scenes. It could therefore improve diagnostic testing and assessment of HAD signal-processing performance by greatly increasing the potential for ecologically valid tests. Furthermore, its small footprint and low cost mean that it could find widespread use, including across low- and middle-income countries. While we used custom HADs with direct microphone access, which aren't currently commercially available, this access to the microphone signals could be gained using Bluetooth Low Energy. Bluetooth Low Energy is already used in most of the latest hearing-assistive devices and is capable of simultaneous multichannel output streaming. This existing technology could readily be adapted to allow audio streaming from device microphones to a VA system. However, the latency of the Bluetooth Low Energy transmission is a potential limitation that should be explored for its use in the system, especially in an adaptive mode of operation. If access to the microphone cannot be gained then additional measurements and equipment would be required (as in Chan et al., 2008; see section "Introduction"), which could be both time consuming and make the system more expensive.

Future work is required to fully establish the efficacy of our proposed approach for use in clinical audiology. While we have demonstrated accurate sound field control at two microphones, it remains to be shown how well the system can control the pressure signals for multiple, closely spaced microphones on a single HAD. Evaluating control at multiple device microphones is important as many modern HADs utilize onboard beamforming algorithms that rely on microphone arrays (e.g., Simon et al., 2020). Future work should evaluate simultaneous sound field control when there are two microphones per device (matching the configuration of many current hearing aids and cochlear implants). Simultaneous control at the ear canals is also desirable, as many HAD users have some degree of residual hearing (Pausch et al., 2018). Proper control at more than two microphones would require the number of loudspeakers to at least match the number of microphones. Thus, at least four loudspeakers would be required to control the sound field at two microphones per device, with two more loudspeakers (six in total) if residual hearing is to also be controlled. A greater number of loudspeakers could have the additional

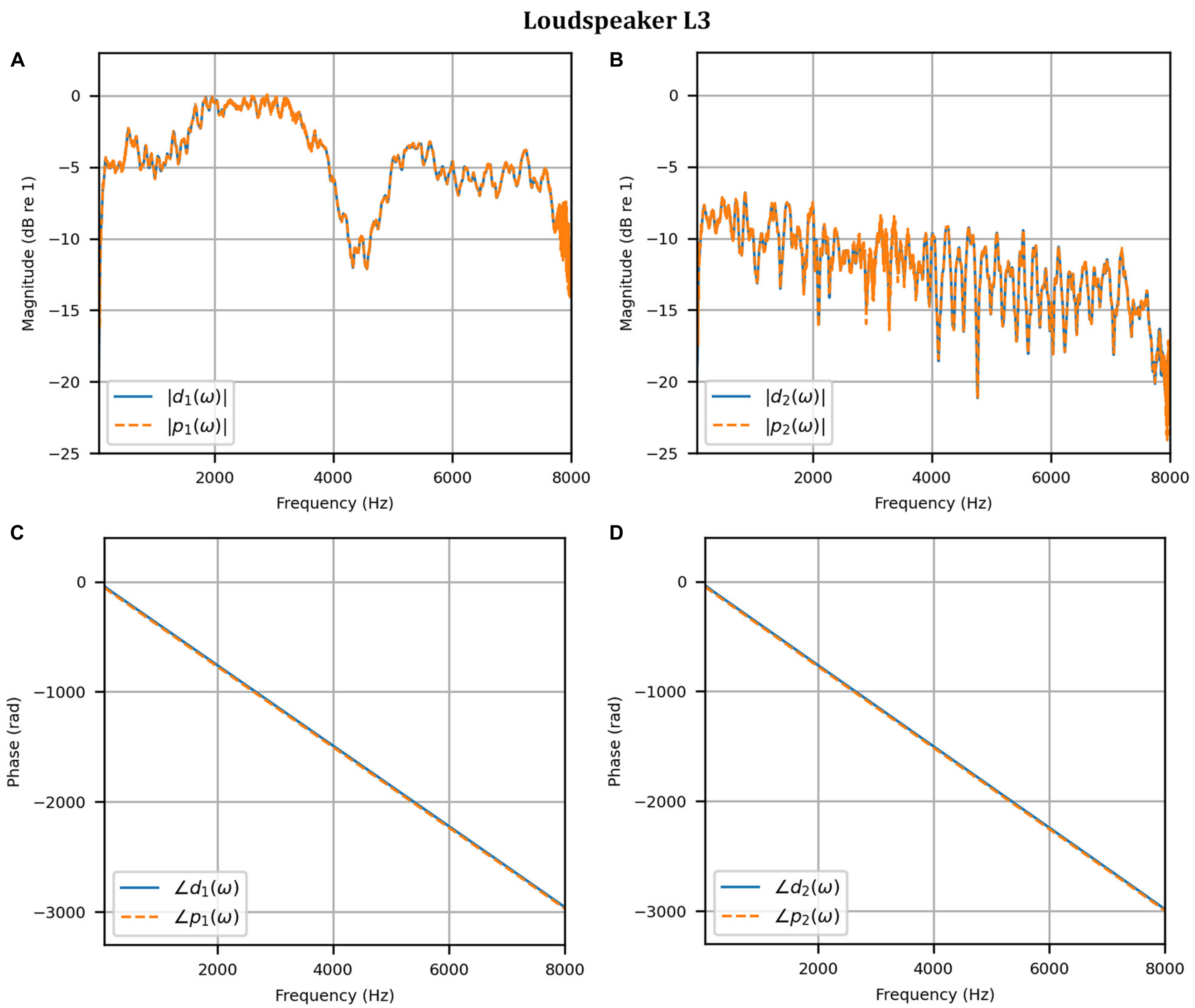


FIGURE 11 | Clinical booth: Target L3 anechoic HADRTF (blue solid) versus reproduced (red dashed) frequency responses: **(A)** Magnitude responses at the left microphone; **(B)** Magnitude responses at the right microphone; **(C)** Phase responses at the left microphone; **(D)** Phase responses at the right microphone.

benefit of allowing sound field control at each site to be improved due to the increased focusing capabilities (Hamdan and Fazi, 2021a,b). While this increases the cost of the system, the VA loudspeakers could be housed in a compact enclosure (e.g., a sound bar) and therefore the system's small footprint could be retained.

A further challenge for the VA system is the influence of visual cues in testing. Since the VA system recreates virtual sound images in directions where there is no loudspeaker or other visual marker, it may be difficult for a listener to properly indicate where different sounds are originating from and performance may be biased toward the VA system loudspeaker array (e.g., Witten and Knudsen, 2005; Mendonça, 2020). There are several potential ways to collect participant responses with the VA system. One would be to use simple markers placed at different locations in the testing room and another would be to deploy head or hand tracking and instruct the participant to direct their head or hand toward the sound source after each trial. To reduce visual biasing effects, VA system loudspeakers could be placed outside of the field of view (e.g., placed laterally) or be disguised (e.g., built into the wall of the testing booth). A more sophisticated approach would be to deploy a virtual or augmented reality headset. This could both give participants a range of visual targets through creation of a custom visual field and allow control of visual biasing effects. These headsets are relatively non-intrusive, low-cost, and would allow the system to maintain a small footprint.

Future work should also establish the link between physical performance and perception of virtual sounds in the clinical environment with HAD users. So far, we have presented

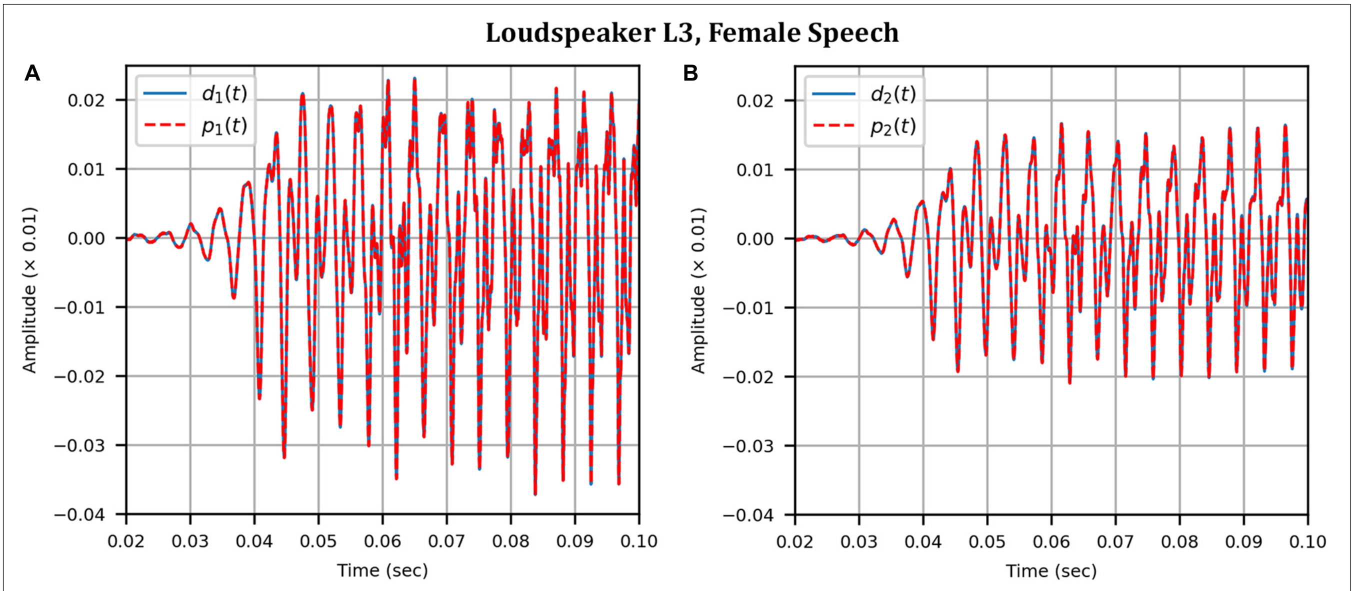


FIGURE 12 | Clinical booth: Target (blue solid) versus reproduced (red dashed) time domain waveforms of a female speech sample played from loudspeaker L3 (originally in the anechoic chamber) reproduced in the audiological booth: **(A)** Left microphone; **(B)** Right microphone.

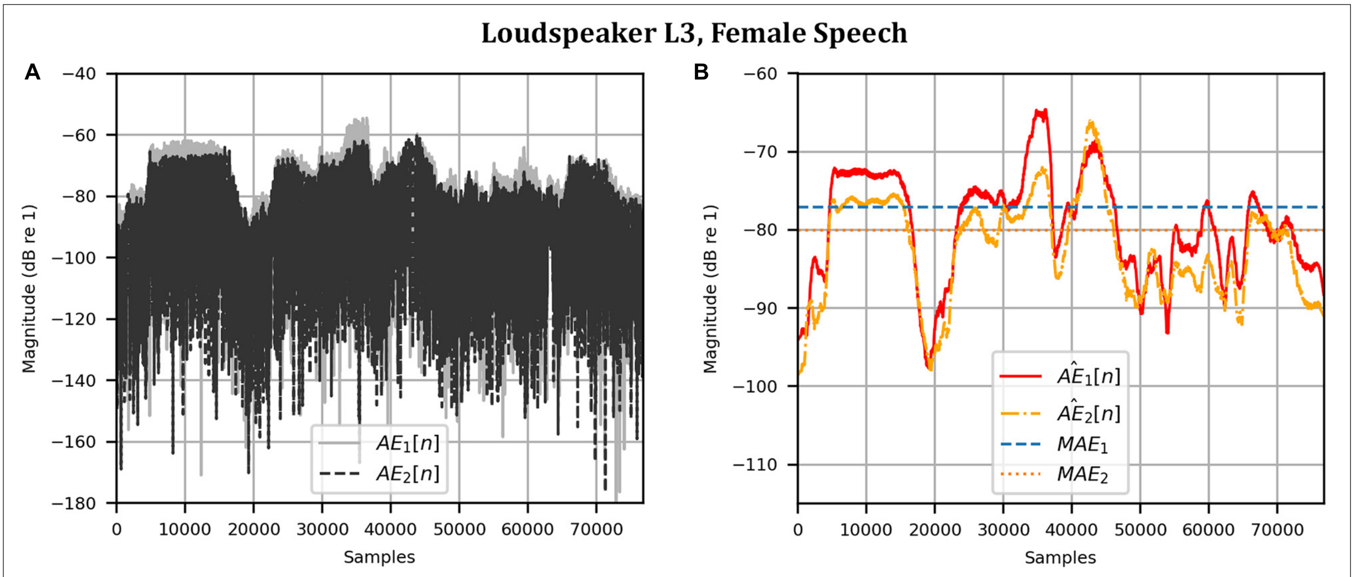


FIGURE 13 | Clinical booth: Error metrics quantifying the difference between the target and reproduced time domain waveforms in **Figure 12**: **(A)** Unsmoothed absolute errors; **(B)** Smoothed absolute errors and mean absolute errors.

a baseline physical performance that was only informally verified perceptually by normal-hearing listeners when listening to the reproduced signals over headphones. Future work should objectively evaluate the perceptual quality of the VA system with HAD users to establish explicit links between the physical and psychoacoustic domains within the intended user groups. Previous work evaluated the amount of channel separation needed for normal-hearing listeners to properly perceive the intended binaural signal using the proposed signal processing approach (e.g., Parodi and Rubak, 2011), however, more information is needed to determine the physical quantities needed for the intended perception in HAD user populations. Future studies should establish, for example, whether virtual sources produce an accurate perception of the intended source location (for both stationary and moving sources), source width, and that unintended spectral coloring doesn't occur for different source locations or different combinations of virtual sound sources. Study of the effectiveness of sound field reproduction for sources behind the listener might be a particular focus, as CTC systems have typically

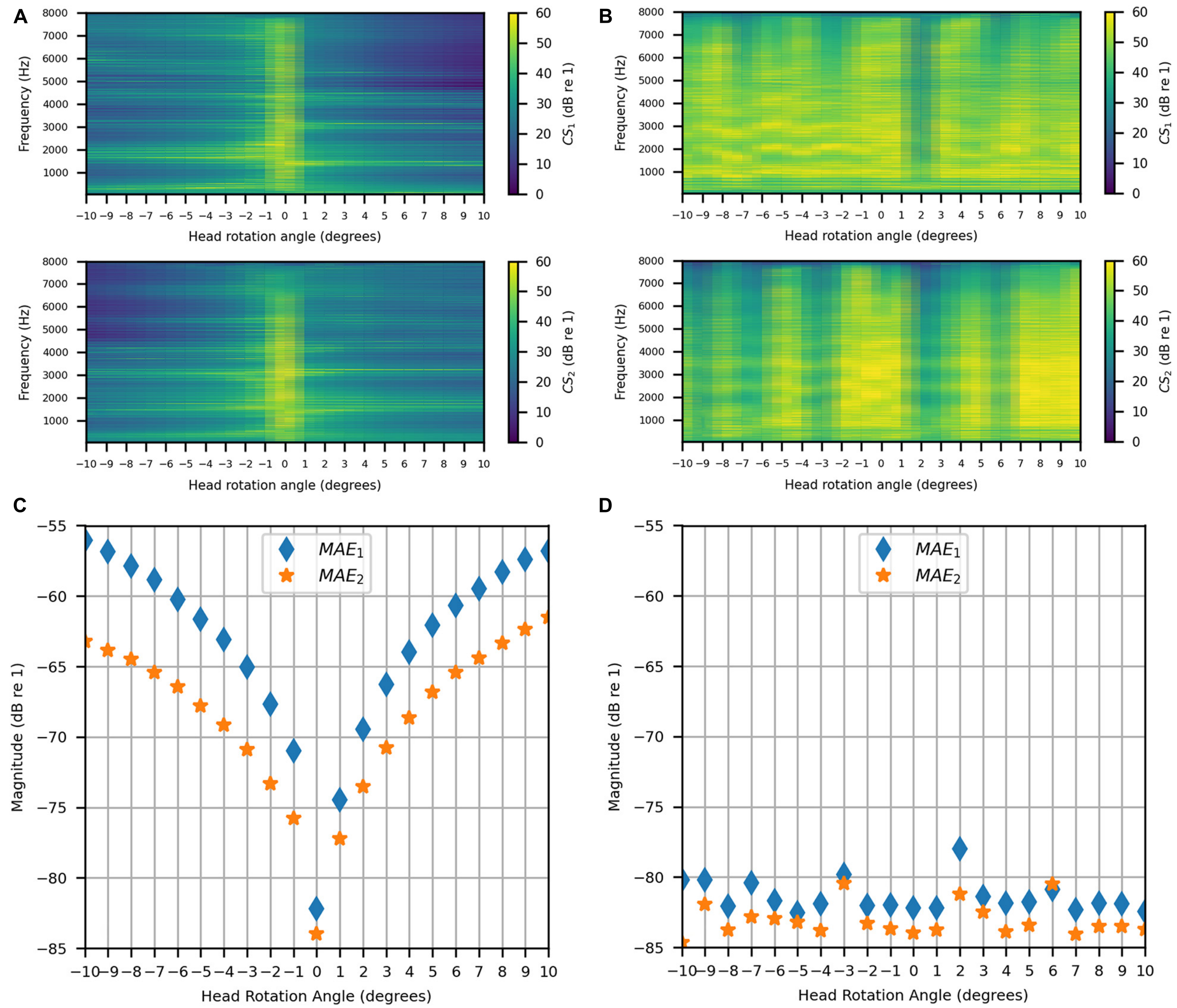


FIGURE 14 | Anechoic chamber: **(A)** Uncompensated channel separation as a function of frequency in Hz and head rotation angle; **(B)** Compensated channel separation as a function of frequency in Hz and head rotation angle; **(C)** Uncompensated mean absolute errors as functions of head rotation angle; **(D)** Compensated mean absolute errors as functions of head rotation angle.

struggled to effectively reproduce such sources (although this reproduction issue may be reduced as the HAD receiver is behind the ear and therefore subject to less extreme front-back spectral differences). Finally, study of the variability of these precepts between individuals will also be critical. In addition to validating the system, study of the link between the physical signal reproduction and the perceived sound source could lead to a more efficient and simple reproduction system if less physical accuracy were required than previously thought.

Finally, future work should assess the performance of the VA system across a wider range of clinical testing facilities, including those without acoustic treatment. The proposed method has the potential to adapt to non-ideal testing environments that have poor acoustic treatment. Advanced machine learning techniques for acoustic scene classification, such as convolutional neural networks, could be explored to aid effective room adaptation (e.g., Valenti et al., 2017; Zhang et al., 2020). The ability to effectively adapt to a wide range of non-ideal settings is likely to be especially important for supporting clinical audiology in low- and middle-income countries, where acoustically treated facilities are often not available. Effective room adaptation might also open the possibility of using the system in people's homes. This could allow more advanced remote audiological testing, training, and rehabilitation, which would be highly timely given the recent

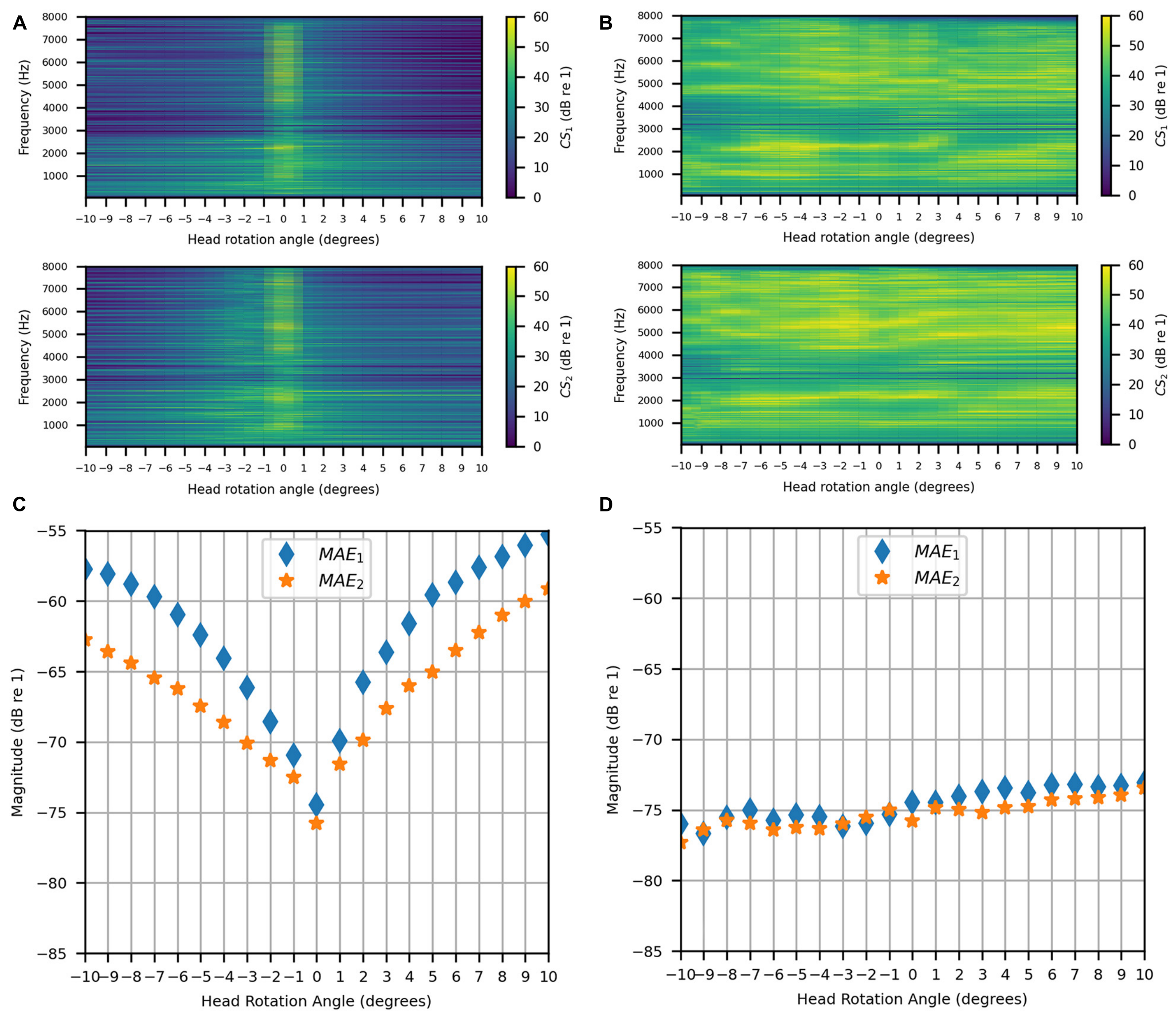


FIGURE 15 | Clinical Booth: **(A)** Uncompensated channel separation as a function of frequency in Hz and head rotation angle; **(B)** Compensated channel separation as a function of frequency in Hz and head rotation angle; **(C)** Uncompensated mean absolute errors as functions of head rotation angle; **(D)** Compensated mean absolute errors as functions of head rotation angle.

surge in interest in telemedicine. Furthermore, if shown to be effective at removing the impact of room acoustics (which can reduce intelligibility) in a home environment, this technology could widen access to media and entertainment for hearing-impaired individuals.

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: doi: 10.5258/SOTON/D2081.

AUTHOR CONTRIBUTIONS

EH: conceptualization, formal analysis, methodology, investigation, manuscript writing, and funding acquisition. MF: conceptualization, methodology, manuscript writing, project administration, supervision, and funding acquisition. Both authors contributed to the article and approved the submitted version.

FUNDING

The salary for both authors and the equipment for this project was funded by the William Demant Foundation. The article processing charge was funded by the Engineering and Physical Sciences Research Council through the University of Southampton UK Research and Innovation block grant for open access publication.

ACKNOWLEDGMENTS

We would like to thank Bradford Backus (Oticon Medical) for providing us with the hearing-assistive devices used in this research and for insightful discussions. We would also like to thank Jacob Hollebon for his help with the experiment equipment, Falk-Martin Hoffmann for his impulse response measurement code, and John Fithyan and the rest of the ISVR Consulting team for their assistance with the anechoic chamber.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.725127/full#supplementary-material

REFERENCES

British Society of Audiology (2017). *Recommended Procedure: Pure-Tone Air-Conduction and Bone-Conduction Threshold Audiometry with and Without Masking*. Bathgate West: British Society of Audiology.

Brown, K. D., and Balkany, T. J. (2007). Benefits of bilateral cochlear implantation: a review. *Curr. Opin. Otolaryngol. Head Neck Surg.* 15, 315–318. doi: 10.1097/MOO.0b013e3282ef3d3e

Chan, J. C. Y., Freed, D. J., Vermiglio, A. J., and Soli, S. D. (2008). Evaluation of binaural functions in bilateral cochlear implant users. *Int. J. Audiol.* 47, 296–310. doi: 10.1080/14992020802075407

Cubick, J., and Dau, T. (2016). Validation of a virtual sound environment system for testing hearing aids. *Acta Acust. United Acust.* 102, 547–557. doi: 10.3813/AAA.918972

Denk, F., Heeren, J., Ewert, S. D., Kollmeier, B., and Ernst, S. M. A. (2017). "Controlling the head position during individual HRTF measurements and its effect on accuracy," in *Proceedings of the Fortschritte der Akustik – DAGA*, Kiel.

Dorman, M. F., Loiselle, L. H., Cook, S. J., Yost, W. A., and Gifford, R. H. (2016). Sound source localization by normal-hearing listeners, hearing-impaired listeners and cochlear implant listeners. *Audiol. Neurotol.* 21, 127–131. doi: 10.1159/000444740

Farina, A. (2000). Simultaneous measurement of impulse response and distortion with a swept-sine technique. *Audio Eng. Soc. Conv.* 108:5093.

Fletcher, M. D., Cunningham, R. O., and Mills, S. R. (2020a). Electro-haptic enhancement of spatial hearing in cochlear implant users. *Sci. Rep.* 10:1621. doi: 10.1038/s41598-020-58503-8

Fletcher, M. D., Song, H., and Perry, S. W. (2020b). Electro-haptic stimulation enhances speech recognition in spatially separated noise for cochlear implant users. *Sci. Rep.* 10:12723. doi: 10.1038/s41598-020-69697-2

Fletcher, M. D., and Zgheib, J. (2020). Haptic sound-localisation for use in cochlear implant and hearing-aid users. *Sci. Rep.* 10:14171. doi: 10.1038/s41598-020-70379-2

Gajecki, T., and Nogueira, W. (2021). Enhancement of interaural level difference for bilateral cochlear implant users. *Hear. Res.* 409:108313. doi: 10.1016/j.heares.2021.108313

Gálvez, M. S., Menzies, D., and Fazi, F. M. (2019). Dynamic audio reproduction with linear loudspeaker arrays. *J. Audio Eng. Soc.* 67, 190–200. doi: 10.17743/jaes.2019.0007

Grimm, G., Ewert, S., and Hohmann, V. (2015). Evaluation of spatial audio reproduction schemes for application in hearing aid research. *Acta Acust. United Acust.* 101, 842–854. doi: 10.3813/AAA.918878

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

Hamdan, E. C., and Fazi, F. M. (2021a). A modal analysis of multichannel crosstalk cancellation systems and their relationship to amplitude panning. *J. Sound Vib.* 490, 115743. doi: 10.1016/j.jsv.2020.115743

Hamdan, E. C., and Fazi, F. M. (2021b). Ideal focusing and optimally-conditioned systems in sound field control with loudspeaker arrays. *J. Sound Vib.* 509:116218. doi: 10.1016/j.jsv.2021.116218

Hirahara, T., Hiroyuki, S., Toshima, I., and Otani, M. (2010). Head movement during head-related transfer function movements. *Acoust. Sci. Technol.* 31, 165–171. doi: 10.1250/ast.31.165

Kabzinski, T., and Jax, P. (2019). An adaptive crosstalk cancellation system using microphones at the ears. *Audio Eng. Soc. Conv.* 147:10307.

Kirkeby, O., Nelson, P. A., Hamada, H., Orduna-Bustamante, F., (1998). Fast deconvolution of multichannel systems using regularization. *IEEE Trans. on Speech and Audio Process.*, 6:189–194. doi: 10.1109/89.661479

Kitterick, P. T., Lovett, R. E., Goman, A. M., and Summerfield, A. Q. (2011). The AB-York crescent of sound: an apparatus for assessing spatial-listening skills in children and adults. *Cochl. Implants Int.* 12, 164–169. doi: 10.1179/146701011x13049348987832

Litovsky, R. Y., Parkinson, A., Arcaroli, J., Peters, R., Lake, J., Johnstone, P., et al. (2004). Bilateral cochlear implants in adults and children. *Arch. Otolaryngol. Head Neck Surg.* 130, 648–655. doi: 10.1001/archotol.130.5.648

Lokki, T., and Savioja, L. (2008). "Virtual acoustics," in *Handbook of Signal Processing in Acoustics*, eds D. Havelock, S. Kuwano, and M. Vorländer (New York, NY: Springer), 761–771.

Lovett, R. E., Kitterick, P. T., Hewitt, C. E., and Summerfield, A. Q. (2010). Bilateral or unilateral cochlear implantation for deaf children: an observational study. *Arch. Dis. Child* 95, 107–112. doi: 10.1136/adc.2009.160325

Mendonça, C. (2020). "Psychophysical models of sound localisation with audiovisual interactions," in *The Technology of Binaural Understanding. Modern Acoustics and Signal Processing*, eds J. Blauert and J. Braasch (Cham: Springer), 289–314.

Meng, R., Xiang, J., Sang, J., Zheng, C., Li, X., Bleeck, S., et al. (2021). Investigation of an MAA test with virtual sound synthesis. *Front. Psychol.* 12:656052. doi: 10.3389/fpsyg.2021.656052

Minnaar, P., Albeck, S. F., Simonsen, C. S., Søndersted, B., Oakley, S. A. D., and Bennedbæk, J. (2013). Reproducing real-life listening situations in the laboratory for testing hearing aids. *Audio Eng. Soc. Conv.* 135:8951.

Moore, B. C. J., Kolarik, A., Stone, M. A., and Lee, Y.-W. (2016). Evaluation of a method for enhancing interaural level differences at low frequencies. *J. Acoust. Soc. Am.* 140, 2817–2828. doi: 10.1121/1.4965299

Olivieri, F., Fazi, F. M., Shin, M., and Nelson, P. (2015). Pressure-matching beamforming method for loudspeaker arrays with frequency dependent selection of control points. *Audio Eng. Soc. Conv.* 138, 9322.

Oreinos, C., and Buchholz, J. M. (2016). Evaluation of loudspeaker-based virtual sound environments for testing directional hearing aids. *J. Am. Acad. Audiol.* 27, 541–556. doi: 10.3766/jaaa.15094

Parodi, L. E., and Rubak, P. (2011). A subjective evaluation of the minimum channel separation for reproducing binaural signals over loudspeakers. *J. Audio Eng. Soc.* 59, 487–497.

Pastore, M. T., Pulling, K. R., Chen, C., Yost, W. A., and Dorman, M. F. (2021). Effects of bilateral automatic gain control synchronization in cochlear implants with and without head movements: sound source localization in the frontal hemifield. *J. Speech Lang. Hear. Res.* 64, 2811–2824. doi: 10.1044/2021_JSLHR-20-00493

Pausch, F., Aspöck, L., Vorländer, M., and Fels, J. (2018). An extended binaural real-time auralization system with an interface to research hearing aids for experiments on subjects with hearing loss. *Trends Hear.* 22:2331216518800871. doi: 10.1177/2331216518800871

Seeber, B. U., Baumann, U., and Fastl, H. (2004). Localization ability with bimodal hearing aids and bilateral cochlear implants. *J. Acoust. Soc. Am.* 116, 1698–1709. doi: 10.1121/1.1776192

Seeber, B. U., Kerber, S., and Hafter, E. R. (2010). A system to simulate and reproduce audio-visual environments for spatial hearing research. *Hear. Res.* 260, 1–10. doi: 10.1016/j.heares.2009.11.004

Simon, L. S., Dillier, N., and Wüthrich, H. (2020). Comparison of 3D audio reproduction methods using hearing devices. *J. Audio Eng. Soc.* 68, 899–909.

Valenti, M., Squartini, S., Diment, A., Parascandolo, G., and Virtanen, T. (2017). "A convolutional neural network approach for acoustic scene classification," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, (Anchorage, AK: Institute of Electrical and Electronics Engineers), 1547–1554. doi: 10.1109/IJCNN.2017.7966035

Williges, B., Jürgens, T., Hu, H., and Dietz, M. (2018). Coherent coding of enhanced interaural cues improves sound localization in noise with bilateral cochlear implants. *Trends Hear.* 22:2331216518781746. doi: 10.1177/2331216518781746

Witten, I. B., and Knudsen, E. I. (2005). Why seeing is believing: merging auditory and visual worlds. *Neuron* 48, 489–496. doi: 10.1016/j.neuron.2005.10.020

Xie, B. (2013). *Head-Related Transfer Function and Virtual Auditory Display*. Plantation, FL: J. Ross Publishing.

Yu, G., Wu, R., Liu, Y., and Xie, B. (2018). Near-field head-related transfer-function measurement and database of human subjects. *J. Acoust. Soc. Am.* 143:194. doi: 10.1121/1.5027019

Zhang, T., Liang, J., and Ding, B. (2020). Acoustic scene classification using deep CNN with fine-resolution feature. *Expert Syst. Appl.* 143:113067. doi: 10.1016/j.eswa.2019.113067

Zotter, F., and Frank, M. (2019). *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. Cham: Springer.

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.

frontiers in Digital Health

ORIGINAL RESEARCH
published: 16 February 2022
doi: 10.3389/fdgth.2022.806076

Check for updates

Preliminary Evaluation of Automated Speech Recognition Apps for the Hearing Impaired and Deaf

Leontien Pragt[1]*, *Peter van Hengel*[1,2], *Dagmar Grob*[3] and *Jan-Willem A. Wasmann*[1]

[1] *Department of Otorhinolaryngology, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center Nijmegen, Nijmegen, Netherlands,* [2] *Pento Audiological Center Twente, Hengelo, Netherlands,* [3] *Department of Medical Imaging, Radboud University Medical Center, Nijmegen, Netherlands*

Objective: Automated speech recognition (ASR) systems have become increasingly sophisticated, accurate, and deployable on many digital devices, including on a smartphone. This pilot study aims to examine the speech recognition performance of ASR apps using audiological speech tests. In addition, we compare ASR speech recognition performance to normal hearing and hearing impaired listeners and evaluate if standard clinical audiological tests are a meaningful and quick measure of the performance of ASR apps.

Methods: Four apps have been tested on a smartphone, respectively AVA, Earfy, Live Transcribe, and Speechy. The Dutch audiological speech tests performed were speech audiometry in quiet (Dutch CNC-test), Digits-in-Noise (DIN)-test with steady-state speech-shaped noise, sentences in quiet and in averaged long-term speech-shaped spectrum noise (Plomp-test). For comparison, the app's ability to transcribe a spoken dialogue (Dutch and English) was tested.

Results: All apps scored at least 50% phonemes correct on the Dutch CNC-test for a conversational speech intensity level (65 dB SPL) and achieved 90–100% phoneme recognition at higher intensity levels. On the DIN-test, AVA and Live Transcribe had the lowest (best) signal-to-noise ratio +8 dB. The lowest signal-to-noise measured with the Plomp-test was +8 to 9 dB for Earfy (Android) and Live Transcribe (Android). Overall, the word error rate for the dialogue in English (19–34%) was lower (better) than for the Dutch dialogue (25–66%).

Conclusion: The performance of the apps was limited on audiological tests that provide little linguistic context or use low signal to noise levels. For Dutch audiological speech tests in quiet, ASR apps performed similarly to a person with a moderate hearing loss. In noise, the ASR apps performed more poorly than most profoundly deaf people using a hearing aid or cochlear implant. Adding new performance metrics including the semantic difference as a function of SNR and reverberation time could help to monitor and further improve ASR performance.

Keywords: automated speech audiometry, (automatic speech recognition), automated speech recognition, (ASR), evaluation metric, hearing impairment, speech-to-text, voice-to-text technology

OPEN ACCESS

Edited by:
Qinglin Meng,
South China University of Technology, China

Reviewed by:
Josef Schlittenlacher,
The University of Manchester, United Kingdom
Huali Zhou,
Shenzhen University, China

***Correspondence:**
Leontien Pragt
leontien.pragt@radboudumc.nl

Specialty section:
This article was submitted to Connected Health, a section of the journal Frontiers in Digital Health

Received: *31 October 2021*
Accepted: *18 January 2022*
Published: *16 February 2022*

Citation:
Pragt L, van Hengel P, Grob D and Wasmann J-WA (2022) Preliminary Evaluation of Automated Speech Recognition Apps for the Hearing Impaired and Deaf. Front. Digit. Health 4:806076. doi: 10.3389/fdgth.2022.806076

INTRODUCTION

Automated Speech Recognition (ASR) has become increasingly sophisticated and accurate as a result of advances in deep learning, cloud computing, and the availability of large training sets (1, 2). The software converts speech into text using artificial intelligence models that have been trained on vast collections of speech containing millions of words. ASR software is widely available on most digital devices, including smartphones, tablets, or laptops. It is primarily used for voice commands (e.g., hey Siri!), at the workplace to create transcripts, or in class for taking notes. Recently, ASR has become available in online meetings (e.g., Microsoft teams) and video recordings (e.g., Google's Youtube) to provide automated captions. Also, several ASR-based speech-to-text apps have been developed for the hearing impaired and deaf, providing live captioning of conversations (2, 3), showing the potential of automation and artificial intelligence for hearing healthcare (4, 5). Early in 2020, we were confronted in our clinic with questions from patients related to the use of ASR apps for daily communication. These questions were especially common among patients with severe to profound hearing loss who visited our outpatient clinic to assess if they were eligible for a Cochlear Implant. Also, patients who had experienced sudden deafness, but had not yet been fitted with hearing aids, made use of an ASR app during their appointments. There was no or little experimental information at the time about the performance and usability of the ASR apps for hearing impaired persons beyond what was shared by developers. Nor did we have clear criteria for which groups of patients we might suggest the ASR apps to.

Background

Since 2017, several ASR systems have claimed speech recognition performance close to that of normally hearing humans (1, 2). The most common metric to express ASR performance, used to underpin these claims, is the word error rate (WER). WER is calculated by adding the number of missing, wrong, and inserted words and dividing this by the total number of words (6). A lower WER score means better performance. The performance of ASR will be best for speech similar to the speech on which it was trained (7). It is therefore important to understand for what specific task an ASR is designed and how it is evaluated. Typically ASRs are evaluated on well-studied large (>100 h) collections of speech, referred to as a corpus. The SwitchBoard corpus and CallHome corpus are well-known collections of conversational phone calls (8), whereas Librispeech is a corpus comprising speech from public domain audiobooks. The SwitchBoard corpus consists of conversations over the phone between strangers about a given topic (9). The CallHome corpus consists of more informal conversations between friends and family (8). None of these corpora are ideal for use in acoustically challenging environments. The SwitchBoard and CallHome were collected under low noise and low reverberation conditions (9), and a large portion of the Librispeech corpus has undergone noise removal and volume normalization (10).

In order to obtain estimates of human speech recognition performance that could be used for comparison with ASR, some researchers have determined the WER among professional transcribers of speech from the SwitchBoard and CallHome corpora. Saon et al. (1) estimated the lowest (best) achievable WER, 5.1% for SwitchBoard and 6.8% for CallHome, based on the best score taken from three professional speech transcribers after a quality check by a fourth speech transcriber. Xiong et al. (2) on the other hand, followed more realistic industry standard procedures, which are similar to how speech is processed by ASR. The reported WERs were 5.9% for SwitchBoard, and 11.3% for CallHome.

For some commonly-used ASR systems, WERs of 5.1% (Microsoft) and 5.5% (IBM) have been reported using the SwitchBoard corpus (11), which is close to the performance of normal hearing professionals reported above (1, 2). Benchmark results of widely used ASR systems tested on the same corpora are not available to our knowledge. Google reported a WER of 4.9%, but used a non-public corpus (11). Koenecke et al. (7) compared the performance of ASR systems from Amazon, Apple, Google, IBM, and Microsoft to transcribe structured interviews using two recent developed corpora (CORAAL and AAVE). However, transcribing a structured interview is a very different task than transcribing a conversation in real-time in acoustically challenging environments. More ecologically valid tasks are needed that take into account the effects of noise, reverberation, talker accent, and slang, for instance, to provide a realistic estimate of ASR performance when used for conversations in daily life under various acoustic conditions.

ASR for Hearing Impaired Listeners

For people with hearing impairments, there are specific user needs to consider when developing ASR apps. For example, these listeners might use both speechreading (12) and text reading of the ASR transcript from a screen. Speechreading conveys important non-verbal cues and nuances not included in a transcript and may enhance speech-in-noise abilities (13). However, without careful design, reading a transcript may interfere with someone's speechreading ability. Speaker identification cues [e.g., by color coding each speaker a feature in AVA (14)] may also direct the reader to the face of an active talker. Other design ideas include the notification of critical environmental sounds [a feature incorporated in Live Transcribe (15)], feedback to the speaker of their intelligibility of the ASR, or feedback to the speaker by making the transcript readable from two sides (e.g., mirrored) so that both the speaker and the listener can check the results [incorporated in Earfy (16)].

The settings where an ASR is used may also differ between individuals with impaired or normal hearing. For example, the settings where people with hearing loss use ASR may be more often in a more homely atmosphere between family members that might use more colloquial language or slang. That situation may be similar to closed caption for video series. The most common complaint of people with hearing loss is the reduced speech perception in complex listening environments including cocktail parties, restaurants, in conversations with their doctor, and family gatherings (15, 16). Adverse acoustic conditions, including low signal-to-noise, make it difficult for normal hearing listeners to understand speech and make the speech incomprehensible for persons with mild to profound hearing loss (17, 18). Finally, the speed of translation to accommodate a fluent conversation and the user interface to make it practical for older users and digitally less proficient users are factors to consider.

A standardized task that fully captures the skills of humans to recognize speech does not yet exist, to our knowledge. Such a task would need to account for factors as background noise, reverberation, accent, and speech impairment. This is needed to verify claims that ASR speech recognition performance is close to humans (1, 2) and should be done using diverse training datasets (7).

Objective

This pilot study aimed to examine the speech recognition performance of ASR apps using audiological speech tests. We normally administer clinical audiology tests in patients from normal hearing to profound hearing loss to assess speech recognition. We tested the hypothesis that our clinical tests might thus provide objective metrics for performance of ASR systems for people with hearing loss, helping us to determine what range of hearing losses could benefit from ASR apps. In addition, we compared ASR results to normal hearing and hearing impaired listeners and evaluated if standard clinical audiological tests provide a meaningful and quick measure of the performance of ASR apps.

METHODS

Four different apps on two smartphones, with various operating systems, were tested on their ability to transcribe speech. For this project, the iOS operating apps were tested using an iPhone 6, and for the Android operating apps, a Samsung A3 was used. Both smartphone devices are widely used. We decided to select inexpensive ASR apps (<$10) for a user-license since they would be most widely used by our patients while the cost for ASR apps is not reimbursed in the Netherlands. The four apps tested were Ava and Earfy that both run on iOS and Android, Speechy iOS only, and Live Transcribe Android only. The tested apps were chosen by searching on the Internet on November 18th, 2019, for the best-known speech recognition apps for the hearing impaired and deaf as well as good reviews on the different app-stores. Also, the apps needed to be suitable to convert English and Dutch speech into text.

The apps were evaluated in similar test conditions used to assess speech reception in human listeners in Dutch Audiology Centers according to best local clinical practice. The smartphones were placed one meter in front of a speaker in a sound treated room compliant with ISO 8253-1 (19). Standard clinical calibration protocols were used for all speech material. The microphone of the smartphone was aimed toward the speaker, which we assumed to be the optimal microphone orientation, at approximately the height of a listener's ears to resemble testing conditions when tested with human listeners (see **Figure 1**). The smartphone screen was facing upwards allowing the experimenter to read the text from the screen. Four different speech reception tests were performed to evaluate the app's ability to convert speech into text.

FIGURE 1 | Set-up of the smartphone in front of the speaker.

First, the apps were tested on speech recognition in quiet by converting a list of single words into text. The standard Dutch speech recognition test for this purpose is the Dutch CNC-test, which consists of phonetically balanced lists of twelve monosyllabic Dutch words in quiet [CNC-list, "Nederlandse Vereniging voor Audiologie;" (20)]. The words were played through a speaker, scored and displayed in a phoneme recognition score. All words consisted of three phonemes with a consonant-nucleus-consonant (CNC) structure. The first word was a test word and was not included in the scoring. A human observer performed the scoring by reading the word from the screen and counting the number of correct phonemes. Inserted phonemes were subtracted from the score according to the clinical scoring procedure (20). If a displayed word changed during the test, the final word was scored. A 100% phoneme recognition score was reached if all 33 phonemes of the 11 words were correct. Several lists were presented at an intensity level of 45, 55, 65, 75, and 85 dB sound pressure level (SPL) and the speech recognition as a function of presentation level (known in human listeners as speech audiogram) is plotted for each app. For comparison, normal hearing listeners achieve 100% phoneme recognition at 45 dB SPL (20).

Second, the Plomp-test (Dutch sentences in noise) was presented (21). The test consists of 13 sentences of 8 to 9 syllables presented in noise with the same averaged long-term spectrum as speech. A sentence was scored to be either correct, if the whole sentence was correctly presented on the screen, or incorrect, which was according to the conventional scoring procedure in clinical practice (22). The speech recognition threshold (SRT) in noise was defined as the signal-to-noise ratio (SNR) expressed in dB where on average 50% of the time the sentences were transcribed correctly, following the adaptive procedure described by Plomp and Mimpen (20, 21). The test was first performed without noise to obtain the SRT in quiet. Afterward, the masking noise level was set 15–20 dB above the SRT of the apps in the quiet situation, which was 70 dB SPL for all apps, to determine the speech reception threshold (SRT) in noise.

Third, a DIN-test (Digits-in-Noise) was performed. Digit triplets (e.g., 1 2 5) were presented in long-term average speech-spectrum noise via a 1-up, 1-down adaptive SNR procedure. SRT was expressed in dB SNR, where a listener can on average recognize 50% of the digit triplets correctly. A test series consisted of 24 triplets. The first four triples were not used to determine the test outcome. The noise level was set at a fixed level of 60 dB with an initial positive SNR of 6 dB. The stepsize to adjust the level of the triplets was 2 dB. The DIN-test has a measurement error in humans of 0.7 dB (23).

Fourth, a fragment of dialogue in Dutch and English at 72.2 dB(A) was presented through the speaker to recreate a more realistic listening setting. The Dutch dialogue was an introduction video of the Radboudumc with a female voice, talking clearly and at a normal pace (https://www.youtube.com/watch?v=zBJBD1-ePRw). For the English dialogue, part of an advanced English tutorial was played. In this video, a conversation could be heard between a male and female voice (https://www.youtube.com/watch?v=JtMgw2rCYSo&t=1s). The Dutch dialogue consisted of 256 words, while the English dialogue consisted of 248 words. After the whole dialogue was played, scoring was performed on the transcript outputted by the app. The number of missing, wrong, and inserted words was counted and expressed in the WER.

In the end, a test-retest was performed to provide insight into the accuracy of the apps. All apps were retested on the CNC-test. The test-retest reliability on the CNC-test was visually assessed using a Bland-Altman graph. The best scoring app on the DIN- and Plomp-test, one for iOS and one for Android, was retested for both speech-in-noise tests. The Root-Mean-Square-Difference (RMSD) was calculated for these results. No retest was performed for the dialogue.

RESULTS

The results for all apps on the Dutch CNC-test (words in quiet) are shown in **Figure 2**. Speech recognition as a function of presentation level was determined per app by interpolating a line using logistic regression on all available-data points (test and retest measurements). A 100% phoneme recognition was reached around 80 dB SPL for all apps except Earfy. Earfy (iOS and Android) scored 90% words correctly around 90 dB SPL. The shape of the app's "speech audiogram" curves look similar to the s-shaped psychometric curve of normal hearing listeners determined by Bronkhorst et al. (24) in 20 normal hearing university students. However, all app's SRT were between 50 and 60 dB SPL, which is 25 to 35 dB poorer than normal hearing listeners who have a SRT around 25 dB SPL (20).

The speech-in-noise results are shown in **Figures 3**, **4**. All apps score a signal-to-noise ratio (SNR) higher than +8 dB on the DIN- and Plomp-test. Live transcribe (Android), and AVA (Android, iOS) achieved the best results on the DIN-test. Earfy on Android performed better than on iOS. Live Transcribe (Android) and AVA (iOS) achieved the best result using the Plomp-test. There was a notable difference between the operating systems for AVA and Earfy when measured with the Plomp-test.

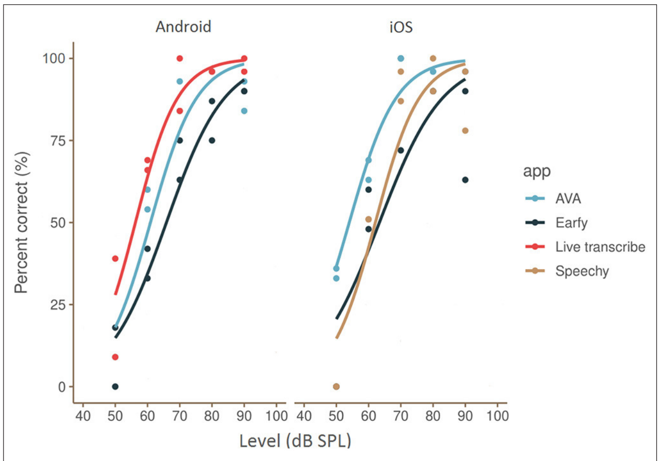


FIGURE 2 | Speech recognition as a function of presentation level (in human listeners known as speech audiogram) of all ASR apps tested on an Android and iOS smartphone. The plotted lines are interpolated using a logistic function through the measured test-retest data-points. Left side, results of the Android apps, right side, results of the iOS apps.

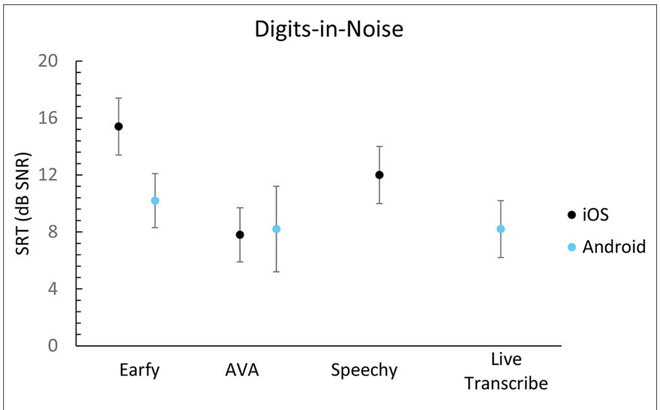


FIGURE 3 | Digit in noise results expressed in SNR per app. A lower score is better. The error bars represent the standard deviation of the response of the app within a single list of triplets.

In **Figure 5**, the WER scores for both the Dutch and English dialogue are shown. Overall, the dialogue in English (WER 19–34%) was more correctly converted into words than the Dutch (WER 25-66%) dialogue. Speechy (iOS) had best matching result for English and Dutch (WER of 19 and 20%). Earfy (iOS) showed the greatest difference between English and Dutch (WER of 19 and 66%).

The test-retest reliability of the CNC-test can be seen in **Figure 6**. Visual inspection of the Bland-Altman plot for the CNC-test-test did not show signs of any systematic bias in the relationships between differences and averages. The test-retest comparison of the CNC-test showed three outliers. Earfy for iOS exhibited large differences between the measurements at 70 and 90 dB and Live transcribe (Android) had a large difference

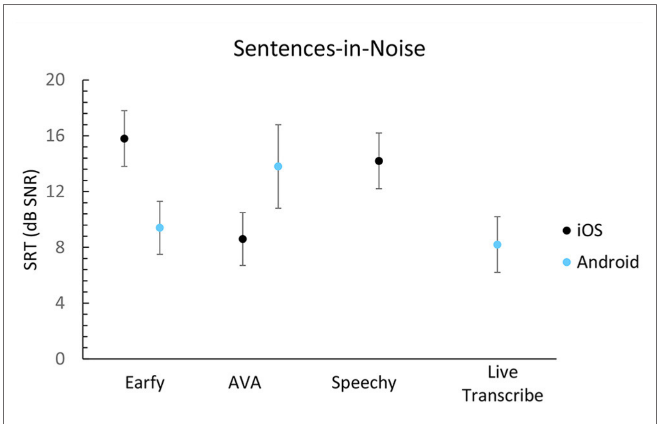


FIGURE 4 | Sentences in noise results expressed in SNR per app. A lower score is better. The error bars represent the standard deviation of the response of the app within a single list of sentences.

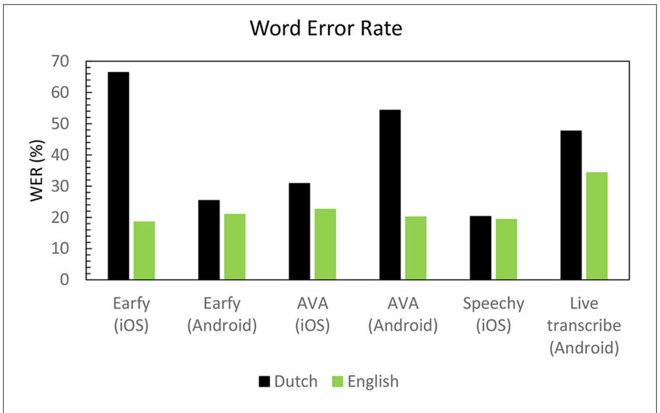


FIGURE 5 | Word error rate in percentage of the dialogue in English and Dutch for the different apps.

between measurements at 50 dB. The test-retest reliability on the DIN- and Plomp-tests was assessed for one Android and one iOS app. The test-retest difference expressed in RMSD on the DIN-test was 0.4 dB iOS Ava and 0.8 dB Android Live Transcribe, which we regard as acceptable since in normal hearing listeners tested monaurally using headphones, 90% of measurements are within 1.4 dB (measurement error is 0.70 dB) for a single list on the DIN-test (23). The RMSD on the Plomp-test was 0.6 dB iOS Ava and 2.0 dB for Android Live Transcribe.

DISCUSSION

Main Results

None of the ASR apps achieved performance close to normal hearing listeners on audiological tests. In quiet, ASR apps performed similarly to listeners with a moderate hearing loss. When transcribing speech-in-noise, the ASR apps performed in the performance range of CI recipients. Sentences-in-noise provided a quick test to assess ASR performance since that test material provided more linguistic cues than digits-in-noise or lists of CNC words.

Performance Compared to Human Listeners

The performance of the ASR apps on speech-in-quiet tests seems comparable to listeners with a moderate conductive hearing loss (30–35 dB threshold shift), which is known as disabling for certain activities in daily life (25). In comparison, Dingemanse and Goedegebure (26) found a mean score of 82% in 50 adult unilateral CI-recipients on the Dutch CNC-test tested in free field at 65 dB SPL, which is the level of conversational speech. This performance may be an overestimation for the average CI user since they excluded participants with a CNC-score below 60%. Kaandorp et al. (27) determined a mean score in free field at 65 dB SPL of 95% while using their preferred device in 24 hearing aid users with a moderate to severe hearing loss and 80% in 24 CI recipients. Only for speech at high-intensity levels, well above the level of conversational speech, do the apps achieve 90 to 100% speech reception. The poor performance at low speech intensity levels may be caused by hardware limitations, as discussed below in the section on hardware. The ASR may score lower due to the lack of contextual information provided in the test. The CNC-test was developed as an auditory test that requires little linguistic skill. The listener can only use the consonant-nucleus-consonant structure and the fact that the lists contain only familiar existing words. The alternative of using nonsense words, or nonsense sentences, would probably further deteriorate ASR performance while being a valid test for assessing auditory function with a lower effect of language skills by the subject (28). Most ASR are trained on sentences of realistic conversations (8). The strength of (deep learning) ASR is based on using contextual information from a natural language processing model (29). That contextual information is not available in word testing.

The performance of the ASR apps on the Digits-in-Noise test was very limited compared to humans. Normal hearing listeners achieve on the DIN-test, monaurally using headphones, an SNR of −8.8 dB (23). CI recipients rated on the same criteria as normal hearing listeners, typically achieve DIN scores ranging from +3 to−6 dB. For instance, Kaandorp et al. (27) found an average SNR of −1.8 (±2.7) dB in a group of 18 adult unilateral CI recipients in free field test conditions. The ASR is at a disadvantage because in the DIN-test, contextual information is lacking and the priors for the ASR and human are not the same. When doing a digits-in-noise test, a human will only report digits. For the ASR it is not a 10-class problem but a problem with several thousand alternatives. The apps tend to construct sentences rather than separate numbers. For conversations where it is important to catch a number, such as the price of an item, the DIN-test might be a useful measure.

The performance of ASR apps on sentences in noise (Plomp-test) was very limited and much poorer than in people with a moderate hearing loss (21). Normal hearing listeners have an SRT at an SNR of −8 to −10 dB (21), while the best ASR apps achieved +8 dB scores. Kaandorp et al. (27) found a mean SRT on Dutch Sentences in noise by scoring keywords of +2.1 dB for 24

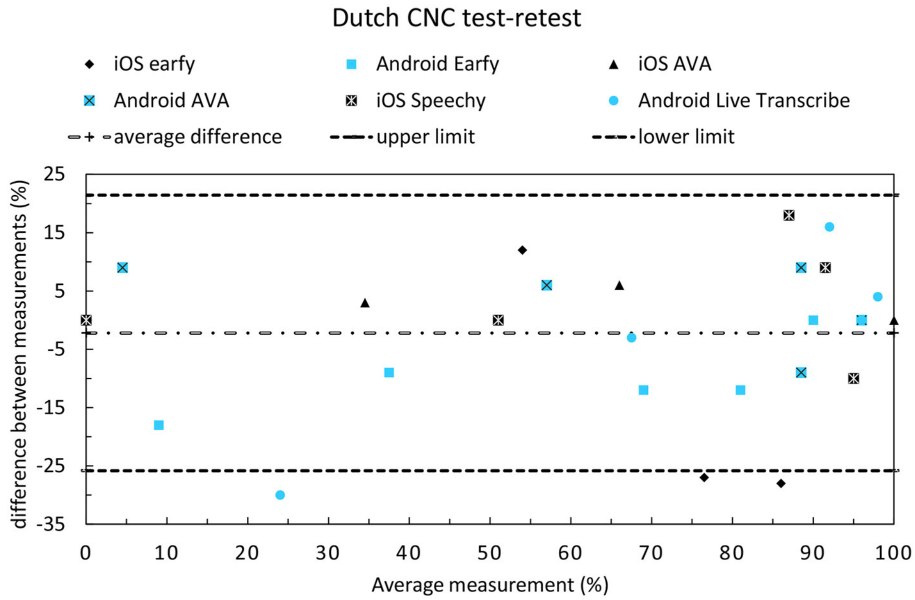


FIGURE 6 | Bland-Altman plot to display the test-retest reliability of the CNC-test.

hearing aid users (tested on their preferred ear) with moderate to severe hearing loss and +8.0 dB for 24 unilateral CI recipients. In CI-recipients, evaluation of speech-in-noise is often performed scoring keywords, instead of full sentences as used in the original procedure by Plomp and Mimpen (20, 21). In another study, Kaandorp et al. (30) found a significant difference of 1.0 dB in favor of a keyword scoring procedure compared to scoring full sentences. However, this 1.0 dB keyword effect does not account for the large difference between the app's performance and the performance of hearing aid users in noise. On the Plomp-test, which provides more linguistic information than the CNC- and DIN-test, the app's performance is far below that of the majority of hearing impaired listeners and similar to the range of outcomes in CI-recipients.

Sentences with and without noise (Plomp-test) could be considered as a performance metric for ASR apps in difficult listening conditions. Possibly with more natural sentences to provide even more linguistic cues. Testing through a loudspeaker has the advantage that it takes the effect of room acoustics into account, making the test condition more realistic. Instead of a sound booth, a room with more representative acoustics for daily situations (e.g., the reverberation time of a classroom or using babble noise instead of speech-shaped noise) would provide even more representative results. The current scoring procedure of the Plomp-test, based on fully correct sentences, leads to very high SNRs that may underestimate the practical value of ASR for hearing impaired persons. For instance, if an ASR in a conversation under noisy conditions provides keywords, it may already benefit the person with hearing impairment. One could easily adopt the Plomp-test by determining the WER score on a fixed SNR level to simulate above example. Or alternatively, accept a higher number of mistakes (compared to none) in the adaptive test by using keywords (30). Besides audiological test outcomes, the systematically collected feedback by groups of users (e.g., a focus group) would be very helpful to further improve the accessibility and usability of ASR apps for hearing impaired listeners.

In longer dialogues, all tested apps provided a running English transcript with a WER around 19–34%. This roughly corresponds to 60–80% correct word (∼1-WER) scores and this is in the same range as for persons with profound hearing loss who use a cochlear implant (31) and better than hearing aid users with a profound hearing loss (32). For these groups, the use of the ASR apps tested here would likely provide benefits.

Hardware and Platform Variability

A possible explanation for the poor performance at low levels could be the smartphone's microphone gain settings and limited dynamic range rendering soft sounds undetectable (33). We chose a microphone orientation, directing it to the speaker that we assumed was optimal for the task. However, we did not check the directionality of the built-in microphones. In actual use, the microphone orientation could be suboptimal, for instance, if a listener positions the device such that it enables better reading of the transcript from the screen. Also in group settings, the user will likely put the device flat on a table and thus not always point the microphone to the talker. We did not investigate the effect of suboptimal microphone orientation. Another explanation for the level dependence in quiet could be pre-processing. Most ASR systems usually normalize the input (34). Potentially the ASR systems classify soft sounds as non-speech or not of interest.

In English, there is not much difference between the apps or between the operating platforms. Therefore, we do not expect differences in the Dutch version to stem from hardware differences between the smartphones (e.g., microphone sensitivity) but from the implementation of the Dutch language in the specific app or the used training data. The difference between iOS and Android was only visible in Dutch. In Dutch, Earfy (iOS) and Ava (iOS) score significantly poorer.

There was no consistent difference favoring either iOS or Android versions of the apps. Earfy performed better on Android, while AVA performed better on iOS. For prospective users, the performance of the app depends on language, and may depend on the platform.

Limitations

The administered tests did not include the effect of accents or speech impairments [e.g., deaf speech; (7, 35)]. The displayed transcripts changed during the dialogue, and the transcript was evaluated at the end of the dialogue instead of in real-time. When reading the transcript in real-time, the performance of the speech recognition apps might be better or worse due to the changing words in real-time to construct a logical sentence.

When measuring performance in noise, an adaptive SNR procedure was used. The effect of noise could be more extensively studied by evaluating ASR by determining the Word Recognition Score (the convention in the field of audiology) or the Word Error Rate (the convention in the field of ASR research) on several fixed SNR levels (e.g.,−5, 0, +5 and +10 dB SNR) that correspond to realistic listening conditions for people using a hearing aid (36). For ecological valid measures, the effect of different fluctuating noise maskers should be considered (37, 38). Babble-noise or traffic noise is much more realistic than (artificial) steady-state speech-shaped noise. In the end, the performance of the ASR must be robust enough that users will put their trust in these apps even in formal situations such as a conversation with their doctor or audiologist.

In this study, only (audiological) speech-to-text performance of the apps was measured. The usability, processing speed, effect on speechreading, and readability of the transcript were not evaluated. Other researchers looked into requirements for speed and user interface and concluded that those are important factors to improve usability (39). We expect that an increasing number of ASR apps will adhere to accessibility guidelines to improve usability for the elderly and people with disabilities as promoted by the Web Accessibility Initiative (40).

The number of apps tested in this study is limited. We did not perform a standardized procedures for literature review (e.g., PRISMA) to find and include ASR apps for this pilot study. In English, more apps may be available than in Dutch and we did not include expensive state-of-the-art (professional) ASR systems.

Other factors to consider not included in this pilot study are the distance between speaker and listener, especially in these times of social distancing and the effect of face masks on a speaker's voice and intelligibility (41). Feedback about voice quality could help the speaker adopt a more intelligible speaker style. The errors made by the ASR may be complementary or redundant to the errors made by persons with hearing loss. We did not study the error patterns. A potential way to determine the complementary effect of ASR could be to evaluate speech-recognition in noise using an audiovisual presentation mode, instead of the audio-only mode that was used in this study, in three distinct aided conditions. (1) participants with hearing loss aided with hearing aid or CI. (2) participants with hearing loss aided with hearing aid or CI and using an ASR app, (3) performance by the ASR app only. Studying the difference between these conditions reveals the added benefit and may penalize ASR systems not designed for simultaneous speechreading and text reading.

Metrics to Evaluate Personalized ASR Performance

Instead of the quick audiological tests we performed here, a more conventional and elaborate evaluation method would be to record several hours of conversations with hearing impaired users (including realistic lexicon and acoustics) via a smartphone while the screen is oriented such that the user can read the transcript. Subsequently, one could create transcripts of the recordings by human transcribers as ground truth, pass the recordings through several ASR apps and determine a performance rating based on WER and other automated metrics such as the semantic distance between the ASR transcript and ground truth (42).

ASR may benefit from domain-specific evaluation tools and have domain-specific applications. For instance, Miner et al. (43) developed a metric based on symptom-focused language in psychotherapy. A domain-specific, or even person-specific factor is that prelingually deaf people often have a speech impairment, leading to lower comprehensibility both for normal hearing listeners who are not accustomed to deaf speech and for ASR apps that are not specifically trained on deaf speech. Fortunately, generic ASR models can be used as a pre-trained model that subsequently is trained on a particular task including a-typical speech, accents, or acoustic conditions without incurring the cost of training a full model (44). Recently, researchers from Google started a project, called Parrotron, to create personalized models which could better convert deaf speech than generic ASR systems. WER dropped from 89.2% for the generic ASR to 32.7% for the finetuned ASR for a single prelingually deaf subject (35). In addition, the Parrotron system can synthesize the speech of a speech impaired person (i.e., voice conversion) to make the speech sound more natural and comprehensible to the untrained ear.

Metrics as, for example, the WER (SNR, RT), or semantic difference (SNR, RT), as functions of signal-to-noise ratio and reverberation time (RT) can provide more ecologically valid estimates of the benefits ASR apps could provide in daily life. Representative SNR values could include −5, +10, +30 (quiet) dB SNR. For ecological valid measures, realistic fluctuating noise maskers should be used (37, 38). Reverberation times typically encountered in daily life to consider are 0, 0.5, and 2.5 s, which corresponds to ideal, classroom (45), and church (46) room acoustics. Presenting the ASR performance using the WER (SNR, RT) reduces the need to study the characteristic of the corpus on which the ASR was trained and or evaluated.

Future Benefits for Audiologists

ASR apps can provide benefits in conversations between patients and their audiologists (47). In addition, ASR technology, when further developed, can play a role in computational approaches to audiology (4). For instance, if personalized ASR apps further develop so that atypical speech is better captured, and if ASR achieves normal hearing performance on audiology tests it may provide another use case: patients could perform self-testing (i.e., automated speech audiometry) by repeating the speech they hear to an ASR system trained on their particular voice replacing or enhancing the task of the professional in the audiology center (48). Manual calculation of complex evaluation metrics is not suitable in clinical settings given the excessive time required and may lead to inter-rater variability (49). Automated speech audiometry using algorithms to score performance can be a valuable complement to automated pure-tone threshold audiometry (50). For example, Venail et al. (48) validated a semi-automatic speech procedure using customized word-lists, in part provided by the subject to include familiar words. The customized word-lists were recorded with the subject's own voice to incorporate personalized acoustic and articulatory parameters. Speech recognition was evaluated on the customized word-list using an algorithm to determine automatically the number of correctly repeated phonemes. In addition, the use of ASR could open venues to improved (automated) scoring methods in audiology tests. Ratnanather et al. (51) demonstrated how one can automate the alignment of phonemes based on the minimum edit distance between the source speech and the utterances of the subject in real time. Visualizing this alignment may provide insights to clinicians about what phonological errors are made.

A factor of variability in rating procedures is that in many speech-in-noise tests, the test is made easier for CI recipients by only scoring correct keywords rather than full sentences (28, 30). Although scoring keywords makes the test accessible to a larger population, it reduces the discriminative power between higher- and lower-educated native listeners (30). An ASR could facilitate an automated scoring procedure that differentiates between errors. For instance, using semantic difference between the ASR transcript and ground truth, errors that lead to semantically similar sentences are weighted favorably, leading to a better outcome metric in terms of how well hearing impaired persons can participate in a conversation under adverse circumstances.

CONCLUSION

None of the ASR apps achieved performance close to normal hearing listeners on audiological tests. No app stood out from the others on performance level. On audiological speech tests in quiet, ASR apps performed similarly to listeners with a moderate hearing loss. When transcribing speech-in-noise, the ASR apps performed in the performance range of CI recipients. Sentences-in-noise provided a quick test to assess ASR performance. Additional performance measures are needed to evaluate ASR apps. Besides the speech material, also type of noise and the presentation mode audio-only vs. audiovisual need to be considered. Adding new performance metrics including the semantic difference as a function of SNR and reverberation time can help to monitor and further improve ASR performance. Clinicians can use benchmarks based on such metrics to counsel prospective users and may benefit from automated procedures. Several hearing impaired listeners, especially CI recipients, report that they benefit from the apps in certain situations (47), which is in accordance with the results of converting a dialogue into text and may stem from complementary error patterns of ASR not investigated here. Personalized ASR could increase the number of listeners enjoying the benefits of ASR.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

LP, J-WW, and PH conceptualized the study. LP, PH, and DG collected the data. LP and J-WW took the lead in drafting the manuscript. All authors contributed to the data interpretation, reviewed the results, and edited the manuscript.

ACKNOWLEDGMENTS

We thank David R. Moore, Paul J. Govaerts, Cas Smits, and J. Tilak Rathanather for their comments on draft versions of the article. From our institution, we thank Cris P. Lanting, Wendy J. Huinck, and Emmanuel A.M. Mylanus for their suggestions over the course of this study. Finally, we thank the reviewers for their comments.

REFERENCES

1. Saon G, Kurata G, Sercu T, Audhkhasi K, Thomas S, Dimitriadis D, et al. *English Conversational Telephone Speech Recognition by Humans and Machines*. ArXiv170302136 Cs (2017). Available online at: http://arxiv.org/abs/1703.02136 (accessed October 8, 2021).
2. Xiong W, Droppo J, Huang X, Seide F, Seltzer ML, Stolcke A, et al. Toward human parity in conversational speech recognition. *IEEEACM Trans Audio Speech Lang Process.* (2017) 25:2410–23. doi: 10.1109/TASLP.2017.2756440
3. Kader SE, Eckert AM, Gural-Toth V. Voice-to-text technology for patients with hearing loss. *Hear J.* (2021) 74:11–4. doi: 10.1097/01.HJ.0000734212.09840.d7
4. Wasmann J-WA, Lanting CP, Huinck WJ, Mylanus EAM, van der Laak JWM, Govaerts PJ, et al. Computational audiology: new approaches to advance hearing health care in the digital age. *Ear Hear.* (2021) 42:1499–507. doi: 10.1097/AUD.0000000000001041
5. Lesica NA, Mehta N, Manjaly JG, Deng L, Wilson BS, Zeng F-G. Harnessing the power of artificial intelligence to transform hearing healthcare and research. *Nat Mach Intell.* (2021) 3:840–9. doi: 10.1038/s42256-021-00394-z
6. Jurafsky D, Martin JH. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. 2nd ed*. Upper Saddle River, NJ: Pearson Prentice Hall (2009).
7. Koenecke A, Nam A, Lake E, Nudell J, Quartey M, Mengesha Z, et al. Racial disparities in automated speech recognition. *Proc*

Natl Acad Sci U S A. (2020) 117:7684–9. doi: 10.1073/pnas.1915768117
8. Cieri C, Miller D, Walker K. The fisher corpus: A resource for the next generations of speech-to-text. In: *Proceedings of the 4th International Conference on Language Resources and Evaluation, LREC 2004*. (2004). p. 69–71. Available online at: http://www.lrec-conf.org/proceedings/lrec2004/
9. Godfrey JJ, Holliman EC, McDaniel J. SWITCHBOARD: Telephone speech corpus for research and development. In: *Acoustics, Speech, and Signal Processing, IEEE International Conference on IEEE Computer Society* (San Francisco, CA). (1992). p. 517–20. doi: 10.1109/ICASSP.1992.225858
10. Panayotov V, Chen G, Povey D, Khudanpur S. Librispeech: an ASR corpus based on public domain audio books. In: *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. (2015). p. 5206–10. doi: 10.1109/ICASSP.2015.7178964
11. Kincaid J. *Which Automatic Transcription Service is the Most Accurate? — 2018*. Descript. (2018). Available online at: https://medium.com/descript/which-automatic-transcription-service-is-the-most-accurate-2018-2e859b23ed19 (accessed October 1, 2021).
12. Bernstein LE, Tucker PE, Demorest ME. Speech perception without hearing. *Percept Psychophys.* (2000) 62:233–52. doi: 10.3758/BF03205546
13. Helfer KS. Auditory and auditory-visual perception of clear and conversational speech. *J Speech Lang Hear Res JSLHR.* (1997) 40:432–43. doi: 10.1044/jslhr.4002.432
14. Coldewey D. Ava Expands its AI Captioning To Desktop And Web Apps, and Raises $4.5M to Scale. TechCrunch (2020). Available online at: https://social.techcrunch.com/2020/12/10/ava-expands-its-ai-captioning-to-desktop-and-web-apps-and-raises-4-5m-to-scale/ (accessed December 12, 2021).
15. Google. *How Google Technology Is Improving Accessibility For Deaf People - Google. About Google*. Available online at: https://about.google/intl/ALL_us/stories/making-conversation-more-accessible-with-live-transcribe (accessed December 11, 2021).
16. Earfy. *New Earfy Functions Based On User Feedback! - Earfy*. (2017). Available online at: https://www.earfy.net/earfy/new-functions-for-earfy-based-on-user-feedback/ (accessed December 11, 2021).
17. Mattys SL, Davis MH, Bradlow AR, Scott SK. Speech recognition in adverse conditions: a review. *Lang Cogn Process.* (2012) 27:953–78. doi: 10.1080/01690965.2012.705006
18. Gatehouse S, Naylor G, Elberling C. Benefits from hearing aids in relation to the interaction between the user and the environment. *Int J Audiol.* (2003) 42:77–85. doi: 10.3109/14992020309074627
19. International Organization for Standardization. ISO 8253-1: 2010. Acoustics: audiometric test methods. Part 1: Pure-Tone Air And Bone Conduction Audiometry. International Organization for Standardization Geneva (2010).
20. Bosman AJ, Smoorenburg GF. Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. *Audiology.* (1995) 34:260–84. doi: 10.3109/00206099509071918
21. Plomp R, Mimpen AM. Speech-reception threshold for sentences as a function of age and noise level. *J Acoust Soc Am.* (1979) 66:1333–42. doi: 10.1121/1.383554
22. Plomp R, Mimpen AM. Improving the reliability of testing the speech reception threshold for sentences. *Audiology.* (1979) 18:43–52. doi: 10.3109/00206097909072618
23. Smits C, Theo Goverts S, Festen JM. The digits-in-noise test: assessing auditory speech recognition abilities in noise. *J Acoust Soc Am.* (2013) 133:1693–706. doi: 10.1121/1.4789933
24. Bronkhorst AW, Bosman AJ, Smoorenburg GF. A model for context effects in speech recognition. *J Acoust Soc Am.* (1993) 93:499–509. doi: 10.1121/1.406844
25. World Health Organization. *World Report On Hearing*. (2021). Available online at: https://www.who.int/publications-detail-redirect/world-report-on-hearing (accessed April 1, 2021).
26. Dingemanse JG, Goedegebure A. The important role of contextual information in speech perception in cochlear implant users and its consequences in speech tests. *Trends Hear.* (2019) 23:2331216519838672. doi: 10.1177/2331216519838672
27. Kaandorp MW, Smits C, Merkus P, Goverts ST, Festen JM. Assessing speech recognition abilities with digits in noise in cochlear implant and hearing aid users. *Int J Audiol.* (2015) 54:48–57. doi: 10.3109/14992027.2014.945623
28. O'Neill ER, Parke MN, Kreft HA, Oxenham AJ. Development and validation of sentences without semantic context to complement the basic English lexicon sentences. *J Speech Lang Hear Res.* (2020) 63:3847–54. doi: 10.1044/2020_JSLHR-20-00174
29. Deng L. Deep learning: from speech recognition to language and multimodal processing. *APSIPA Trans Signal Inf Process*. (2016) 5. doi: 10.1017/ATSIP.2015.22
30. Kaandorp MW, De Groot AM, Festen JM, Smits C, Goverts ST. The influence of lexical-access ability and vocabulary knowledge on measures of speech recognition in noise. *Int J Audiol.* (2016) 55:157–67. doi: 10.3109/14992027.2015.1104735
31. Blamey P, Artieres F, Başkent D, Bergeron F, Beynon A, Burke E, et al. Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: an update with 2251 patients. *Audiol Neurotol.* (2013) 18:36–47. doi: 10.1159/000343189
32. Flynn MC, Dowell RC, Clark GM. Aided speech recognition abilities of adults with a severe or severe-to-profound hearing loss. *J Speech Lang Hear Res.* (1998) 41:285–99. doi: 10.1044/jslhr.4102.285
33. Faber BM. Acoustical measurements with smartphones: Possibilities and limitations. *Acoust Today.* (2017) 13:10–6.
34. Jakovljević N, Janev M, Pekar D, Mišković D. Energy normalization in automatic speech recognition. In: *International Conference on Text, Speech and Dialogue*. Springer (2008). p. 341–7. doi: 10.1007/978-3-540-87391-4_44
35. Biadsy F, Weiss RJ, Moreno PJ, Kanevsky D, Jia Y. *Parrotron: An End-To-End Speech-To-Speech Conversion Model And Its Applications To Hearing-Impaired Speech And Speech Separation. ArXiv Prepr ArXiv190404169* (Graz). (2019). doi: 10.21437/Interspeech.2019-1789
36. Christensen JH, Saunders GH, Havtorn L, Pontoppidan NH. Real-world hearing aid usage patterns and smartphone connectivity. *Front Digit Health.* (2021) 3:722186. doi: 10.3389/fdgth.2021.722186
37. Festen JM, Plomp R. Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J Acoust Soc Am.* (1990) 88:1725–36. doi: 10.1121/1.400247
38. Francart T, Van Wieringen A, Wouters J. Comparison of fluctuating maskers for speech recognition tests. *Int J Audiol.* (2011) 50:2–13. doi: 10.3109/14992027.2010.505582
39. Glasser A, Kushalnagar K, Kushalnagar R. Deaf, hard of hearing, and hearing perspectives on using automatic speech recognition in conversation. In: *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*. New York, NY, USA: Association for Computing Machinery (2017). p. 427–32. (ASSETS'17). Available online at: doi: 10.1145/3132525.3134781 (accessed September 20, 2021).
40. Initiative (WAI) WWA. *Home. Web Accessibility Initiative* (WAI). Available online at: https://www.w3.org/WAI/ (accessed December 16, 2021).
41. Yi H, Pingsterhaus A, Song W. Effects of wearing face masks while using different speaking styles in noise on speech intelligibility during the COVID-19 pandemic. *Front Psychol.* (2021) 12:682677. doi: 10.3389/fpsyg.2021.682677
42. Kim S, Arora A, Le D, Yeh C-F, Fuegen C, Kalinli O, et al. *Semantic Distance: A New Metric For Asr Performance Analysis Towards Spoken Language Understanding. ArXiv Prepr ArXiv210402138.* (2021). doi: 10.21437/Interspeech.2021-1929
43. Miner AS, Haque A, Fries JA, Fleming SL, Wilfley DE, Wilson GT, et al. Assessing the accuracy of automatic speech recognition for psychotherapy. *NPJ Digit Med.* (2020) 3:1–8. doi: 10.1038/s41746-020-0285-8
44. Huang W-C, Hayashi T, Wu Y-C, Kameoka H, Toda T. *Voice Transformer Network: Sequence-to-Sequence Voice Conversion Using Transformer with Text-to-Speech Pretraining. ArXiv191206813 Cs Eess*. (2019). Available online at: http://arxiv.org/abs/1912.06813 [accessed December 11, 2021).
45. Knecht HA, Nelson PB, Whitelaw GM, Feth LL. Background noise levels and reverberation times in unoccupied classrooms: predictions and measurements. *Am J Audiol.* (2002) 11:65–71. doi: 10.1044/1059-0889(2002/009)
46. Desarnaulds V, Carvalho AP, Monay G. Church acoustics and the influence of occupancy. *Build Acoust.* (2002) 9:29–47. doi: 10.1260/135101002761035726
47. Berenger M. *Hearing Australia. New App From National Acoustic Laboratories Improves Communication At Hearing Health Clinics*. Available online at:

https://www.hearing.com.au/About-Hearing-Australia/Hearing-news/New-app-from-National-Acoustic-Laboratories-improv (accessed December 23, 2021).

48. Venail F, Legris E, Vaerenberg B, Puel J-L, Govaerts PJ, Ceccato JC. Validation of the French-language version of the OTOSPEECH automated scoring software package for speech audiometry. *Eur Ann Otorhinolaryngol Head Neck Dis.* (2016) 133:101–6. doi: 10.1016/j.anorl.2016.01.001

49. Smith M, Cunningham KT, Haley KL. Automating error frequency analysis via the phonemic edit distance ratio. *J Speech Lang Hear Res.* (2019) 62:1719–23. doi: 10.1044/2019_JSLHR-S-18-0423

50. Wasmann J, Pragt L, Eikelboom R, Swanepoel DW. Digital approaches to automated and machine learning assessments of hearing: scoping review. *J Med Internet Res*. (2022) 24:e32581. doi: 10.2196/32581

51. Ratnanather J, Wang L, Bae S-H, O'Neill E, Sagi E, Tward D. Visualization of speech perception analysis via phoneme alignment: a pilot study. *Front Neurol.* (2021) 12:724800. doi: 10.3389/fneur.2021.724800

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.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 21 February 2022
doi: 10.3389/fnins.2022.779048

OPRA-RS: A Hearing-Aid Fitting Method Based on Automatic Speech Recognition and Random Search

Libio Gonçalves Braz[1†], Lionel Fontan[2*†], Julien Pinquier[1], Michael A. Stone[3] and Christian Füllgrabe[4]*

[1] IRIT, CNRS, Université Paul Sabatier, Toulouse, France, [2] Archean LABS, Montauban, France, [3] Manchester Centre for Audiology and Deafness, School of Health Sciences, University of Manchester, Manchester, United Kingdom, [4] School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, United Kingdom

Hearing-aid (HA) prescription rules (such as NAL-NL2, DSL-v5, and CAM2) are used by HA audiologists to define initial HA settings (e.g., insertion gains, IGs) for patients. This initial fitting is later individually adjusted for each patient to improve clinical outcomes in terms of speech intelligibility and listening comfort. During this fine-tuning stage, speech-intelligibility tests are often carried out with the patient to assess the benefits associated with different HA settings. As these tests tend to be time-consuming and performance on them depends on the patient's level of fatigue and familiarity with the test material, only a limited number of HA settings can be explored. Consequently, it is likely that a suboptimal fitting is used for the patient. Recent studies have shown that automatic speech recognition (ASR) can be used to predict the effects of IGs on speech intelligibility for patients with age-related hearing loss (ARHL). The aim of the present study was to extend this approach by optimizing, in addition to IGs, compression thresholds (CTs). However, increasing the number of parameters to be fitted increases exponentially the number of configurations to be assessed. To limit the number of HA settings to be tested, three random-search (RS) genetic algorithms were used. The resulting new HA fitting method, combining ASR and RS, is referred to as "objective prescription rule based on ASR and random search" (OPRA-RS). Optimal HA settings were computed for 12 audiograms, representing average and individual audiometric profiles typical for various levels of ARHL severity, and associated ASR performances were compared to those obtained with the settings recommended by CAM2. Each RS algorithm was run twice to assess its reliability. For all RS algorithms, ASR scores obtained with OPRA-RS were significantly higher than those associated with CAM2. Each RS algorithm converged on similar optimal HA settings across repetitions. However, significant differences were observed between RS algorithms in terms of maximum ASR performance and processing costs. These promising results open the way to the use of ASR and RS algorithms for the fine-tuning of HAs with potential speech-intelligibility benefits for the patient.

Keywords: random search (RS), automatic speech recognition (ASR), hearing aids (HAs), prescription rule, age-related hearing loss (ARHL), insertion gains, compression thresholds

OPEN ACCESS

Edited by:
Jing Chen,
Peking University, China

Reviewed by:
Yi-Wen Liu,
National Tsing Hua University, Taiwan
Xiang Li,
University of Cambridge, United Kingdom

****Correspondence:***
Libio Gonçalves Braz
libio.goncalves@irit.fr
Lionel Fontan
lfontan@archean.tech

† These authors share first authorship

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *17 September 2021*
Accepted: *13 January 2022*
Published: *21 February 2022*

Citation:
Gonçalves Braz L, Fontan L, Pinquier J, Stone MA and Füllgrabe C (2022) OPRA-RS: A Hearing-Aid Fitting Method Based on Automatic Speech Recognition and Random Search. Front. Neurosci. 16:779048. doi: 10.3389/fnins.2022.779048

1. INTRODUCTION

The aim of hearing-aid (HA) prescription rules is to provide an appropriate level of amplification to restore audibility to hearing-impaired (HI) listeners while avoiding uncomfortable loudness levels. Established prescription rules—such as NAL-NL2 (Keidser et al., 2011), DSL-v5 (Scollie et al., 2005), and CAM2 (initially referred to as CAMEQ2-HF and commercialized in two software variants, CAM2A, and CAM2B; Moore et al., 2010b)—incorporate theoretical models of speech intelligibility and loudness perception. The amount of signal amplification is determined by frequency-specific insertion gains (IGs). As people with age-related hearing loss (ARHL) show a reduced dynamic range due to the elevation of hearing thresholds and loudness recruitment, compression of the signal amplitude is applied. When the input level exceeds a given threshold (referred to as the compression threshold, CT), the amount of amplification applied by the HA decreases more or less abruptly depending on the compression ratio (CR), defined as the increase in input level required for a 1-dB increase in output level. To determine IGs, in addition to the audiogram, prescription rules additionally take into account the number of HA channels, the maximum CR, the CTs, and the compression speed.

Initial HA fittings based on prescription rules generally lead to satisfactory clinical outcomes (Moore, 2008), but do not allow to address specific needs of the individual patient (Søgaard Jensen et al., 2019). As a consequence, audiologists often have to adjust HA settings for a given patient over several visits. During this fine-tuning stage, the benefits of HA settings are usually assessed using speech intelligibility tests. Since these tests can be lengthy, their administration may occupy a considerable part of the consultation. Also, the patient's performance can be affected by fatigue and loss of motivation (Sorin and Thouin-Daniel, 1983). Finally, given that speech intelligibility is influenced by the familiarity with the speech material (Hustad and Cahill, 2003) and that most speech intelligibility tests are composed of a fairly small set of items, the number of HA settings that can be assessed is limited.

In an effort to address these issues, Fontan et al. (2020b) demonstrated that, when combined with ARHL simulation, automatic speech recognition (ASR) can be used to assess the speech intelligibility benefits of specific IG functions in older HI patients listening through simulated HAs. An ASR system was used to quantify the benefits in speech intelligibility associated with IGs systematically varied relative to the CAM2 prescription by 0, ±3, or ±6 dB. Single-word recordings were amplified by an HA simulator, and then processed to simulate two of the perceptual consequences of ARHL based on the patient's audiogram, namely the elevation of hearing thresholds and loudness recruitment (Nejime and Moore, 1997). Finally, the recordings were fed to an ASR system to compute speech-identification scores. The IG function yielding the highest ASR performance and CAM2 gains were implemented in a simulated HA. Higher human speech-identification scores and subjective ratings of speech pleasantness were observed when speech was amplified with the ASR-based IG functions. The method used to determine these IG functions was named OPRA, an acronym for Objective Prescription Rule based on ASR.

To reduce processing costs, Fontan et al. (2020b) used only a large stepsize to vary IGs across a limited range and within four frequency bands, while keeping compression parameters fixed. These choices might have limited the observed amount of benefits associated with OPRA.

The aim of the present study was to extend previous work by using a broader range of possible gain values in five frequency bands, and a smaller stepsize for the variation of IGs. In addition, not only IGs, but also CTs were optimized. Given the number of parameters and possible values, a systematic assessment of all possible HA configurations, as done by Fontan et al. (2020b), would have been computationally extremely costly (in the present study, 7.77×10^{18} possible HA configurations would have to be assessed). The solution adopted in the present study was to use a random search (RS) approach, testing only a subset of all possible HA configurations. RS algorithms can be applied to a wide range of optimization problems, using different approaches, such as tabu search, ant-colony optimization, cross-entropy, multi-start and clustering, or genetic algorithms (for an overview, see Blum and Roli, 2001). While these approaches differ in their search procedure, they have in common to use probabilities for the exploration of the search space. The basic idea of the genetic RS algorithms used in this study is to vary randomly the value of several variables (here, IGs and CTs), and to quantify the result of this selection on the outcome variable (here, ASR performance). This process is repeated for *N* iterations. After each iteration, the new result is compared to the previous result. In case of an improvement, the search range is centered around the new configuration. From one iteration to the next, the search range is reduced by a constant factor. This results in the RS algorithm gradually converging on an optimal configuration. Three RS algorithms were tested: (i) one that tuned simultaneously CTs and IGs in all HA channels; (ii) one that tuned the CT and IGs for each HA channel, one after the other; and (iii) one that tuned in all HA channels first CTs, then IGs. As was done for OPRA (Fontan et al., 2020b), the current study exclusively focused on the optimization of speech intelligibility in quiet. It should however later be determined if the same method can be applied to speech in noise, and, in case of positive results, how it can interact with the signal-processing schemes currently developed to reduce the effects of background noise on speech intelligibility (for example, see the current Clarity challenge; Graetzer et al., 2021). More precisely, the current study aimed at addressing the following questions:

(i) As already observed for OPRA, does OPRA-RS yield higher ASR scores than CAM2?
(ii) Are there differences between the three RS algorithms in terms of ASR scores and speed (i.e., the number of iterations needed to reach a target ASR score)?
(iii) How reproducible are the outcomes of OPRA-RS, in terms of ASR scores, IG functions, and CTs?

2. MATERIALS AND METHODS

2.1. Description of the OPRA-RS Processing Chain

Figure 1 details the different components of the processing chain used to generate the OPRA-RS-based HA settings for a given input audiogram.

First, the RS algorithm (implemented in Python; for a further description, see section 2.2) randomly defines CTs for each of the five frequency channels of the HA simulator used later in the processing chain. The audiometric thresholds and CTs are then inputted to the CAM2B-v2 software (Cambridge Enterprise, 2014) for the calculation of CAM2 IGs. This software also requires information about the frequency ranges of the HA channels (here, 0.1–0.7 kHz, 0.7–1.4 kHz, 1.4–2.8 kHz, 2.8–5.6 kHz, and 5.6—8 kHz) and the maximum CR allowed in the HA simulator (here, 10). In the present study, CAM2B-v2 was configured for an experienced HA user wearing a completely-in-the-canal HA, and assuming that the reference microphone for real-ear measurements was positioned near the tragus.

Second, IGs are defined by CAM2 for the two input levels for speech of 65 and 85 dB SPL (referred to as IGSP65 and IGSP85, respectively) at 11 center frequencies (0.125, 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4, 6, and 8 kHz), and fed back to the RS algorithm.

Third, the RS algorithm defines, for each of the parameters to be tuned, the range of values to be explored. Default search ranges were defined at the initialization of the algorithm (for more details, see section 2.3.2). At each iteration, the algorithm centers the ranges around the values that yield the best ASR performance, reduces the search ranges (by a constant factor), and assigns random values to each parameter within these ranges. For each channel, the algorithm then calculates the CR based on the IGSP65 and IGSP85 at the channel center frequency, following the equation:

$$\mathrm{CR} = \frac{\Delta \mathrm{input}}{\Delta \mathrm{output}} = \frac{85 - 65}{85 + \mathrm{IGSP85} - (65 + \mathrm{IGSP65})} \tag{1}$$

Next, the algorithm checks that the CR falls within the range 1–10 (for more details, see section 2.3.2.3). If that is not the case, the following adjustments are applied until the CR falls within the desired range:

- If $\mathrm{CR} < 0$, IGSP85s are increased by 0.5 dB;
- If $0 \leq \mathrm{CR} < 1$, IGSP65s are increased by 0.5 dB;
- If $\mathrm{CR} > 10$, IGSP65s are decreased by 0.5 dB.

Then, the Speech Intelligibility Index (SII; American National Standards Institute, 1997) is calculated for the input audiogram and speech amplified according to (i) IGSP65s recommended by CAM2 and (ii) IGSP65s modified by OPRA-RS. The difference between the two SII values, ($SII_{\mathrm{CAM2}} - SII_{\mathrm{OPRA\text{-}RS}}$), is calculated. Depending on the sign of the difference, the current OPRA-RS IGs are either increased or decreased by 0.1 dB, and the difference in SII values is re-calculated. This process is repeated until the difference cannot be further reduced. This adjustment was implemented as the ASR system used by OPRA-RS normalizes the input signal, which means that its performance is not affected by changes in overall level. Only non-linear amplification (i.e., amplification that modifies the shape of the speech spectrum) impacts the ASR system's ability to recognize spoken words. As a consequence, using only ASR performance to guide the search for the best HA configuration might lead to a setting that would be inappropriate for actual patients in terms of audibility or loudness. The systematic SII-based adjustments of OPRA-RS IGs ensure that the audibility of the amplified speech is close to that of speech amplified according to CAM2, which aims at maximizing speech audibility while taking into account the overall loudness of processed speech (Moore et al., 2010b).

Fourth, an HA simulator (described in Moore et al., 2010a) is used to amplify 50 speech recordings of an adult male native-French speaker according to the HA settings received from the RS algorithm, for an input level of 65 dB SPL. Each recording consisted of the French definite article "le" followed by a disyllabic noun (e.g., "le parfum"—"the perfume.") The speech material corresponded to five ten-word lists of the speech-intelligibility test developed by Fournier (1951), and which is commonly used in France for speech audiometry (Rembaud et al., 2017). The HA simulator, which is implemented in MATLAB® (Mathworks, Natick, MA, USA), uses two dynamic range compressors implemented in series (the second one acting as a limiter; for more details, see Fontan et al., 2020b) in each of the five following frequency channels: 0.1–0.7 kHz, 0.7–1.4 kHz, 1.4–2.8 kHz, 2.8–5.6 kHz, and 5.6–8 kHz.

Fifth, the HL simulator, developed by Nejime and Moore (1997) and implemented in MATLAB®, is used to degrade the amplified speech material. The speech level at the output of the HA was used as the input level for the HL simulator. The simulator mimicks two of the perceptual consequences of ARHL: elevation of hearing thresholds (achieved by using linear filtering) and loudness recruitment (achieved by raising the signal envelope to a power; Moore and Glasberg, 1993). Loss of frequency selectivity was not simulated as it has been shown to decrease the strength of the correlation between ASR and human identification scores for speech in quiet (Fontan et al., 2020a).

Sixth, an ASR system is used to assess the intelligibility of the (amplified and degraded) speech material. The system is based on the ASR engine Julius 4.4.2 (Lee and Kawahara, 2009), and uses Gaussian Mixture Models (GMMs) and Hidden Markov Models (HMMs). Its acoustic models were trained using the Hidden Markov Model toolkit (HTK, version 3.4.1; Young, 1994) on the corpora ESTER (Galliano et al., 2006) and ESTER 2 (Galliano et al., 2009), consisting of approximately 100 h of recordings of radio broadcast news. The language model used by the ASR system is a finite state grammar designed to recognize the 50 noun phrases (i.e., article and noun) used in the study. Since the same article "le" is used in all noun phrases, ASR performance is calculated based on the recognition of final words (i.e., nouns). For each recording, the five words with the highest log-likelihood (a measure of the goodness-of-fit of the acoustic and language models to the speech signal) are returned by the ASR system. If the target word is included in the list, it is considered as recognized by the ASR system. Based on the processing of all recordings, two performance measures are computed: (i) the ASR score, which corresponds to the percentage of recognized

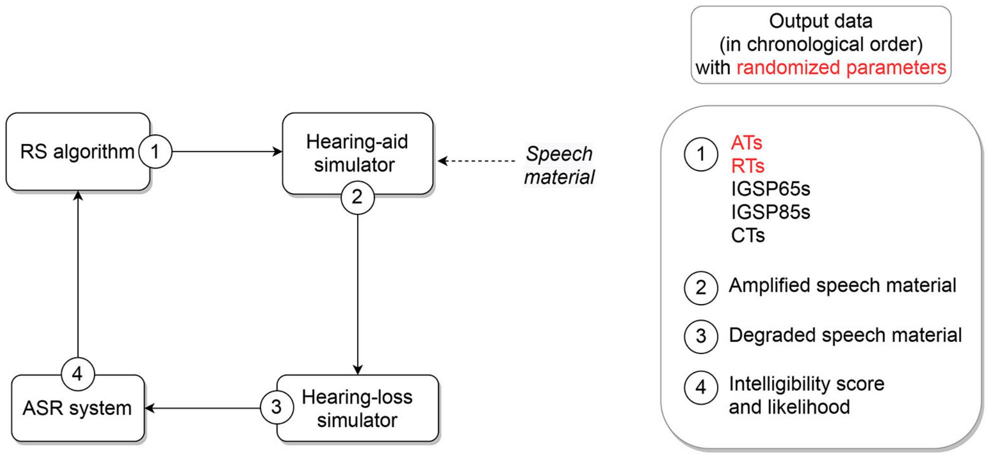


FIGURE 1 | Overview of the components of the OPRA-RS processing chain and associated output data. HA parameters randomized by the RS algorithm appear in red.

words, and (ii) the average log-likelihood of all words that were recognized by the ASR system.

In summary, the input/output function of the whole OPRA-RS processing chain takes as its only input the audiogram. Three outcome HA settings are eventually provided by the system: the IGSP65s, IGSP85s, and CTs associated with the highest ASR performance.

2.2. Description of the Random-Search Algorithms

The present study used three genetic RS algorithms which are based on biologically-inspired operators, such as mutations (i.e., random variations) and selection (Goldberg, 1989). One algorithm simultaneously tuned CTs and IGs in all HA channels. For this algorithm, the number of possible HA configurations is very large, and this might compromise the convergence on optimal settings. Hence, two other algorithms were designed to reduce the number of possible HA configurations. The second algorithm simultaneously tuned the CT and IGs for one HA channel at a time, from channel 1 to channel 5. The final algorithm tuned, in all HA channels, first CTs and then IGs. Each of the three algorithms used a total of 750 iterations to optimize HA settings. Preliminary tests with the first algorithm showed that this number of iterations was sufficient to achieve near-ceiling ASR performance. The three algorithms, referred to as GEN1, GEN2, and GEN3, are further described in the following sections.

2.2.1. Description of GEN1

The optimization process implemented in GEN1 is shown in **Figure 2**. In order to test a larger number of configurations, GEN1 is composed of four independent search threads, run in parallel. At initialization, each thread selects one HA configuration by assigning random values (picked from a uniform distribution) to all CTs and IGs. The IGs of this random configuration are then adjusted according to the constraints in terms of CR and SII. The configuration is sent to the other components of the processing chain (HA and HL simulators and ASR system in **Figure 1**). The ASR performance is returned to the RS thread and used together with the associated configuration as the baseline for the iterative search of the optimal configuration.

At each iteration i and for each parameter P to be tuned, a variation in the range from $-\Delta_P$ to Δ_P is randomly selected and applied to P. The value of Δ_P is initialized to half of the search range and decreased after each iteration following the equation:

$$\Delta_{P_{i+1}} = \Delta_{P_i} - \frac{\text{initial } \Delta_P - \text{stepsize}_P}{750} \tag{2}$$

where stepsize_P corresponds to the step used to vary P (for details, see section 2.3.2). Following this equation, the final search range (i.e., the range explored during the 750th iteration) will equal stepsize_P. The new configuration is then assessed. In the case where it yields a higher ASR score than the best configuration found so far, it becomes the new baseline for the application of random variations. Otherwise, a rollback is applied to return to the best configuration. In case that both configurations yield identical ASR scores, the log-likelihoods associated with the ASR scores are used to decide which configuration is retained. This process is repeated for 750 iterations in each of the four threads, and the best final configuration across threads is retained.

2.2.2. Description of GEN2

In contrast to GEN1, GEN2 tunes the CT and IGs for each HA channel one after the other (**Figure 3**). For the optimization of CT and IGs in a given channel, four threads are run in parallel. At initialization, random values are assigned by each thread to all parameters and, following CR and SII adjustments, the resulting HA configuration is evaluated. Contrary to GEN1, GEN2 uses the best configuration found across all threads as the baseline for the next iteration.

At each iteration, the search range for each parameter is decreased around the best value (i.e., the value yielding highest

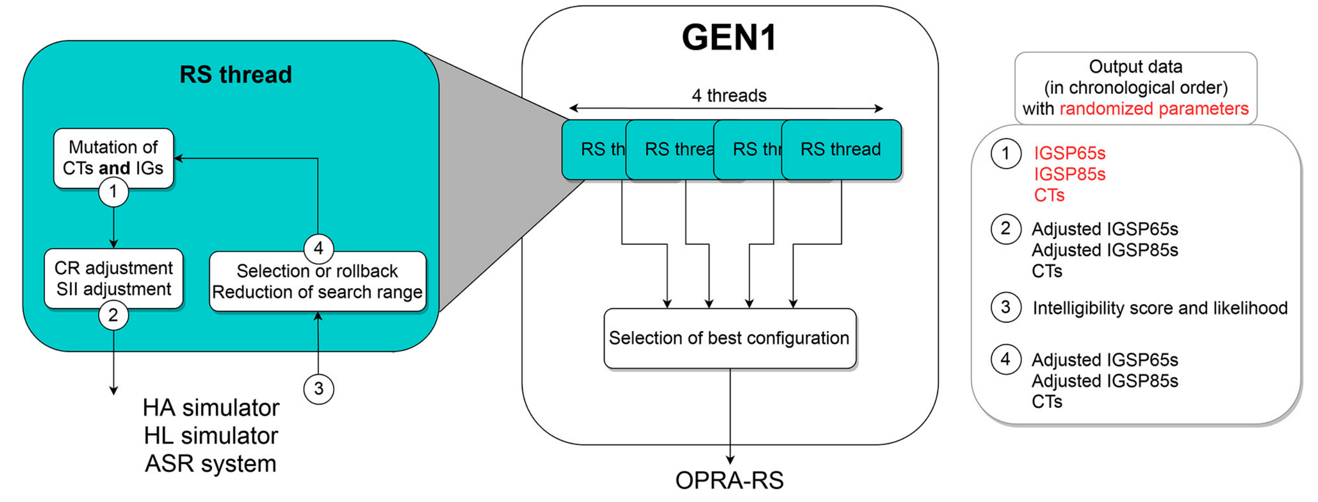


FIGURE 2 | Schematic representation of GEN1, which simultaneously optimizes CTs and IGs in all HA channels.

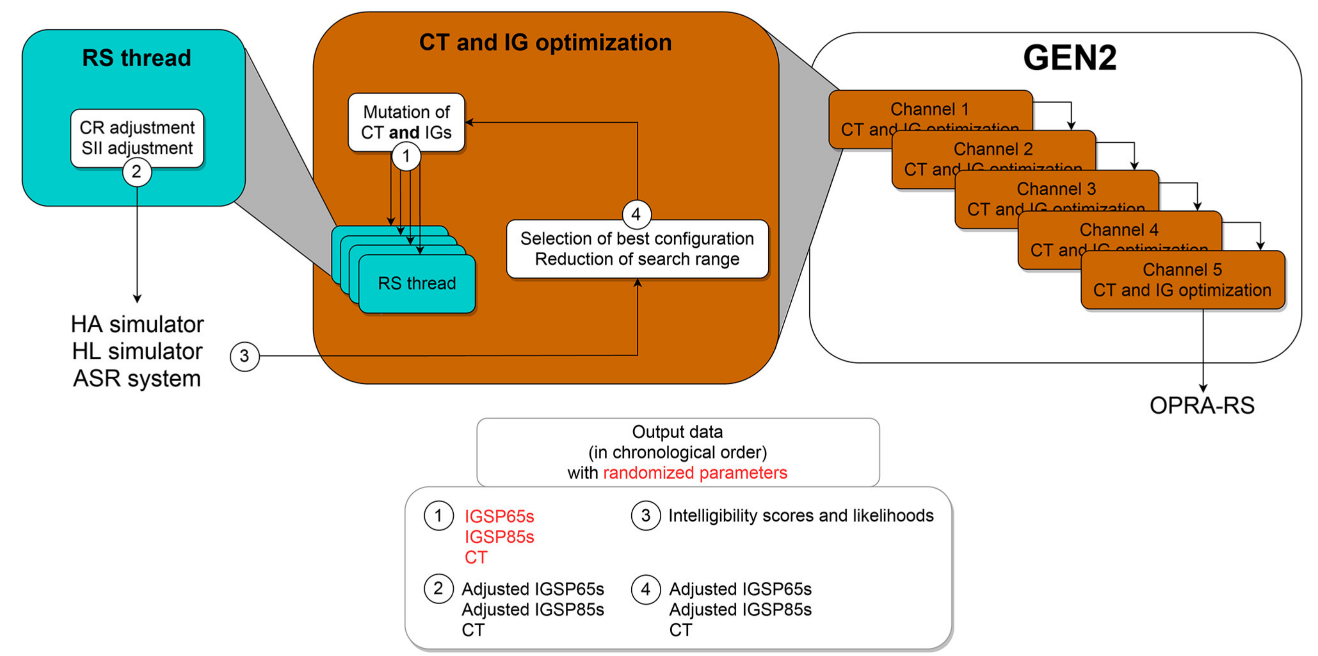


FIGURE 3 | Schematic representation of GEN2, which simultaneously optimizes the CT and IGs for one HA channel at the time, one after the other.

ASR performance) found so far. This new search range is then [best value $-\ \Delta_{P_i}$; best value $+\ \Delta_{P_i}$] with:

$$\Delta_{P_i} = (750 - i - 1) \times \frac{\text{half-range}_P}{750} \tag{3}$$

with half-range$_P$ corresponding to the initial search range used for the parameter P, divided by two.

Once again, random variations are applied to all parameters within the target search ranges, and the resulting configurations are evaluated. In the case that the configuration yielding the highest ASR performance across all threads is better than the best configuration found up to the current iteration, it is used as a baseline for the next iteration. This process is repeated sequentially for 150 iterations in each of the five channels and the best final configuration across threads is selected.

2.2.3. Description of GEN3

Similar to GEN1, GEN3 tunes parameters in all channels at the same time. However, GEN3 tunes CTs and IGs

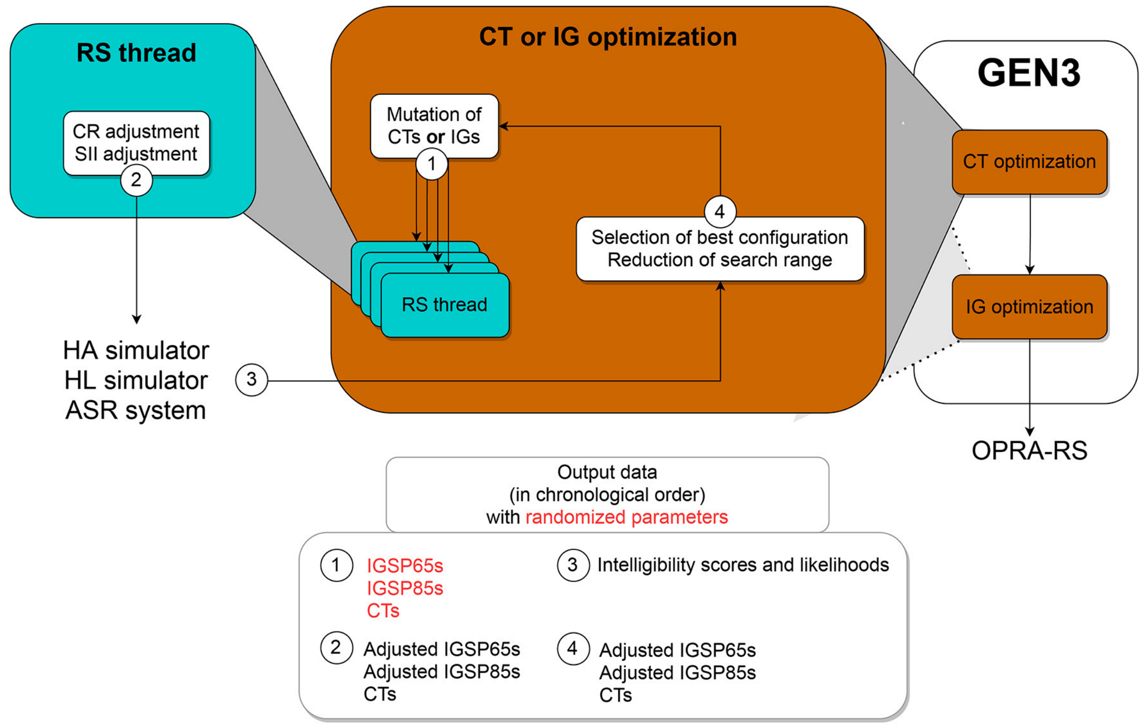


FIGURE 4 | Schematic representation of GEN3, which optimizes first CTs and then IGs in all HA channels.

sequentially (first CTs in all channels, then IGs in all channels; see **Figure 4**).

Four threads are used for the search. As for the initialization of GEN1 and GEN2, at start each thread of GEN3 assigns random values to all parameters. After adjusting the IGs of this random configuration according to the constraints in terms of CR and SII, the configuration is evaluated. If it yields better ASR performance than the best configuration found so far, then it is retained and used by all threads as the new baseline for the search. Otherwise, a rollback is applied. After each iteration, the search range is reduced using the same equation than in GEN2 (see Equation (3)). This process is repeated 250 times for the tuning of CTs, and then 500 times for the tuning of IGs.

2.3. Experiment Protocol

2.3.1. Input Audiograms

Figure 5 shows the four mean and eight individual audiograms that were used in the study. The mean audiograms correspond to the audiometric data reported by Humes (2021) for levels 4 to 7 of the Wisconsin Age-Related Hearing Impairment Classification Scale (WARHICS; Cruickshanks et al., 2020). As Humes (2021) did not report all hearing thresholds required for the HA and HL simulations, the missing values at frequencies of 0.125, 0.25, 0.75, and 1.5 kHz were intra- or extrapolated using 3rd-order least-squares polynomial fits.

The eight individual audiograms were selected based on them complying with the maximum frequency-specific threshold values defined by the WARHICS scale for levels 4 to 7. Two audiograms were selected for each of the four WARHICS levels. The individual audiograms came from eight older patients diagnosed with sensorineural HL (mean age: 70 years; age range: 63–78 years). Audiograms corresponding to different levels of HL severity were used in order to verify that OPRA-RS works for a wide range of HLs. Indeed, Fontan et al., 2020a observed that ASR performance is reduced for simulated severe-to-profound HLs. This might impact the search for the optimal HA configuration.

As can be observed in **Figure 5**, the mean audiograms show monotonic decline with increasing frequency, as is typical of ARHL. The individual audiograms follow the same general tendency, but are more erratic, which could impact the performance and outcome of the RS. The pure-tone average (PTA) of all individual audiograms bar one is identical to, or higher than the mean audiogram corresponding to the same WARHICS level.

2.3.2. Random-Search Ranges and Variation Stepsizes

2.3.2.1. Compression Thresholds

For the variation of CTs, 1-dB steps were used, and the lower and upper limits of the search range were 20 and 50 dB SPL, respectively. These choices were based on values used in previous studies implementing HA settings recommended by CAM2 in the same HA simulator as used here (Moore et al., 2010a, 2011; Moore and Sek, 2016).

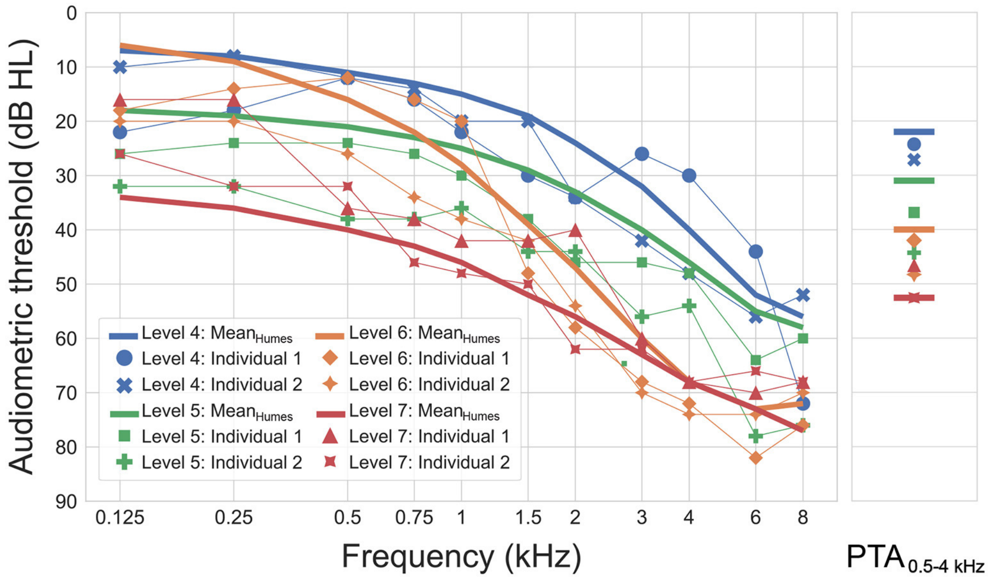


FIGURE 5 | Individual and mean audiograms for different WARHICS levels used as input to the OPRA-RS processing chain (left panel). Pure-tone average (PTA) for frequencies between 0.5 to 4 kHz are shown for each audiogram in the right panel.

2.3.2.2. Insertion Gains

A step size of 0.1 dB was used to vary IGs. The lower and upper limits for OPRA-RS IGs were set to ±10 dB relative to the IGs prescribed by CAM2. This choice was based on previous studies that compared initial prescriptions to final HA settings (i.e., after fine-tuning based on self-adjusted user preferences). Søgaard Jensen et al. (2019) reported that IGs after fine-tuning differed from initially applied IGs by ±9.6 dB. Similar differences (of ±10 dB) were observed by Boothroyd and Mackersie (2017).

2.3.2.3. Compression Speed and Maximum Compression Ratio

For people with severe HLs, it may be theoretically useful to use high CRs (> 3) to restore audibility at a comfortable level (Moore, 2008; Moore and Sek, 2016). In the case of fast compression, such high CRs can lead to a loss of intelligibility due to distortions of the signal envelope (Verschuure et al., 1996; Souza, 2002). As this study included audiograms corresponding to mild-to-moderately-severe HLs, only slow compression speeds were used, with CRs allowed to vary from 1 to 10, as done by Moore and Sek (2016). More precisely, for the tuning of CTs and IGs, attack times (ATs) were set to 200, 100, 100, 100, and 100 ms, and release times (RTs) to 2000, 1500, 1200, 1000, and 1000 ms, for HA channels 1 to 5, respectively.

2.3.3. Procedure

The OPRA-RS processing chain was run on the OSIRIM platform (http://osirim.irit.fr/site/en), a cluster of 928 central processing units and 28 graphical processing units. Six Intel® Xeon® Gold 6136 processors were used for the computation of OPRA-RS settings. Each RS algorithm was run twice to optimize CTs and IGs for each of the 12 input audiograms. A single run of any of the algorithms required an average processing time of 38 h.

2.3.4. Statistical Analyses

The significance of overall differences in data distribution were investigated through linear mixed models, followed by paired comparisons. T-tests were used, except when the assumptions for parametric tests were not met, in which case Wilcoxon tests were used. In case of multiple comparisons, the Holm-Bonferroni correction was applied. To investigate the association between PTA and ASR scores or convergence speed, Spearman correlations were used since data were not normally distributed. Data from the two repetitions of the RS algorithms were used to assess the reproducibility of OPRA-RS outcomes; for all other analyses, only data from the first repetition were used. Only IGs specified for a 65-dB-SPL speech input level were used, as this was the level set in the HA simulator. For these analyses R (R Core Team, 2021), with *lmerTest* and *emmeans* packages, and SPSS Statistics version 23 (IBM, Chicago, IL), were used.

3. RESULTS

3.1. Comparison of OPRA-RS and CAM2

To compare the ASR scores associated with OPRA-RS and CAM2, the optimal CTs selected by OPRA-RS algorithms were inputted to CAM2Bv2, as this software does not provide CTs. **Figure 6** shows the distribution of ASR scores for the 12 audiograms with the IGs selected by OPRA-RS or recommended by CAM2. In all conditions, OPRA-RS prescriptions yielded ASR scores in excess of 90%, with median values of 98% for GEN1 and 96% for GEN2 and GEN3. By comparison, ASR scores associated with CAM2 prescriptions are more broadly distributed, and their median values are lower (88% for GEN1, 86% for GEN2, and 94% for GEN3). Wilcoxon signed-rank tests show that the differences

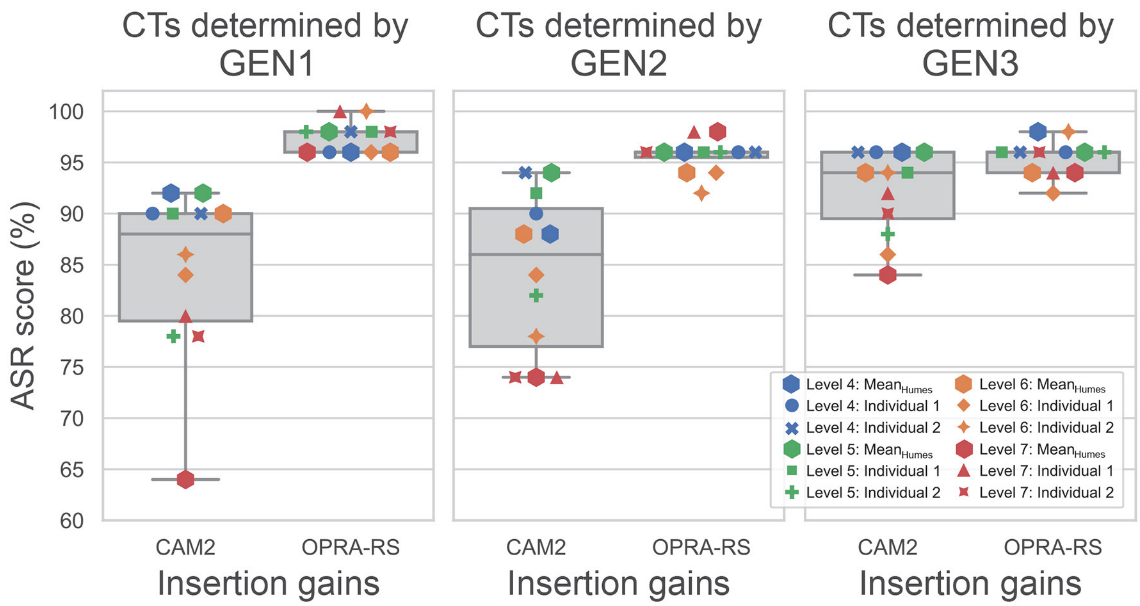


FIGURE 6 | Distribution of ASR scores based on IGs selected by OPRA-RS or recommended by CAM2, for each of the three RS algorithms. The horizontal lines inside the boxes represent median ASR scores. Whiskers and horizontal limits of the boxes represent, from bottom to top, the 0, 25th, 75th, and 100th percentiles. In the case of OPRA-RS, the medians and 75th percentiles overlap.

between ASR scores obtained with the two prescription rules (10 percentage points for GEN1 and GEN2, and 2 percentage points for GEN3) are statistically significant (for GEN1: $Z = 3.07; p = 0.002$; for GEN2: $Z = 3.06; p = 0.002$; for GEN3: $Z = 2.54; p = 0.011$).

As the same CTs were used by OPRA-RS and CAM2, the observed differences in ASR scores are due to the IGs prescribed by the two rules. **Figure 7** shows the IG functions averaged across audiograms, for OPRA-RS and CAM2, as a function of the type of RS algorithm. The IG functions for the two prescription rules are very similar (mean absolute difference of 1.7 dB), except for GEN1 for which, consistent with Fontan et al. (2020b), OPRA-RS tended to prescribe higher IGs (average absolute difference of 6.8 dB) for the low frequencies (0.125–0.5 kHz).

3.2. Influence of RS Algorithm Type, HL Severity, and Audiogram Type on CAM2

Differences between ASR scores associated with CAM2 can be observed across the three RS algorithms, with ASR scores higher for GEN3 than for GEN1 and GEN2. The statistical significance of these differences was assessed by a linear mixed model with the type of RS algorithm as a fixed effect and the input audiogram as a random effect. The results show that the type of RS algorithm is highly significant ($F(2, 22) = 16.7; p < 0.001$). Corrected *post-hoc* comparisons indicate highly significant differences between GEN1 and GEN3 ($t(22) = -5.0; p < 0.001$), as well as between GEN2 and GEN3 ($t(22) = -5.1; p < 0.001$). The difference between GEN1 and GEN2 is not significant ($t(22) = 0.1; p = 0.915$).

The lowest ASR scores are obtained for audiograms corresponding to level 7 of the WARHICS scale (i.e., the most severe HLs), whereas the highest ASR scores are obtained for audiograms corresponding to WARHICS levels 4 and 5. This suggests a negative association between HL severity and ASR scores obtained with CAM2. To assess the statistical significance of this relationship, Spearman's correlation was computed between ASR scores and the pure-tone average (PTA) for frequencies 0.5, 0.75, 1, 1.5, 2, 3, and 4 kHz. For all three RS algorithms, significant strong negative correlations are found (for GEN1: $\rho = -0.88; p < 0.001$; for GEN2: $\rho = -0.85; p < 0.001$; for GEN3: $\rho = -0.82; p = 0.001$).

Finally, ASR scores do not seem to be affected by the more erratic nature of individual audiograms. The distribution of ASR scores for mean and individual audiograms is rather similar, with lowest and highest scores observed in both cases.

3.3. Influence of RS Algorithm Type, HL Severity, and Audiogram Type on OPRA-RS

Only small differences are observed between ASR scores associated with OPRA-RS across the RS algorithms. Highest scores are found for GEN1 and only slightly lower scores were found for GEN2 and GEN3. The statistical significance of these differences was assessed using a linear mixed model with the type of RS algorithm as a fixed effect and the audiogram as a random effect. The results show that the type of RS algorithm has a significant effect on the ASR scores ($F(2, 22) = 5.6; p = 0.011$). Corrected *post-hoc* comparisons indicate significant differences between GEN1 and GEN2 ($t(22) = 2.8; p = 0.023$) and between GEN1 and GEN3 ($t(22) = 3.0; p = 0.020$). The difference

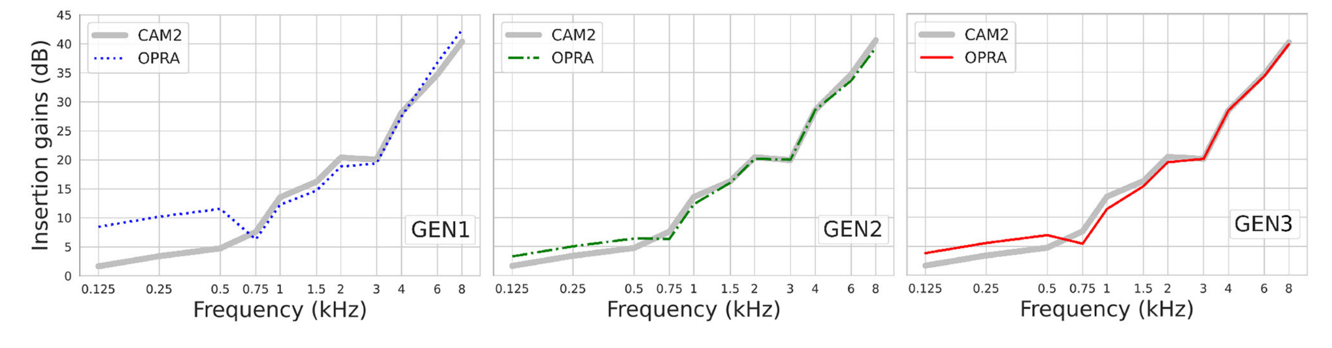


FIGURE 7 | IG functions prescribed by OPRA-RS and CAM2, averaged across audiograms.

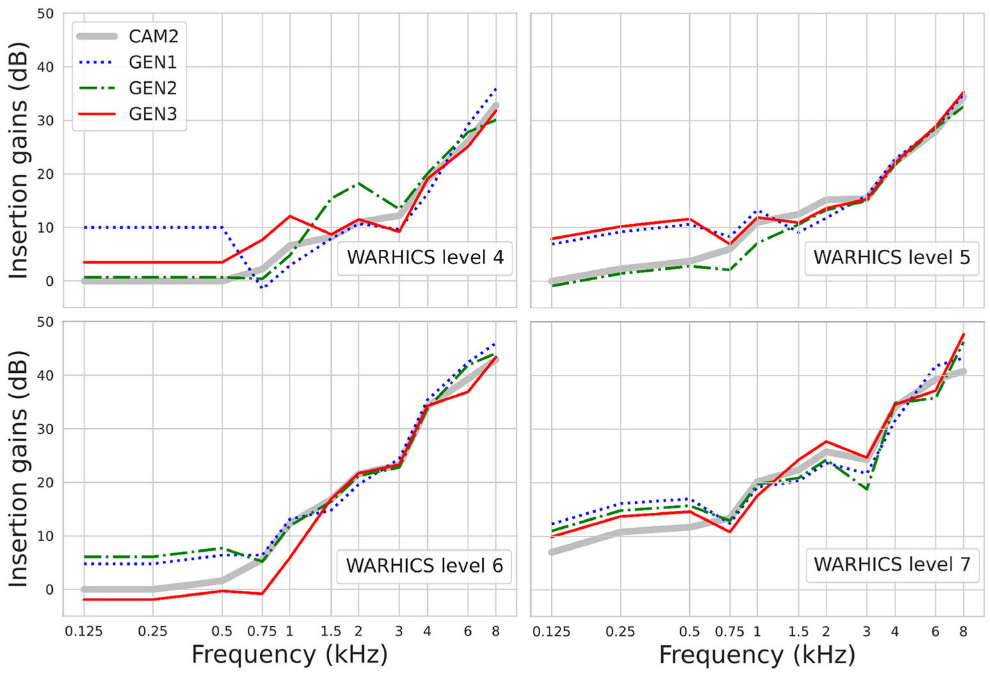


FIGURE 8 | Insertion gains prescribed by GEN1, GEN2, GEN3, and CAM2 using Humes (2021)'s average audiograms corresponding to WARHICS levels 4 to 7.

between GEN2 and GEN3 is not significant ($t(22) = 0.3$; $p = 0.81$).

Figure 8 shows the IGs prescribed by the three RS algorithms for Humes (2021)'s average audiograms corresponding to WARHICS levels 4 to 7. The largest difference observed between the IGs selected by the three RS algorithms is 9.3 dB. IG differences are often larger at low frequencies than at higher frequencies (see for example the IGs selected by GEN1 and GEN2 for the mean audiogram corresponding to WARHICS level 4). The results show that highest gains are not systematically prescribed by the same RS algorithm.

Spearman correlations do not indicate any association between ASR scores and PTA: for all three RS algorithms, correlation coefficients are weak (all $\rho \leq 0.35$) and non-significant (all $p \geq 0.28$).

As for CAM2, the distribution of ASR scores suggests that the more erratic nature of individual audiograms did not have a negative impact on the outcomes of the RS algorithms, with four out of the six highest ASR scores being associated with individual audiograms. This effect cannot be explained by better PTAs in individual audiograms, which are generally worse than the mean audiograms reported by Humes (2021, see right panel of **Figure 5**).

As ASR scores varied as a function of the input audiogram and the RS algorithm, the highest ASR score achieved by all three algorithms (ASR_{Common}) for the same audiogram was used to compare their speed of convergence. **Figure 9** shows the median and individual number of iterations needed by each algorithm to reach ASR_{Common} for each of the 12 audiograms. ASR_{Common} is sometimes achieved very early during the RS: for GEN1,

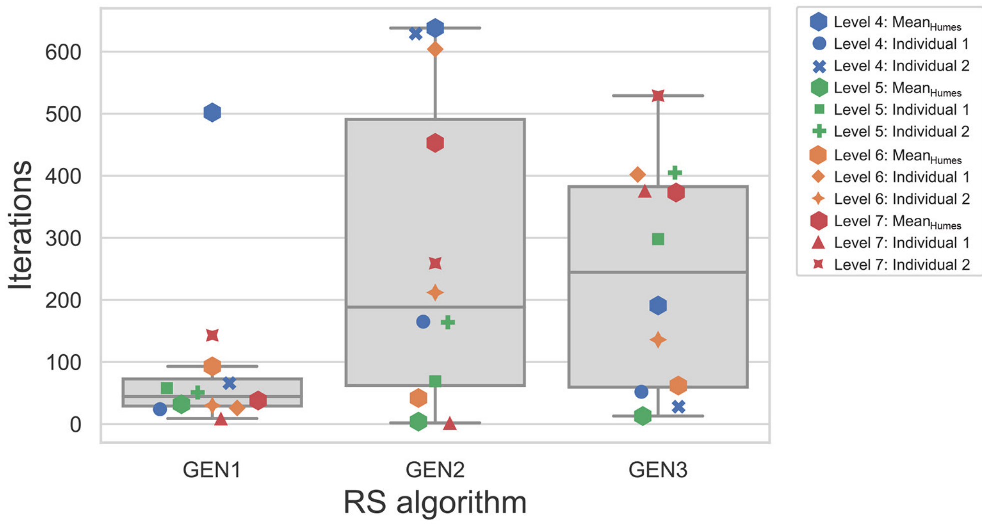


FIGURE 9 | Distribution of the number of iterations needed to reach the highest ASR score that was reached by the three RS algorithms. Otherwise as in **Figure 6**.

GEN2, and GEN3, the minimum number of iterations that were needed to achieve ASR_{Common} are 9, 2, and 13, respectively. GEN1 is generally faster and reaches, for ten out of the 12 audiograms, ASR_{Common} in less than 100 iterations. An outlier is however observed for the mean audiogram corresponding to WARHICS level 4, for which 504 iterations are used by GEN1. The convergence speed for GEN2 and GEN3 is more broadly distributed than for GEN1. A linear mixed model, with the type of RS algorithm as a fixed effect and the audiogram as a random effect, was used to assess the significance of these differences. The results indicate that the type of RS algorithm has a significant effect on convergence speed [$F(2, 22) = 3.6; p = 0.043$], and corrected *post-hoc* tests confirm that significant differences exist between GEN1 and GEN2 [$t(22) = -2.5; p = 0.039$] and between GEN1 and GEN3 [$t(22) = -2.1; p = 0.049$].

To assess the existence of an association between convergence speed and severity of HL (as measured by the PTA), Spearman correlations were computed. For GEN1 and GEN2, this relationship is not statistically significant (both $\rho \geq -0.16$, both $p \geq 0.63$). In contrast, a significant positive correlation is observed for GEN3 ($\rho = 0.65; p = 0.022$). The convergence speed does not seem to depend on the type of audiogram, as fast and slow convergence speeds are observed for both mean and individual audiograms.

3.4. Reproducibility of OPRA-RS Outcomes

The reproducibility of OPRA-RS outcomes was assessed in terms of ASR scores, as well as IGs and CTs, by comparing the outcomes of the two repetitions for each RS algorithm. ASR scores obtained after the second repetition of the RS algorithms (data not shown) are again very high, with median scores equal to those obtained during the first repetition: 98% for GEN1, and 96% for GEN2 and GEN3. Wilcoxon tests showed that there was no significant difference between ASR scores across repetitions for any of the RS algorithms (for GEN1: $Z = -1.3; p = 0.18$; for GEN2: $Z = 0.00; p = 1$; for GEN3: $Z = -0.22; p = 0.83$).

TABLE 1 | Pearson correlations for IG functions yielded by the two repetitions of each RS algorithm.

	GEN1	GEN2	GEN3
r_{min}	0.89	0.84	0.90
r_{max}	0.99	0.99	0.99
r_{mean}	0.97	0.95	0.97

For each algorithm, 12 correlations were computed (one correlation per audiogram). Minimum (r_{min}), maximum (r_{max}), and average (r_{mean}) correlation coefficients are reported.

Table 1 shows the Pearson correlation coefficients computed for IG functions outputted by the three RS algorithms after each repetition. For a given RS algorithm, each repetition yielded 12 IG functions (one for each audiogram), each composed of 11 frequency-specific IGs. Twelve correlations were thus computed for each RS algorithm, for a total of 36 correlations. Across RS algorithms, correlation coefficients range from 0.84 to 0.99 and are all highly significant (all $p \leq 0.002$). Average correlation coefficients are near perfect, ranging from 0.95 (GEN2) to 0.97 (GEN1 and GEN3), indicating that, for each RS algorithm, very similar IG functions were found from one repetition to the next. This consistency builds up gradually, with the correlation coefficient between the best IG function selected at the same point of the RS process for the two repetitions increasing as a function of iteration number (see **Supplementary Figure**),

Figure 10 shows the IGs prescribed by the three RS algorithms for Humes (2021)'s average audiograms corresponding to WARHICS levels 4 and 6. Results are reported for these audiograms as they correspond to the audiograms that, respectively, yielded the weakest and strongest correlation between IG functions found in each of the two repetitions.

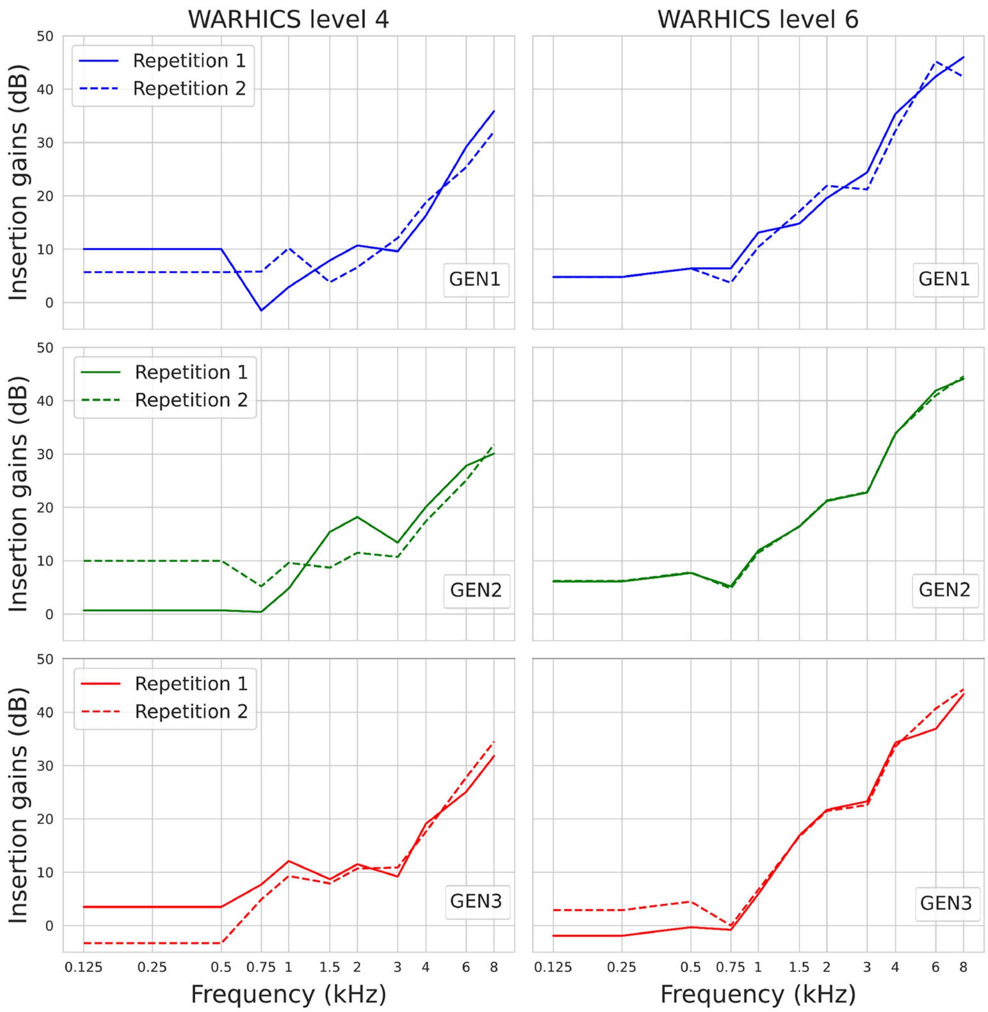


FIGURE 10 | IGs prescribed by OPRA-RS implementing each of the three RS algorithms (see rows) for Humes (2021)'s average audiograms corresponding to WARHICS levels 4 and 6 (see columns).

To assess to which extent the optimal IG functions found by each RS algorithm differ from the ones found by the other two RS algorithms, Pearson correlations were computed between the IG functions yielded by the different RS algorithms for each audiogram. Twelve correlations were computed for each pair of algorithms (i.e., one correlation per audiogram). Minimum, maximum, and mean correlation coefficients are shown in **Table 2**. All correlation coefficients are highly significant (all $p \leq 0.002$), and range from 0.82 to 0.99.

Figure 11 shows, for each RS algorithm, the absolute differences in CTs between the two repetitions as a function of HA channel. Across channels, median differences range from 2 to 12 dB, with a maximum individual difference of 27 dB.

TABLE 2 | Pearson correlation coefficients for IG functions obtained by the three RS algorithms for the 12 audiograms.

	GEN2	GEN3
GEN1	$r_{min} = 0.82$ $r_{max} = 0.99$ $r_{mean} = 0.95$	$r_{min} = 0.84$ $r_{max} = 0.99$ $r_{mean} = 0.95$
GEN2		$r_{min} = 0.89$ $r_{max} = 0.99$ $r_{mean} = 0.96$

For each pair of algorithms, minimum (r_{min}), maximum (r_{max}), and average (r_{mean}) correlation coefficients are reported.

4. DISCUSSION AND CONCLUSION

This study extends the previous work of Fontan et al. (2020b), in which only the IGSP65s recommended by CAM2 were varied in a very limited range of possible values and in four frequency bands to maximize ASR scores for different audiometric profiles. Here, the use of RS algorithms allowed to vary more parameters (namely, IGSP65s, IGSP85s, and CTs) in more (i.e., five) frequency bands and to use a broader range of possible IGs.

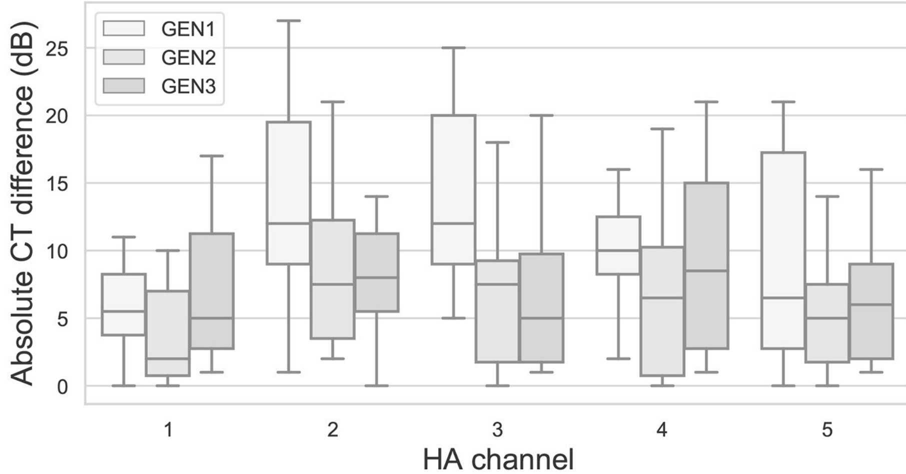


FIGURE 11 | Absolute differences in CTs between repetitions for the three RS algorithms as a function of HA channel. Otherwise as in **Figure 6**.

For the three RS algorithms that were used, the ASR scores yielded by optimized IGs were significantly higher than those obtained with the IGs recommended by CAM2. However, the observed differences were small, corresponding to the recognition of one (GEN3) to five (GEN1 and GEN2) out of the 50 words used in the study. It is possible that this small effect size is partly due to a ceiling effect. The observed benefits were indeed larger for the most severe HLs, for which CAM2 yielded lower ASR scores, therefore leaving more room for improvement. In future studies, the use of a larger set of speech stimuli, as well as of an ASR system with a larger lexicon (making the ASR system more prone to confusions), should be explored in order to avoid such ceiling effects.

The IGSP65s prescribed by CAM2 and OPRA-RS were very similar, most likely due to the SII-based equalization. The fact that, despite this similarity, OPRA-RS yielded higher ASR scores than CAM2 might indicate that the other parameters tuned by the RS algorithms (i.e., CTs and IGSP85s) were also determinant for the maximization of ASR scores.

The analysis of the association between PTA and ASR performances revealed that the ASR scores yielded by CAM2 were negatively impacted by HL severity. This might be explained by CAM2 prescriptions not being designed to restore fully audibility for severe cases of HL (Moore et al., 2010b). The fact that, when using OPRA-RS IGs, very high ASR scores were obtained even for the most severe HLs might be related to the fact that OPRA-RS used lower CRs than those recommended by CAM2 in such cases. An additional analysis conducted on OPRA-RS and CAM2 IGs confirms that OPRA-RS CRs were lower (by 1.7 points on average) than those defined by CAM2 for the audiograms corresponding to the most severe HLs used in the study (i.e., audiograms corresponding to the WARHICS level 7). Future research would therefore be warranted to check if OPRA-RS IGs do improve speech intelligibility for actual listeners with severe HLs (in which case the CRs recommended by CAM2 might be regarded as too high) or not (in which case one could put into question the representativeness, for the most severe audiograms, of the ARHL simulation used by OPRA-RS).

As, when simulating severe HLs, the ASR performance can be very low (Fontan et al., 2020a), it was hypothesized that the RS algorithms would need more iterations, and therefore maybe yield lower ASR scores in such cases. The results showed that only the convergence speed of GEN3 was affected by the severity of the simulated HL. As GEN3 only used the IGs prescribed by CAM2 during the first 250 iterations, it is possible that the ASR scores yielded by GEN3 remained very low at this stage, which would explain a slower convergence rate that with GEN1 and GEN2. Contrary to CAM2, the ASR scores yielded by all three RS algorithms were not affected by the HL severity.

Another hypothesis was that the convergence speed would be slower for individual audiograms, whose shape is more erratic than that of the mean audiograms, and that the ASR scores would maybe be lower too for this type of audiograms. This was not the case for any of the three RS algorithms.

GEN1 outperformed the two other RS algorithms in terms of ASR scores and convergence speed, and thus might be the best candidate for future investigations. It is possible that its performances could be further improved if after each iteration the best HA configuration identified across threads would be used as a baseline for the next iteration by all threads, instead of using independent threads.

Finally, the comparison of the outcomes of the RS algorithms across repetitions showed that ASR performance and optimized IGSP65s were reproducible. Some variability was however observed across the IGSP65s selected by the different algorithms, especially at lower frequencies ($\leq$ 0.5 kHz). The fact that, as already observed by Fontan et al. (2020b), ASR scores were impacted by variations in low-frequency IGs, should be interpreted with caution. Indeed, the ability of real HAs to apply

IGs in frequencies below 0.2 kHz is very limited, even when wearing ear molds, due to several reasons related to transducer coupling, background noise intrusion, and resulting upward spread of masking (Moore, 2008). It is thus possible that benefits in ASR scores induced by higher gains in lower frequencies would not translate into "real-life" situations. Contrary to the IGSP65s, the CTs showed a high variability across repetitions of the RS algorithms. This may indicate that in the present study the ASR scores were more impacted by IGs and CRs than by CTs, probably because a single presentation level of 65 dB SPL was used.

Taken together, the results obtained in the present study are encouraging and open the way to the use of ASR, combined with RS, to assess very large numbers of possible HA configurations, and to identify the settings yielding maximal speech intelligibility for specific individual audiometric profiles. In the previous study of Fontan et al. (2020b), a similar but simpler optimization chain was used and the optimized settings (which yielded improvements in ASR scores comparable to those yielded by GEN1 and GEN2) led to significant improvements of speech intelligibility and perceived quality in actual older HI persons. The improvements due to this new optimization chain, varying more parameters using smaller setpsizes, might thus be even higher. Moreover, as RS algorithms allow to test very large numbers of conditions, the present study could be extended in several regards. First, only one speech presentation level was used here. Future research should investigate the possibility to use OPRA-RS with speech presented at different levels (e.g., from 50 to 85 dB SPL); this would allow the use of OPRA-RS in real-life scenarios. The speech material should be diversified in order to be more representative of realistic situations (e.g., including speakers of different genders and ages, and different background noises). Finally, other HA parameters such as compression speed, that also impact speech intelligibility, could be investigated using OPRA-RS (see for example Fontan et al., submitted).

DATA AVAILABILITY STATEMENT

The datasets generated and analyzed for this study can be obtained from the corresponding authors for any research purpose.

AUTHOR CONTRIBUTIONS

LF initiated the idea. LG designed and implemented the random-search algorithms and all modifications applied to the previous OPRA prescription chain, under the supervision of JP. MS provided scientific advice about fitting algorithms, and the hearing-aid and hearing-loss simulations. LG, LF, MS, and CF analyzed and interpreted the data. LG, LF, and CF wrote the manuscript. All authors approved the final version of the manuscript.

ACKNOWLEDGMENTS

The experiments presented in this paper were carried out using the OSIRIM platform that is administered by the Institut de Recherche en Informatique de Toulouse (IRIT) and supported by the French National Center for Scientific Research (CNRS), the Occitanie Region, the French Government, and the European Regional Development Fund (ERDF). Author MS was funded by the NIHR Manchester Biomedical Research Centre (United Kingdom). The authors would like to thank A. A. Zarrabi and Dr. S. Jmel for their help with statistics. The authors are also indebted to V. De Fino and L. Cances for their help during the development and testing of OPRA-RS.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.779048/full#supplementary-material

REFERENCES

American National Standards Institute (1997). *American National Standard: Methods for Calculation of the Speech Intelligibility Index*. New York, NY: American National Standards Institute.

Blum, C., and Roli, A. (2001). Metaheuristics in combinatorial optimization: overview and conceptual comparison. *ACM Comput. Surveys* 35, 268–308. doi: 10.1145/937503.937505

Boothroyd, A., and Mackersie, C. (2017). A "Goldilocks" problem approach to hearing-aid self-fitting: User interactions. *Am. J. Audiol.* 26, 430–435. doi: 10.1044/2017_AJA-16-0125

Cambridge Enterprise (2014). *CAM2B-v2* [Software]. Cambridge: Cambridge Enterprise.

Cruickshanks, K. J., Nondahl, D. M., Fischer, M. E., Schubert, C. R., and Tweed, T. S. (2020). A novel method for classifying hearing impairment in epidemiological studies of aging: the Wisconsin Age-Related hearing impairment classification scale. *Am. J. Audiol.* 29, 59–67. doi: 10.1044/2019_AJA-19-00021

Fontan, L., Cretin-Maitenaz, T., and Füllgrabe, C. (2020a). Predicting speech perception in older listeners with sensorineural hearing loss using automatic speech recognition. *Trends Hear.* 24:2331216520914769. doi: 10.1177/2331216520914769

Fontan, L., Gonçalves Braz, L., Pinquier, J., Stone, M., and Füllgrabe, C. (submitted). Using automatic speech recognition to optimize hearing-aid compression speed. *J. Acoust. Soc. America* 148:EL227.

Fontan, L., Le Coz, M., Azzopardi, C., Stone, M., and Füllgrabe, C. (2020b). Improving hearing-aid gains based on automatic speech recognition. *J. Acoust. Soc. Am.* 148, EL227–EL233. doi: 10.1121/10.0001866

Fournier, J.-E. (1951). *Audiométrie Vocale : Les épreuves d'intelligibilité et leurs Applications au Diagnostic, à l'Expertise et à la Correction Prothétique des Surdités—[Speech Audiometry: Speech-Intelligibility Tests and Their Application for the Diagnosis, Survey and Hearing-aid Rehabilitation of Hearing Losses]*. (Paris: Maloine).

Galliano, S., Geoffrois, E., Gravier, G., Bonastre, J.-F., Mostefa, D., and Choukri, K. (2006). "Corpus description of the ESTER evaluation campaign for the rich transcription of French broadcast news," in *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC)* (Genoa: European Language Resources Association), 885–888.

Galliano, S., Gravier, G., and Chaubard, L. (2009). "The ESTER 2 evaluation campaign for the rich transcription of French radio broadcasts," in *INTERSPEECH 2009, 10th Annual Conference of the International Speech Communication Association* (Brighton: ISCA), 2583–2586.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. (New York, NY: Addison-Wesley).

Graetzer, S. N., Barker, J., Cox, T. J., Akeroyd, M., Culling, J. F., et al. (2021). "Clarity-2021 challenges: machine learning challenges for advancing hearing aid processing," in *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH)* (Brno), 686–690.

Humes, L. E. (2021). Further evaluation and application of the Wisconsin Age-Related hearing impairment classification system. *Am. J. Audiol.* 30, 359–375. doi: 10.1044/2021_AJA-20-00207

Hustad, K. C., and Cahill, M. A. (2003). Effects of presentation mode and repeated familiarization on intelligibility of dysarthric speech. *Am. J. Speech Lang. Pathol.* 12, 198–208. doi: 10.1044/1058-0360(2003/066)

Keidser, G., Dillon, H., Flax, M., Ching, T., and Brewer, S. (2011). The NAL-NL2 prescription procedure. *Audiol. Res.* 1:e24. doi: 10.4081/audiores.2011.e24

Lee, A., and Kawahara, T. (2009). "Recent development of open-source speech recognition engine Julius," in *Proceedings of the 2009 Annual Summit and Conference of the Asia-Pacific Signal and Information Processing Association (APSIPA ASC 2009)* (Sapporo), 131–137.

Moore, B. C. J. (2008). *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues*, 2nd Ed. Chichester: Wiley.

Moore, B. C. J., Füllgrabe, C., and Stone, M. A. (2010a). Effect of spatial separation, extended bandwidth, and compression speed on intelligibility in a competing-speech task. *J. Acoust. Soc. Am.* 128, 360–371. doi: 10.1121/1.3436533

Moore, B. C. J., Füllgrabe, C., and Stone, M. A. (2011). Determination of preferred parameters for multichannel compression using individually fitted simulated hearing aids and paired comparisons. *Ear Hear.* 32, 556–568. doi: 10.1097/AUD.0b013e31820b5f4c

Moore, B. C. J., and Glasberg, B. R. (1993). Simulation of the effects of loudness recruitment and threshold elevation on the intelligibility of speech in quiet and in a background of speech. *J. Acoust. Soc. Am.* 94, 2050–2062. doi: 10.1121/1.407478

Moore, B. C. J., Glasberg, B. R., and Stone, M. A. (2010b). Development of a new method for deriving initial fittings for hearing aids with multi-channel compression: CAMEQ2-HF. *Int. J. Audiol.* 49, 216–227. doi: 10.3109/14992020903296746

Moore, B. C. J., and Sek, A. (2016). Preferred compression speed for speech and music and its relationship to sensitivity to temporal fine structure. *Trends Hear.* 20:2331216516640486. doi: 10.1177/2331216516640486

Nejime, Y., and Moore, B. C. J. (1997). Simulation of the effect of threshold elevation and loudness recruitment combined with reduced frequency selectivity on the intelligibility of speech in noise. *J. Acoust. Soc. Am.* 102, 603–615. doi: 10.1121/1.419733

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rembaud, F., Fontan, L., and Füllgrabe, C. (2017). L'audiométrie vocale en France: état des lieux [Speech audiometry in France: current clinical practices]. *Cahiers de l'Audition* 6, 22–25.

Scollie, S., Seewald, R., Cornelisse, L., Moodie, S., Bagatto, M., Laurnagaray, D., et al. (2005). The desired sensation level multistage input/output algorithm. *Trends Amplific.* 9, 159–197. doi: 10.1177/108471380500900403

Søgaard Jensen, N., Hau, O., Bagger Nielsen, J. B., Bundgaard Nielsen, T., and Vase Legarth, S. (2019). Perceptual effects of adjusting hearing-aid gain by means of a machine-learning approach based on individual user preference. *Trends Hear.* 23:2331216519847413. doi: 10.1177/2331216519847413

Sorin, C., and Thouin-Daniel, C. (1983). Effects of auditory fatigue on speech intelligibility and lexical decision in noise. *J. Acoust. Soc. Am.* 74, 456–466.

Souza, P. E. (2002). Effects of compression on speech acoustics, intelligibility, and sound quality. *Trends Amplific.* 6, 131–165. doi: 10.1177/108471380200600402

Verschuure, J., Maas, A. J., Stikvoort, E., de Jong, R. M., Goedegebure, A., and Dreschler, W. A. (1996). Compression and its effect on the speech signal. *Ear Hear.* 17, 162–175. doi: 10.1097/00003446-199604000-00008

Young, S. J. (1994). *The HTK Hidden Markov Model Toolkit: Design and Philosophy*. Technical Report, Engineering Department, Cambridge University.

Conflict of Interest: This study is part of the development of a future product/service by Archean LABS intended for hearing-aid audiologists. CF acted as a scientific consultant.

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 interests.

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.

frontiers in Neuroscience

BRIEF RESEARCH REPORT
published: 17 March 2022
doi: 10.3389/fnins.2022.779062

Check for updates

Using Automatic Speech Recognition to Optimize Hearing-Aid Time Constants

Lionel Fontan[1†], Libio Gonçalves Braz[2*†], Julien Pinquier[2], Michael A. Stone[3] and Christian Füllgrabe[4]*

[1] Archean LABS, Montauban, France, [2] IRIT, CNRS, Université Paul Sabatier, Toulouse, France, [3] Manchester Centre for Audiology and Deafness, School of Health Sciences, University of Manchester, Manchester, United Kingdom, [4] School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, United Kingdom

OPEN ACCESS

Edited by:
Qinglin Meng,
South China University of Technology,
China

Reviewed by:
Jonathan Matthew Vaisberg,
The University of Western Ontario,
Canada
Zhenyu Guo,
South China University of Technology,
China

***Correspondence:**
Lionel Fontan
lfontan@archean.tech
Libio Gonçalves Braz
libio.goncalves@irit.fr

†These authors share first authorship

Specialty section:
This article was submitted to
Auditory Cognitive Neuroscience,
a section of the journal
Frontiers in Neuroscience

Received: 17 September 2021
Accepted: 14 February 2022
Published: 17 March 2022

Citation:
Fontan L, Gonçalves Braz L,
Pinquier J, Stone MA and Füllgrabe C
(2022) Using Automatic Speech
Recognition to Optimize Hearing-Aid
Time Constants.
Front. Neurosci. 16:779062.
doi: 10.3389/fnins.2022.779062

Automatic speech recognition (ASR), when combined with hearing-aid (HA) and hearing-loss (HL) simulations, can predict aided speech-identification performances of persons with age-related hearing loss. ASR can thus be used to evaluate different HA configurations, such as combinations of insertion-gain functions and compression thresholds, in order to optimize HA fitting for a given person. The present study investigated whether, after fixing compression thresholds and insertion gains, a random-search algorithm could be used to optimize time constants (i.e., attack and release times) for 12 audiometric profiles. The insertion gains were either those recommended by the CAM2 prescription rule or those optimized using ASR, while compression thresholds were always optimized using ASR. For each audiometric profile, the random-search algorithm was used to vary time constants with the aim to maximize ASR performance. A HA simulator and a HL simulator simulator were used, respectively, to amplify and to degrade speech stimuli according to the input audiogram. The resulting speech signals were fed to an ASR system for recognition. For each audiogram, 1,000 iterations of the random-search algorithm were used to find the time-constant configuration yielding the highest ASR score. To assess the reproducibility of the results, the random search algorithm was run twice. Optimizing the time constants significantly improved the ASR scores when CAM2 insertion gains were used, but not when using ASR-based gains. Repeating the random search yielded similar ASR scores, but different time-constant configurations.

Keywords: random search, automatic speech recognition, hearing aids, age-related hearing loss, compression speed, attack time, release time

INTRODUCTION

Recent studies have shown that automatic speech recognition (ASR) can be used, in combination with signal-processing algorithms mimicking the effects of hearing loss (HL), to predict the speech-identification performances of older hearing-impaired (OHI) listeners [see Schädler et al. (2015) and Fontan et al. (2020a) for a discussion of the advantages of ASR-based metrics by comparison to other objective measures of speech intelligibility]. This was demonstrated for unaided (Kollmeier et al., 2016; Schädler et al., 2018; Fontan et al., 2020a) and aided (using simulated or real hearing aids; Fontan et al., 2020b; Schädler et al., 2020) speech perception.

Based on these findings, it has been speculated that ASR-based prediction systems could also be used to assess speech-intelligibility benefits resulting from various hearing-aid (HA) configurations. Recently, Fontan et al. (2020c) used an ASR system to evaluate and to improve the insertion gains recommended by the CAM2 HA fitting rule (Moore et al., 2010b). For each of their hearing-impaired (HI) participants, 625 gain functions (corresponding to systematic variations of CAM2 gains by 0, ± 3, or ± 6 dB) were assessed. Each gain function was applied to speech stimuli using an HA simulator. The amplified speech material was then degraded using the HL simulator developed by Nejime and Moore (1997). Based on each participant's audiogram, both the elevation of hearing thresholds and loudness recruitment were mimicked. Spectral smearing, which is also implemented in the original HL simulator to mimic the loss of frequency selectivity, was not used used by Fontan et al. (2020c), since its simulation resulted in weaker correlations between ASR scores and human speech intelligibility (Fontan et al., 2020a). Finally, the amplified and degraded stimuli were fed to the ASR system for computing recognition scores. Fontan et al. (2020c) compared the benefits associated with the insertion-gain function yielding the highest ASR scores (the "optimized" gains yielding a mean improvement of 13 percentage points) to those obtained with CAM2 gains in a group of OHI participants. Significantly higher human speech-identification scores were observed for speech amplified with optimized gains than for speech amplified according to the gains recommended by CAM2. These significant improvements were observed both for word and sentence materials.

Gonçalves Braz et al. (2022) extended this work and combined ASR with several random-search (RS) algorithms to optimize not only insertion gains but also compression thresholds. This approach is referred to as OPRA-RS, which stands for "Objective Prescription Rule based on ASR and Random Search." Using slow time constants for the compressor of the simulated HA, optimized insertion gains and compression thresholds were determined for 12 audiometric profiles corresponding to different levels of HL severity. ASR scores yielded by the optimized parameters were significantly higher than those obtained with CAM2 (mean improvements ranged from 2 to 10 percentage points for the different RS algorithms). Significant differences were observed between RS algorithms in terms of ASR score and convergence speed.

A limitation of Gonçalves Braz et al.'s (2022) study is that only one set of time constants was used. However, aided speech intelligibility depends on the attack and release times of the HA compressor (Moore et al., 2011; Hopkins et al., 2012). Small time constants (i.e., "fast" compression) help perceiving rapid changes in loudness, such as those occurring when a weak speech sound (e.g., a consonant) precedes or follows a speech sound with higher energy (e.g., a vowel; Souza, 2002; Hopkins et al., 2012). At the same time, when fast compression is implemented in a multi-channel HA that processes each frequency channel independently, it tends to reduce spectral contrasts (i.e., by flattening the speech spectrum) and may thus have a deleterious effect on the perception of speech formants, which are crucial for the identification of vowels (Bor et al., 2008). By causing rapid variations of the signal amplitude at the onset and offset of speech sounds, fast compression speeds can also distort the signal envelope and therefore negatively impact speech intelligibility (Stone and Moore, 1992, 2008; Stone et al., 2009). These distortions are more likely to happen when high compression ratios are used (Verschuure et al., 1996). Despite their impact on speech intelligibility, there is currently no consensus as to the best time constants that should be used: time constants used clinically and commercially in hearing aids vary broadly, with attack and release times ranging from 0.5 to 2,000 ms and 10 to 5,000 ms, respectively (Moore and Sęk, 2016).

The present study extends the work of Gonçalves Braz et al. (2022) by investigating whether OPRA-RS can also be used to optimize time constants. Attack and release times were optimized for HA configurations that corresponded to the compression thresholds and/or the insertion gains recommended either by OPRA-RS or by CAM2 for the same 12 audiometric profiles as used in Gonçalves Braz et al. (2022). As these HA configurations sometimes involved high compression ratios (>3), and that in such cases, fast compression can distort the signal envelope and thus affect speech intelligibility (Souza, 2002), only "slow" compression speeds were used. To assess ASR performance, speech stimuli were first amplified using an HA simulator and then degraded to mimic the perceptual consequences of the elevation of hearing thresholds and loudness recruitment. The resulting speech signals were eventually fed to an ASR system for recognition. The optimization of compression speed was carried out twice in order to assess the reproducibility of the outcomes in terms of ASR scores and optimized time constants.

METHODS

Overview of the Optimization Chain

Figure 1 describes the processing chain used to optimize time constants for a given input audiogram. At initialization, the RS algorithm randomly selects attack and release times within two ranges of possible values. These time constants, as well as the compression thresholds and the insertion gains prescribed by OPRA-RS or CAM2 for a 65- and 85-dB-SPL speech input level (IGSP65s and IGSP85s, respectively; see **Supplementary Datasheet** for more details), are transmitted to an HA simulator. The HA simulator amplifies 50 speech stimuli corresponding to five 10-word lists of the speech intelligibility test of Fournier (1951), which is the test most often used by French audiologists for speech audiometry (Rembaud et al., 2017). The amplified speech signals are then degraded by the HL simulator, according to the input audiogram. The resulting speech signals are finally processed by an ASR system developed for the French language.

A total of *N* iterations are used to assess *N* time-constant configurations. After each iteration, the ASR score and average log-likelihood of recognized words yielded by the current time constants are compared to those obtained with the best time-constant configuration found up to the current iteration. If the

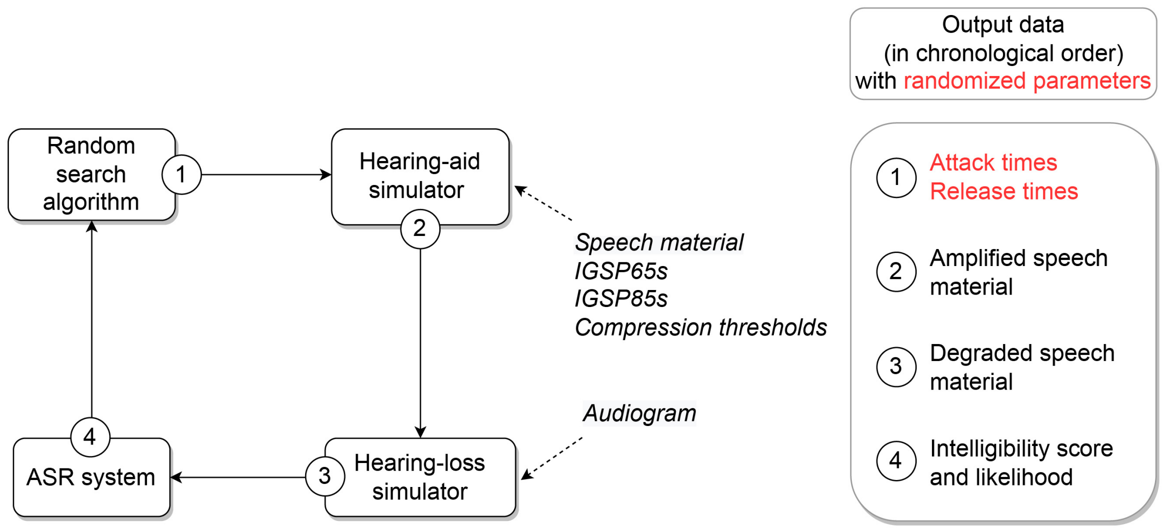


FIGURE 1 | Components of the OPRA-RS optimization chain, with associated input data (in italics) and output data (right panel). The parameters randomized by the RS algorithm are highlighted in red.

current configuration yields a higher ASR score (or the same ASR score but with a higher log-likelihood) than the previous best configuration, the current configuration is used as a baseline for the next iteration. Otherwise, the best previous configuration serves as a baseline for the next iteration.

Based on the current number of iterations *i*, the search ranges are reduced around the baseline time constants for the next iteration, following the equation:

$$\textit{search range}\ (i+1) = \textit{search range}\ (i) - \frac{\textit{initial search range} - (2 \times \textit{stepsize})}{N} \quad (1)$$

where *stepsize* corresponds to the step (in ms) used to define possible values within the search range.

Simulation of Hearing-Aid Processing

A 5-channel HA simulator implemented in MATLAB$^{\mathrm{TM}}$ (Moore et al., 2010a) was used to amplify the speech signals. The frequency ranges of the five HA channels were 0.1–0.7, 0.7–1.4, 1.4–2.8, 2.8–5.6, and 5.6–8 kHz. In each channel, the simulator used two dynamic range compressors placed in series: the wide dynamic range compression function was applied in the first compressor, while the second compressor was used as a limiter. For further details about the implementation of the HA simulator, see Fontan et al. (2020c).

Simulation of Hearing Loss

The functioning of the HL simulator, also implemented in MATLAB$^{\mathrm{TM}}$, is detailed in Nejime and Moore (1997). As done in Gonçalves Braz et al. (2022), the simulator was used to mimic two of the perceptual consequences of age-related HL: Based on the input audiogram, a linear filter simulated the elevation of hearing thresholds, while loudness recruitment was simulated by raising the signal envelope (Moore and Glasberg, 1993).

Automatic Speech Recognition System

The ASR system used in the study consisted of Hidden Markov Models and Gaussian Mixture Models. It was implemented using the Julius ASR engine (Lee and Kawahara, 2009). The acoustic models were trained on approximately 100 h of French radio broadcast news. These speech recordings were not processed to mimic HA amplification or HL and did not include the 50 word recordings used in the study to evaluate time constants. The lexicon used by the ASR system only comprised the 50 target words. A more detailed description of the ASR system is given in Gonçalves Braz et al. (2022).

Test Procedure

The processing chain was used to optimize time constants for 12 audiograms, using the compression thresholds selected by OPRA-RS and the insertion gains prescribed either by OPRA-RS or by CAM2. The audiograms, shown in **Figure 2**, represented mean or individual audiometric thresholds falling into levels 4–7 of the Wisconsin Age-Related Hearing Impairment Classification Scale (WARHICS; Cruickshanks et al., 2020). The audiograms corresponded to mild-to-moderately severe losses, with thresholds generally increasing as a function of frequency, as is typical of age-related HL. The mean audiograms were based on the data collected by Humes (2021). Some of the hearing thresholds required by the HL simulator (corresponding to the frequencies 0.125, 0.25, 0.75, and 1.5 kHz) were not included in the mean audiograms reported by Humes (2021). Those missing thresholds were intra- or extrapolated using third-least-squares polynomial regressions. The individual audiograms corresponded to older patients (mean age: 70 years; age range: 63–78 years) with sensorineural HL. For each of the four WARHICS levels, one mean audiogram and two individual audiograms were used.

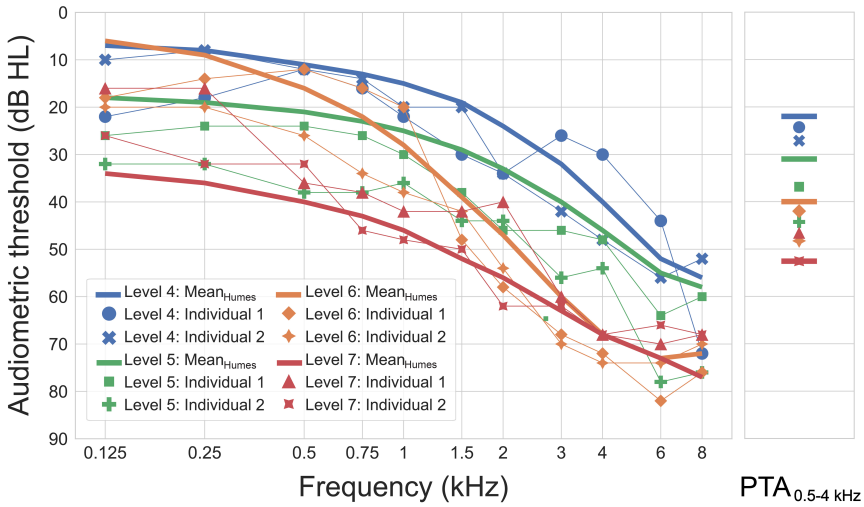


FIGURE 2 | Audiograms used as an input for the simulation of hearing loss. Corresponding pure-tone averages (PTAs) for frequencies between 0.5 and 4 kHz are shown in the right panel. Figure reproduced from Gonçalves Braz et al. (2022).

The RS algorithm that yielded the highest ASR performance in Gonçalves Braz et al. (2022) was used in the present study. This algorithm tunes all parameters (here, time constants) in all HA channels simultaneously. As in Gonçalves Braz et al. (2022), four independent RS threads were run in parallel. Each thread consisted of 1,000 iterations, during which time constants were randomly varied within predefined search ranges, using 10-ms steps. At the start of the RS, the search ranges were 10–500 ms for attack times, and 300–2,000 ms for release times. These ranges correspond to those generally associated with a slow compression system (Moore, 2008a,b; Moore et al., 2010a; Moore and Sęk, 2013). For each audiogram, the final time-constant configuration yielding the highest ASR performance across the four search threads was selected. In what follows, unless explicitly mentioned, only data from the first repetition of the RS algorithm are used.

RESULTS

Figure 3 compares the ASR scores achieved with default and optimized time constants, using either the insertion gains recommended by CAM2 (left panel) or those calculated by OPRA-RS (right panel). The default time constants correspond to the fixed compression speeds used by Fontan et al. (2020c) and Gonçalves Braz et al. (2022). Those were 200, 100, 100, 100, and 100 ms for attack times, and 2,000, 1,500, 1,200, 1,000, and 1,000 ms for release times for HA channels 1–5, respectively.

With the insertion gains recommended by CAM2, it can be noticed that ASR scores tended to be higher after the optimization of time constants (median ASR score: 92%) than with default time constants (median ASR score: 88%). As Kolmogorov-Smirnov tests indicated that the ASR scores were not normally distributed ($p \leq 0.044$ in both conditions), a Wilcoxon signed-rank test was used to assess the significance of the observed difference. The results show that ASR scores are significantly higher after the optimization of time constants ($Z = 2.8$; $p = 0.005$). In contrast with this general trend, for two out of the 12 audiograms, all time-constant configurations tested during the RS yielded lower ASR scores than those obtained with the default constants. The improvements due to the optimization of time constants seem to be larger for the most severe HLs than for milder HLs. For example, for audiograms corresponding to level 7 of the WARHICS scale, the ASR score improved by 10 percentage points on average, whereas an average improvement of 1.3 percentage point is observed for audiograms corresponding to level 4 of the WARHICS scale. A Spearman correlation was computed to assess the existence of a significant association between HL severity, represented by the pure-tone average (PTA) for frequencies of 0.5, 0.75, 1, 1.5, 2, 3, and 4 kHz, and the improvement in terms of ASR score due to the optimization of time constants. The results indicate a significant positive relationship between the two variables ($\rho = 0.62$; $p = 0.03$), that is, the higher the PTA, the larger the benefit due to the optimization of time constants.

Contrary to the ASR scores obtained with CAM2 gains, no improvement was observed after the optimization of time constants when using the gains recommended by OPRA-RS. For six out of the 12 audiograms, all time-constant configurations tested during the RS yielded lower ASR scores than those obtained with default time constants.

The reproducibility of the ASR scores was assessed by comparing the outcomes of the two repetitions of the RS algorithm. For CAM2, the median ASR score achieved during the second repetition of the algorithm (91%) was very close to the score achieved during the first repetition (92%); a Wilcoxon test revealed that no significant difference existed between the ASR scores yielded by each of the repetitions ($Z = -1.7$; $p = 0.10$). For OPRA-RS, all ASR scores remained equal across repetitions of the RS algorithm.

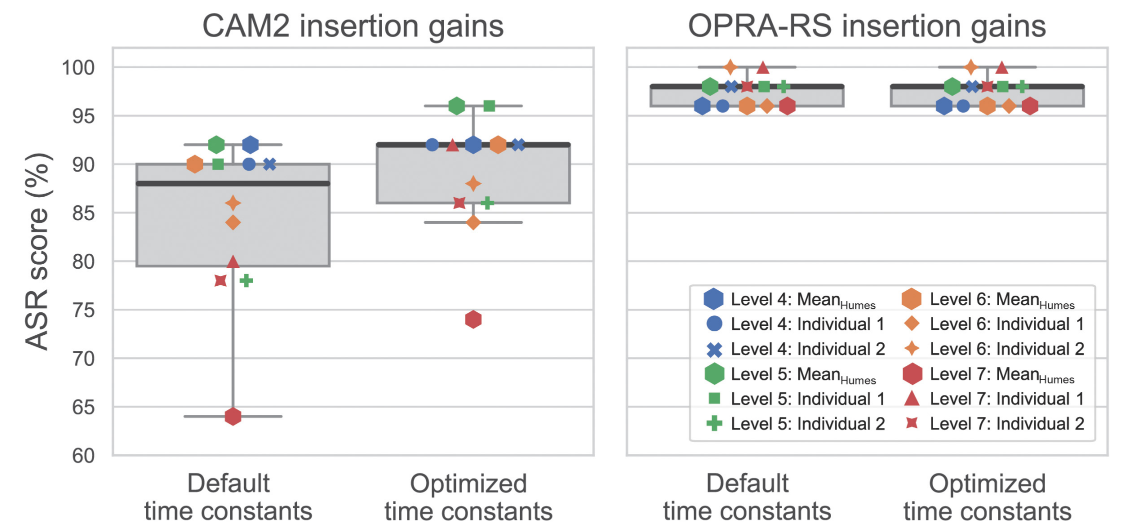


FIGURE 3 | ASR scores with or without optimization of time constants, using the insertion gains recommended by CAM2 **(left panel)** or OPRA-RS **(right panel)** for the 12 audiograms. Horizontal, thick dark lines inside the boxes represent median values. The 0 and 100th percentiles are represented by the bottom and top whiskers, while the bottom and top limits of the boxes represent the 25th and 75th percentiles.

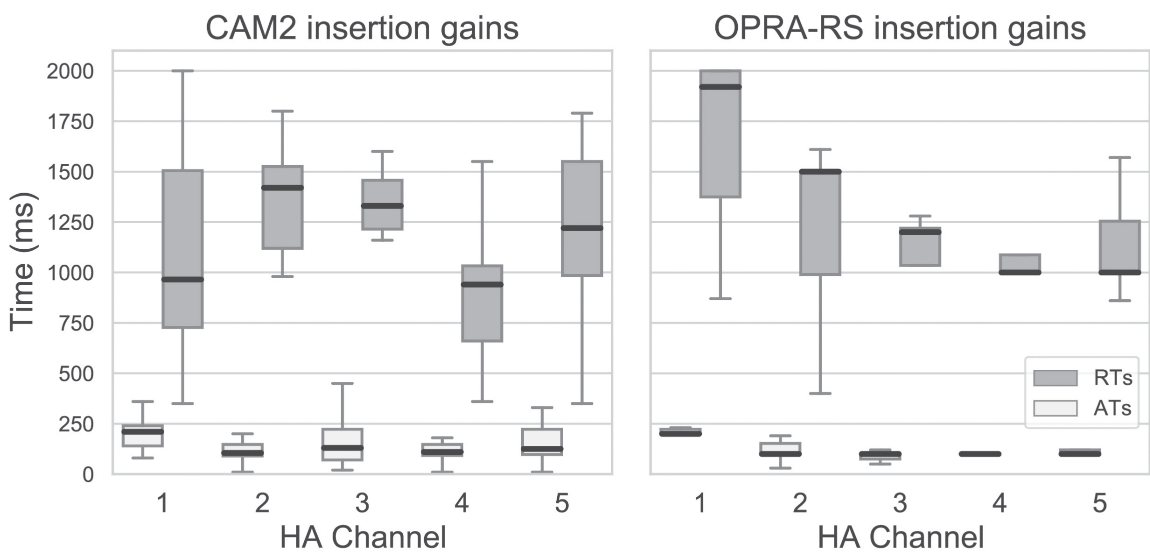


FIGURE 4 | Distribution of the attack times (ATs) and release times (RTs) yielding the highest ASR performances when using the insertion gains recommended by CAM2 **(left panel)** or OPRA-RS **(right panel)** for the 12 audiograms. Otherwise as **Figure 3**.

Figure 4 shows the distribution of attack and release times yielding the highest ASR performances for the 12 audiograms with the insertion gains recommended by CAM2 (left panel) or by OPRA-RS (right panel). In the cases for which better ASR scores were achieved with the default time constants used by Gonçalves Braz et al. (2022), these default values were retained as best configurations.

Median attack and release times across channels are 135 and 1,190 ms, respectively, for CAM2 gains, and 100 and 1,200 ms, respectively, for OPRA-RS gains. Contrary to attack times, optimized release times span the entire possible range of values. Optimized time constants seem less variable for OPRA-RS than for CAM2. This is at least partially due to the fact that, for OPRA-RS, a larger proportion of the optimized time constants correspond to the default constants used by Gonçalves Braz et al. (2022).

Finally, the time constants obtained during the two repetitions of the RS algorithm were compared. As default time constants corresponded to fixed values, the HA configurations for which the best ASR scores were achieved with default time constants were excluded from this analysis. For the remaining HA configurations ($N = 16$), the median absolute differences across repetitions were 85 and 355 ms for attack and release times, respectively. The minimum and maximum absolute differences were 0 and 420 ms for attack times, and 20 and 1,640 ms for release times.

DISCUSSION

This study provides proof of concept that RS can be used for the optimization of HA time constants for a given audiometric profile. This approach might prove particularly useful since there is currently no consensus as to the time constants that should be used to maximize speech intelligibility for a HI individual (Moore and Sęk, 2016). It has been shown that knowledge of the HA user's cognitive abilities might help to choose slow or fast compression (Gatehouse et al., 2003; Souza and Sirow, 2014), but the results of studies addressing the relationship between hearing abilities and optimal time constants are heterogeneous (Hopkins et al., 2012; Moore and Sęk, 2016). Within this context, OPRA-RS represents a novel approach that, given the audiometric profile of the HA user, can be used to systematically explore a large number of time-constant configurations and assess their impact in terms of speech intelligibility.

ASR scores for the optimized time constants were reproducible across repetitions of the RS algorithm, but were associated with different combinations of time constants. This is possibly due to an interaction between attack and release times, as the two parameters were optimized simultaneously. Future studies should optimize each parameter independently to assess their reproducibility. It might also be interesting to extend in future studies the search ranges used for attack and release times, which were limited in the present study to values generally associated with slow compression.

For the mild-to-moderately-severe HLs used in the this study, the optimization of time constants yielded significant improvements in ASR scores for CAM2, but not for OPRA-RS. In addition, the improvements observed for CAM2 were small (4 percentage points, corresponding to 2 out of the 50 words used in the study). These observations are likely due to ceiling effects in the two test conditions, even before the optimization of time constants. Indeed, it was observed that more severe HLs, yielding the lowest ASR scores with CAM2 and default time constants, were associated with higher improvements after the optimization of time constants. To limit such ceiling effects and thus to assess if clinically significant benefits can be obtained, future studies should use more challenging experimental conditions (e.g., speech materials that are shorter and/or presented in noise).

Finally, it should be determined if, as shown by Fontan et al. (2020c) for the fine-tuning of insertion gains, the benefits observed in ASR performance due to the fine-tuning of time constants translate into speech-intelligibility benefits for actual listeners with age-related HL, and if these benefits are clinically relevant. Also, in the present study, CAM2 as a baseline prescription since it was used in the previous experiments on ASR-based optimization of HA parameters (Fontan et al., 2020c; Gonçalves Braz et al., 2022). It should be determined if significant improvements are also observed for those prescription rules that are more widely used in clinical practice, such as NAL-NL2 (Keidser et al., 2011).

DATA AVAILABILITY STATEMENT

The datasets generated and analyzed for this study can be obtained from the corresponding authors for any research purpose.

AUTHOR CONTRIBUTIONS

LF initiated the idea. LG designed and implemented the random-search algorithms, under the supervision of JP. MS provided scientific advice about fitting algorithms, and the hearing-aid and hearing-loss simulations. LF, LG, MS, and CF analyzed and interpreted the data. LF and CF wrote the manuscript. All authors approved the final version of the manuscript.

ACKNOWLEDGMENTS

The experiments presented in this manuscript were carried out using the OSIRIM platform (http://osirim.irit.fr/) that is administered by the Institut de Recherche en Informatique de Toulouse (IRIT) and supported by the French National Center for Scientific Research (CNRS), the Occitanie Region, the French Government, and the European Regional Development Fund (ERDF). MS was funded by the NIHR Manchester Biomedical Research Centre (United Kingdom).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.779062/full#supplementary-material

REFERENCES

Bor, S., Souza, P., and Wright, R. (2008). Multichannel compression: effects of reduced spectral contrast on vowel identification. *J. Speech Lang. Hear. Res.* 51, 1315–1327. doi: 10.1044/1092-4388(2008/07-0009)

Cruickshanks, K. J., Nondahl, D. M., Fischer, M. E., Schubert, C. R., and Tweed, T. S. (2020). A novel method for classifying hearing impairment in epidemiological studies of aging: the Wisconsin Age-Related Hearing Impairment Classification Scale. *Am. J. Audiol.* 29, 59–67. doi: 10.1044/2019_AJA-19-00021

Fontan, L., Cretin-Maitenaz, T., and Füllgrabe, C. (2020a). Predicting speech perception in older listeners with sensorineural hearing loss using automatic speech recognition. *Trends Hear.* 24:4522331216520914769. doi: 10.1177/2331216520914769

Fontan, L., Farinas, J., Segura, B., Stone, M. A., and Füllgrabe, C. (2020b). "Using automatic speech recognition to predict aided speech-in-noise intelligibility," in *Proceedings of the Speech-In-Noise Workshop*, (Toulouse), doi: 10.13140/RG.2.2.27785.19043

Fontan, L., Le Coz, M., Azzopardi, C., Stone, M. A., and Füllgrabe, C. (2020c). Improving hearing-aid gains based on automatic speech recognition. *J. Acoust. Soc. Am.* 148, EL227–EL233. doi: 10.1121/10.0001866

Fournier, J.-E. (1951). *Audiométrie Vocale : Les Épreuves D'intelligibilité Et Leurs Applications Au Diagnostic, À L'expertise Et À La Correction Prothétique Des Surdités [Speech Audiometry: Speech-Intelligibility Tests and their Application for the Diagnosis, Survey and Hearing-Aid Rehabilitation of Hearing Losses]*. Paris: Maloine.

Gatehouse, S., Naylor, G., and Elberling, C. (2003). Benefits from hearing aids in relation to the interaction between the user and the environment. *Int. J. Audiol.* 42, S77–S85. doi: 10.3109/14992020309074627

Gonçalves Braz, L., Fontan, L., Pinquier, J., Stone, M. A., and Füllgrabe, C. (2022). OPRA-RS: a hearing-aid fitting method based on automatic speech recognition and random search. *Front. Neurosci.* 16:7779048. doi: 10.3389/fnins.2022.779048

Hopkins, K., King, A., and Moore, B. C. J. (2012). The effect of compression speed on intelligibility: simulated hearing-aid processing with and without original temporal fine structure information. *J. Acoust. Soc. Am.* 132, 1592–1601. doi: 10.1121/1.4742719

Humes, L. E. (2021). Further evaluation and application of the Wisconsin Age-Related Hearing Impairment Classification System. *Am. J. Audiol.* 30, 359–375. doi: 10.1044/2021_AJA-20-00207

Keidser, G., Dillon, H., Flax, M., Ching, T., and Brewer, S. (2011). The NAL-NL2 prescription procedure. *Audiol. Res.* 1, 88–90. doi: 10.4081/audiores.2011.e2

Kollmeier, B., Schädler, M. R., Warzybok, A., Meyer, B. T., and Brand, T. (2016). Sentence recognition prediction for hearing-impaired listeners in stationery and fluctuation noise with FADE: empowering the attenuation and distortion concept by Plomp with a quantitative processing model. *Trends Hear*. 20:233121651665579. doi: 10.1177/2331216516655795

Lee, A., and Kawahara, T. (2009). "Recent development of open-source speech recognition engine Julius," in *Proceedings of the Annual Summit and Conference of the Asia-Pacific Signal and Information Processing Association (APSIPA ASC 2009)*. Sapporo 131–137.

Moore, B. C. J. (2008a). *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues*. Hoboken, NJ: Wiley, doi: 10.1121/1.2982418

Moore, B. C. J. (2008b). The choice of compression speed in hearing aids: theoretical and practical considerations and the role of individual differences. *Trends Amplif.* 12, 103–112. doi: 10.1177/1084713808317819

Moore, B. C. J., Glasberg, B. R., and Stone, M. A. (2010b). Development of a new method for deriving initial fittings for hearing aids with multi-channel compression: CAMEQ2-HF. *Int. J. Audiol.* 49, 216–227. doi: 10.3109/1499202090329674

Moore, B. C. J., Füllgrabe, C., and Stone, M. A. (2010a). Effect of spatial separation, extended bandwidth, and compression speed on intelligibility in a competing-speech task. *J. Acoust. Soc. Am.* 128, 360–371. doi: 10.1121/1.3436533

Moore, B. C. J., Füllgrabe, C., and Stone, M. A. (2011). Determination of preferred parameters for multichannel compression using individually fitted simulated hearing aids and paired comparisons. *Ear Hear*. 32, 556–568. doi: 10.1097/AUD.0b013e31820b5f4c

Moore, B. C. J., and Glasberg, B. R. (1993). Simulation of the effects of loudness recruitment and threshold elevation on the intelligibility of speech in quiet and in a background of speech. *J. Acoust. Soc. Am.* 94, 2050–2062. doi: 10.1121/1.40747

Moore, B. C. J., and Sęk, A. (2013). Comparison of the CAM2 and NAL-NL2 hearing aid fitting methods. *Ear Hear*. 34, 83–95. doi: 10.1097/AUD.0b013e3182650adf

Moore, B. C. J., and Sęk, A. (2016). Preferred compression speed for speech and music and its relationship to sensitivity to temporal fine structure. *Trends Hear.* 20:2331216516640486. doi: 10.1177/2331216516640486

Nejime, Y., and Moore, B. C. J. (1997). Simulation of the effect of threshold elevation and loudness recruitment combined with reduced frequency selectivity on the intelligibility of speech in noise. *J. Acoust. Soc. Am.* 102, 603–615. doi: 10.1121/1.41973

Rembaud, F., Fontan, L., and Füllgrabe, C. (2017). L'audiométrie vocale en France: état des lieux [Speech audiometry in France: current clinical practices]. *Cah. Audition* 6, 22–25.

Schädler, M. R., Hülsmeier, D., Warzybok, A., and Kollmeier, B. (2020). Individual aided speech-recognition performance and predictions of benefit for listeners with impaired hearing employing FADE. *Trends Hear.* 24:2331216520938929. doi: 10.1177/2331216520938929

Schädler, M. R., Warzybok, A., Hochmuth, S., and Kollmeier, B. (2015). Matrix sentence intelligibility prediction using an automatic speech recognition system. *Int. J. Audiol.* 54, 100–107. doi: 10.3109/14992027.2015.1061708

Schädler, M. R., Warzybok, A., and Kollmeier, B. (2018). Objective prediction of hearing aid benefit across listener groups using machine learning: speech recognition performance with binaural noise-reduction algorithms. *Trends Hear*. 22:233121651876895. doi: 10.1177/2331216518768954

Souza, P. E. (2002). Effects of compression on speech acoustics, intelligibility, and sound quality. *Trends Amplif*. 6, 131–165. doi: 10.1177/108471380200600402

Souza, P. E., and Sirow, L. (2014). Relating working memory to compression parameters in clinically fit hearing aids. *Am. J. Audiol*. 23, 394–401. doi: 10.1044/2014_AJA-14-0006

Stone, M. A., and Moore, B. C. J. (1992). Syllabic compression: effective compression ratios for signals modulated at different rates. *Br. J. Audiol*. 26, 351–361. doi: 10.3109/03005369209076659

Stone, M. A., and Moore, B. C. J. (2008). Effects of spectro-temporal modulation changes produced by multi-channel compression on intelligibility in a competing-speech task. *J. Acoust. Soc. Am.* 123, 1063–1076. doi: 10.1121/1.2821969

Stone, M. A., Moore, B. C. J., Füllgrabe, C., and Hinton, A. C. (2009). Multichannel fast-acting dynamic range compression hinders performance by young, normal-hearing listeners in a two-talker separation task. *J. Audio Eng. Soc.* 57, 532–546.

Verschuure, J., Maas, A. J., Stikvoort, E., de Jong, R. M., Goedegebure, A., and Dreschler, W. A. (1996). Compression and its effect on the speech signal. *Ear Hear*. 17, 162–175. doi: 10.1097/53500003446-199604000-00008

Conflict of Interest: This study is part of the development of a future product/service by Archean LABS intended for hearing-aid audiologists. CF acted as a scientific consultant.

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.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership