Toward AI-Assisted Teleaudiology

Jan-Willem Wasmann

DONDERS SERIES

RADBOUD UNIVERSITY PRESS

Radboud Dissertation Series

Toward AI-Assisted Teleaudiology

Jan-Willem Wasmann

The research presented in this thesis was conducted at the Radboud university medical center and Donders Institute for Brain, Cognition and Behavior.

Parts of the research presented in this thesis were carried out with financial support from Cochlear Ltd and the European Industrial Doctorate project, funded by the European Union's Horizon 2020 framework program for research and innovation under the Marie Sklodowska-Curie Grant Agreement No. 860718. Printing of the thesis was financially supported by the Radboud University and the Donders Institute for Brain, Cognition and Behavior.

Author: Jan-Willem Wasmann
Title: Toward AI-Assisted Teleaudiology

Radboud Dissertations Series
ISSN: 2950-2772 (Online); 2950-2780 (Print)

Published by RADBOUD UNIVERSITY PRESS
Postbus 9100, 6500 HA Nijmegen, The Netherlands
www.radbouduniversitypress.nl

Design: Proefschrift AIO | Annelies Lips
Cover: Proefschrift AIO | Guntra Laivacuma
Printing: DPN Rikken/Pumbo

ISBN: 9789493296657
DOI: 10.54195/9789493296657
Free download at: www.boekenbestellen.nl/radboud-university-press/dissertations

RADBOUD
UNIVERSITY
PRESS

Toward AI-Assisted Teleaudiology

Proefschrift ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.M. Sanders,
volgens besluit van het college voor promoties
in het openbaar te verdedigen op

dinsdag 18 maart 2025
om 14.30 uur precies

door

Jan-Willem Anne Wasmann
geboren op 21 november 1984
te Utrecht

Promotor:
Prof. dr. E.A.M. Mylanus

Copromotoren:
Dr. ir. C.P. Lanting
Dr. W.J. Huinck

Manuscriptcommissie:
Prof. dr. A.C.H. Geurts
Prof. dr. H.E. Cullington (University of Southampton, Verenigd Koningrijk)
Prof. dr. ir. J.C.M. Smits (Amsterdam UMC)

Toward AI-Assisted Teleaudiology

Dissertation to obtain the degree of doctor
from Radboud University Nijmegen
on the authority of the Rector Magnificus prof. dr. J.M. Sanders,
according to the decision of the Doctorate Board
to be defended in public on

Tuesday, March 18, 2025
at 2.30 pm

by

Jan-Willem Anne Wasmann
born on November 21, 1984
in Utrecht (the Netherlands)

Supervisor:
Prof. dr. E.A.M. Mylanus

Co-supervisors:
Dr. ir. C.P. Lanting
Dr. W.J. Huinck

Manuscript Committee:
Prof. dr. A.C.H. Geurts
Prof. dr. H.E. Cullington (University of Southampton, United Kingdom)
Prof. dr. ir. J.C.M. Smits (Amsterdam UMC)

TABLE OF CONTENTS

1

General Introduction & Thesis Outline

GENERAL INTRODUCTION

The Global Health Challenge: Inequality, Aging, and the Rise of Chronic Care

The world's aging population presents a significant challenge to healthcare worldwide. By 2050, the old-age economic dependency ratio, which is the ratio between the number of individuals aged ≥65 years and those of working age (20-64 years), is projected to reach 55% in the European Union, and on average, 52.7% in the member countries of the Organization for Economic Co-operation and Development (OECD), indicating a trend likely to extend globally (OECD, 2024). This demographic shift implies that an increased demand for chronic care needs to be provided by a relatively small working population (OECD, 2024). The healthcare challenge intensifies when considering the sustainable development goals that address historic inequalities (United Nations, 2023). These inequalities are evident in low and middle-income countries where underserved populations lack access to primary healthcare and have lower general well-being (Kruk et al., 2018).

These challenges are also prominent in hearing healthcare, meaning that both the growth in the number of older adults with hearing loss and the necessity to distribute hearing healthcare more equally worldwide will significantly increase the overall pressure on our hearing healthcare system (Haile et al., 2021). Budgets need to be distributed more fairly, and fewer personnel will be available relative to the number of people requiring audiological care.

The Global Burden of Hearing Loss

An estimated 1.5 billion people worldwide are directly affected by hearing loss, projected to increase by another billion by 2050 (World Health Organization, 2021). Additionally, family members and friends of people with hearing loss are indirectly affected since hearing loss hampers interpersonal communication and psychosocial well-being. The estimated global annual cost of untreated hearing loss exceeds 750 billion US dollars (McDaid et al., 2021). For children, untreated hearing loss hampers language development and limits educational potential (Lieu et al., 2020). In adults, it results in higher unemployment rates, missed workdays, social isolation, and a lower quality of life (Shield, 2019). Hearing healthcare faces various obstacles, including lack of awareness, lack of motivation and support, stigma, financial limitations, competing comorbidities, and, in some regions, limited infrastructure and workforce shortages (Barnett et al., 2017; Kamenov et al., 2021). For instance, in parts of Africa, there are fewer than one hearing health professional per million people (Kamenov et al., 2021). This results in a gap between the

400 million people worldwide who might benefit from hearing aids and the 68 million (17%) using them (Orji et al., 2020). Existing audiological services and technologies cannot address the current need nor the growing need to address aging and the rise of chronic care. New approaches are needed to overcome the global obstacles in hearing healthcare.

Computational Audiology to Address the Global Hearing Health Challenges

Artificial intelligence (AI) can be part of the solution to alleviate the anticipated pressure on hearing healthcare. AI is the simulation of human intelligence in machines that are designed to think, learn, and perform tasks autonomously. It encompasses various technologies, including machine learning, natural language processing, and robotics, which are used to solve complex problems and enhance decision-making (S. J. Russell & Norvig, 2016). AI could help us perform tasks and complement our capabilities (H. J. Wilson & Daugherty, 2018). For the application of complex models in hearing healthcare practice, we coined the term "computational audiology," defined as "the branch of audiology that employs techniques from mathematics and computer science to improve clinical treatments and scientific understanding of the auditory system." Audiology and hearing science have always been data-driven, mainly involving psychophysical measurements based on models that capture facets of hearing. Dick Lyon, for example, provides in his book, "Human and Machine Hearing: Extracting Meaning from Sound," an extensive history of the applications of models to describe and mimic aspects of the auditory system from the first theories of hearing, the use of computers for modeling, up to shallow neural networks (Lyon, 2017). More recent examples of computational audiology include enhancing speech processing and intelligibility in hearing aids and cochlear implants (Bramsløw et al., 2018; Goehring et al., 2019; Y.-H. Lai et al., 2018; D. Wang, 2017) as well as deep learning applied to auditory system modeling (Baby et al., 2021; A. J. E. Kell et al., 2018). AI has the potential to assist in clinical practices using efficient testing and simulation platforms (Heisey et al., 2020; Hendrikse et al., 2024; Lesica et al., 2021; Meyer et al., 2023) and to increase accessibility by applying AI in inexpensive mobile devices (Slaney et al., 2020). Chapter 2 of this thesis provides further examples of computational audiology to illustrate how it can be used in the diagnosis of hearing loss, hearing rehabilitation, and hearing research. In this chapter, a vision is sketched of how computational approaches combined with affordable devices, standardization, evidence-based principles, and shared data applied within a network of distributed expertise can improve access to hearing healthcare. However, there is still a long way to go before implementing such approaches at a large scale among audiology clinics worldwide due to barriers,

including the complexity of dealing with privacy issues, data ownership, and lack of interoperability (Lehne et al., 2019). In addition, there is fear among clinicians of de-skilling and a need for proper training to use AI (Oremule et al., 2024). AI-based patient support tools that are more user-centered instead of clinic-centered seem to be easier to start with before approaches in clinics take off (Slaney et al., 2020).

AI-based Patient Support Tools: AI chatbots and Automated Speech Recognition for People with Hearing Loss

The most dramatic developments in AI in the last decade have been in image and speech recognition and, more recently, in large language models, which are now visibly impacting various sectors. To guide the reader through the relevant recent progress in AI in hearing healthcare, a brief description of the history of automated speech recognition (ASR) and chatbot technology is given below. ASR tools might benefit people with hearing loss and are already available at low cost, even without a solution for the interoperability and data exchange challenges that medical applications face within clinics.

ASR is a technology that enables computers to interpret and process human speech into written text (Jurafsky & Martin, 2009). The development of ASR required a relatively long ramp-up. In the 1980s, IBM made progress in ASR by introducing Hidden Markov Models, which enhanced the accuracy and reliability of recognizing spoken words (Rabiner, 1989). This development laid the foundation for modern speech recognition technology. In addition, discriminative learning and deep learning methodologies played a significant role in advancing ASR (Deng & Li, 2013). Progress was slow, and early applications were limited to predictable, well-controlled environments using a closed set of options. For example, systems like AT&T's "How May I Help You?" (HMIHY), deployed in the mid-1990s, were among the first to incorporate speech recognition for handling customer service calls. These systems were designed to process a limited set of spoken requests in order to connect callers to the correct services or information (Gorin et al., 1997). Speech recognition remained a very hard task for machines for a very long time (Juang & Rabiner, 2005). However, in 2016, major tech companies, including Microsoft and Google, claimed their machines achieved human parity in conversational speech recognition (Saon et al., 2017; Xiong et al., 2017). This accomplishment was underpinned by word error rates similar to those of humans in phone call transcriptions.

Also, in AI chatbots, technology has progressed slowly toward human performance, with significant improvements in recent years. An AI chatbot is a software

application that simulates human conversation through text or voice interactions (Adamopoulou & Moussiades, 2020). The earliest known chatbot, ELIZA, was developed in 1966 to act as a psychotherapist, responding to user inputs by rephrasing them as questions (Weizenbaum, 1966). Decades later, the A.L.I.C.E. chatbot, created by Richard Wallace in 1995, incorporated natural language processing to improve interaction capabilities (Wallace, 2009). Significant progress in chatbot development occurred with the introduction of Transformer deep learning models (Vaswani et al., 2017), which formed the architecture for large language models, including BERT and GPT-3 (Brown et al., 2020). By 2021, OpenAI launched ChatGPT/GPT-3, a large language model based on an astounding 175 billion parameters, quickly becoming renowned for its human-like text generation capabilities.

What is required to adopt new technologies such as ASR, and what benefits might they bring? Researchers have estimated that a word recognition accuracy of at least 80-90% is a critical threshold for ASR systems to be widely adopted and to offset the drawbacks, such as manually correcting errors in poor transcriptions (Jurafsky & Martin, 2009). Surpassing this minimum accuracy threshold has led to a significant user base (i.e., achieving a critical mass of users), making ASR commercially viable. Consequently, this has led to a virtuous cycle of increased data and further improvement. The evolution of ASR has shown that:

- An enormous set of training data, an improved digital infrastructure, and better algorithms led to a significant increase in accuracy, making ASR viable in real applications (Baker et al., 2009). The data were collected from video and speech recordings, including those with subtitles on platforms like YouTube, encompassing various voices, accents, and acoustics. By the 2010s, this led to a jump in quality similar to that seen in image recognition (Xiong et al., 2017), resulting in popular applications, including Google Live Transcribe.
- Certain users, particularly those with profound hearing loss who lacked access to adequate hearing aids or cochlear implant care, found immediate value in these advancements (Berke, 2017). As the technology improved, the potential user base expanded, and our evaluation of ASR apps in 2020, described in Chapter 4, demonstrated their capability to transcribe speech to support communication for many individuals with severe to profound hearing loss (Pragt et al., 2022).
- There are broader applications, as similar technology can remove other communication barriers. Deep learning models can also be trained to recognize people's hand gestures used in sign language and convert them to text or speech to communicate more effectively with people who are not proficient in

sign language (Bharathi et al., 2021; Jin et al., 2023). Such models can be trained on prelingual deaf speech or other atypical speech patterns that are challenging to capture for those not accustomed to them (i.e., those not trained in it (Biadsy et al., 2019).

- In cases where ASR solutions failed to meet the minimum required accuracy, human interpreters could enhance the transcript by manually correcting it (Gaur et al., 2015) or completely taking over the speech transcription process. This development made high-quality speech-to-text solutions accessible and affordable for a broader population since machine transcription is cheaper than human transcription. Today, users can opt for free transcription for everyday tasks, including meeting notes, and choose higher accuracy options (paid) for critical communications, such as legal agreements, to minimize miscommunication risks.

Advancements in AI are progressing rapidly and have significant societal consequences. AI might accelerate the adoption of new technologies needed to address the ongoing challenges in hearing healthcare. Free ASR apps that run on smartphones and tablets are currently used by people with hearing loss in various communicative settings. These apps help lower communication barriers and enable better participation at work, in groups, or in private settings (Loizides et al., 2020). AI chatbots may play a role in accessing and digesting health information and supporting people with hearing loss, clinicians, and hearing researchers (Chapter 3). Another approach to increase access to hearing healthcare is the development of remote hearing health care. Building digital infrastructure using the devices people already own is less expensive, although standardized protocols and methods to collect data across clinics and countries are also lacking in the digital realm (D'Onofrio & Zeng, 2022).

Teleaudiology and Self-evaluation of Hearing Status as a Driver for AI in Audiology

There are various terms that refer to remote hearing health care, for example, "connected hearing health care," "teleaudiology," or "e-audiology" (Young et al., 2022). The term "remote" seems to be the most obvious choice. However, at this moment, it only refers to being "remote" from the perspective of the healthcare provider and overlooks the proximity of the end-user. The term "connected" hearing healthcare is also not preferable. It may be misinterpreted as doing care via Bluetooth only. The term "teleaudiology" appears to be the least ambiguous and is, therefore, the preferred term in the general introduction and discussion of this thesis. Teleaudiology technologies facilitate the online or at-home delivery of

hearing care services, encompassing self-administered evaluations, user-directed fitting adjustments and fine-tuning, and video consultations (Bennett et al., 2024).

Teleaudiology can reduce barriers to seeking care. Travel inconvenience, time, and cost significantly hinder healthcare access, particularly in rural and developing areas. Bush et al. (2013) found that patients in rural areas travel on average 100 miles to audiology clinics compared to 15 miles for patients living in urban areas, with distance correlating with delays in hearing aid and cochlear implant interventions. Coco et al. (2016) identified that transportation, motivation, and financial constraints are major barriers in urban areas, suggesting that teleaudiology could mitigate these challenges by providing services closer to patients' homes.

Beside the benefits for end-users, teleaudiology may also support professionals. Self-administered hearing tests reduce the workload for professionals, making the diagnostic process less dependent on clinical capacity. Applying Bayesian active learning techniques that select the most informative stimuli instead of a fixed number of stimuli can reduce the testing time of self-administered tests and provide room to add other measurements (Heisey et al., 2020). Consequently, self-fitting and remote fitting of hearing aids and cochlear implants are increasingly being introduced (Mashmous, 2022; Schepers et al., 2019). Based on a systematic review, Mashmous (2022) concluded that remote fitting of hearing aids can provide results similar to face-to-face programming in the clinic, even for people with no previous experience with hearing aids. In addition to all the advantages of teleaudiology, it is also important to be aware of the disadvantages. One significant drawback of teleaudiology is the loss of direct human contact. When improving teleaudiology tools, one must consider the loss of direct personal interaction along with the potential barriers for less digitally proficient people.

Teleaudiology might be a catalyst for a much broader use of AI in audiology for several reasons:

- More data: teleaudiology enables data collection directly from the end-user, providing amassed data including telemetry, automated audiometry, ecological momentary assessments, but also usage data, and users' environment (Christensen, Saunders, Porsbo, et al., 2021). This data gathering enables a much-refined analysis than possible with current data collection in clinics, possibly leading to even more personalized care;
- More autonomy for the end-user: teleaudiology gives the end-user easy access to support, which AI chatbots could facilitate. Examples are frequent feedback

on the technical integrity of the device and warnings on possible declines in device performance;
- Faster adoption: Teleaudiology helps researchers and developers reach out to early adopters around the globe who already find value in preliminary AI solutions and involve them in further improving such solutions that may help address the challenges stated at the beginning of this introduction.

Teleaudiology in Cochlear Implant Care

This thesis focuses on the application of teleaudiology within cochlear implant (CI) care. The journey of a CI user is well-suited for exploring teleaudiology due to the relatively standardized technology employed and the clinics' involvement in every aspect of the journey from diagnosis, indication, and implantation to life-long care. Teleaudiology is well-suited to meet the CI users' demand for 24/7 support and troubleshooting. It is now possible for CI users to perform audiometry tests at-home as part of regular care (Maruthurkkara et al., 2022).

A CI restores sound perception in people with severe to profound hearing loss who do not receive sufficient benefit from powerful hearing aids. A CI system consists of an external part worn on the ear or head, referred to as a processor, that includes microphones, a sound processor that converts sounds into patterns of electric pulses, and a coil to transmit data and power to the internal part. The internal part, located under the skin, consists of a coil, stimulator, and electrode array; the latter is located in the cochlea. The internal part provides electric pulses in the cochlea that, by exciting nerve fibers (i.e., ganglion cells), lead to the perception of sound. Several well-tested and validated strategies exist to convert sound into electric patterns, i.e., Continuous Interleaved Sampling (CIS; B. S. Wilson et al., 1993) and derivations such as ACE (Skinner et al., 2002) and HiRes (Brendel et al., 2008).

Sound is converted into electric patterns in the following way. First, sound is captured by the microphone(s) and passed through a filter bank, where each filter band corresponds to an electrode in use. The extracted envelopes per filter band are presented as amplitude-modulated fixed-rate pulse trains at each electrode surface. This process takes advantage of the tonotopic organization of the auditory pathway at the level of the basilar membrane because the electrode surfaces on the electrode array are located at different positions along the basilar membrane. Charge injected from electrodes deep in the cochlea stimulates nerve fibers that typically transmit responses to low-frequency sounds. In contrast, charge injected by electrodes close to the base of the cochlea stimulates nerve fibers that typically

transmit responses to high-frequency sounds. Pulses on different electrodes are interleaved in time to prevent electric field interactions between electrodes.

Even within clinically validated strategies to convert sound into pulses, there are numerous parameters that can be adjusted to the individual CI user. This is called the CI fitting procedure. The complexity of CI fitting and the variation in patient outcomes has led to the development of one of the first medical expert systems using AI (Meeuws et al., 2017). One way forward is to add more automated audiometry tests to gain more insights into individual errors. If that reveals patterns, it could be used in a learning cycle to improve fitting algorithms and increase individual CI performance (Battmer et al., 2015; Meeuws et al., 2017; Opstal & Noordanus, 2023).

Summary of Key Points and Future Directions

The growing demands of an aging population are driving significant capacity challenges in hearing healthcare, and the development of teleaudiology will play a critical role in meeting these needs. To address these challenges, new technologies such as AI and affordable devices are emerging as potential solutions to unmet needs in enhancing communication abilities for people with hearing loss (Slaney et al., 2020). In the longer term, more integrated solutions as depicted in Figure 1, may become part of the clinical pathway. Here, AI can serve as an expert system to support clinicians (Meeuws et al., 2017), a tool for diagnosis (Heisey et al., 2020), an interface with patients (Swanepoel et al., 2023), a tool for signal enhancement (Goehring et al., 2017), and a method to provide precision medicine (Barbour, 2018), among many more applications within hearing healthcare (AlSamhori et al., 2024). Teleaudiology services empower patients by enabling them to conduct hearing tests and fit devices remotely, either independently or with professional oversight, thereby increasing efficiency and reducing dependency on in-person visits (Convery et al., 2019). In addition, better self-management of hearing aid care is associated with higher health literacy (Caposecco et al., 2016). Therefore, feedback from patients and clinicians on the effectiveness and usability of these new technologies is crucial for further improvement and adoption. For responsible adoption on a large scale, ethical considerations and potential drawbacks of implementing AI and teleaudiology need to be addressed to ensure responsible use. For instance, concerns about data privacy, equitable access to technology, biases, and how to regulate AI within healthcare must be considered (Gilbert et al., 2023; Maddox et al., 2019).

Although the adoption of AI-aided teleaudiology remains in its early stages, rapid developments are occurring. The potential for AI-aided teleaudiology is considerable, particularly since teleaudiology is entirely digital. Some centers have already integrated digital tools and remote consultations (Ratanjee-Vanmali et al., 2020; Siggaard et al., 2023), setting the stage for more inclusive, accessible, and affordable audiology care on a global scale.

Figure 1. *DALL·E created artwork. Prompt "A thought-provoking and inspiring digital art representation of the future of audiology. The image shows a futuristic cityscape with advanced technology seamlessly integrated into healthcare. In the foreground, a diverse group of audiologists and computational scientists are gathered around a digital interface, analyzing interconnected data from various sources. In the background, people of different ages and ethnicities are using wearable devices and smartphones for hearing assessments and treatments, representing accessibility and equity in healthcare. The sky is filled with abstract visualizations of data streams and algorithms, highlighting the power of computational sciences. The scene conveys a sense of hope, innovation, and the transformative potential of technology in advancing healthcare."*

AIM AND OUTLINE OF THIS THESIS

This thesis, titled "Toward AI-Assisted Teleaudiology," explores various issues confronting hearing healthcare in the digital age. It is investigated whether accessibility and quality of hearing healthcare can be improved using teleaudiology and AI. This work only covers the initial steps, focusing on teleaudiology in CI care as a kind of microcosm (or model) for hearing healthcare at large. Although CI users are only a small percentage of the people with concerns about their hearing, teleaudiology and AI-aided tools are wider applicable. The chapters on diagnosis, ASR, and chatbots are relevant beyond CI care and may lower barriers to hearing healthcare.

In **Chapter 2** the concept of computational audiology is explored, assessing its potential in resource-limited hearing healthcare settings and setting out guidelines and ethical considerations for its implementation. It contains practical advice for policymakers and clinicians. This chapter addresses the global challenge of hearing loss in today's digital era and investigates the prospects of artificial intelligence, big data, and automation. The concept of computational audiology is introduced, which, in brief, involves applying complex models to hearing healthcare practice. The potential of computational audiology to improve hearing healthcare in terms of precision and efficiency is emphasized while acknowledging the challenges and risks inherent in this digital transition. The need for a responsible implementation within a framework prioritizing patient safety and autonomy is underscored. Since this chapter was written in 2019, it only covers some developments that have emerged since then. However, the surprisingly quick development of AI chatbots is covered in **Chapter 3**.

Chapters 3 and 4 focus on AI-based patient support in hearing healthcare. **Chapter 3** explores the rise of AI chatbots, mainly focusing on large language models (LLMs). More than just providing automated counseling, AI chatbots have the potential to enhance healthcare accessibility, improve patient outcomes, and support research by automating various tasks, including administering questionnaires. However, AI chatbots also pose risks of producing inaccurate outputs known as "hallucinations." **Chapter 4** assesses the accuracy of ASR apps in transcribing speech tokens from conventional audiological speech tests compared to human listeners by evaluating the Word Error Rates (WERs). Additionally, by comparing ASR apps' performance between people with normal hearing and those with hearing loss, the populations that could benefit most from ASR technology in bridging communication gaps are identified.

Chapter 5 discusses self-administered assessments of hearing status. This chapter reviews recent approaches in automated and machine-learning hearing assessments, focusing on pure-tone audiometry. Building upon a previous systematic review, the potential of novel self-administered hearing assessments is assessed according to the guidelines for a scoping review. Automated audiometry's accuracy, reliability, and time efficiency for clinical applications are also evaluated. **Chapter 6** investigates the impact of factors such as time of day, listener fatigue, and motivation on test outcomes in experienced CI users who perform tests at home. The variability of self-administered tests conducted via smartphones and tablets is analyzed, providing insights into their usability in cochlear implant care.

Chapter 7 addresses self-evaluation-guided CI fitting adjustments. It explores whether the data from self-administered assessments can inform clinical action. It investigates an approach based on phoneme identification errors for fitting cochlear implants. Specific phoneme confusions are targeted, and CI settings are adjusted at the electrode level to reduce those confusions. The potential and limitations of this approach are also discussed.

In the concluding **Chapter 8,** the general discussion of this thesis is presented, making a case for how AI can assist teleaudiology, leading to more accessible and affordable hearing healthcare. This chapter synthesizes the results of AI-aided teleaudiology, highlighting its implications for hearing healthcare and providing recommendations for future research and applications.

2

Computational Audiology: New Approaches to Advance Hearing Health Care in the Digital Age

Jan-Willem Wasmann, Cris Lanting, Wendy Huinck, Emmanuel Mylanus, Jeroen van der Laak, Paul Govaerts, De Wet Swanepoel, David Moore, Dennis Barbour

Wasmann, J.-W. A., Lanting, C. P., Huinck, W. J., Mylanus, E. A. M., van der Laak, J. W. M., Govaerts, P. J., Swanepoel, D. W., Moore, D. R., & Barbour, D. L. (2021). Computational Audiology: New Approaches to Advance Hearing Health Care in the Digital Age. Ear and Hearing, 42(6), 1499–1507. https://doi.org/10.1097/AUD.0000000000001041

ABSTRACT

The global digital transformation enables computational audiology for advanced clinical applications that can reduce the global burden of hearing loss. In this article, we describe emerging hearing-related artificial intelligence applications and argue for their potential to improve access to, precision, and efficiency of hearing healthcare services. Also, we raise awareness of risks that must be addressed to enable a safe digital transformation in audiology. We envision a future where computational audiology is implemented via interoperable systems using shared data and healthcare providers adopt expanded roles within a network of distributed expertise. This effort should occur in a health care system where privacy, responsibility of each stakeholder, and patients' safety and autonomy are all guarded by design.

Keywords: Artificial intelligence, Big data, Computational audiology, Computational infrastructure, Digital hearing health care, Hearing loss, Machine learning.

INTRODUCTION

The estimated number of individuals suffering from disabling hearing loss has been growing ever since global reporting began (Vos et al., 2016; World Health Organization, 2019), with WHO projections reaching 900 million by 2050 (World Health Organization, 2019). Besides effects on interpersonal communication, psychosocial well-being, and quality of life, hearing loss has a substantial socio-economic impact (Olusanya et al., 2014; World Health Organization, 2017). Conservative estimates suggest that the overall global annual cost of unaddressed hearing loss is 750–790 billion US dollars (World Health Organization, 2017). In children, hearing loss restricts language development, often resulting in a lasting effect on social and cultural engagement and unfulfilled educational potential. In adults, hearing loss leads to higher unemployment, missed workdays, and social isolation (Kramer et al., 2006). Hearing loss is further associated with more rapid cognitive decline and increased occurrence of dementia-like symptoms (Livingston et al., 2017). Evidence is growing that timely intervention, including hearing aids, can reduce many of these consequences (Maharani et al., 2018).

The actual problem could be even greater, stressing the need for the computational approaches we introduce below, since mild hearing loss (20-34 dB Hearing Level or HL), which is 2-3 times more prevalent than moderate or more severe loss (>35 dB HL), has recently been recognized as an adverse factor in daily life (according to the new GBD 2010 classification on grades of hearing loss; Shield, 2019; B. S. Wilson et al., 2017). Hearing loss is arguably the most prevalent of all impairments in years lived with disability (YLDs; Vos et al., 2016) if we include all known pathologies that currently have no clinical consequences for rehabilitation. Examples include slight or minimal hearing loss (15-20 dB HL; Moore et al., 2020), extended high-frequency loss (8-20 kHz; Motlagh Zadeh et al., 2019), and suprathreshold deficits related to understanding speech in noisy situations (Kollmeier & Kiessling, 2018).

Existing audiological services cannot address the global burden of hearing loss due to inherent barriers, including a dearth of trained professionals, equipment costs, and required expertise (Swanepoel & Clark, 2019). New approaches that transcend current models of practice are essential to overcome global access challenges. Computational augmentation, enhancing and complementing human capabilities by digital tools (H. J. Wilson & Daugherty, 2018), is an essential strategy given the lack of enough qualified human experts in ear and hearing care worldwide (World Health Organization, 2013), the large number of people suffering from hearing loss that is currently underserved, and the growing complexity of high-quality diagnostics and therapeutics.

Computational approaches are enabled by significant global developments, including growing computational power, data storage, and artificial intelligence; a paradigm shift referred to as the fourth industrial revolution (Schwab, 2016). An essential enabler for this digital transformation is the exponential growth in internet connectivity in almost every country, exemplified by the broadband subscription penetration in Africa (currently 81%; Jonsson et al., 2019). Continued growth is expected worldwide as 4G and 5G mobile networks become increasingly available. Another catalyst is the tech companies entering the medical market, applying expertise from algorithms and big data to health problems. There is also a trend towards the "quantified self," which encourages the continuous use of personal tracking devices and stimulates the development of future generations of personal (in-ear) electronics that monitor stress, mental effort, and mental well-being (Crum, 2019).

Other clinical disciplines have implemented computational approaches to parts of the clinical care pathway, but this has not yet resulted in a paradigm shift in health care (Rajkomar et al., 2019). To give a few examples, the field of ophthalmology has adopted the use of automated diagnostic data collection hardware (Bizios et al., 2011). Radiology has begun adopting computational image segmentation for automated diagnoses (Hosny et al., 2018). Genotype information is standardized to evaluate patient health and effective cancer treatment (Benson et al., 2012). Also, mobile phones are becoming standard tools in many disciplines, including diabetes management (Thabit & Hovorka, 2016) and dermatologic diagnoses (Ashique et al., 2015), among many other applications. These are examples of computational approaches for diagnosis, self-evaluation, and treatment. Unfortunately, all the different components identified have developed across different fields – there is no clear indication that all have been applied to a single field. Therefore, if clinical audiology adopts most of the principles defining computational audiology, it can generally become a standard-bearer for modern clinical care delivery. In this perspective paper, we sketch out how computational approaches may further develop audiology and illustrate fundamental advances in diagnosis, therapy, and rehabilitation that could become essential elements in a comprehensive digital transformation of clinical audiology.

Definition and examples of computational audiology that may improve precision

Audiology is an exceptionally strong candidate for computational augmentation and may benefit from the current and novel power of computational science because of its strong mechanistic theory, numerical nature, measurement-driven

procedures, and the multitude of clinical decisions to be made. Here, we introduce the term *computational audiology*, which we define as:

computational audiology

- The approach to diagnosis, treatment, and rehabilitation in audiology that uses algorithms and data-driven modeling techniques, including machine learning and data mining, to generate diagnostic and therapeutic inferences and to increase knowledge of the auditory system;
- leverages current biological, clinical and behavioral theory and evidence;
- provides or augments actionable expertise for patients and care providers.

The readily quantifiable nature of audiological procedures makes audiology well suited for modern machine learning and data collection techniques. Translational reasons to apply computational techniques in audiology include (i) improved accuracy, increased speed, wider application of (diagnostic) tests and evaluation (applied to, e.g. audiometry; Schlittenlacher et al., 2018b); (ii) objective and consistent interventions, outcomes, and decisions across clinicians and clinics (applied to, e.g. CI-fitting; Meeuws et al., 2017). Over time, algorithms can become more sophisticated and take over tasks now performed by humans or take on tasks that are currently not performed due to a lack of resources, time, or clinical consequences, including screening for milder forms of hearing impairment. Computational audiology can improve care by dealing with multifactorial data, including indices of psychosocial well-being, quality of life, co-morbidity, and patient-centered, individual descriptors of complaints and symptoms. For example, Palacios et al. (2020) used an unsupervised learning approach to study heterogeneity of patients suffering from tinnitus by analyzing the complaints and symptoms described in an online patient forum. In addition to deterministic methods, it also facilitates the use of probabilistic methods that include uncertainty and likelihood to cope with the wide variability across people with hearing loss.

The application of algorithms in audiology is not new. Historically, it has been restricted mainly to cohort-level inference, for example, in understanding the incidence and degree of hearing loss in the general population (Mościcki et al., 1985), and the prescription of sound-amplification for different types and degrees of hearing loss (Byrne et al., 2001). Individual refinement based on learning systems could be a promising way forward but raises many challenges to perform in an evidence-based manner (Barbour, 2018).

Diagnostics. In general, diagnostic procedures in audiology consist of a sequence of psychometric and physiologic tests. Clinicians may benefit from computational augmentation because they need to deal with uncertainty, time constraints for testing, and the individual features of the patient. Clinical experts will typically evaluate test results visually and from summary statistics (e.g. average HL), which requires skill and experience but also introduces subjective variability in interpretation, restricts estimates on the certainty of the overall outcome, and impedes more advanced (multifactorial) analysis, which is difficult for humans (Kahneman, 2011).

Limited time for testing is arguably the most significant constraint in collecting high-quality multi-dimensional data for an individual patient. However, machine learning allows, in principle, for flexible, efficient estimation tools that do not require excessive testing time. In an approach known as active learning, new computational tools actively determine which stimuli would be most valuable to deliver to converge onto an accurate estimate rapidly. Active learning was recently applied to diagnostic tests including basic audiometry (Barbour, Howard, et al., 2019; Schlittenlacher et al., 2018), determination of the edge frequency of a high-frequency dead region in the cochlea (Schlittenlacher et al., 2018a), and hearing aid personalization (Nielsen et al., 2014). Also, when multiple factors that share some relationship are available, an active learning method can learn and exploit the relationships in real-time. For instance, data from the National Institute for Occupational Safety and Health (NIOSH) database (Masterson et al., 2013) has been deployed as Bayesian "prior beliefs" to assess the similarity between ears of 1 million participants. A bilateral audiogram procedure that uses these priors speeds up testing considerably (Barbour, Howard, et al., 2019).

Principles of computational audiology may be applied to current research and clinical issues. For example, machine learning approaches to image analyses of otoscopy of the eardrum demonstrate the potential to supplement audiological tests with a diagnosis of potential outer and middle ear pathology (Cha et al., 2019; Myburgh et al., 2018). With a reported accuracy of between 81 to 94% and options for capturing and receiving diagnosis using mobile phone-based otoscopy, these approaches provide direct feedback to the clinician and therefore could allow point-of-care interventions and optimize current care (Cha et al., 2019; Myburgh et al., 2018).

Combining self-reported difficulty and genetic data may lead to potential candidate genes for hearing loss. Such a procedure, applied to the data from 250,000 people, identified 44 new genetic loci potentially associated with hearing loss (Wells et

al., 2019). Individual, patient-centered (hearing) health care could become more comprehensive by collecting more extensive hearing profiles combined with other patient characteristics beyond the audiogram (Sanchez Lopez et al., 2018). For example, the genetic profile (Hildebrand et al., 2009) can be used to differentiate better various underlying causes leading to hearing loss (Dubno et al., 2013). A probabilistic interpretation of a patient profile can be further refined using auditory modeling (Verhulst et al., 2018) and AI and, among other applications, form the basis for prognosis. It is paramount to know the underlying pathology to determine a specific target therapy or rehabilitation strategy. By combining these examples, audiology may become a prime example of precision medicine.

Rehabilitation. When fitting cochlear implants or hearing aids, machine learning may help clinicians optimize parameters by minimizing a cost function. A recently developed clinical decision support system calculates a utility function based on a weighted combination of outcome measures (Meeuws et al., 2017). The utility function is continuously updated as the system learns from previous outcomes. The system also incorporates active learning by determining which of the collected outcomes are most clinically useful. Such a system can oversee the effect of considerably more fitting parameters than those commonly adjusted by audiologists. It can be used to make more accurate predictions of the expected outcome, enable cost-benefit evaluation by reducing the time needed by a trained professional to perform tests, and facilitate a more standardized CI fitting (Meeuws et al., 2017). In the future, the system might be extended to individualized cochlear implant surgery based on high-resolution medical images of the cochlea (Heutink et al., 2020). Also, users' preferences can be collected to make data-driven, individual adjustments to their cochlear implant or hearing aid. The internet of things provides suitable interfaces for users to provide feedback under ecologically valid circumstances (e.g. Ecological Momentary Assessments, EMA; Wu et al., 2015), but also provides tools that monitor behavior that could serve as a proxy to derive user preferences (Johansen et al., 2018).

Another example of computational approaches to improve rehabilitation is applying neural networks to enhance speech-in-noise understanding in cochlear implant (CI) users (Goehring et al., 2019). Noisy speech signals were decomposed into time-frequency units, extracting a set of psychophysically-verified features, fed into a neural network to select frequency channels with a higher signal-to-noise ratio. This pre-processing of the input signal significantly increased speech understanding, even of unfamiliar speakers (i.e. not used to train the network). The

developers limited the required computational power and memory for their model to make it implementable on mobile devices.

Hearing research. Machine learning techniques could also lead to better models of human auditory behavior and a better understanding of the auditory system. Recently, Ausili (2019) used a neural network to model experience-dependent sound localization for different hearing impairments. Deep neural networks are achieving parity with humans for some tasks, and it is possible that these networks could mimic aspects of representation and functional organization of the human brain (Güçlü & van Gerven, 2017; Huang et al., 2018; Kell & McDermott, 2019).

We can conclude that the trend of applying computational approaches in audiology could lead to more individualized hearing care and new services, as illustrated in Example 1. We base this claim on above-cited examples in diagnosis, rehabilitation, and hearing research, and on computational approaches in audiology already employed by digital hearing health technologies around the world (Swanepoel & Hall, 2020). A part of these new services could be provided by companies that traditionally did not specifically target customers with hearing loss. For example, speech-to-text apps provide new functionality to people with hearing loss (Pragt et al., 2020), and AirPods Pro are nearing the functionality of hearing aids (Bailey, 2020) but do not yet fulfill all FDA requirements and fall short in terms of amplification for the rehabilitation of people with moderate to severe hearing loss.

Example 1 rehabilitation service (based on Crum, 2019).

A person tests her hearing with an app to find out that her hearing profile is similar to that of 1.7 million other people in a global database who reported good results using hearing aids. She buys two hearing aids and signs up for a service, an app that sends programming instructions and settings to the hearing aids and asks for feedback to ascertain audibility and judge sound quality. Indications of momentary and remaining hearing problems, including expressions like "excuse me" or "what did you say?" are detected using automatic speech recognition. After a couple of weeks, the system provides fine-tuning based on her needs and similarity to other cases. It automatically determines that when entering her local subway station, substantial echo cancellation is needed. After a few years, the system detects specific changes in the spectral quality and patterns of sounds when she speaks. After tracking this trend for several months, the system suggests scheduling an appointment with a physician because these changes can correlate with heart disease.

How could computational approaches improve access to hearing health care?

Hearing health care is challenging to deliver in Low- and Middle-Income Countries (LMICs) because it currently requires specialized equipment and trained professionals. Smartphone-mediated telehealth holds great promise to lower many of these barriers (Swanepoel & Clark, 2019). Smartphone penetration now exceeds 80% in LMICs (Jonsson et al., 2019), and low-cost equipment and robust test procedures are becoming available to perform audiometric (Potgieter et al., 2018; Swanepoel & Clark, 2019) and otologic (Chan et al., 2019) diagnostic measures with acceptable levels of quality and reproducibility. We foresee a considerable growth in mobile app usage for self-administered hearing tests (Hazan et al., 2020; Swanepoel et al., 2019) and self-adjustment by hearing aid users (Søgaard Jensen et al., 2019) that in turn could lead to self-fitted hearing aids. In the simplest form of telehealth, the caregiver and patient are physically separated, and technology facilitates interaction. However, telehealth can be expanded by distributing expert knowledge across the health care delivery system, with clinical expertise incorporated into algorithms employed on devices used by patients or by local caregivers, making hearing health care possible and affordable in remote and underserved areas where experts are lacking, as illustrated in Example 2.

Example 2: hearing screening in early childhood (based on Barbour et al., 2019; Chan et al., 2019; Swanepoel & Clark, 2019).

1. Children in LMICs typically do not have access to hearing screening. However, a community-based project relying on AI assistance offers screening, diagnosis, and referral in underserved communities.
2. Screening is conducted via an automated pure-tone-screening test facilitated on a smartphone for children from 3 to 4 years.
3. Test quality is monitored locally on the smartphone and regionally via uploaded data on a cloud-based data management portal.
4. If a child fails the screening test, an automated report is generated from the cloud-based data management portal and sent to caregivers by text message or email.
5. If the child fails the screening a second time, automated threshold pure-tone audiometry facilitated by an operator and AI-supported middle-ear function assessment is carried out. A clinical decision support system assists local caregivers in diagnosing hearing loss and referring to specialized care.

If screening and diagnosis of hearing loss can be improved in LMICs, the next requirement is to provide specialized care and affordable hearing loss rehabilitation. Global awareness for hearing loss has recently been spurred by the formation of a Lancet Commission examining strategies to reduce the burden of hearing loss (B. S. Wilson et al., 2019). Recommendations include stimulating the development of low-cost hearing prostheses, leveraging smartphone technologies for use as hearing assistive devices, and equipping a small number of specialist centers for medical and surgical management of ear disease. Computational audiology as an emerging field is uniquely positioned to combine inexpensive, ubiquitous hardware and software (e.g. smartphones with apps) and sophisticated multifactorial (meta) data modeling. By transforming cheap hardware and equipping it with (AI-based) software, LMICs can benefit from advanced automated diagnostic tools and interventions to address hearing loss. The overall cost of devices and services incurred per user will drop, which is expected to compensate for the resources needed for building and maintaining the computational infrastructure, defined here as all hardware, software, protocols, practices, and regulation needed to apply computational approaches on an international scale (O'Brien, 2020). An interesting (but solvable) question is how governments, companies, health care providers, and users will together bear the cost of computational infrastructure, R&D, IP, licenses, devices (e.g. smartphones), and other indirect costs. How to align the involved stakeholders together with the potential risks, privacy issues, and technical requirements are the topics that we consider in the next sections.

Ethical considerations and technical requirements concerning computational approaches in hearing health care

Whereas AI applications in audiology outlined previously should be considered an improvement, they may also involve some additional risks.

1. *Unauthorized or undesirable use*. For example, AI researchers recently introduced new lip-reading technology to facilitate speech understanding in people with a hearing impairment. They trained their algorithm on TV footage, and it outperformed expert lip-readers. This solution could, in theory, allow people with hearing loss to augment their speech understanding (Shillingford et al., 2018). However, the technique could also be used for other purposes, including mass surveillance (Metz, 2018). Footage from closed-circuit TV could be fed into the algorithm to track conversations of unknowing citizens, invading their privacy. A similar privacy issue may apply to devices that incorporate tracking technology. GPS can be used to track a smartphone on a rideshare journey, but it can also track smart hearing aids. Current hearing aids can log users' preferences in

particular environments, monitor adjustments users make in each place, log those preferences, use GPS to detect when they return to those places, and automatically or manually reactivate the preferred settings (Wolfgang, 2019). In courts, tracking the whereabouts of personal devices has already led to erroneous criminal accusations (Valentino-DeVries, 2019).

2. *Bias in the data used to train an AI-system*. Buolamwini (2017), for example, uncovered large gender and racial biases in face recognition systems sold by tech giants IBM and Microsoft. Errors in gender identification were substantially lower for lighter-skinned men (1% error rate) than darker-skinned women (35% error rate). One explanation was that the face recognition systems were trained on data sets containing many more men with light skin than women with dark skin. This example shows that real-world biases may translate to inherent biases in the outcome of AI systems, whether we are aware of those biases or not. As a result, it might be a risk to apply data collected in, for example, Western countries to solutions for non-Western regions with other ethnic characteristics, including race and lifestyle.

3. *Violation of privacy.* Privacy protection has begun to be taken seriously in recent years, resulting in the EU's GDPR (Regulation (EU) 2016/679, 2016). In addition to general privacy issues, one article of the GDPR explicitly states that individuals should not be subjected to a decision based on automatic processing, including profiling, except when explicit consent is given (Goodman & Flaxman, 2017). Manufacturers of hearing devices and cochlear implants are already collecting large bodies of data (data profiles) beyond the view of (independent) publicly funded hearing health care providers and researchers. Clinicians use that data for counseling purposes, for instance, to evaluate hearing aid usage based on data-logging (Saunders et al., 2020). However, Mellor et al. (2018) reported that a hearing aid manufacturer did share a large dataset but did not share possibly relevant commercially sensitive information, which may limit insights drawn by researchers from the data. Automatic processing could be problematic with machine learning and big-data designs, even using anonymous data only. When a database uses many types of data from individual subjects, it will increase the likelihood that data can be traced back to individuals (re-identification; Leese, 2014; Rocher et al., 2019). Privacy concerns and the sheer amount of data have led to the development of distributed learning, an approach that allows for decentralized training (Konečný et al., 2016). For example, in federated learning, models are trained locally on a local device (e.g. a smartphone connected to a hearing aid; Szatmari et al., 2020), and only aggregate meta-data (updated priors) travel from central databases to users and back.

4. Restricted access and control over data. All human stakeholders must have access to relevant information to make the right decision about the diagnosis, treatment, or rehabilitation that affects a patient's health. Data from which relevant information could be extracted is currently scattered across databases residing with different stakeholders (i.e. companies, hospitals, research institutions). The data are collected for distinct purposes and might have a particular status, for example, proprietary or open. In effect, data are vital for so many processes that control over them may lead to a strategic advantage in business, clinical care, or science. Companies might collect data to improve products (proprietary data) or evaluate services, but also because of legal requirements or for quality assurance. It is mandatory for health care professionals to keep a medical record that contains all information needed to provide accountable care according to good clinical practices[1] (article 454 WGBO; Eijpe, 2014). The Health Insurance Portability and Accountability Act (HIPAA) in the US and General Data Protection Regulation (GDPR) in the EU provide the legislative framework that enables patients and care providers access and control over personal data (Forrest, 2018; Individuals' Right under HIPAA to Access Their Health Information, 2016). An individual can request access to his/her data stored by a health care provider (HIPAA) or any organization (GDPR). Therefore, in theory, it is possible to create a global system that can access patients' health history. However, in reality, appropriate data-exchange practices are lacking, which seriously hampers patients' control over their data. The (re)use of proprietary data can be restricted and is subject to trade secrets, patents, copyrights, or licenses (e.g. see for legal rights governing research data; Carroll, 2015; and for property rights; Stepanov, 2020). Vested interests, a motivation to influence factors for your benefit, is a considerable barrier to the reuse of proprietary data. Without access to relevant information, patients cannot make informed (shared) decisions. Clinicians will lack insight into decision support systems, regulators will be unable to inspect and audit, and researchers will be unable to appraise outcomes and methods critically.

5. Liability. For anyone working with new AI paradigms, it needs to be clear who is responsible if anything goes wrong. Is it the scientist who made the algorithm, is it the health care professional, or is it the patient who is ultimately responsible for their own decisions? For example, how can a clinician (or a patient) ascertain that an algorithm's outcome is correct and valid? An explicit example of a potentially invalid test result is an auditory steady-state response (ASSR) exam performed on a restless neonate that results in measurement conditions markedly different from the conditions on which the algorithm was trained (Sininger et al., 2018). The test result may not be accurate, but this shortcoming might not be noticeable to the clinician.

---

[1] For the following examples we chose to apply Dutch law to illustrate a legal framework.

Oversight and regulation (in general for medicine; Maddox et al., 2019) for hearing-related AI also needs to be in place. The level of this oversight will need to be increased in cases of highly autonomous and self-learning clinical decision support systems operating in highly complex environments and circumstances that have severe consequences for erroneous actions. Furthermore, AI-based clinical decision support systems need to be transparent to inspection and audit, and robust for application in a specified context (in general for health technology; Shuren & Califf, 2016).

The role of computational audiology in personalized hearing health care

AI, automation, and remote care will become more widespread and better available in the coming years. Redesigning the clinical workflow, implementing AI technology, and changing the clinician's role should become a top priority (Rajkomar et al., 2019). Below we discuss what role clinicians and other stakeholders might play in the digital transition and its meaning for patients. Already, remote care has become more mainstream due to the COVID-19 pandemic that has provided an unprecedented impetus to develop and employ hearing health solutions that reduce physical contact (De Sousa, Smits, et al., 2020; Swanepoel & Hall, 2020). This situation has demonstrated that clinicians can adapt if appropriate benefits are clear; for example, keeping practice doors open.

Clinicians' role. Hearing health care professionals, including audiologists, have valuable insights needed to implement these new approaches successfully. For instance, algorithmic bias is reduced if a system is trained in a situation comparable to where it is employed. Therefore, early involvement by hearing care professionals in the design of algorithms could lead to products that better fit the clinical pathway. In a concept mapping study, a structured method to produce a conceptual representation, clinicians from Canada reported that structural training on implementation and best-practices of remote care is needed (Davies-Venn & Glista, 2019). Also, the application of AI requires clinicians to have appropriate training to use AI tools and to be aware of their validity and limitations as well as how to use them. Clinicians should also use their position (e.g. in a collective) to advocate for necessary user requirements, including transparency and clarity, so that, as professionals, they can take responsibility for actions and decisions supported by those systems.

Not everything valuable in hearing health care is quantifiable and automatable. Machines do not easily replace a clinician in aspects of care based on clinical

judgment, soft skills, and the personal touch that help the clinician understand the patient's needs. Clinicians need to see the patient's perspective while offering knowledge, creating realistic expectations, providing rehabilitation, and collecting feedback. They are also the mediators that counsel patients in using remote care options, translating outcomes to individual cases, and interpreting results from AI approaches.

Automation of routine diagnostic procedures might free up clinician time to design more elaborate therapeutic interventions, rehabilitation strategies, or even patient engagement/education initiatives. Technical tasks, including hearing tests and hearing aid fitting, will benefit from best practices for accuracy and efficiency standardized in automated routines. One example could be visual reinforcement audiometry (VRA) for infants, which currently requires two clinicians to implement: one that conditions the child while the other selects each stimulus and the timing of its delivery. Suppose the stimulus selection is optimized through active learning. In that case, a single clinician could condition the child while also registering responses and selecting the timing of delivery with a handheld remote. The result would be more accurate test results with half the labor, potentially enabling a practice to double its patient throughput. In considering such scenarios, clinician concerns about becoming marginalized in the face of automation deserve consideration. AI technology can eventually standardize best practices of efficiency and effectiveness for all clinicians while preserving the necessary human element of care that only a person can provide. In no way are these ideas intended to take clinicians out of the loop or diminish their contribution. On the contrary, their new ability to reach more patients and provide better care is expected to expand their clinical impact.

Collaboration among stakeholders. This paper attempts to start the dialogue needed to create a shared vision among stakeholders regarding computational audiology, one of the first steps towards effective collaboration. As examples, one could think of health care decision-making and advocacy groups including health departments; non-governmental organizations including WHO and patient associations; but also hearing health care professionals including medical doctors and audiologists; device manufacturers, insurance companies, and researchers in audiology. The way to get there could be by stakeholder collaboration, for which Sekhri et al. (2011) provide successful examples within medicine, which we regard as a necessary step to implement the current advances in computational audiology on a large scale. Besides a shared vision, we also need to think about aligning the interests of above stakeholders. By putting patients' interests first and creating the proper incentives,

i.e. rewards that encourage people or organizations to do something, we may overcome professional inertia, defined here as the resistance to change. For this, we need to assess and create awareness about vested interests that hamper innovation (e.g. reimbursement policies, Davies-Venn & Glista, 2019) and find common ground so that by collaboration, we can jointly overcome the barriers and all benefit fairly from the forthcoming advances.

An opportunity to further improve diagnostic and therapeutic procedures is to make anonymous data openly available so that algorithms can train on larger populations. All stakeholders involved who collect data should apply privacy guarded-by-design, which requires built-in safety measures to protect patients' privacy (A&L Goodbody, 2016). These measures should require all stakeholders to assume responsibility for their specified share within the system. A prerequisite for collaboration is the standardization of clinical procedures and how data is stored and annotated within a computational infrastructure. Only then is pooling of high-quality data possible. The time of small-scale research with small (uniform) samples should be consigned to the past. Here, we may learn from other fields. For example, in neuroimaging and genetics, research groups started a consortium to facilitate data aggregation and sharing on a scale unprecedented in audiology (Bis et al., 2012).

Standardization would help clinicians collect evidence and create independent outcome measures to assess new tools and comparing them with established and validated methods and it ensures that clinicians are talking about the same thing when operating within a network of distributed expertise. Besides, by enabling interoperability between manufacturers and clinics, clinical procedures can be more readily adopted. Interoperable systems in combination with licenses to protect proprietary data will reduce risk and costs for companies (e.g. missing out on a standard, maintaining a platform, adhering to regulatory requirements). These systems keep the option open to compete and excel, and tackle the problem of vendor lock-in that currently limits freedom of choice for clinics and patients.

What does computational audiology mean to patients? For many people worldwide, the access to screening and diagnosis of hearing impairment will improve. The complexity of a patient's hearing problem and his / her self-reliance will determine the required degree of professional guidance. A large group with mild and moderate hearing loss may be helped with relatively simple devices and may even apply forms of self-care. More intensified professional help is needed for

more complex fittings or for people who cannot apply self-care (e.g. those with specific co-morbidities).

We believe it is still a significant challenge to make self-care by people with hearing loss possible even for those with sufficient autonomy and health literacy, for reasons including lack of trust in the transition and how digital information is presented and exchanged between patients, clinicians, and companies. If information is not clear to the patient, how can he/she act upon it? Clinicians will play an essential role in maintaining patient trust in the transition and adapting to new practices. Hearing health care may evolve to the point where parts of care are organized remotely, for instance, screening of hearing loss, monitoring the status quo, and making adjustments to rehabilitation depending on the patient's situation.

The future of audiology

Modernization of audiology towards greater quality, accessibility, and equity will benefit immensely from the emerging power of computational sciences. We envision a future where patient well-being is promoted by judicious evaluation of data shared between interoperable systems of public or private origin. Health care providers will adopt expanded roles within a network of distributed expertise that continually updates best practices as they are accumulated and quantified. Clinicians will be empowered to reach more patients by offloading decisions about data collection to supportive tools while reserving complex and rare clinical decisions for human experts. In the next decade, we foresee that widely available devices, including smartphones, will catalyze the democratization of audiology and benefit millions of people who suffer from the disabling effects of hearing loss by helping evaluate and treat them with support and guidance from advanced algorithms. For this to happen, we must join forces with experts in computational sciences, agree on global standards and evidence-based procedures, and carefully consider the possible challenges of big data and AI technology.

3

The Rise of AI Chatbots in Hearing Health Care

De Wet Swanepoel, Vinaya Manchaiah, Jan-Willem Wasmann

Swanepoel, D. W., Manchaiah, V., & Wasmann, J.-W. A. (2023). The Rise of AI Chatbots in Hearing Health Care. *The Hearing Journal*, *76*(04), 26. https://doi.org/10.1097/01.HJ.0000927336.03567.3e

ABSTRACT

In 2022 and 2023, the world witnessed the advent of AI chatbots driven by large language models (LLMs). These AI chatbots, such as ChatGPT, promise human-like interactions but are primarily regarded as text generators, lacking a deep understanding of healthcare intricacies. This brief perspective paper explores the potential and limitations of AI chatbots in hearing health care, focusing on their role in providing personalized patient support and aiding hearing healthcare researchers. Despite their inherent limitations, we speculate that AI chatbots can enhance healthcare accessibility, improve patient outcomes, and support research.

Contrary to the conventional belief that AI chatbots are limited to text generation, our findings illuminate their broader healthcare potential. In addition, they can offer 24/7 health advice, potentially reducing the need for in-person consultations and contributing valuable data to healthcare research. The emergence of AI chatbots signifies a paradigm shift in healthcare accessibility and data acquisition. Collaboration among healthcare professionals, researchers, and policymakers is essential to maximize their benefits while ensuring ethical design, awareness of potential biases, and responsible use. This collaborative effort is vital for the ethical, efficient, and safe implementation of AI chatbots in hearing health care.

Key words: AI Chatbots, Hearing Health Care, Large Language Models (LLMs), Healthcare Accessibility, Referral and Consultation

Acknowledgements

We would like to acknowledge the contribution of ChatGPT, an AI chatbot trained by OpenAI using a large language model (LLM), in providing valuable insights and guidance for this article. We experimented with prompt engineering and had conversations with ChatGPT playing the role of patient and clinician to get a first impression of what AI chatbots, such as ChatGPT, could offer and not offer.

INTRODUCTION

One of the most exciting recent technological innovations has been the deployment of artificial intelligence (AI) chatbots based on large language models (LLM). AI chatbots are a type of generative AI that can generate text. Other examples of generative AI create pictures (e.g., DALL-E or Stable Diffusion; See Figure 1 and 2 for examples) or music (e.g., Jukebox). In November 2022, OpenAI launched ChatGPT publicly, an AI chatbot that can engage in conversations in response to questions from the user, so-called prompts, generating responses to users' questions (i.e., prompts) that are almost indistinguishable from those of humans. The launch of ChatGPT represents a technological revolution, one that could change the face of healthcare as we know it, including hearing healthcare. ChatGPT is not an isolated example but part of a global race of who can develop the most compelling AI chatbot. Besides OpenAI, which is financed by Microsoft, other large corporations such as Meta, Google, and Tencent have launched similar proprietary products based on LLM (e.g., LaMDA).

Figure 1. *DALL·E created artwork. Prompt "The rise of AI chatbots in hearing healthcare, digital art."*

AI Chatbots and Healthcare

AI chatbots are computer programs that use natural language processing (NLP) to communicate with humans. They are trained on large collections of language (e.g., all written books and most of the internet) to predict what response is most likely to a wide range of queries. For a human user, it may appear as if the system understands the question and can provide personalized advice, recommendations, and support. In reality, chatbots have no understanding of the world around them nor of the human body and its health status. Still, the potential applications for AI chatbots in healthcare are broad, with use cases for patients, clinicians, researchers, and training students (T. H. Kung et al., 2023).

Table 1. *AI chatbots in hearing health care - applications, risks and research priorities for patients, clinicians and researchers*

Target users	Potential Applications	Potential Risks	Examples of Research Priorities
Patients	Initial screening and recommendation of interventions	Inaccurate or misleading information	Efficacy of AI chatbots in providing accurate and reliable information to patients
	Education and support	Over-reliance on chatbots for decision making	Impact of AI chatbots on patient outcomes and satisfaction with care
	Reminders and follow-up	Loss of human touch and emotional support	Effectiveness of chatbots in improving adherence to treatment plans
	Tele-audiology services	Potential for technical issues or difficulties with communication	Feasibility and acceptability of tele-audiology services assisted by chatbots
Clinicians	Data collection and analysis	Loss of empathy and understanding in patient care	Integration of AI chatbots with existing healthcare systems
	Decision support	Misdiagnosis or delayed diagnosis due to chatbot errors	Evaluation of chatbot accuracy and reliability
	Patient triage and referral	Incomplete patient information leading to improper triage or referral	Feasibility and effectiveness of chatbots in improving patient triage and referral
Researchers	Data collection and analysis	Incomplete or inaccurate data collection	Development of standardized protocols for chatbot data collection Open source systems so that one can better judge how data is used?
	Participant recruitment	Potential for selection bias in participant recruitment	Comparison of chatbot-assisted and traditional research methods
	Cognitive testing and assessment	Limitations of chatbots in capturing complex cognitive processes	Evaluation of the validity and reliability of chatbot-assisted cognitive testing and assessment

The broad trend for the use of AI chatbots in healthcare is to increase accessibility (to medical knowledge) and affordability of care. Chatbots can provide 24/7 access to healthcare advice and support, reducing the need for in-person consultations, and potentially improving patient outcomes. Additionally, AI chatbots could potentially provide valuable insights and data to healthcare professionals, allowing them to make more informed decisions about patient care. More transparency on the data these chatbots have access to and use to produce their output is important and has been raised as a concern with regards to existing systems (Van Dis et al., 2023).

In hearing healthcare, chatbots could be used to support patients, clinicians, and researchers (Table 1).

Patients and AI Chatbots

Patients can benefit from AI chatbots in hearing healthcare in various ways. One potential application is for initial screening and the recommendation of interventions. For example, a patient could interact with a chatbot that asks about their symptoms and hearing history and provides recommendations for self-management of symptoms, further evaluation, or treatment based on the patient's responses (Wasmann & Swanepoel, 2023). This could be particularly useful in cases where patients are unsure whether or not they are experiencing hearing loss, or are hesitant to seek medical attention, or where a profound hearing loss inhibits a conversation with a clinician. Chatbots can also serve as educational resources, self-management tools and screening tools for comorbidities, including social needs (Kocielnik et al., 2020). Patients can receive information about hearing health, prevention tips, and advice on how to manage hearing conditions. Chatbots can provide information on the use of management options such as hearing aids, how to change batteries, and troubleshooting common issues. However, a potential risk is that chatbots may not provide accurate recommendations, leading to delayed diagnosis or inappropriate treatment.

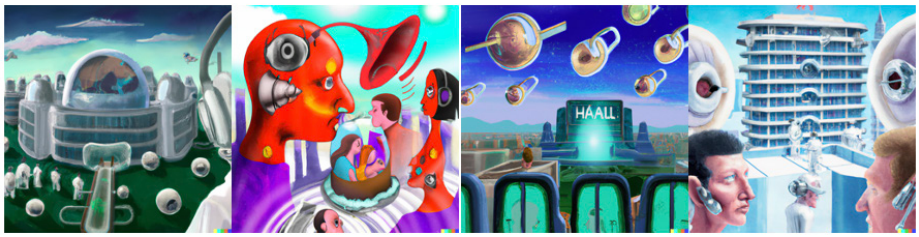

***Figure 2**. DALL·E created artwork. Prompt "A futuristic illustration of the "Planet of the AI chatbots in hearing health care, a movie about invading an ear that is also an hospital."*

Clinicians and AI Chatbots

Clinicians can benefit from AI chatbots in hearing healthcare in various ways. Chatbots can assist with data collection and analysis by collecting data on patients' hearing health, such as self-reported symptoms or hearing aid usage. Chatbots can provide summary reports or visualizations to help clinicians make treatment decisions, such as providing a summary report of a patient's hearing test results, highlighting areas of concern, and providing recommendations for further evaluation or treatment. Another potential application is to assist with decision-making and treatment planning. For medical applications, Google and Deep Mind developed Med-Palm, a LLM that incorporated clinical knowledge that has been evaluated using newly developed benchmarks (Singhal et al., 2023). Chatbots that unlock clinical knowledge could suggest treatment options based on a patient's hearing health history and symptoms and provide information on the benefits and risks of each option. For instance, chatbots could suggest a specific type of treatment based on a patient's hearing test results and preferences. Chatbots can also support clinicians in their communication of information in more accessible and person-centered ways.

A potential risk is that chatbots may not provide the same level of clinical judgment and decision-making as a human healthcare professional. Additionally, there is a risk that the data collected by chatbots may be inaccurate, incomplete, biased, or dated, which could lead to misdiagnosis or inappropriate treatment.

Hearing Researchers and AI Chatbots

Researchers can benefit from AI chatbots in hearing healthcare in various ways. Chatbots can collect large amounts of data from diverse populations, providing researchers with valuable insights into the prevalence and impact of hearing loss. For instance, chatbots can potentially collect data on the prevalence of tinnitus in different countries or regions. Another potential application is to facilitate clinical trials and research studies. Chatbots can screen potential participants for eligibility, collect informed consent, and administer study protocols (Kocielnik et al., 2020). For example, chatbots can collect self-reported data on hearing aid usage and satisfaction in large-scale clinical trials.

However, a potential risk is that the data collected by chatbots may be incomplete or biased, particularly if the chatbots are only accessible to certain populations or if the questions asked by the chatbots are not culturally sensitive or appropriate for all participants (Van Dis et al., 2023). Additionally, chatbots may inadvertently

exclude certain populations from research studies, such as individuals who do not have access to technology or who are not comfortable using it.

Current Priorities

There is an urgent priority to investigate the (clinical) application of AI chatbots in hearing health care. General guidelines for the appropriate use of AI chatbots by researchers are being developed in this rapidly changing landscape. Academic journals have broadly agreed that chatbots may not be co-author on research papers since they cannot take responsibility for their work (Stokel-Walker, 2023; Van Dis et al., 2023). In terms of hearing research applications, priority should be given to evaluate the validity and reliability of chatbots in collecting and analyzing hearing health data. Researchers and clinicians need to ensure that chatbots can provide accurate recommendations and treatment options and that the data collected by chatbots is reliable.

Usability is another important research priority to ensure that chatbots are user-friendly and accessible to as many patients as possible, regardless of their age or technological literacy. Cultural sensitivity is also important to ensure that chatbots are culturally sensitive and appropriate for all populations. There are also important ethical considerations for using chatbots in hearing healthcare, including issues related to informed consent, data privacy, and data security. Researchers will also need to assess long-term outcomes of using chatbots in hearing healthcare. This includes evaluating the impact of chatbots on patient outcomes such as quality of life, satisfaction, and adherence to treatment. Overall, the research priorities for AI chatbots in hearing research should focus on ensuring that chatbots are accurate, reliable, accessible, and culturally sensitive.

Guidelines for appropriate use of AI chatbots by clinicians or patients are not yet available. As the language models have been trained largely by using text from the internet, they are likely to have the same general opinions, stereotypes and biases that are present on the internet. For this reason, we see a task for specialists and patient organizations to test what prompts yield the best results and provide the guidelines to avoid misuse or misunderstandings.

CONCLUSION

The rise of AI chatbots (based on LLMs) represents a significant technological advancement that has the potential to revolutionize hearing health care. AI chatbots have the potential to provide personalized advice and support to patients while also providing valuable insights and data to healthcare professionals. However, it is important to consider the potential risks and benefits of AI chatbots and to prioritize further research to ensure that these technologies are used ethically, effectively, and safely in hearing health care.

4

Preliminary Evaluation of Automated Speech Recognition Apps for People with Hearing Loss

Leontien Pragt, Peter van Hengel, Dagmar Grob, Jan-Willem Wasmann

Pragt, L., van Hengel, P., Grob, D., & Wasmann, J.-W. A. (2022). Preliminary Evaluation of Automated Speech Recognition Apps for the Hearing Impaired and Deaf. *Frontiers in Digital Health*, *4*, 806076. https://doi.org/10.3389/fdgth.2022.806076

*The wording of this published text has been adapted to make it more inclusive, using terms as “people with hearing loss” avoiding as much as possible medicalized words including "impairment" and "deficit".

ABSTRACT

Objective: Automated speech recognition (ASR) systems have become increasingly sophisticated, accurate, and deployable on many digital devices, including smartphones. 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 with people with normal hearing and people with hearing loss 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 apps' 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. AVA and Live Transcribe had the lowest (best) signal-to-noise ratio +8 dB on the DIN-test. 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 those with moderate hearing loss. The ASR apps performed more poorly in noise than most people with profound hearing loss who use 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.

Key words: automated speech audiometry (automatic speech recognition), automated speech recognition, (ASR), evaluation metric, hearing loss, speech-to-text, voice-to-text technology

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 (Saon et al., 2017; Xiong et al., 2017). 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 automatic captions. Also, several ASR-based speech-to-text apps have been developed for people with hearing loss, providing live captioning of conversations (Kader et al., 2021; Xiong et al., 2017), showing the potential of automation and artificial intelligence for hearing healthcare (Lesica et al., 2021; Wasmann et al., 2021). 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 people with hearing loss beyond what was shared by developers. Nor did we have clear criteria for which groups of patients we might suggest the ASR apps to.

Since 2017, several ASR systems have claimed speech recognition performance close to that of people with normal hearing (Saon et al., 2017; Xiong et al., 2017). 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 (Jurafsky & Martin, 2009). 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 (Koenecke et al., 2020). It is therefore important to understand for what specific task an ASR is designed for and how it is evaluated. Typically ASRs are evaluated on well-studied large (>100 hours) collections of speech, referred to as a corpus. The SwitchBoard corpus and CallHome corpus are well-known collections of conversational phone calls (Cieri et al., 2004), 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 (Godfrey et al., 1992). The CallHome corpus consists of more informal conversations between friends and family (Cieri et al., 2004). 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 (Godfrey et al., 1992), and a large portion of the Librispeech corpus has undergone noise removal and volume normalization (Panayotov et al., 2015).

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. (2017) 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 (Saon et al., 2017). Xiong et al. (2017) on the other hand, followed more realistic industry standard procedures, which are similar to how speech is processed by ASR (Xiong et al., 2017). 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 (Kincaid, 2018), which is close to the performance of professionals with normal hearing reported above (Saon et al., 2017; Xiong et al., 2017). 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 (Kincaid, 2018). Koenecke et al. (2020) 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 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.

For people with hearing loss, there are specific user needs to consider when developing ASR apps. For example, these listeners might use both speechreading (Bernstein et al., 2000) and text reading of the ASR transcript from a screen. Speechreading conveys important nonverbal cues and nuances not included in a transcript and may enhance speech-in-noise abilities (Helfer, 1997). 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; Coldewey, 2020) 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; Google, 2021), 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; Earfy, 2017).

The settings where an ASR Is used may also differ between individuals with hearing loss 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. Adverse acoustic conditions, including low signal-to-noise, make it difficult for people with normal hearing to understand speech and make the speech incomprehensible for people with mild to profound hearing loss. 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 (Saon et al., 2017; Xiong et al., 2017) and should be done using diverse training datasets (Koenecke et al., 2020).

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 people with normal hearing and people with hearing loss 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) 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 people with hearing loss 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 and inexpensive (less than $10 for a license).

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 (International Organization for Standardization, 2010). Standard clinical calibration protocols were used for all speech material. The microphone of the smartphone was aimed towards the speaker, which we assumed to be the optimal microphone orientation, at approximately the height of a listeners' ears to resemble testing conditions when tested with human listeners. 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 apps' ability to convert speech into text.

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 was the Dutch CNC-test, which consists of phonetically balanced lists of twelve monosyllabic Dutch words in quiet (CNC-list, 'Nederlandse Vereniging voor Audiologie'; Bosman & Smoorenburg, 1995). 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 (Bosman & Smoorenburg, 1995). 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, people with normal hearing achieve 100% phoneme recognition at 45 dB SPL (Bosman & Smoorenburg, 1995).

Figure 1. *Set-up of the smartphone in front of the speaker.*

Second, the Plomp-test (Dutch sentences in noise) was presented (Plomp & Mimpen, 1979b). The test consists of 13 sentences of 8 to 9 syllables presented in noise with the same averaged long-term spectrum as the 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 (Plomp & Mimpen, 1979a). 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 (1979). The test was first performed without the 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 a 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 (Smits et al., 2013).

Finally, a fragment of dialogue in Dutch and English at 72 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. The 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 apps' "speech audiogram" curves look similar to the s-shaped psychometric curve of people with normal hearing determined by Bronkhorst et al. (1993) in 20 university students with normal hearing (Bronkhorst et al., 1993). However, all apps' SRT were between 50 and 60 dB SPL, which is 25 to 35 dB poorer than people with normal hearing who have a SRT around 25 dB SPL (Bosman & Smoorenburg, 1995).

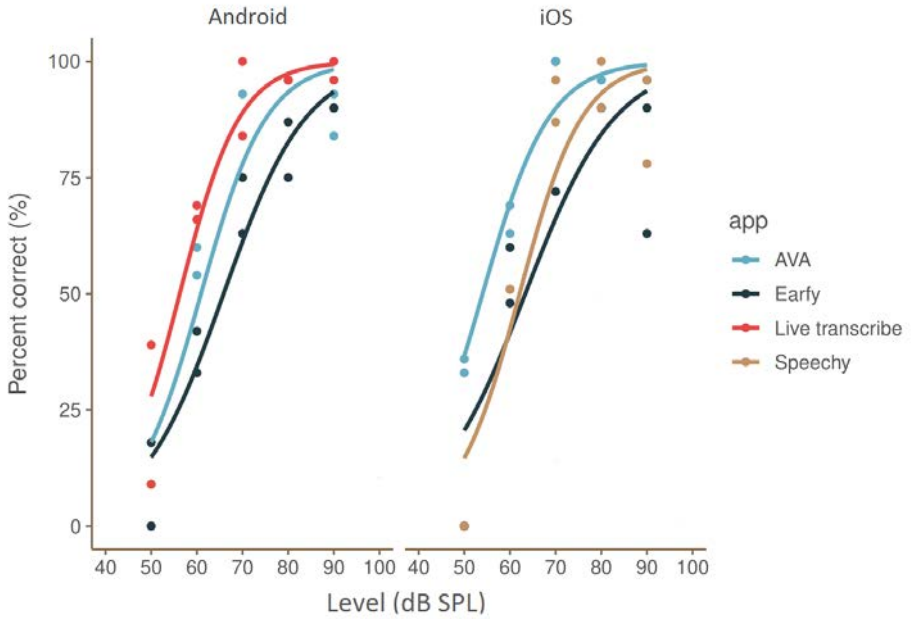

Figure 2. *Speech recognition as a function of presentation level (in human listeners known as speech audiogram) of all ASR apps tested on an Android and iOS smartphone. The plotted lines are interpolated using a logistic function through the measured test-retest data-points. Left side: results of the Android apps, right side: results of the iOS apps.*

The speech-in-noise results are shown in Figures 3 and 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.

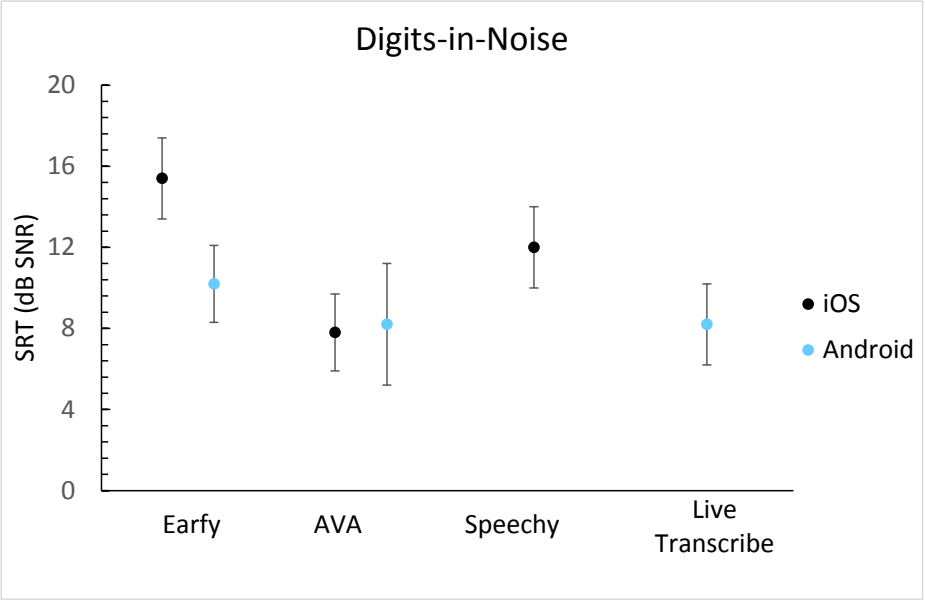

Figure 3. *Digits-in-Noise results expressed in SNR per app. A lower score is better. The error bars represent the standard deviation of the response of the app within a single list of triplets.*

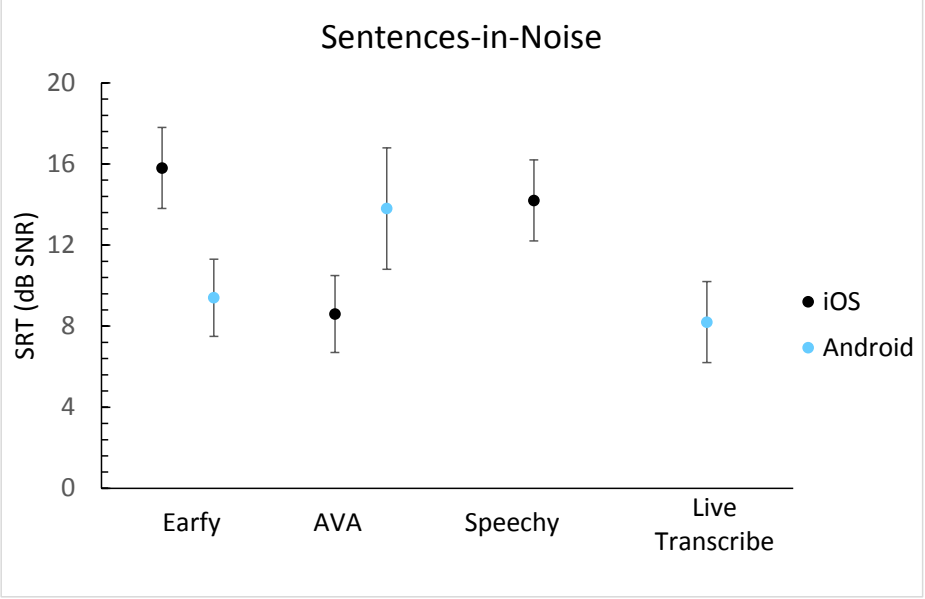

Figure 4. *Sentences in noise results expressed in SNR per app. A lower score is better. The error bars represent the standard deviation of the response of the app within a single list of sentences.*

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%).

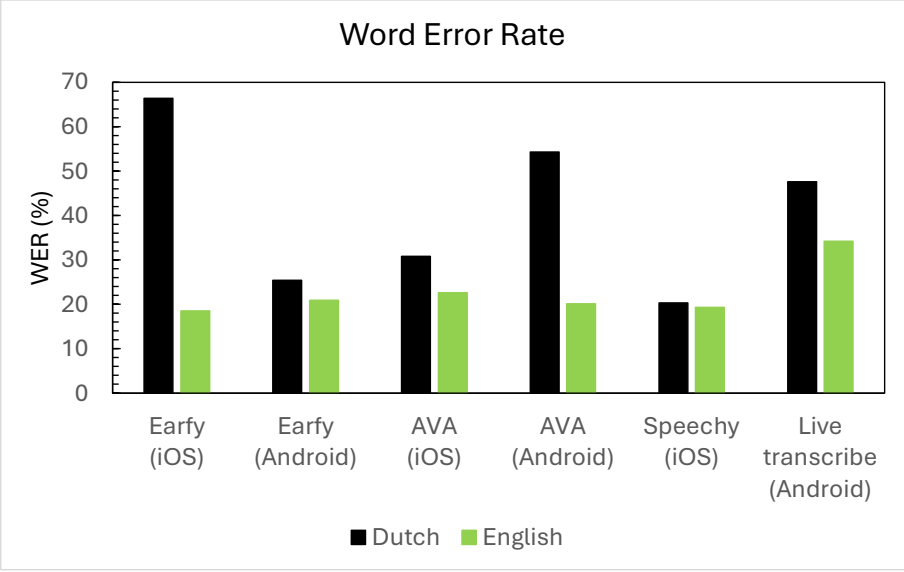

Figure 5. *Word error rate in percentage of the dialogue in English and Dutch for the different apps. The test-retest reliability of the CNC-test can be seen in figure 6.*

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 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 Root-Mean-Square-Difference on the DIN-test was 0.4 dB iOS Ava and 0.8 dB Android Live Transcribe, which we regard as acceptable since in people with normal 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 (Smits et al., 2013). The Root-Mean-Square-Difference on the Plomp-test was 0.6 dB iOS Ava and 2.0 dB for Android Live Transcribe.

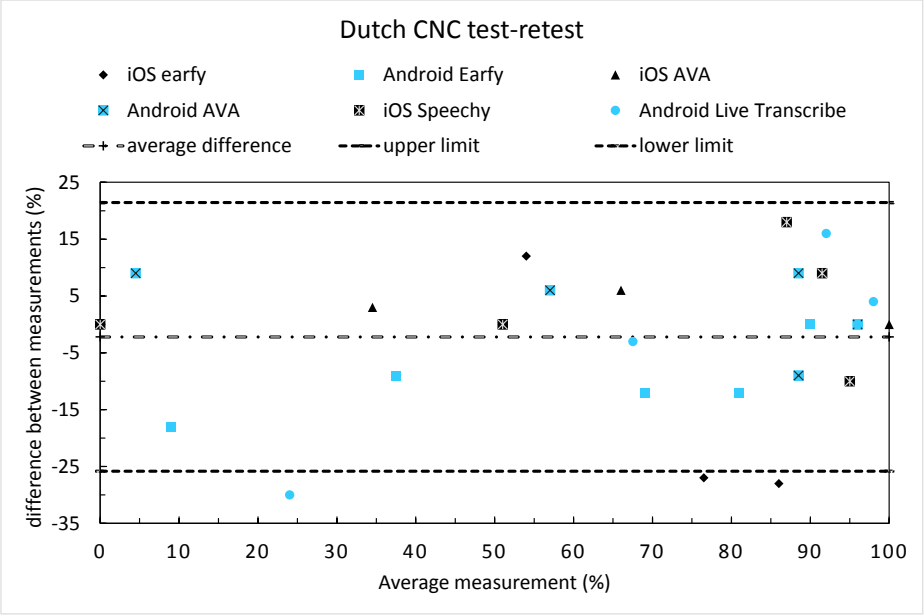

Figure 6. *Bland-Altman plot displaying the test-retest reliability of the CNC-test.*

DISCUSSION

Main results

None of the ASR apps achieved performance close to people with normal hearing on audiological tests. In quiet, ASR apps performed similarly to people 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 people

The performance of the ASR apps on speech-in-quiet tests seems comparable to people with a moderate conductive hearing loss (30-35 dB threshold shift), which is known as disabling for certain activities in daily life (World Health Organization, 2021). In comparison, Dingemanse & Goedegebure (2019) 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 users since they excluded participants with a CNC-score below 60%. Kaandorp et al. (2015) 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-vowel-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 (O'Neill et al., 2020). Most ASR are trained on sentences of realistic conversations (Cieri et al., 2004). The strength of (deep learning) ASR is based on using contextual information from a natural language processing model (Deng, 2016). 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. People with normal hearing achieve on the DIN-test monaurally using headphones an SNR of –8.8 dB (Smits et al., 2013). CI recipients rated on the same criteria as people with normal hearing, typically achieve DIN scores ranging from +3 to -6 dB. For instance, Kaandorp et al. (2015) 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 the ASR apps on sentences in noise (Plomp-test) was very limited and much poorer than in people with a moderate hearing loss (Plomp & Mimpen, 1979b). People with normal hearing have an SRT at an SNR of –8 to –10 dB (Plomp & Mimpen, 1979b), while the best ASR apps achieved +8 dB scores. Kaandorp et al. (2015) found mean SRT on Dutch Sentences in noise by scoring keywords of +2.1 dB for 24 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 (1979; 20). In another study, Kaandorp et al. (2016) 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 apps' performance is far below that of the majority of people with hearing loss 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 people with hearing difficulties. 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 (Kaandorp et al., 2016). 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 people with hearing difficulties.

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 people with profound hearing loss who use a cochlear implant (Blamey et al., 2013) and better than hearing aid users with a profound hearing loss (Flynn et al., 1998). 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 (Faber, 2017). 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 speaker. 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 (Jakovljević et al., 2008). 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; Biadsy et al., 2019; Koenecke et al., 2020). 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 (Christensen, Saunders, Havtorn, et al., 2021). For ecological valid measures, the effect of different fluctuating noise maskers should be considered (Festen & Plomp, 1990; Francart et al., 2011). 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 (Glasser et al., 2017). 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 (Initiative (WAI), 2021).

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 speakers' voice and intelligibility (Yi et al., 2021). 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 people 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 people (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 transcriptions 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 (Kim et al., 2021).

ASR may benefit from domain-specific evaluation tools and have domain-specific applications. For instance, Miner et al. (2020) 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 people with normal hearing 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 (W.-C. Huang et al., 2019). 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 (Biadsy et al., 2019). In addition, the Parrotron system can synthesize the speech of a speech impaired person with impaired speech (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 masker should be used (Festen & Plomp, 1990; Francart et al., 2011). Reverberation times typically encountered in daily life to consider are 0, 0.5, and 2.5 seconds, which corresponds to ideal, classroom (Knecht et al., 2002), and church (Desarnaulds et al., 2002) 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 (Berenger, 2021). In addition, ASR technology, when further developed, can play a role in computational approaches to audiology (Wasmann et al., 2021). 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 (Venail et al., 2016). Manual calculation of complex evaluation metrics is not suitable in clinical settings given the excessive

time required and may lead to inter-rater variability (Smith et al., 2019). Automated speech audiometry using algorithms to score performance can be a valuable complement to automated pure-tone threshold audiometry (Wasmann et al., 2022). For example, Vinail et al (2016) 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. (2021) 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 (Kaandorp et al., 2016; O'Neill et al., 2020). Although scoring keywords makes the test accessible to a larger population, it reduces the discriminative power between higher- and lower-educated native listeners (Kaandorp et al., 2016). 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 measure in terms of how well people with hearing loss can participate in a conversation under adverse circumstances.

CONCLUSION

None of the ASR apps achieved performance close to people with normal hearing on audiological tests. No app stood out from the others on performance level. On audiological speech tests in quiet, ASR apps performed similarly to people 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 versus 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 people with hearing loss, especially CI recipients, report that they benefit from the apps in certain situations (Berenger, 2021), 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 people enjoying the benefits of ASR.

5

Digital Approaches to Automated and Machine Learning Assessments of Hearing: Scoping Review

Jan-Willem Wasmann, Leontien Pragt*, Rob Eikelboom, De Wet Swanepoel*

Wasmann, J.-W. A., Pragt, L., Eikelboom, R., & Swanepoel, D. W. (2022). Digital Approaches to Automated and Machine Learning Assessments of Hearing: Scoping Review. *Journal of Medical Internet Research*, *24*(2), e32581. https://doi.org/10.2196/32581

ABSTRACT

Background: Hearing loss affects 1 in 5 people worldwide and is estimated to affect 1 in 4 by 2050. Treatment relies on the accurate diagnosis of hearing loss; however, this first step is out of reach for >80% of those affected. Increasingly automated approaches are being developed for self-administered digital hearing assessments without the direct involvement of professionals.

Objective: This study aims to provide an overview of digital approaches in automated and machine-learning assessments of hearing using pure-tone audiometry and to focus on the aspects related to accuracy, reliability, and time efficiency. This review is an extension of a 2013 systematic review.

Methods: A search across the electronic databases of PubMed, IEEE, and Web of Science was conducted to identify relevant reports from the peer-reviewed literature. Key information about each report's scope and details was collected to assess the commonalities among the approaches.

Results: A total of 56 reports from 2012 to June 2021 were included. From this selection, 27 unique automated approaches were identified. Machine learning approaches require fewer trials than conventional threshold-seeking approaches, and personal digital devices make assessments more affordable and accessible. Validity can be enhanced using digital technologies for quality surveillance, including noise monitoring and detecting inconclusive results.

Conclusions: In the past 10 years, an increasing number of automated approaches have reported similar accuracy, reliability, and time efficiency as manual hearing assessments. New developments, including machine learning approaches, offer features, versatility, and cost-effectiveness beyond manual audiometry. Used within identified limitations, automated assessments using digital devices can support task-shifting, self-care, telehealth, and clinical care pathways.

Key words: audiology; automated audiometry; automatic audiometry; automation; digital health technologies; digital hearing health care; machine learning; remote care; self-administered audiometry; self-assessment audiometry; user-operated audiometry; digital health; hearing loss; digital hearing; digital devices; mobile phone; telehealth.

INTRODUCTION

Hearing loss affects 1.5 billion persons globally and is expected to increase by another billion by 2050 (Haile et al., 2021; World Health Organization, 2021). Hearing testing is the first step towards appropriate and timely treatment. Unfortunately, most affected persons are unable to access hearing assessments with less than one hearing health professional for every million people in regions such as Africa (Kamenov et al., 2021; World Health Organization, 2021). Increasingly automated approaches (all aspects of the method associated with automated audiometry), including machine learning, are being developed and made available that provide self-administered hearing assessment. The term automated audiometry refers to all hearing tests that are self-administered from the point the test starts. More specifically in this review, we define automated audiometry as calibrated pure-tone threshold audiometry in any setting (i.e. hearing health care, occupational health and community settings) that is self-administered from the point the test starts. Machine learning refers to model based approaches that learn from examples (data) instead of being programmed with rules (Rajkomar et al., 2019). Since professionals' direct involvement is not required, automated approaches enable health care pathways with the potential to increase accessibility, efficiency and scalability. Digital (health) technologies, including apps, smartphones, tablets and wearables, can acquire data remotely, expand the reach and precision of clinicians, and facilitate more personalized hearing health care within a network of distributed expertise (K. I. Taylor et al., 2020; Wasmann et al., 2021). Recent examples of automated hearing assessments include clinical- and consumer-grade applications (Swanepoel et al., 2019). General global health trends suggest that increased availability of diagnostic tools could lower healthcare costs while improving quality of life (World Economic Forum, 2021). For example, in Parkinson's disease, remote care based on wearables provides ecologically valid methods for monitoring and evaluating symptoms (Bloem et al., 2019; Gatsios et al., 2020). In tuberculosis screening in low-resource settings, automated diagnosis can increase sensitivity of identifying persons at risk while reducing cost (Philipsen et al., 2019). Self-assessment using eHealth vision tools improves access to diagnosis and facilitates timely diagnosis, although consistent criteria for referring to the clinical pathway and validity and reliability of eHealth tools are still a concern (W. K. Yeung et al., 2019).

Timely detection and treatment of hearing loss is essential to enable optimal outcomes and quality of life across the life course (World Health Organization, 2021). Untreated hearing loss restricts language development and educational potential in children and is associated with more rapid cognitive decline in adults

(B. S. Wilson et al., 2017). It may lead to social isolation, lower socioeconomic status, increased social disparities and decreased health, resulting in lower quality of life at the individual level and substantial costs at the community level (McDaid et al., 2021; Tsimpida et al., 2021). Importantly, treating hearing loss in mid-life has been identified as the largest potentially modifiable risk factor for developing dementia in later life (Livingston et al., 2017). The global annual cost of untreated hearing loss is $980 million (McDaid et al., 2021). Global health investment models indicate a significant return of investment both in hearing diagnosis and treatment (World Health Organization, 2021). The capacity of the entire clinical pathway should be increased since a bottleneck looms if the accessibility of diagnosis is increased faster than the availability of affordable treatment and rehabilitation.

Automated self-test options are important for detecting and diagnosing hearing loss to direct timely and appropriate treatments. The overwhelming majority of treatments are for permanent age-related and noise-induced hearing loss, but a significant portion of the population requires medical treatment for hearing loss (Haile et al., 2021). The onset of COVID-19 has further emphasized the importance of self-test approaches (Manchaiah et al., 2021; Saunders & Roughley, 2021). Automation on digital devices is a powerful enabler for alternative diagnostic pathways that can include home-based testing, low-touch service models outside traditional clinic settings, and decentralized community-based models that rely on task-shifting to minimally trained facilitators (Eksteen et al., 2019).

Automation in hearing assessment is not a new concept and dates back more than seven decades (Békésy, 1947). In recent years, it has resurged with the convergence of digital technologies and machine learning approaches. The primary tool for hearing assessment is pure-tone audiometry which describes the degree of hearing loss relative to normal hearing persons expressed in decibels Hearing Level (dB HL) across specific frequencies (125 - 8000 Hz). Pure-tone audiometry can also differentiate the type of hearing loss, i.e. sensorineural or conductive, when bone conduction and air conduction transducers are used. Machine learning-based threshold seeking approaches, known as Bayesian active learning, have demonstrated their potential to optimise efficiency and increase automated hearing assessments' precision (Barbour, Howard, et al., 2019). The increased efficiency comes from these methods' ability to target trials to those areas of the frequency space where the estimation has greatest uncertainty (Gardner et al., 2015; Schlittenlacher et al., 2018b).

In 2012, a systematic review that included 29 reports on automated audiometry showed that automated procedures have comparable accuracy as manual procedures when performing air conduction audiometry. However, few validated automated procedures that included automated bone conduction audiometry had been reported, machine learning based audiometry approaches were not reported yet, and approaches were rarely validated in children or hard-to-test populations (Mahomed et al., 2013). Since 2012, there has been significant work and innovation in this area, which calls for an update and extension of the previous review. This scoping review aims to provide the current status of automation and machine learning approaches in hearing assessment using validated pure-tone audiometry with potential indicators of accuracy, reliability and efficiency of these approaches.

METHOD

We conducted a systematic scoping review of peer-reviewed literature on automated and machine learning approaches to validated pure-tone threshold audiometry using digital technologies, by considering accuracy, reliability, and efficiency. This review followed the methodological framework outlined in Arksey & O'Malley (2005).

Identifying potentially relevant records

A search across the electronic databases from PubMed, IEEE, and Web of Sciences was conducted to identify relevant reports from peer-reviewed literature. Complementary and redundant search terms were applied to ensure thorough coverage and cross-checking of search findings. In the PubMed database, medical subject headings and relevant keywords were collected to determine all records relating to the study aim. The following synonyms of, and closely related terms to, automated audiometry were used: automatic audiometry, self-administered audiometry, self-assessment audiometry, and user-operated audiometry. The complete set of terms and applied search strategy are provided in the supplemental materials, Table 1. The IEEE database is engineering-oriented, and only relevant keywords based on audiometry were used since it was assumed that any result in audiometry would be highly associated with automated audiometry. The Web of Science database is known to index the PubMed and IEEE databases and was explored using similar search terms as for the PubMed search. After preliminary explorations to identify appropriate keywords, we conducted a search on 8 July 2020 and updated it on 12 January 2021 and 6 July 2021. The search included all reports that meet the inclusion criteria published from 1 January 2012 until 30 June 2021. The start date was chosen as we regard this scoping review as an extension

and generalisation of a previous (systematic) review by Mahomed et al. (2013), which included studies up to 20 July 2012.

Selecting relevant records

There were three inclusion criteria the reports had to meet; (i) the report had to be about automated or machine learning, pure-tone frequency-specific threshold audiometry; (ii) written in English; (iii) the automated threshold audiometry had to be compared against the gold standard or reasonable standard. The gold standard is defined as manual audiometry in a sound booth according to ISO standards. The automated audiometry also needed to be performed inside a sound booth, and results needed to be compared to the gold standard. A reasonable standard for validation was defined as either a within-subject comparison between the gold standard and automated audiometry in an unconventional setting (for example, a quiet room), or a within-subject comparison between a validated automated audiometry approach and an experimental approach of audiometry in the same unconventional setting.

We excluded reports on screening audiometry (e.g. gave pass/refer as an outcome) rather than threshold audiometry, review papers and studies reporting approaches that were not compared to the gold or reasonable reference standard.

The first phase of the screening was based on title. If the title indicated that content was within the scope of the research question (i.e. automated or machine learning approaches in diagnostic hearing assessment), the report was included into the second screening phase. In the second phase, the remaining reports' abstracts were assessed using the inclusion and exclusion criteria stated above.

Two authors (LP and JWW) conducted the abstract screening. They were blinded from each other to avoid confirmation bias. After the screening, the authors discussed any disagreements to reach an agreement. When in doubt, the report was admitted to the third, full-text review phase. In this phase, all remaining reports were reviewed in full to determine if the inclusion criteria were met. As can be seen in the PRIMSA flow diagram, Figure 1, the resulting selection of reports was complemented by additional reports. After some reports were clustered as having identical approaches (explained under the heading "*Collating approaches, summarizing, and reporting the results*"), additional reports were added to avoid missing validation data of these clustered approaches. These reports were published before the inclusion date criteria (from before 1 January 2012) or did not appear in the search and were added based on the reference lists of the already included reports.

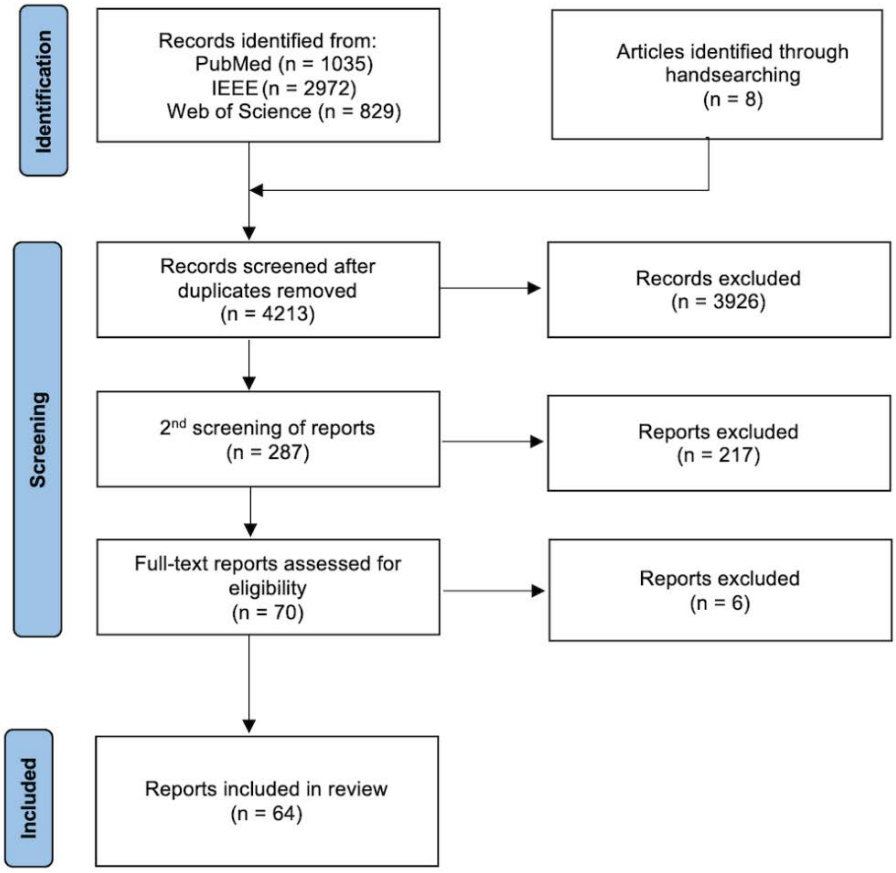

Figure 1. *PRISMA flow diagram of the screening process.*

Extracting data items

A template for grading the reports was agreed upon by all authors (supplementary materials, Table 2). Two authors (LP and JWW) independently extracted the information directly relevant to the scoping review question. In cases of disagreement, a consensus was reached after discussion between these two authors. The compulsory data fields were: test frequency and intensity range; response method; test equipment including the type of transducers; calibration; hardware; test quality control; accuracy; reliability; efficiency; validation; and test population. In the report of Mahomed et al. (2013), accuracy and reliability of manual and automated approaches demonstrated equivalent performance. Time efficiency had primarily been reported by comparing the testing time of manual and automated audiometry (Heisey et al., 2020; Swanepoel et al., 2010; van Tonder et al., 2017). The reports on machine learning audiometry explicitly use the

number of trials/stimuli needed to converge to a certain precision (e.g. 5 dB) as a performance outcome (Heisey et al., 2020; Schlittenlacher et al., 2018b). Therefore, we added time-efficiency as a necessary parameter. Where available, accuracy and reliability were expressed in decibels (dB) using the overall root-mean-square-deviation (RMSD) between the automated approach and the gold (or reasonable) standard. Based on the work by Margolis et al. (2010) and the minimum acceptable accuracy recommended by clinical guidelines (American Speech-Language-Hearing Association, 2005), a RMSD of 6 dB and 10 dB was chosen as criteria for desired and minimal accuracy, respectively. In order to establish a benchmark for acceptable test duration, the mean testing time for conventional manual bilateral audiometry (air seven and bone five frequencies) was estimated (see supplementary material Table 4). For manual bilateral air conduction, a mean testing time of 5-10 minutes was considered acceptable, and 10-20 minutes for manual bilateral air & bone conduction. If testing times exceeded these ranges by more than 5 minutes, time-efficiency was assessed as a potential issue.

The data collected from the reports provided key information about each report's scope and details, enabling the authors to assess commonalities between the approaches.

Collating approaches, summarizing, and reporting the results

When multiple reports described the same underlying approach, these reports were pooled in one approach-cluster. The first report describing an approach and subsequent studies that validated or extended that approach were included. The name of the approach, citations to the initial report and/or common authorships were used to cluster the reports. The grading table was completed for each cluster separately to provide a structure for the subsequent content analysis. In the last part of the grading table, under the heading "validation approach," all validation studies are described together. For every approach-cluster a key contribution for the audiological field was derived from the associated report(s). A key contribution is a finding or claim made by the author(s) significant for the approach in general, stated in either the conclusion or discussion of a report in accordance with their objective.

RESULTS

A total of 64 reports were included in this study. Fifty-six of the 64 reports were included according to the inclusion- and exclusion criteria, and eight reports were added to the approach-clusters. After clustering identical approaches, 27 approach-clusters remained including two that used machine learning. Extracted data items and grading of results on approaches are provided in the supplemental material, Table 3. Specifications of the reported accuracy, reliability, and time-efficiency are described in Table 1.

Table 1. *Review of the accuracy, test-retest reliability, and time-efficiency for automated and machine learning audiometry approaches 2012 – 2021.*

Type of transducer	***Accuracy***		***Reliability (test-retest)***		***Time-Efficiency***	
	N	***Reported finding***	***N***	***Reported finding***	***N***	***Reported Finding***
Air conduction (n=23 approach-clusters)	4 7 9 3	RMSD < 6 dB RMSD < 10 dB Statistical equivalence No statistical equivalence	4 1 9 9	RMSD < 6 dB RMSD < 10 dB Statistical equivalence Not reported	10 2 1 1 9	Acceptable testing time per (partial) audiogram Acceptable testing time and number of trials per audiogram Acceptable testing time and number of trials per frequency Testing time potential burden Not reported
Bone conduction (n=1 approach-cluster)	1	Statistical equivalence	1	Test-retest not reported	1	Not reported
Both air and bone conduction (n=3 approach-clusters)		*Air conduction*		*Air conduction*		*Air conduction*
	2	RMSD < 6 dB	1	RMSD < 6 dB	2	Acceptable testing time per audiogram
	1	RMSD < 10 dB	2	RMSD < 10 dB		*Air and bone conduction*
		Bone conduction		*Bone conduction*		
	1	RMSD < 10 dB	1	RMSD < 6 dB	1	Acceptable testing time per audiogram
	2	Statistical equivalence	2	Test-retest not reported		

Accuracy

The accuracy is represented as a comparison against the gold standard or reasonable standard. The majority of the automated techniques (n=14) expressed the accuracy in RMSD. Other types of analyses used average differences and standard deviation (n=10), average thresholds and standard deviation (n=1; Patel et al., 2021), linear regression and correlation coefficients (n=1; Dewyer et al., 2019) and ANOVA analysis (n=1; Corry et al., 2017). The types of analysis used can be seen in supplemental material, Table 5.

Test-Retest Reliability

Test-retest reliability were reported for some automated and machine learning audiometry approaches. Seventeen of the twenty-seven approaches did not report on test-retest reliability. Seven approaches expressed it in RMSD. Other statistical methods used were average differences and standard deviation (n=6), Pearson Product moment correlation coefficients (n=2; Colsman et al., 2020; Manganella et al., 2018), standard of variance (n=1; Schmidt et al., 2014) and repeated ANOVA (n=1; Corry et al., 2017).

Test efficiency

Seventeen approaches reported a measure for test efficiency based on test duration. Test efficiency expressed in testing time seems to be a standard metric, similar across studies defined as the time from presenting the first stimulus until the final response of the subject, expressed in seconds or minutes. However, there was no agreement between reports on what to include in the measurement and what groups to use as a reference. Reported time-efficiency measures included the recorded time per frequency, recorded time per unilateral or bilateral air conduction audiogram (between 2 to 7 frequencies) in normal hearing or hearing impaired persons, or full air and bone conduction audiograms in hearing impaired persons. Thirteen approach-clusters reported an acceptable testing time. Three approach-clusters also indicated the number of trials in addition to the testing time for either a bilaterally masked air audiogram (Heisey et al., 2020), unilateral air audiogram (Schlittenlacher et al., 2018b) or per frequency (Vinay et al., 2015). One approach-cluster that applied Bekesy tracking did report the testing time, but was not in the acceptable range (Poling et al., 2016). Ten approach-cluster did not report anything about test-time.

Test parameters and specifications

Tests were all self-administered from the point the test started. Four approaches had the option to switch to a manual audiometry mode. Table 2 summarises an overview of test parameters and specifications of the 27 approach-clusters, and Table 3 highlights key contributions. Most of the approaches used adaptive procedures that rely on the previous response only (here referred to as partially adaptive procedures). The most common example was the (modified) Hughson-Westlake staircase procedure (n=20), which is based on the classical method of limits (Gescheider, 2013). Other partially adaptive procedures applied the method of adjustment, such as the Bekesy tracking method (Poling et al., 2016) or the ‘coarse-to-fine focus’ algorithm (Chen et al., 2019). There was a single report of an approach that did not define the threshold seeking method, but had a built in

protocol to alternate between ears during testing (Manganella et al., 2018). Fully adaptive procedures, in contrast, use the complete set of all previous responses. Examples include Bayesian active learning procedures (also referred to as machine learning audiometry; n=2; Barbour, Howard, et al., 2019; Schlittenlacher et al., 2018b) and maximum likelihood estimation (n=2; Schmidt et al., 2014; Vinay et al., 2015). All the machine learning audiometry methods applied active Bayesian model selection, which is a type of shallow machine learning that uses individual models. They apply supervised learning since every data point is labelled by the subject (Gardner et al., 2015).

Most of the approaches used conventional calibration (20/27) according to ISO-standards. Six approaches used an unconventional calibration technique. Patel et al. (2021) determined a reference equivalent threshold level (RETSPL) for air conduction for the specific phone-headphone combination using manual audiometry as a reference. Masalski et al. (2016) used reference levels for calibration for smartphone and transducer combinations, collected in uncontrolled conditions in normal-hearing persons. Other calibration techniques set the volume of the device to 50% (Szudek et al., 2012), comparing and adjusting the output level to the input using a sound level meter (Corry et al., 2017; Foulad et al., 2013) or using Thévenin-equivalent probe calibration (Poling et al., 2016).

Twenty-two approaches were validated on normal-hearing and hearing-impaired people. Four studies were performed using normal-hearing subjects (Colsman et al., 2020; Corry et al., 2017; Vinay et al., 2015). One approach-cluster was only validated in a hearing-impaired population, using hearing aids as transducers (Chen et al., 2019). Automated audiometry was applied across a range of populations. All approaches were applied to adults except for Patel et al. (2021) who only included children in their study. Eight approaches were validated in children, including four approaches that designed a child-friendly user-interface (B. Kung et al., 2021; Margolis et al., 2011; Patel et al., 2021; J. C. Yeung et al., 2015). Other test populations were elderly (Maclennan-Smith et al., 2013), veterans (Margolis et al., 2016), and persons exposed to occupational noise (Henriksen et al., 2014) or ototoxic substances (Jacobs et al., 2012). Automated audiometry has also been applied as an alternative in low-resource environments for traditional manual audiometry (Kelly et al., 2018; Sandström et al., 2020; Visagie et al., 2015). The user-interface plays an important role in making self-testing feasible in all populations and may require an iterative design process (including clinical pilot-studies) (Sandström et al., 2016, 2020).

Table 2. *Description of test parameters and specifications for automated audiometry approaches 2012 – 2021.*

Test parameters & specifications	**Descriptions of approach clusters (n=27)**
Threshold seeking method *Underlying algorithm to determine the thresholds*	20 (modified) Hughson-Westlake 2 machine learning 1 Bekesy tracking 4 other method
Test range *Limits of the frequency that can be tested*	18 full frequency range 125 – 8000 Hz 4 extended high frequencies 125 – 16000 Hz 5 reduced frequency range
Test range *Limits of intensity that can be tested*	14 Intensity range 0 – 100 dB HL 10 reduced intensity range 3 intensity range not reported
Masking *Needed to prevent responses from the non-test ear and obtain the true threshold of the test ear*	9 automated masking 1 manual masking 13 no masking 4 masking not reported
Response method *Method of recording subject's responses to test stimuli*	9 forced-choice 13 single response 3 forced-choice and single response 2 not reported
Transducers *Method of presenting stimuli, e.g. insert phone or supra or circumaural headphone*	23 air conduction transducers 3 air- and bone conduction transducers 1 only bone conduction transducer
Calibration *Unconventional calibration methods are explained in the text*	20 conventional calibration 6 unconventional calibration 1 calibration not reported
Digital devices *Reported Hardware needed to run the test.*	2 portable audiometer 9 computer-based 1 web-based (requires connectivity) 15 smartphone- or tablet-based
Quality control measures *Indicators of the reliability of the test*	5 detect false-responses 6 have noise control 7 detect false-responses and have noise control 9 quality control measures not reported
Validation *Highest level of validation reported for each approach-cluster*	22 gold standard 4 reasonable standard 1 proof-of-concept
Test Population *Hearing status*	3 Normal hearing only 1 Hearing loss only 23 Normal hearing & Hearing loss
Test Population *Age*	17 Adults only 1 Children only 9 Adults & Children

Table 3. *key contributions of the automated and machine learning approaches to the audiological field; a smartphone-based hearing test application, not yet commercialized; b automated hearing test commercialized by hearX™ group.*

Approach-Cluster (first author, year, publications, approach name)	**Key contribution(s) to the field**
Bean et al., 2021, **OtoKiosk**	Has the potential to be used in test environments like exam rooms as a clinical tool for identifying hearing loss via air conduction separating persons with normal and impaired hearing.
Chen et al., 2019, **SHSA**	A demonstration smartphone-based hearing self-assessment (SHSA) that runs on a hearing aid which has statistical equivalence to manual audiometry
Colsman et al., 2020	Portable devices that use calibrated headphones result in much higher accuracies than the uncalibrated devices.
Corry et al., 2017	The reliability of audiometer apps should not be assumed. Issues of accuracy and calibration of consumer headphones need to be addressed before such combinations can be used with confidence.
Dewyer et al., 2019, **Earbone**	A proof of concept for smartphone-based bone conduction threshold testing.
Foulad et al., 2013; Kelly et al., 2018; Saliba et al., 2017, **Eartrumpet**	An iOS-based software application for automated pure-tone hearing testing without need for additional specialized equipment, yielding hearing test results that approach those of conventional audiometry.
Jacobs et al., 2012; Dille et al., 2013, **Oto-ID**	Automated (remote) hearing tests to provide clinicians information for ototoxicity monitoring.
B. Kung et al., 2021, **Kids Hearing Game**	Tablet based audiometry using game-design elements that can be used to test and screen for hearing loss in children who may not have adequate access to resources for traditional hearing screening.
Liu et al., 2015	An interactive hearing self-testing system, consisting of a notebook computer, sound card, and insert earphones is a valid, portable and sensitive instrument for hearing thresholds self-assessment.
Manganella et al., 2018, **Agilis**	An application that detects increased levels of ambient noise, when it is programmed to stop the testing.
Margolis et al., 2007; 2010; Eikelboom et al., 2013; Margolis & Moore, 2011; Margolis et al., 2011, **AMTAS**	Designed to fit into the clinical care pathway including air and bone conduction, and incorporates a quality assessment method (QUALIND®) that predicts the accuracy of the test.
Margolis et al., 2016, 2018; Mosley et al., 2019. **Home Hearing Test**	Developed and well suited to provide increased access to hearing testing and support home telehealth programs.
Masalski & Kręcicki, 2013; Masalski et al., 2016, 2018	An automated method that uses smartphone model-specific reference sound levels for calibration in the app. Biological reference sound levels were collected in uncontrolled conditions in normal-hearing persons.

Table 3. *continued*

Approach-Cluster (first author, year, publications, approach name)	**Key contribution(s) to the field**
Meinke et al., 2017; Magro et al., 2020, **WHATS**	A mobile wireless automated hearing-test system in occupational audiometry for obtaining hearing thresholds in diverse test locations without the use of a sound booth
Patel et al., 2021, **HearTest**[a]	A novel subjective test-based approach was used to calibrate a smartphone-earphone combination with respect to the reference audiometer.
Poling et al., 2016	Specific Bekesy tracking patterns are identified in subjects who experienced difficulty converging to a reliable threshold.
Schlittenlacher et al., 2018	Bayesian active-learning methods provide an accurate estimate of hearing thresholds in a continuous range of frequencies.
Schmidt et al., 2014	A user-operated two-alternative forced-choice in combination with the method of maximum likelihood does not require specific operating skills and the repeatability is acceptable and similar to conventional audiometry.
Song et al., 2015; Barbour et al., 2019; Heisey et al., 2018, 2020, **MLAG**	Bayesian active-learning method which determines the most informative next tone leading to fast audiogram procedure and threshold estimation in a continuous range of frequencies with potential to measure additional variables efficiently.
Sun et al., 2019	Active noise control technology to measure outside the sound booth.
Swanepoel et al., 2010; Brennan-Jones et al., 2016; Govender & Mars, 2018; Maclennan-Smith et al., 2013; Storey et al., 2014; Swanepoel & Biagio, 2011; Swanepoel de et al., 2015; Visagie et al., 2015, **KUDUwave**	An automated portable diagnostic audiometer using improved passive attenuation and real-time environmental noise monitoring making audiometry possible in unconventional settings.
Swanepoel et al., 2014; Bornman et al., 2019; Brittz et al., 2019; Corona et al., 2020; Rodrigues et al., 2020; Sandström et al., 2016, 2020; van Tonder et al., 2017, **HearTest**[b]	A smartphone-based automated hearing test applicable in low resource environments.
Szudek et al., 2012; Handzel et al., 2013; Khoza-Shangase & Kassner, 2013, **Uhear**	An approach that is applicable to the initial evaluation of patients with sudden sensorineural hearing loss before a standard audiogram is available.
Van Tasell & Folkeard, 2013	Method of adjustment and the Hughson-Westlake method embedded in automated audiometry can be considered equivalent in accuracy to conventional audiometry
Vinay et al., 2015; Henriksen et al., 2014, **NEWT**	The New Early Warning Test (NEWT), which is incorporated inside an active communication earplug, serves as a reliable and efficient method to measure auditory thresholds, especially in the presence of high background noise

Table 3. *continued*

Approach-Cluster (first author, year, publications, approach name)	**Key contribution(s) to the field**
Whitton et al., 2016	A proof of concept study of a several self-administered, automated hearing measurements at home, showing statistical equivalency to conventional audiometry in the clinic.
J. C. Yeung et al., 2015; Bastianelli et al., 2019; Thompson et al., 2015; Vijayasingam et al., 2020; Yalamanchali et al., 2020; J. Yeung et al., 2013, **Shoebox**	A method for threshold hearing assessments outside conventional sound booths and with an interface suitable for children.

DISCUSSION

In 2012 evidence for automated audiometry demonstrated similar reliability and accuracy as manual audiometry, but especially for children and bone conduction the number of reports was limited (Mahomed et al., 2013). In less than a decade, twenty-two novel approaches and developments across five existing approaches had appeared in 56 publications, adding to the 29 published prior to 2012. Promising new developments include machine learning techniques for more time-efficient hearing assessment (n=2), use of tablets or smartphones as audiometer interface (n=15), and child-friendly user-interfaces (n=4) including game-design elements. The number of approaches that include bone conduction is still limited (n=4), only two more than were reported in 2012 (Mahomed et al., 2013).

Accuracy

The required accuracy, reliability, and efficiency depend on the clinical aims and consequences. The ultimate aim for automated hearing assessment is to deliver clinically actionable estimates of hearing status (i.e. the clinician or patient acts appropriately for treatment given the diagnostic test results). In fully adaptive procedures, the level of precision and confidence needed to conclude the assessment can be set to any level by choosing the proper termination criteria, resulting in different trade-offs. Schmidt et al. (J. H. Schmidt et al., 2014), for instance, aimed for high accuracy and reliability, whereas Heisey et al. (2020) aimed with their machine learning audiometry for high efficiency. Overall, a shift in the type of analysis to demonstrate accuracy is observed. In this review, the two major type of analysis included were RMSD (n=14) and average differences and standard deviation (n=10). In the report by Mahomed et al. (2013), the accuracy was primarily expressed in average differences or thresholds and standard deviation (both n=11).

In our view, RMSD is the preferred indicator for accuracy with clinical relevance (American Speech-Language-Hearing Association, 2005), assuming it has already been demonstrated that there is no bias between the automated and manually determined hearing thresholds (e.g. signed differences). In traditional clinical terms, automation is equal in accuracy to manual audiometry if the difference is within 6 dB RMSD. Six of the twenty-seven automated approaches meet this strict accuracy criterion. For many applications, however, the less strict 10 dB RMSD criterium is sufficient, which was achieved by seven additional automated approaches.

For bone conduction measurements, the accuracy was inherently lower than air conduction measurements due to conductor placement (Margolis et al., 2010). However, this accuracy is typically sufficient to address the clinical question of whether conductive/mixed hearing loss is present and to choose and evaluate appropriate treatment. The technical feasibility of bone conduction assessments outside of a clinical setting (sound booth) remains difficult. Alternatively, this clinical question can be addressed with other tests, including tympanometry, otoscopy, or a combination of air conduction thresholds for tone and speech stimuli (De Sousa, Swanepoel, et al., 2020). At least 13 automated techniques had accuracy comparable to traditional manual air conduction audiometry as expressed in RMSD. Eighteen of the twenty-seven approaches did not report on test-retest accuracy or used a measure of statistical equivalence that does not allow us to assess the accuracy.

One limit to the impact of achieved test accuracy is the high variation in the interpretation of audiograms by clinicians, regardless if those audiograms are determined using an automated or manual approach (Brennan-Jones et al., 2018). Automation can assist clinicians and patients to interpret the measurement by data-driven automated reporting of accuracy and reliability (including signalling for suspicious outcomes) such as QUALIND® (Margolis et al., 2007) or by automated classification for diagnostic purposes (including type and degree of hearing loss). Examples of automated classification include AMCLASS (Margolis & Saly, 2008), Autoaudio (Crowson et al., 2020), and data-driven audiogram classification (Charih et al., 2020).

Reliability

RMSD is also increasingly used as a measure in test-retest reliability. Seven approaches used RMSD as a measure, whereas in 2012 this was only used in two studies. Advances in automated audiometry that increase reliability included procedures to identify invalid responses (n=5), monitoring environmental

noise (n=6), or both (n=7) to warn for invalid test conditions, making these tests applicable in more populations and environments. The reliability can be increased, for instance, by alternative response methods including the forced choice paradigm (J. H. Schmidt et al., 2014) or by using machine learning to account for lapses of attention (Schlittenlacher et al., 2018b). Digital (health) technologies, including smartphones and tablets, lend themselves to quality control measures for increased reliability with the host of integrated sensors (K. I. Taylor et al., 2020).

Efficiency

A fair indicator of efficiency is the overall time required to conduct a test. Most approaches (n=20) used the modified Hughson-Westlake procedure, of which seven showed a similar test duration to manual audiometry. Maximum likelihood procedures demonstrated a 45% reduction in test time in normal hearing persons (Vinay et al., 2015). Bayesian active learning methods can be extended by adding variables that share some interrelationship using a conjoint estimator that exploits nonlinear interactions between the variables (Barbour, DiLorenzo, et al., 2019). The resulting machine learning based automated procedures had demonstrated a 30-70% reduction in test time compared to manual audiometry for air conduction audiograms both in normal and hearing impaired persons (Heisey et al., 2020). No machine learning approaches had incorporated bone conduction yet. Therefore, time-efficiency gains compared to full audiogram procedures are not available but one can assume these will yield similar time-efficiency gains. Another indicator for test efficiency is the number of stimuli required to reach the desired accuracy. This indicator is helpful to optimise the threshold seeking part of the approach. Reporting the equivalent time gains under operational conditions is recommended since this can be readily compared to other efficiency gains, including the reduced travelling time if a visit to the outpatient clinic can be replaced for an at-home test, or time savings by automating other parts of the clinical care pathway such as interpretation of the outcome. Other aspects of efficiency beyond time that should be considered are the cost reductions when enabling task-shifting of professionals or the ability to test outside the sound booth.

Future developments required

To get an overall indicator of the technical maturity of an approach, developers should be encouraged to use the technology readiness level (TRL) to report the development phase of a technology. Technology readiness levels were initially developed in aerospace to estimate the maturity of technology from basic concept to flight-proven product (Héder, 2017). To apply TRLs to automated audiometry, one could make further adjustments to fit the hearing healthcare sector to the version of

biomedical TRLs created by the United States Army Medical Research and Materiel Command (Office of the Director, Defense Research and Engineering (DDR&E), 2009). For those approaches that are ready for operational use, certification (e.g. CE and FDA) can further stimulate clinical adoption and iterative improvements based on clinical feedback. In order to be cost-effective, timely and responsive, certification for digital self-care approaches may need to be less stringent than for clinical care. W. K. Yeung et al. (2019) proposed alternative procedures for (fast) certification to keep up with the rapidly developing field of visual eHealth tools, for example. Their recommendations might also be applicable to automated hearing assessments, including a rating by health agencies or NGOs (e.g. a repository of trusted approaches, see Psyberguide as an example of mental health apps reviewed by experts (Garland et al., 2021)) or adopting the Clinical Laboratory Improvement Amendments (CLIA) model to ensure that approaches comply with basic requirements of usability, privacy, and security (W. K. Yeung et al., 2019). Following similar certification procedures in the visual and auditory domain may facilitate diagnosis across medical domains. In addition, standards on minimum quality, and consensus on what meta-data are needed in health applications to describe the test conditions and facilitate interpretation are currently missing.

Limitations

This scoping review included peer-reviewed reports taken from widely used and recognized scientific databases. A potential limitation is that some of the commercialized automated approaches may have been developed without peer-reviewed reports. Some automated approaches could therefore be more mature than reported. There is no gold standard for reporting audiometry validation studies, which limits a consistent comparison among approaches. Lastly, automated procedures may well be embraced by early adopters first, which could lead to projections on suitability that are overly optimistic for users with poorer digital proficiency.

Conclusion and recommendations

Since 2012 an increasing number of automated audiometry approaches on digital devices demonstrate similar accuracy, reliability and time-efficiency as conventional manual audiometry. New developments offer features, versatility, and cost-effectiveness beyond manual audiometry. Fully adaptive procedures, including machine learning techniques, seek hearing thresholds more efficiently. Inexpensive digital devices such as smartphones can be turned into audiometers, increasing accessibility and availability. Higher reliability is achievable by signalling invalid test conditions, and child-friendly user-interfaces offer a solution to the hard to test

population. These approaches can be implemented in the clinical care pathway, remote or virtual hearing healthcare, community-based services, and occupational healthcare to address the global need for accessible hearing loss diagnosis.

For successful adoption, standardized measures of accuracy, reliability, and efficiency are needed for comparative purposes. Certification and independent reviews may help prospective users in selecting trustworthy approaches. Further reliability can be achieved by determining which difficult to test populations may not be appropriate for automated testing and how to detect and then triage these patients to specialized centres. More user-friendly and failsafe procedures that include remote surveillance and quality control can support automated hearing assessment at-scale in specific populations and in concert with diagnostic assessments in other medical domains, including visual health and mental wellbeing (Garland et al., 2021; W. K. Yeung et al., 2019). Further contextual information, e.g. standardized meta-data, is needed to help clinicians interpret test outcomes' context and limitations. If researchers and clinicians deal carefully with its limitations, automated hearing assessments can be designed such that they form an effective part of service delivery for many people who have or are at risk of hearing loss. Automated audiometry can be part of existing care pathways but also enable new service models, including task-shifting to community health workers delivering decentralized care, virtual hearing healthcare, and over-the counter or direct-to-consumer hearing aid dispensing.

6

Remote Cochlear Implant Assessments: Validity and Stability in Self-Administered Smartphone-Based Testing

Jan-Willem Wasmann, Wendy Huinck, Cris Lanting**

** is for shared last authorship*

Wasmann, J.-W. A., Huinck, W. J., & Lanting, C. P. (2024). Remote Cochlear Implant Assessments: Validity and Stability in Self-Administered Smartphone-Based Testing. *Ear and Hearing*, *45*(1), 239–249. https://doi.org/10.1097/AUD.0000000000001422

ABSTRACT

Research Question: What is the stability of remote testing in cochlear implant care? We studied the influence of time-of-day, listener fatigue, and motivation on the outcomes of the aided threshold test (ATT) and digit triplets test (DTT) in cochlear implant (CI) recipients using self-tests at-home on a smartphone or tablet.

Design: A single-center repeated measures cohort study design (n = 50 adult CI recipients). The ATT and DTT were tested at-home ten times, with nine of these sessions planned within a period of eight days. Outcomes were modeled as a function of time-of-day, momentary motivation, listeners' task-related fatigue, and chronotype (i.e., someone's preference for morning or evening due to the sleep-wake cycle) using linear mixed models. Additional factors included aided monosyllabic word recognition in quiet, daily-life fatigue, age, and CI experience.

Results: Out of 500 planned measurements, 407 ATTs and 476 DTTs were completed. The ATT determined thresholds and impedances were stable across sessions. The factors in the DTT model explained 75% of the total variance. Forty-nine percent of the total variance was explained by individual differences in the participants' DTT performance. For each 10% increase in word recognition in quiet, the DTT speech reception threshold improved by an average of 1.6 dB. DTT speech reception threshold improved, on average, by 0.1 dB per repeated session and correlated with the number of successful DTTs per participant. There was no significant time-of-day effect on auditory performance in at-home administered tests.

Conclusions: This study is one of the first to report on the validity and stability of remote assessments in CI recipients and reveals relevant factors. CI recipients can be self-tested at any waking hour to monitor performance via smartphone or tablet. Motivation, task-related fatigue, and chronotype did not affect the outcomes of ATT or DTT in the studied cohort. Word recognition in quiet is a good predictor for deciding whether the DTT should be included in an individual's remote test battery. At-home testing is reliable for cochlear implant recipients and offers an opportunity to provide care in a virtual hearing clinic setting.

Key words: Automated audiometry, Chronotype, Circadian rhythm, Cochlear implant, Fatigue, Hearing impairment, Self-test, Speech-in-noise, Remote care, Time-of-day effect, Virtual hearing clinic.

INTRODUCTION

Severe hearing loss is a chronic disability requiring lifelong care (Haile et al., 2021; World Health Organization, 2021). The accessibility and affordability of high-quality care for chronic diseases are challenging for traditional care provider models that are typically organized within specialized centers (Abegunde et al., 2007). The need for specialized centers is a barrier to care delivery and a source of inequity for those who cannot regularly travel to specialized care centers far from home. New service models that use remote monitoring are required to reduce clinical visits and overcome capacity and accessibility problems. Remote monitoring has several advantages, including greater autonomy and lower medical costs for the patient, and more frequent data collection (K. I. Taylor et al., 2020). It could lead to fewer clinical visits, less travel time, and a lower carbon footprint (L. B. Russell et al., 2008; Swanepoel & Hall, 2020; Wasmann & Laat, 2022).

Cochlear implants (CIs) can improve hearing in individuals with severe hearing loss who cannot be sufficiently rehabilitated with conventional hearing aids. The number of cochlear implantations is growing worldwide (Sorkin & Buchman, 2023), leading to a need for increased capacity in specialized aftercare. Aftercare includes equipment maintenance, performance monitoring, and, when needed, adjustment of the CI fitting (i.e., the setting of the CI). Most CI recipients are well-suited for remote testing during the aftercare phase due to the controlled and calibrated streaming of audio signals from a smartphone to their processor. Additionally, the limited residual hearing in most CI recipients renders the impact of environmental sounds negligible, making remote testing a viable option and potentially paving the way for the development of virtual hearing clinics.

Recently, a Remote Check app was developed and validated by Cochlear Ltd. (Sydney, Australia). The app facilitates a clinician to remotely monitor the long-term auditory performance of CI recipients, examine implant status, and query whether the CI recipient experiences any issues in daily usage (Maruthurkkara et al., 2021, 2022). The Remote Check app includes the aided threshold test (ATT), the digit triplets test (DTT), and two questionnaires. The aided thresholds indicate a CI recipients' ability to hear soft everyday sounds. While, DTT outcome indicates how well the CI recipient understands speech in more adverse listening situations and is considered a suprathreshold measure of auditory functioning, complementary to the ATT. Furthermore, remote testing technology also permits two distinct functions for CI recipients during the aftercare phase: evaluating the impedance, which refers to the electrical conductance of the electrode contacts, and taking a

photo of the implant area to check skin integrity around the implant area. In a pilot study carried out by Cochlear Ltd. in Australia and the UK, it was found that the app identified 94% of issues encountered during a clinical visit and could be used for triaging in-clinic visits (Maruthurkkara et al., 2021). Almost all issues were revealed via the questionnaires.

In at-home monitoring applications, CI recipients often choose the time-of-day they want to perform the test. This prompts the question of whether aided threshold or speech-in-noise tests are affected by the time-of-day, in which they are performed. Maruthurkkara et al. (2022) did not consider time-of-day effects in their Remote Check follow-up validation study, and started evaluations with in-clinic sessions. Although the literature lacks consensus regarding time-of-day effects on auditory performance, research suggests that such effects are minimal during office hours and instead may be linked to factors including aging and chronotype. Chronotype refers to an individual's preference for morning or evening based on their circadian rhythm, which is characterized by fluctuations in physiological markers, including core body temperature and melatonin, commonly known as the “sleep hormone” (C. Schmidt et al., 2007). The importance of chronotype may become particularly relevant in the context of at-home testing.

A previous study found age-related differences in auditory performance and time-of-day effects, with older normal-hearing participants scoring better in gap detection (silent intervals within a sound) at 9:00 am than at 11:30 am or 3:30 pm. However, time-of-day did not affect cognitive performance, including memory tasks (Ezzatian et al., 2010). Another study found time-of-day effects in younger adults with normal hearing (20–28 years) who performed better on a speech-in-noise task in the evening (i.e., after 5:00 pm) than in the morning (i.e., before 10:00 am), while no effect was found in older participants (66–78 years; Veneman et al. 2013).

The aim of this study was to investigate the potential impact of time-of-day effect and chronotype on the outcomes of at-home measurements in cochlear implant (CI) recipients. While performing tests outside of regular office hours can be advantageous, it is currently unclear whether time-of-day effect or chronotype affects test outcomes. Time-of-day effects could potentially be influenced by various factors, including chronotype, attention, task-related fatigue, and long-term fatigue or learning effects (Hornsby et al., 2016; Pichora-Fuller et al., 2016). In this study, daily-life fatigue refers to chronic or long-term fatigue, while task-related fatigue refers to short-term fatigue experienced during or directly after an activity (Y. Wang et al., 2018). Listening effort and fatigue may be more pronounced in CI recipients compared to normal hearing

listeners. This is because adults with hearing loss often report higher levels of fatigue compared to individuals with normal hearing (Alhanbali et al., 2017). However, no significant differences were found in self-reported daily-life fatigue (Y. Wang et al., 2018) or in objective indicators of fatigue, such as cortisol levels in groups with different hearing status (Dwyer et al., 2019). Additionally, motivation to perform the test could also impact the results of self-administered tests. In clinical settings, the tester can observe lack of motivation and intervene when necessary.

This study hypothesized that aided thresholds and speech-in-noise test performance in self-testing at-home could be impacted by various factors, including momentary motivation, task-related fatigue, and chronotype, in light of unknowns regarding time-of-day effects. The objective of this study was to improve the accuracy and reliability of at-home monitoring applications for CI recipients by identifying the optimal time-of-day for self-testing. The findings of this study may have implications for remote care implementation, particularly in monitoring CI recipients during the aftercare phase.

MATERIALS AND METHODS

The study was set up following a single-center repeated measures cohort study design. The ATT and DTT were tested ten times, including nine sessions within eight days. For benchmarking and long-term stability purposes, participants performed the first and final sessions at times they chose freely. The other eight sessions followed a strict schedule, as detailed in Table 1, to ensure an appropriate selection of times-of-day were tested.

Participants

Fifty participants were recruited from the outpatient clinic of the Radboud university medical center in Nijmegen, the Netherlands (Radboudumc). Participants were recruited via general notices and during regular visits at the Radboudumc audiology center. The inclusion criteria were: i) a minimum age of 16 years, ii) a Cochlear Ltd. (Sydney, Australia) cochlear implant (excluding the CI24M, CI24R, and N22 implants as these are not compatible with Remote Check), and iii) an N7 or Kanso2 processor. Further, iv) participants had to have access to an appropriate iPhone, iPod touch, or iPad, v) at least six months of experience with the CI, and vi) an aided monosyllabic word recognition score (Consonant Nucleus Consonant; CNC) greater than 40% was required. The CNC scores of the participants at 65 dB SPL at 12 months post-implantation were collected from their medical records. If a

12-month evaluation was missing from the medical history, the last measurement from the medical records was selected. Table 2 provides demographic details of the participants.

Table 1. *Test Schedule for the study. Sessions are indicated by T0 – T9.*

Time-of-Day	Day 1	Day 2	Day 3	Day 4	Day 8	3 Months
<2 Hour of awakening	T_0	T_1 Morning T_2 Morning		T_4 Morning	T_6 Morning	T_9
3 Hour awake				T_5 Noon	T_7 Noon	
<2 Hour before bedtime			T_3 Night		T_8 Night	
Test battery	Full Remote Check + MM + Chronotype + CIS	Short Remote Check + MM				Same as T_{0+} end-evaluation

Day 1 was always scheduled on a Monday. The full Remote Check consisted of photographs, questionnaires (SSQ-12), the aided threshold test (ATT) and digit triplet test (DTT) and impedance measurement. At each session, participants completed the Momentary Motivation (MM) questionnaire. The short Remote Check included only ATT, DTT, and impedance measurements. Additionally, at T0 and T9, participants completed the Checklist of Individual Strength (CIS) questionnaires. The end evaluation at T9 included a questionnaire to assess satisfaction with the Remote Check.

Table 2. *Demographic characteristics of the participants.*

Number of participants	**50**
Gender	
Male	21
Female	29
Mode of hearing	
Unilateral CI	45
Bilateral CI	5
Cause of deafness	
Unknown	17
Genetic (confirmed)	11
Meningitis / Encephalitis	9
Sudden deafness	7
Otosclerosis	3
NF2 / Acoustic neuroma	2
Mastoid fracture	1
Age at testing (yrs, mean median, range)	58.2, 67 (18-79)
Experience with CI (yrs mean median, range)	7.8, 4 (0.5-31)

The research protocol was submitted to the Medical Research Ethics Committee at the Radboud university medical center Nijmegen, and they considered the study not subject to the Medical Research Involving Human Subjects Act. The study was performed in accordance with good clinical practice and signed informed consent forms were obtained from all participants.

Technical errors and problems reported were collected from the conversations over the messenger system between the participants and the researchers, the warnings in the Cochlear clinician's portal, and by inspecting the test results.

Study procedures

Participants carried out all listening tests at-home using the commercial Remote Check implementation (a module of the Nucleus Smart App), and the CI settings they used in daily life. Most participants had already installed the app as a remote control, either independently or with assistance from their clinician. During the ATT and DTT, direct streaming of the signal to the processor resulted in electric stimulation of the implanted cochlea, eliminating the possibility of crossover. Bilateral CI recipients completed the questionnaires once and did the listening tests one ear at a time, while the processor of the contralateral ear was inactive. Testing always started on the right side. The time taken to perform a Remote Check was approximately 10-20 minutes per ear. Only the results of the best ear of bilateral CI recipients were included in the analysis.

A long-term follow-up session was conducted three months after the initial nine measurements, which were performed within eight days. Some of the first ten participants had technical issues that jeopardized the strict test schedule. Therefore, from the 11th participant onwards, a practice Remote Check was provided one week before the start of the test schedule to ensure that participants understood the test procedure and that technical issues were addressed in advance. A detailed description of the Remote Check test battery can be found in a previous report (Maruthurkkara et al., 2022).

Sessions

Participants were instructed to strictly follow testing at the time-of-day indicated on the test schedule (Table 1). T_1-T_8 were time-of-day sensitive sessions. At sessions T_0 and T_9, the participants were allowed to test at any time they preferred. The morning sessions (T_1, T_2, T_4, and T_6) were scheduled between 6:00 and 10:00 am and had to be done within two hours after waking up. The noon sessions (T_5, T_7) started after a subject was at least three hours awake, typically between 10:00 am and 12:00

noon, and had to be started before 1:00 pm. The afternoon session was scheduled close to noon because some research suggests a decline in performance level on cognitive tasks and a dip in alertness after lunchtime (Abdullah et al., 2016; Monk et al., 1997). The night sessions (T_3, T_8) started two hours before (individual) bedtime, typically between 8:00 pm and 12:30 am. On day 2, two consecutive sessions were acquired for test-retest purposes, and T_2 was measured 10-15 minutes after T_1, allowing participants to take a short break in between.

Once a session started, the participant had to complete it without breaks. If a participant was unable to carry out the session according to the schedule, the session was rescheduled to another day at the same time-of-day as initially planned. The Remote Check was not rescheduled if data were missing due to a failed test. All tests were initiated remotely, and (synchronous) support and troubleshooting were performed by J.-W.A.W & D.B. via a messenger system or by phone. They acted as the troubleshooting support team and received technical support from Cochlear Ltd. when needed. Here, synchronous support means the researcher was directly available online to help the participant, while asynchronous support means the researcher provided feedback before or after the participant performed the tasks. Participants received a reminder at the beginning of the day. They had to indicate to the researchers when they planned to do the test in order to organize remote support availability.

Aided Threshold Test (ATT)

For a complete description of the ATT, the reader is referred to the previous validation study (Maruthurkkara et al., 2022). In short, a two-alternative forced-choice paradigm was used to determine the aided auditory threshold. Participants were required to indicate whether a sound was heard by swiping to the right (heard) or left (not heard). They controlled the presentation of the stimuli by pushing a button in the middle of the screen.

The threshold search algorithm followed a staircase procedure with a stepsize that decreased from 8 to 1 dB near the threshold. The test order of frequencies was 1000, 2000, 4000, 6000, 250, 500, and again 1000 Hz. The presentation level decreased when a participant correctly heard the tone and increased if the participant did not. The algorithm presented silent trials in 33% of the trials. This percentage was adjusted depending on the false-positive rate of the participant's previous responses. When a false-positive response was given, the participant received a message that there was no sound in some trials, and that the participant should swipe to the left. The algorithm stopped with a conclusive threshold when the

difference between the lowest audible and highest non-audible level converged to 1 dB. The algorithm stopped with an inconclusive threshold when either the maximum of 30 trials was reached or when there had been three negative responses at the maximum presentation level of 62 dB HL or three positive responses at the minimum level of 10 dB HL.

Digit Triplets Test (DTT)

In this study, the Dutch version of the DTT was used. The implementation of this test is derived from the work by de Graaff et al. (2016) and Smits et al. (2013). For a complete test description, the reader is referred to the previous validation study (Maruthurkkara et al., 2022). These authors describe the English implementation, which is similar to the Dutch version except for the digits included and the masking noise accompanying the triplets. In the Dutch version, all digits (i.e., 0 – 9) are included. The steady-state noise has the same long-term spectrum as the included Dutch triplets. The signal-to-noise ratio (SNR) expresses the ratio between the sound level of the triplets and the noise. The combined sound level of the triplets and noise remained fixed at 65 dB SPL. At an SNR of '0' dB, the triplets presented at 62 dB SPL were combined with noise at 62 dB SPL. As the doubling of the power increased the level by 3 dB, the combined presentation level was 65 dB SPL. The signal level was lowered at lower SNRs, and the noise level was increased. For example, when the SNR was reduced to -2 dB, the triplets were presented at a lower level of 60.9 dB SPL and the noise at 62.9 dB SPL.

During testing, the participant was asked to enter the three numbers heard on the numerical keypad of their device. If the response was correct for all three items, the SNR was reduced by 2 dB; otherwise, the SNR was increased by 2 dB. The test began at an SNR of-6 dB. The Speech Reception Threshold (SRT), which is the SNR with 50% correct responses, was determined based on the average DTT of two blocks of 8 triplets after the first correct response. This DTT SRT score was displayed in the Cochlear clinician's portal. Subsequently, if the standard deviation of the SNR across both blocks exceeded 3 dB, the test was flagged as unreliable; in that case, the test was repeated once. If the standard deviation exceeded 3 dB in the second run, the DTT test was stopped, and the result "Participant's responses unreliable" was displayed in the Cochlear clinician's portal. Each time a participant started the DTT test, a short practice session was performed, during which the participants received visual feedback to indicate whether the responses were correct. In addition, the participants had to correctly identify all digits at least once before the actual test started. Otherwise, the test was aborted, and the next part of the test battery was presented.

Questionnaires

In addition to the ATT and DTT, participants completed a number of questionnaires. Daily-life fatigue was administered at T_0 and T_9 with the Checklist Individual Strength (CIS) questionnaire to evaluate daily-life fatigue (Vercoulen et al., 1994). During these sessions, participants also filled out the 12-item version of the Speech, Spatial, and Qualities of Hearing scale (SSQ-12 questionnaire) as part of the complete Remote Check. The Momentary Motivation (MM) questionnaire was used to assess short-term task-related fatigue in every test session (see supplemental appendix A for details of the MM questionnaire, including the list of questions). The Chronotype, CIS, and MM questionnaires were filled out online by the participants using their iPhone, iPad, or PC (Castor EDC, 2022).

Chronotype

The chronotype category expresses whether participants consider themselves a morning person, evening person, or neither. Participants were asked at T_0 and T_9 to rate themselves on a Visual Analogue Scale (VAS) with a 0-10 range as morning-type (0-3), neither (4-6), or evening-type (7-10). They also indicated at what time during the week they typically woke up and went to bed. This sleep pattern was used to plan the morning (T_1, T_2, T_4, T_6) and night (T_3, T_8) tests. At the start of each morning test, participants had to fill out what time they had woken up. Test moments were labelled as 1 = morning, 2 = noon, 3 = night. The interaction between the test moment and chronotype was tested to assess whether the self-rated chronotype affected the outcome at a specific moment. For instance, do morning persons perform better in the morning than the evening?

Checklist Individual Strength questionnaire

The Checklist Individual Strength (CIS) is a validated questionnaire comprising 20 statements concerning daily-life fatigue (Vercoulen et al., 1994). For each statement, the participant rated on a seven-point scale (1-7) how well it applied to their situation and state of mind over the last two weeks. This study used an aggregate score collected at the start and end of the survey (range: 20–140). Scores greater than 76 were indicative of chronic fatigue.

Momentary Motivation Questionnaire

The momentary motivation (MM) is a six-item questionnaire developed for this study to determine motivation, hearing status, and perceived listening effort at each moment of testing (see Appendix A for the list of questions). Four questions reflected the restedness, motivation, and hearing status before testing, and two questions reflected the difficulty and tendency to give up. Participants indicated

their mood directly before and after the measurement on a VAS (range: 0–10). After four questions, participants were instructed to start a Remote Check via the Cochlear app. Directly after completing the Remote Check, participants answered the remaining two questions. As was done for the CIS, an aggregate score was used (range: 0-60).

Analysis

Data from the online questionnaires were pseudo-anonymously stored in Castor EDC (Castor EDC, 2022). The Remote Check data were stored within the Cochlear clinician's portal. The measurements from these two databases were merged using the Statistical Package for Social Sciences (SPSS) [version 28] and manually verified.

To investigate the impact of time-of-day and fatigue on the ATT, DTT SRT, and impedances, three separate linear mixed models (LMMs) were developed. LMMs are ideal for analyzing repeated measures while considering shared variance within subjects and modeling between-subject differences. The models featured Subject ID as a random factor and the other factors as fixed factors, with early morning, noon, and night categorized as time-of-day. To fit the models, the lmer function in RStudio's LME4 package [version 1.1-30] was used in RStudio [version 2022.07.1+554].

Since separate LMMs were created, potential collinearity that might result from a simultaneous decrease in hearing performance reflected in ATT, DTT SRT, and impedance was excluded from the analysis. The factors (abbreviated in parenthesis) included in the models were: Session, Momentary Motivation (MM), Test Moment (TM), Chronotype Category (CC), Consonant Nucleus Consonant word recognition score (CNC), Checklist of Individual Strength (CIS), Age, and CI experience (CI exp). Random intercepts were incorporated to account for individual differences in absolute ATT, DTT SRT, and impedances.

The significance of factors was assessed by inspecting the effect size, t-values, and R-squared explained variance. Multicollinearity within the model (i.e., dependence between factors) was determined by computing the generalized variance inflation factor (GVIF) using the car package [version 3.1-0] in R. As a rule of thumb, values of GVIF < 5 indicate acceptable independence of factors (Tsagris & Pandis, 2021). Effect sizes and confidence intervals were based on the restricted maximum likelihood (REML) estimates (Bates et al., 2015) and were plotted using the sjPlot package in R [version 2.4.1.9000] (Lüdecke, 2018).

RESULTS

Hearing performance of CI recipients was evaluated at-home by repeatedly administering the streamed ATT and DTT using the commercial version of the Remote Check app on an iPhone or iPad. A total of 519 Remote Checks were collected from 50 participants during sessions T_0 – T_9. To investigate the impact of time-of-day and fatigue on the ATT, DTT SRT, and impedance, three separate linear mixed models (LMMs) were developed. After removing double entries (incomplete first entries removed) due to rescheduled measurements, 500 Remote Checks were analyzed.

Reported (technical) difficulties

The main (technical) issues encountered during testing are listed in Table 3. In 37% (183/500) of the tests, technical difficulties resulted in failure to complete (parts of) the test. Critical issues, defined as those resulting in significant delays, missing data, or requiring additional actions by the participant or researcher, were detected in 151 out of 183 technical difficulties encountered. Delays were considered critical if they resulted in a starting time more than 2 hours later than intended.

Table 3. *Reported issues encountered while completing the Remote Check.*

Reported technical issues	Number reported	Relative (%)
Connectivity & delay	31	31 out of 500 (6.2%)
Inability to take a photo and / or to proceed to the next test	28	28 out of 100 (28%)
Inability to complete the DTT	31	31 out of 500 (6.2)
Inability to complete the ATT on all tested frequencies	93	93 out of 500 (18%)
Identified as critical issues (e.g excluding photos and brief delays)	151	168 out of 500 (30%)
Total number of issues	183	183 out of 500 (37%)

ATT Analysis

The mean ATT threshold was calculated using the mean across all determined thresholds (at 250, 500, 1000, 2000, 3000, 4000, and 6000 Hz). The mean ATT threshold per participant per session is shown in Supplemental Figure B2. Only complete ATT measurements, defined as having thresholds completed on all seven test frequencies, were included (n = 407). Figure 1 (left) shows example measurements.

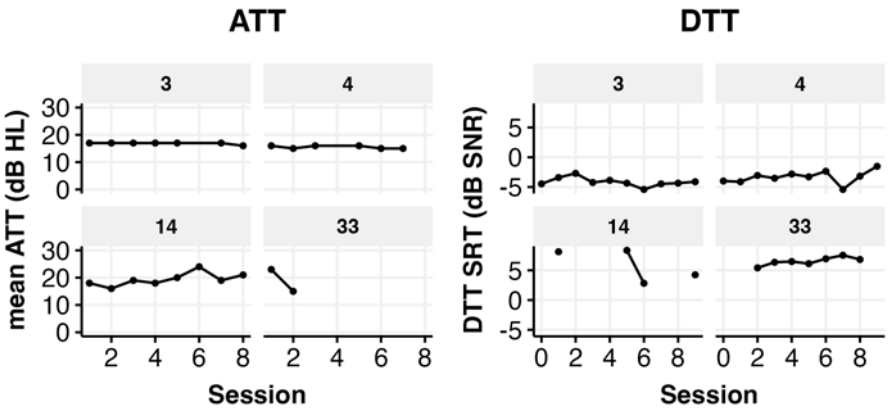

Figure 1. *Mean aided thresholds and digit triplet test speech reception threshold (DTT SRT) per session for a selection of participants. The participant number is indicated at the top of each graph. The left panel displays the mean aided thresholds across all tested frequencies determined via the aided threshold test (ATT), while the right panel illustrates the DTT SRT.*

The mean aided thresholds were stable across sessions. The linear mixed model showed no significant association between ATT and session (b = -.04, SE =.03, t = -1.28, p =.2); see Figure 4 middle panel. In 22 of the 50 participants, all ATT tests were completed on all seven test frequencies. If one clinically accepts audiograms with reliable responses on at least five out of seven frequencies, then 31 participants completed all ATT accordingly. The overall success rate at each measured frequency was 89%. The success rate per clinical audiogram was 82% (453 out of 550). All participants, except participant #31, were able to successfully complete the ATT test at least once. Ten participants produced at least two clinically incomplete audiograms, meaning they completed fewer than five frequencies. Additionally, four participants, identified as #15, #31, #33, and #37 generated clinically incomplete audiograms in at least half of their attempts.

The test-retest accuracy for the ATT was determined by calculating the overall root mean square deviation (RMSD) per complete aided audiogram at different sessions. The mean RMSD repeated within the morning (T_2 versus T_1) was 3.6 dB, and that between sessions in the morning (T_4-T_1) was 3.1 dB. The reliability falls within the 6 dB RMSD criterion recommended for clinically validated automated audiometry approaches (Wasmann et al., 2022). Thus, no clinically relevant differences were observed between and within the sessions.

Impedance Analysis

The mean CI electrode impedance across the array (referenced to the common ground) per participant per session was inspected (see Supplemental Figure B4). No effect of time-of-day was observed. The linear mixed model showed no significant associations between impedance and session (b =.01, SE =.01, t =.78, p =.4) or test moment (b = -.02, SE =.06, t = -.39, p =.7), see Figure 4 right panel.

DTT versus CNC, ATT, and Incomplete Tests

For inspecting the DTT results, the DTT SRT per participant per session is displayed in supplemental Figure B3. Example measurements are shown in Figure 1 (right). Participants with higher (worse) DTT SRTs showed significantly more missing DTT data, as shown in Figure 2 (right). Poorer CI performers (participants with aided CNC scores < 70%) had more difficulties in passing the practice session and could not complete the DTT more often. Based on this, at least one missing DTT score was expected, see e.g., Figure 2 (left panel).

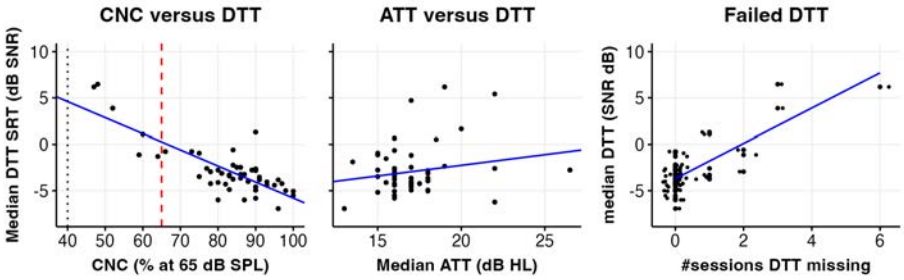

Figure 2. *The relationship between median digit triplet test speech reception threshold (DTT SRT) plotted against aided consonant-nucleus-consonant (CNC) speech reception score, median aided threshold, as well as the number of sessions with missing DTT measurements. All median scores are calculated per participant. The left panel displays the long-term mean aided CNC scores in relation to the median DTT SRT scores (b = -0.17, p <.001), with the vertical dotted black line indicating the audiological inclusion criterion for participation in the study and the vertical dashed red line projecting the median CNC score at 0 decibel signal-to-noise ratio (dB SNR). The middle panel displays the median ATT versus DTT SRT for each participant (b = 0.23, p <.001). The right panel shows the number of sessions with incomplete DTT measurements plotted against the median DTT SRT per participant (b = 1.9, p <.001).*

The median DTT SRT was predicted using a simple linear regression model based on the mean aided CNC results. The median DTT SRTs were significantly related to the mean CNC score taken from the medical records (Figure 2, left panel), and improved by an average of 1.6 dB for each 10% increase in CNC score, see equation 1.

$$DTT\ (dB) = 10.36 - 0.16 \times CNC(\%) \qquad (1)$$

For excellent CI performers with word reception scores of 100% at normal conversation level, the maximum score was -5.6 dB. The predicted DTT SRT upper limit suffered from a ceiling effect in CNC scores. The lower limit had to be extrapolated for two reasons: 1) It is impossible to identify digits without speech recognition, and 2) an inclusion criterion was applied during participant recruitment (greater than 40% CNC), which, according to the linear regression model, corresponds to a DTT SRT of +4.0 dB.

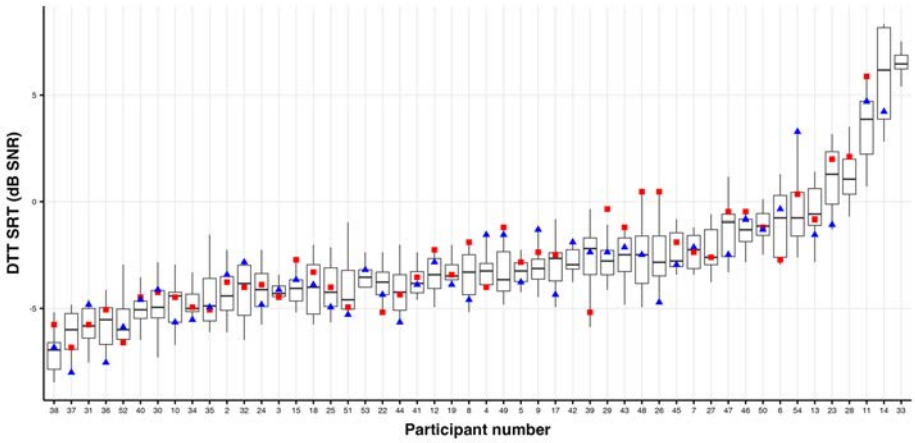

Figure 3. *Box plots of the digit triplet test reception threshold (DTT SRT) for each participant. The outcomes are arranged in ascending order of median DTT SRT, including all ten sessions. The red dots denote the first session (T0) and the blue dots indicate the last session (T9).*

The median DTT SRT across all participants and sessions was -3.3 dB SRT. According to the Shapiro-Wilk normality test, the overall (p =.001) and individual (<.001) DTT SRT scores were not normally distributed. The difference between the best (-7 dB) and poorest (+8 dB) scoring participants, was larger than the difference within participants (Figures 2-3). The individual interquartile range, that is, the difference between the first and third quartile, varied from approximately 1 to 5 dB with a median value of 1.6 dB. The mean individual total range of the DTT [min – max] was 3.9 dB, with 90% of this range within 3.4 dB (here calculated as p_{95}-p_5).

The median absolute test–retest differences between DTT tests repeated within the morning (T_2 versus T_1) was 1.3 dB (IQR 0.6-2.1). The median absolute test–retest differences between sessions in the morning (T_4-T_1) was 1 dB (IQR 0.6-2). Similarly, the mean absolute test–retest difference between DTT tests repeated between sessions around noon (T_7-T_5) was 1.4 dB (IQR 0.7-2), and between night (T_8-T_3) was 1.5 dB (IQR 0.5-2.4). The test-retest differences were not normally distributed

due to outliers. The Shapiro-Wilk normality test was significant for all test-retest differences. When a subset of the 37 best performers (CNC scores at 65 dB SPL ≥ 80%) was analyzed, the test-retest distributions remained non-normal.

Checklist Individual Strength

The Checklist Individual Strength (CIS) questionnaire was used to measure daily-life fatigue at two timepoints (T_0 and T_9). The median CIS score at T_0 was 51 (IQR 33-66). At T_9, the mean score was 50 (IQR 35-69). According to the normative data, the scores were within the normal range at both timepoints (Vercoulen et al., 1994). There was a subgroup with scores around 20 (indicating no daily-life fatigue) and a subgroup with scores around 60 (indicating moderate daily-life fatigue), as shown in Supplemental Figure B1. At baseline, seven out of 50 participants had scores greater than 76, indicative of chronic fatigue. The linear mixed model below revealed a correlation between CIS and DTT SRT, but it was not deemed clinically relevant since a significant change of 8 points in CIS score would result in only a 0.16 dB change in DTT SRT. The linear mixed models showed no significant associations between CIS scores and ATT or Impedance.

Momentary Motivation

The aggregate Momentary Motivation (MM) score was stable across sessions (see supplemental Figure B5). Of the factors comprising the Momentary Motivation (i.e., restedness, motivation, hearing status before the test, effort, and tendency to give up during the test), only restedness changed with time-of-day. As expected, the participants felt tired when they performed the night session just before bedtime. This demonstrates that the effect of fatigue was elicited using the test schedule. Based on the CIS and MM questionnaires, it was concluded that the cohort was representative in terms of daily-life fatigue and that daily-life fatigue varied between test moments and, therefore, could be studied as a factor that impacted the results.

Chronotype and Test Moment

Based on the self-reported VAS scores, participants were grouped into three categories: 1) morning persons (Chronotype score 1-3, N = 14, 28%), 2) neither morning nor evening persons (Chronotype score 4-6, N = 24, 48%), and 3) evening persons (Chronotype score 7-10, N = 12, 24). The outcomes at the three tested times-of-day (also referred to as test moment) were compared to assess its effect on test outcome, taking into account the participants' self-rated chronotype. The timestamps of the completed tests verified participants' adherence to the pre-determined test moments. No evident differences were observed in outcomes

in the morning, noon, or night, nor were differences between the groups with different chronotypes (see Figure 4 for the effect size and confidence interval of all factors, including test moment).

Description of Linear Mixed Models

The effect estimates of the three separate LMMs are plotted in Figure 4. Given the absence of significant predictors in the ATT and impedance models, the analysis below centers on the DTT model, in which the factors explained 75% of the total variance (conditional R^2). The marginal R^2, which is the proportion of variance explained by the fixed effect (Nakagawa & Schielzeth, 2013), explained 49% of the variance. In the DTT model, the marginal R^2 corresponds to the individual differences in participants' DTT SRT performance (conditional R^2 =.75, marginal R^2 =.49, Observations = 301, $N_{participants}$ = 48, p =.003). Due to missing data, only 301 of 500 observations and 48 of the 50 participants were included.

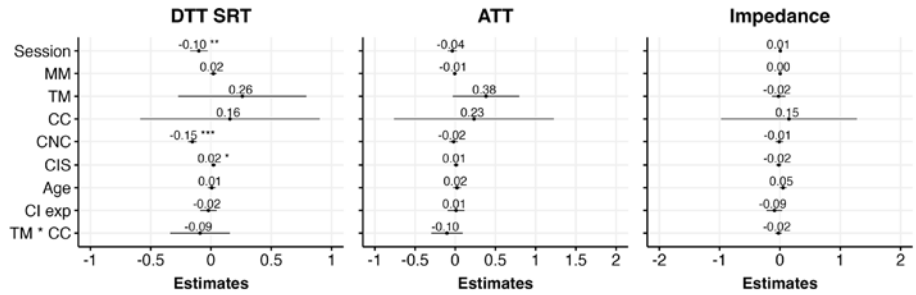

Figure 4. *Effect estimates of the linear mixed models predicting DTT SRT, ATT and Impedance based on repeated sessions T1 – T8. The included factors (abbreviated in parenthesis) are: Session, Momentary Motivation (MM), Test Moment (TM), Chronotype Category (CC), Consonant Nucleus Consonant word recognition score (CNC), Checklist of Individual Strength (CIS), Age, and CI experience (CI exp). Factors that increase the outcome are represented in blue, while those that decrease the outcome are represented in red. Significance is indicated by asterisks (*p =.038, **p =.007, ***p <.001).*

The DTT model showed a small but significant effect of session on DTT (b = -.1, SE =.04, t = -2.7, p =.007), with the DTT SRT improving on average by 0.1 dB with each subsequent session. This was a small but significant learning effect. Removing sessions T_0 or T_9 did not affect the results; there was a gradual improvement in DTT over time (see Figure 4). The DTT model also indicated a strong effect of CNC scores (b = -.15, SE =.02, t = -8.38, p <.001), as a 10% increase in CNC was associated with a 1.5 dB improvement in the DTT SRT (see, e.g., Equation 1), and a small effect of the Checklist Individual Strength (CIS) score (b =.02, SE =.01, t = 2.09, p =.038). The

GVIF of session, CNC, and CIS was < 1.5, indicating negligible collinearity within the model. None of the other factors were significant predictors of the DTT SRT, including time-of-day, chronotype, its interaction with time-of-day, momentary motivation, daily-life fatigue, impedances, aided threshold, and CI experience.

DISCUSSION

The primary goal of this study was to investigate the effect of time-of-day, fatigue, motivation, and chronotype on auditory performance in at-home measurements in CI recipients, focusing on the thresholds (ATT), impedances, and the speech perception in noise (DTT). Contrary to the hypothesis, the results indicated that time-of-day effects, chronotype, momentary motivation, and task-related fatigue do not significantly impact the aided thresholds, impedances, and speech-in-noise test performance during self-testing at home. The data collected showed that Remote Checks provide a reliable snapshot of auditory performance, exhibiting acceptable test-retest characteristics. Stable results were observed over time for participants; differences of ≥ 4 dB in DTT SRT rarely occurred within the studied cohort. Thus, a change exceeding 4 dB from baseline in DTT may indicate the need for additional assessment, aligning with the clinically significant difference of 3.1 dB identified in prior research (Maruthurkkara et al., 2022).

Remote Checks can be performed remotely at-home, at work, or even while traveling abroad, without problems or the need for synchronous support from a clinician. Using the ATT, determining streamed aided thresholds with the CI was straightforward. Only four out of the 50 participants had problems conducting the test. Approximately one in five ATTs was not fully completed, possibly because the algorithm aborts the test when the participant responds to stimuli below 10 dB HL. In cases in which the electric thresholds (i.e., T-levels) are fitted at levels higher than the actual thresholds for CI stimulation, (very) soft stimuli may become audible. Consequently, the algorithm might erroneously terminate the test if the thresholds of streamed stimuli fall below the 10 dB HL limit. This phenomenon could explain the failed ATT tests observed in participants with otherwise good hearing (for example participants #31 and #33, as illustrated in Supplemental Figure B3). To address test length and participant concentration issues, future improvements can be made by adapting the test paradigm. For instance, utilizing a fully adaptive threshold-seeking algorithm can be particularly advantageous. Such algorithm takes into account all previous responses across frequencies to infer thresholds, unlike the current partially adaptive procedure that relies on responses from a

single test frequency (Wasmann et al., 2022). Six of the 50 participants had multiple incomplete DTT measurements. Poorer DTT scores were associated with a higher number of failed DTTs. Almost all participants in the studied cohort had streamed aided thresholds of 20 dB HL or better, averaged across sessions (see middle panel of Figure 2). Unfortunately, there are too few measurements with an ATT greater than 20 dB HL to assess whether increased aided thresholds leads to poorer DTT, as was predicted in the model by de Graaff et al. (2020). Please note that in the model by de Graaff et al. (2020), the thresholds were determined using a different threshold-seeking method and stimuli were presented via free field instead of streaming.

While hearing tests are useful tools, they may not capture all the nuances or issues that CI recipients may experience (e.g. loudness imbalance between frequency bands, poor battery capacity, difficulties to engage at meetings at work, etc.) and that could be identified during face-to-face conversations with trained clinicians during aftercare (Maruthurkkara et al., 2021). Based on interactions the researchers had via the messenger system, it is advisable to add to the hearing tests questionnaires that query personal matters (including open questions such as "What matters to you?" or "Did you experience important changes in your life") and to provide resources for online counseling (for Q&A and troubleshooting) leading to more personalized and engaging remote care in the future.

The linear mixed model approach showed that the DTT SRT varied substantially between participants; 49% of the total variance was explained by baseline differences between participants. A learning effect was observed in repeated DTTs. Unlike naive normal-hearing listeners, CI recipients showed a prolonged training effect beyond the initial sessions. In contrast to the study by (Smits et al., 2013), most participants in our study had prior experience with the DTT. Kropp et al. (2021) suggest that additional test runs can reduce procedural learning effects, despite the potential fatigue it may incur. Moreover, an auditory training effect similar to that described by Oba et al. (2011) might occur with repeated DTT administration, possibly improving CI recipients' speech discrimination in noise. Oba et al. (2011) found that extensive auditory training, using a closed-set digit recognition task, significantly improved speech recognition and retained benefits for up to a month post-training. Thus a prolonged learning effect may be explained by a combination of above factors.

The CNC and CIS scores were the other significant factors determining the DTT score; for each 10% increase in CNC score, the DTT SRT improved on average by 1.6

dB, and for each 5-point increase in CIS score, the DTT deteriorated by 1 dB. The CNC score is a good indicator of whether a CI recipient will be able to perform the DTT successfully (see Figure 2). Based on the findings of this study, it is recommended (in a Dutch setting) to use the Dutch DTT only in CI recipients when one expects a DTT SRT of 0 dB or better (corresponding to a CNC ≥ 65% at 65 dB SPL) to avoid frustration in (relatively poorly performing) CI recipients who are unable to pass the test successfully. The ≥ 65% criterion recommended based on in this study is more strict than the ≥ 40% CNC score previously recommended by Kaandorp et al. (2015). The main reason for this discrepancy may be that poorer-performing participants had difficulty passing the practice session during the Remote Check, and their DTT SRTs were more frequently flagged as unreliable (greater than 3 dB standard deviation across the trials within the session). Unfortunately, the commercial Remote Check implementation does not display all test details, and the test log files are only temporarily stored on the participant's device. Therefore, it was not possible to more thoroughly investigate the root causes of the problems encountered when performing the ATT and DTT.

Limitations of this study

Firstly, the pre-determined timeslots might not have been ideal for task performance due to individual chronotypes. Morning sessions may have been too early for "morning-type" participants, while the test moment for "evening-type" participants could have been too close to bedtime. This assumption is supported by reports from evening-type participants, who mentioned feeling tired during the evening tests. To improve accuracy, it is recommended to consider participants' peak moments and adjust the test schedule accordingly. Secondly, the design was planned to balance the testing order, thereby nullifying potential learning effects by always having the morning session (e.g., T_6) before the night session (T_8). However, it did not control for a potential learning effect in noon sessions, which were always after a morning session. Nor did the design control for a distinct learning effect in bilateral CI participants; the number of bilateral CI recipients was too small to assess this specific form of learning. Thirdly, to limit the burden of questionnaire completion for participants, a brief self-reported chronotype instead of the established Morning-Evening type Questionnaire (MEQ) was used. Nevertheless, the distribution of chronotype categories was similar to that of the general population, of which 60% had no particular preference between the morning and evening (Merikanto et al., 2012). Fourthly, the study ran from January 2021 until April 2022, while COVID-19 restrictions were in place, including advice to stay at-home as much as possible, work remotely, curfew, and a reduced number of visitors one could receive at-home. The COVID-19 restrictions, but also seasonal changes due to variations in daylight

hours, may have affected the participants' daily routines and sleep patterns. No effect of time-of-day on DTT was found, nor was an interaction between the time-of-day and chronotype. Therefore, alternative methods to determine chronotype are not expected to lead to a different conclusion. Fifthly, only subjective measures were used to assess workload or fatigue. This may have led to outliers in the DTT caused by daily-life or task-related fatigue that participants were unaware of and therefore did not report in the subjective MM. However, the influence of subjective (or objective) task-related fatigue may be counteracted by increased attention and listening effort during the relatively short task. Notably, the cohort of adult CI recipients reported similar daily-life fatigue as normal or hearing-impaired listeners on the CIS questionnaire (Vercoulen et al., 1994; Y. Wang et al., 2018). The median scores were comparable to those reported previously and did not significantly differ from those of normal hearing or hearing-impaired (Y. Wang et al., 2018). CIS and DTT SRT did not correlate with self-reported listening effort. Therefore, although only self-reported listening effort or fatigue was determined, these factors did not affect auditory performance over time. Auditory tasks other than the DTT are needed to study the effect of listening effort and fatigue at-home. Sixth, an early commercial version of the Remote Check app was used. Technical issues or difficulties occurred in 20-40% of the measurements. Initial in-clinic instructions might have prevented some of the challenges experienced by the participants. Fortunately, many of the technical issues were solved during this study. The photograph feature of the app was not essential for the purposes of this study. However, because we used an early commercial version of the Remote Check app, it was not possible to exclude photos from the Remote Check test battery. Meanwhile, a more flexible and improved version of the Remote Check has become available for clinical use in both iOS and Android devices.

Despite the COVID-19 pandemic, when in-clinic visits were not possible, a large amount of data was collected in a relatively short period of time. Our findings suggest that remote assessments can be performed successfully anywhere and anytime. This may lower barriers to large-scale data collection and the creation of data lakes, allowing for the detection of patterns that may not be evident in smaller monocenter cohorts. However, other important factors, such as how to deal with algorithmic bias, risk of re-identification, data ownership, what data to transfer when referring (including when to refer), and how to share data appropriately, were beyond the scope of this work but need particular attention in future (see for a discussion of those topics Wasmann et al. 2021).

CONCLUSION AND RECOMMENDATIONS

This study is one of the first to examine the validity and stability of remote assessments in CI recipients and identify important factors (e.g., speech perception in quiet at 65 dB SPL) and those that are not (e.g., time of day, fatigue, motivation, and chronotype). A Remote Check is a relevant tool for CI recipients to monitor their auditory performance over time. For clinicians, automated procedures to flag suspicious results will become essential to organize their workflow efficiently. Future steps involve establishing referral criteria for when performance declines, and defining the appropriate support for specific inquiries, such as assistance from the clinic, the CI manufacturer, or a hearing-aid dispenser partnering with the CI manufacturer. The ultimate step is to empower CI recipients with their data.

CI recipients are able to initialize remote at-home measurements to monitor their performance anywhere and anytime. Motivation, daily-life fatigue, or chronotype do not affect the outcomes of the ATT or DTT, and the test-retest falls within clinically acceptable limits. If needed, deviating tests can be repeated (after additional instruction) and interpreted in concert with accompanying questionnaires before making clinical decisions. Based on the findings of this study, it can be concluded that at-home testing is reliable for cochlear implant recipients and offers an opportunity to provide care in a virtual hearing clinic.

7

Feasibility of a Cochlear Implant Fitting Approach Based on Phoneme Confusions: Lessons Learned from the AuDiET Study

Jan-Willem Wasmann*, Enrico Miglioni*, Nikki Philpott, Birgit Philips, Bas van Dijk, Wendy Huinck

Manuscript in preparation: Wasmann, J.-W. A., Migliorini, E., Philpott, N., Philips, B., van Dijk, B., & Huinck, W. (2024). *Feasibility of a Cochlear Implant Fitting Approach Based on Phoneme Confusions: Lessons Learned from the AuDiET Study*. https://doi.org/10.31219/osf.io/k8f2y

ABSTRACT

Background: Traditional speech recognition testing in Cochlear Implant (CI) care primarily captures aggregate speech recognition performance, often overlooking detailed phoneme identification errors. This feasibility study introduces a fitting approach focusing on individual CI users' phoneme difficulties identified through self-testing paradigms.

Methods: Twenty-three postlingually deaf, experienced CI users underwent fitting adjustments based on Phoneme Recognition in Quiet test outcomes. A basic fitting check was followed by advanced fitting adjustments that ranged from generic (7 out of 23) to specific adjustments targeting specific phonemes (16 out of 23).

Results: The new MAP was preferred by 74% (18 out of 23) of participants, yet the aggregate phoneme identification performance showed no significant change between the pre- and post-fitting visits. However, a positive trend in targeted phoneme identification was noted ($t(22) = -2.3$, $p = .03$), approaching but not reaching conventional significance after Bonferroni-Holm correction (adjusted $p = .09$). A significant improvement in targeted phoneme identification was observed in the subgroup that adhered to a targeted fitting ($t(11) = -3.3$, $P = .006$, adjusted $p = .03$, Cohen's $d = .88$).

Conclusion: Using phoneme identification evaluations in the CI fitting process in experienced adult CI users is feasible.

Keywords: Cochlear Implant, Speech Perception, Cochlear Implant Fitting, Phoneme Identification, Self-Administered Testing

INTRODUCTION

Background

Improving speech recognition is an important goal for adult cochlear implant (CI) users, reflecting their desire to achieve better communication in their daily lives. A CI offers enhanced audibility of everyday sounds post-fitting, which involves fine-tuning the parameters of the CI system followed by a period of acclimatization and auditory training. Typically, it provides improved speech recognition in both quiet and (in practice to a lesser extent) noisy environments, often leading to a notable enhancement in quality of life (McRackan et al., 2018). However, it is worth noting that a significant degree of variability in performance among CI users has been documented (Holden et al., 2013). To limit this variability, acquiring detailed data on performance is a prerequisite. In most clinical care centers, only the aggregate scores in speech recognition performance tests are recorded, without specifying which phonemes are difficult to identify (Buchman et al., 2020). The variability and lack of detailed insights underscore a critical knowledge gap: the challenge in adequately determining whether each CI user has achieved the most optimal rehabilitation that suits their potential.

Clinical CI fitting practices

In this paper, we focus on the Cochlear™ Nucleus® system so naming conventions and fitting procedures may be biased towards that system, however, the general principles remain valid for all cochlear implant systems. Fitting involves setting electrical T-levels (just audible levels) and C-levels (most comfortable levels) per electrode, typically through methods including threshold-seeking, loudness scaling, or loudness balancing across electrodes (Skinner et al., 1995). Threshold Neural Response Telemetry (T-NRT) levels offer another fitting approach. Although the correlation between T-NRT and individually set C-levels is moderate to low, the *shape* of the T-NRT profile appears to correlate with the *shape* of the C-level profile (Botros & Psarros, 2010; W. K. Lai et al., 2009). Individual variability in T- and C-levels stems from factors such as cochlear anatomy and electrode positioning, affecting the spread of excitation (Stickney et al., 2006). Audiologists also consider the CI user's hearing history and behavioral aspects, which, along with the personal preferences of both the CI user and the audiologist, can potentially influence fitting outcomes (Vaerenberg et al., 2014). In most CI centers, parameters other than T- and C-levels are typically set to "default" (Vaerenberg et al., 2014) or are adjusted based on "informal" feedback from the CI user in terms of sound quality, auditory and non-auditory sensations (e.g., facial nerve stimulation, pain or dizziness).

Based on surveys among clinicians, Vaerenberg et al. (2014) and Browning et al. (2020) found a high degree of variability in CI fitting procedures among clinicians and centers. There is still no consensus on what constitutes good clinical practice (Wathour et al., 2021), although efforts to reach consensus and create international guidelines are progressing (Buchman et al., 2020). Given the variability in outcomes with CI, it may be important to include other input than just single channel loudness to guide T- and C-level fitting procedures, as described in the next section. Additionally, evidence-based procedures on how to fit those parameters are needed.

Speech-based Fitting strategies

Instead of fitting based on T-and C-levels, some fitting approaches explicitly use speech recognition outcomes to guide fitting. Baskent et al., (2007) applied genetic algorithms to optimize settings in hearing aid and CI fitting based on participants' rating of speech intelligibility. Holmes et al. (2012) used the "*Clarujust optimization*" method of CI fitting. The exact details of Clarujust were proprietary, but they describe that they systematically varied pulse rate, loudness growth and Frequency Allocation Table (FAT) based on the outcome of vowel-consonant-vowel stimuli tests and evaluated the effect on speech in quiet (CVC) and sentences in noise (BKB-SIN) test outcomes. The method of Holmes et al. (2012) did improve outcomes, reporting that optimal parameters varied between individuals without a clear pattern on how to adjust parameters, meaning it is based on trial and evaluation. Noteworthy is that distinct combinations of parameter settings resulted in similar outcomes in speech recognition.

The *Fitting to Outcome eXpert* system (FOX; Otoconsult NV, Antwerp, Belgium) is an Artificial Intelligence (AI) based decision support system for fitting cochlear implants. It calculates a utility function based on a weighted combination of outcome measures, including aided thresholds, loudness growth, phoneme discrimination, and speech recognition (CVC), to predict and evaluate fitting settings (Meeuws et al., 2017; Wathour et al., 2023). The exact details of the FOX system, including how the outcome measures are weighted, are proprietary. On average, FOX system-based CI fittings lead to clinical outcomes comparable to traditional fitting approaches (Waltzman & Kelsall, 2020) and may reduce variability across CI centers (Battmer et al., 2015). However, it is hard to pinpoint how a specific speech perception error leads to a specific fitting intervention (traceability of how the utility function is defined and updated) or how each fitting intervention influences speech perception errors on a phoneme level. So, while fitting approaches exist based on speech outcomes, it remains unclear how exactly a change of fitting parameters leads to a change in speech perception and how fitting procedures can be improved based on these data.

Our Fitting Approach and Hypothesis

In this feasibility study, we systematically examined which phonemes experienced CI users find difficult to identify. We adjusted the fitting based on these findings and evaluated the effect by utilizing self-testing paradigms. At the start of the study, participants' performance on the Phoneme Recognition in Quiet (PRQ) test (Migliorini et al., 2024) was measured using their preferred CI fitting, i.e., MAP. Based on this, up to three of the largest systematic vowel and/or consonant confusions were selected for each participant. A confusion is a {*stimulus*, *response*} phoneme pair, indicating that a specific phoneme {*stimulus*} was misidentified as phoneme {response}. The data were collected and visualized using tools specifically developed for this study.

We aimed to explore the effect of fitting adjustments based on individualized phoneme errors made in the PRQ test. We hypothesized that the perception of the most systematically confused phoneme pairs could be improved by adjusting stimulation patterns within our control, i.e., by adjusting the MAP. First, a fitting check was performed following standard clinical care practices. Subsequently, advanced fitting procedures were performed to improve the fitting based on the individual phoneme error patterns of each participant. Those advanced fitting procedures ranged from generic adjustments (not targeting a specific phoneme confusion such as a Pulse Rate change) to more specific adjustments targeting the largest systematic phoneme confusions. Immediately after adjusting the MAP, the effect per newly created MAP was measured using a reduced version of the PRQ test. Based on the test outcome and participants' willingness to use the new settings, a preferred MAP was selected (referred to as take-home MAP) and used for two weeks in daily life after which a new evaluation followed. If advanced fitting approaches alter PRQ performance, this might lead to future fitting approaches that potentially improve PRQ results and subsequently lead to better speech recognition performance.

METHODS

Study Design and Setting

This feasibility study was conducted using a single-center pre-post interventional design and performed at the Radboud university medical center's outpatient clinic in Nijmegen, The Netherlands (Radboudumc). The local ethics committee of the Radboudumc approved the study (METC, file number 2022-13495). This study is part of the Auditory Diagnostics and Error-based

Treatment (AuDiET) trial, additional information on that study can be found at https://clinicaltrials.gov/study/NCT05307952.

Participants

Twenty-seven adults (14 males, 13 females with a mean age of 69±11 years) with a post-lingual onset of severe hearing loss were included in the AuDiET study. Of these participants, twenty-three completed the fitting part of the study. Participants were recruited via general notices, via email, and during regular visits at the Radboudumc audiology center. Recruitment and testing took place between 2022 and 2023, with the first visit on the 30th of May 2022 and the last visit on the 4th of December 2023. Inclusion criteria were a minimum age of 18 years and at least one year of experience with a Cochlear® Ltd. CI implant (model CI422, CI512, CI522, CI532, CI24M, CI24R, or CI24RE). Exclusion criteria were abnormally formed cochleae, severe pre-implantation ossification, severe cognitive disorders, intense facial nerve stimulation, unaddressed tip fold over, or more than four deactivated electrodes. Participants' aided audiometric thresholds, baseline phoneme scores measured using CVC words, and demographic details are detailed in Table 1 by Migliorini et al. (2024).

Equipment, Calibration / Experimental Setting

The experimental setup consisted of a laptop (Lenovo Thinkpad T440, Hong Kong) connected to an RME Fireface UC audio card (Fireface UC, RME intelligent audio solutions, Haimhausen, Germany) via USB2.0. The audio output from the Fireface UC was presented via a Direct Audio Cable. The participants used a Cochlear™ Nucleus® 6 loaner processor during the session. The equipment was calibrated by monitoring the Digital Signal Processor (DSP) input levels of the Nucleus 6 processor using a proprietary tool from Cochlear Ltd. This calibration procedure involved comparing the audio streamed to the processor with reference levels obtained from a similar processor placed on a mannequin in a calibrated free-field environment within a soundproof room. The calibration ensured that test signals via the soundcard were delivered at a 65 dBA equivalent level. Since stimuli used in the PRQ and CVC tests were directly streamed to the participants' speech processor (audio-input only) and participants had limited or absent bilateral residual hearing, acoustically shielded rooms were not deemed necessary.

During the PRQ test, participants listened to Dutch triphones streamed directly to the Nucleus 6 processor. The triphones had a /hVt/ or /aCa/ structure, where V denotes a vowel (ɑ, a, ɑu, ɛ, e, ɛi, ø, ɪ, l, ɔ, u, o, ʏ, œy, y) and C a consonant (b, d, f, ɣ, h, j, k, l, m, n, p, r, s, t, v, w, z). The participants were instructed to select on the laptop screen what triphone they heard from all possible options, with vowels

and consonants being assessed separately. The next stimulus was presented as soon as the participant responded. Therefore, participants could not correct their responses. The stimuli were presented in randomized order, with each consonant being presented eight times and each vowel six times. In the reduced version of the test, both vowels and consonants were presented only twice. The average testing time for the full PRQ was approximately ten minutes, while the reduced version of the test took about three minutes to complete. The test software, specifically designed for this study, was developed in Python 3, and further details can be found in Migliorini et al. (2024).

In the full CVC test, the participants listened to fifteen word lists. Due to constraints in overall testing time and listening effort, a reduced CVC test was created in which only two lists were presented. Each list contained twelve meaningful Dutch CVC words from an NVA word list in randomized order (Bosman & Smoorenburg, 1992). Participants knew they listened to short existing meaningful words. They were instructed to type what word was heard and, in case of doubt, make their best guess. The responses were automatically parsed into triplets of phonemes to allow for a detailed analysis of phonemic errors. The software for the CVC test, developed by Cochlear Ltd., has been previously validated in a clinical setting, as reported by de Graaff et al. (2018).

All tests were conducted in consultation rooms at the ENT department normally used for CI and hearing aid fitting, conforming to NVKF norms (Dingemanse et al., 2023). Pure tone audiometry was tested using narrow-band noise in a free field setup in a non-soundproof room using a speaker placed 1 meter in front of the participant, tested in the CI-only condition using the Nucleus 6 loaner processor. The contralateral hearing aid was switched off and the contralateral ear was covered with an earmuff in case of significant residual hearing in that ear.

Initial Assessment

Participants underwent a comprehensive test battery during the first visit using their daily MAP (from now on referred to as original MAP) on a Nucleus 6 processor with preprocessing switched off (e.g. autosensitivity control (ASC) and adaptive dynamic range optimization (ADRO)) and in audio-input only mode. The relevant part of the test battery for assessing the CI fitting included: aided pure tone audiometry using a Hughson-Westlake staircase procedure and the self-test versions of the full PRQ and full CVC test. Based on our goal of using phoneme identification and the results of the Visit 1 analysis (Migliorini et al., 2024), it was decided only to use PRQ as our primary endpoint.

Confusion matrices

The outcomes of the PRQ test were summarized in confusion matrices (Miller & Nicely, 1955; Remus et al., 2007). These square matrices plot the target phonemes against the responses, with a 17x17 matrix for consonants and a 15x15 matrix for vowels. Within these matrices, cell values represent the response ratio out of the total times a phoneme was presented, with diagonal entries reflecting the correct identification rates per phoneme. As an illustrative example, the confusion matrices of Participant 25 based on the PRQ test at visit 1 are shown in the results section in Figure 4.

The confusion matrices provided a quick way for the audiologist to assess phoneme confusions. Using the confusion matrix, up to three of the largest systematic vowel and/or consonant confusions were selected for further analysis based on the error distribution. Systematic errors typically involve confusion between a phoneme and one or two alternatives, suggesting they stem from distinct perceptual features potentially modifiable through fitting adjustments.

Electrodograms

The differences in CI activity for vowel and consonant confusions were analyzed by comparing electrodograms. These are electrode-output visualizations of the electrical activity per electrode over time of the sounds captured by the CI processor. The electrodograms were created by processing WAV files of the phonemes from the PRQ test using the Nucleus® Matlab Toolbox (NMT; Swanson & Mauch, 2006) in Matlab R2022a. The NMT software simulated the signal processing of the participant's implants using the following parameters: ACE™ strategy, default FAT, PR=900 pps, maxima (n=10).

By comparing the electrodograms of the presented versus responded phonemes, the differences at the electrode level for default CI parameters were assessed. Before the advanced fitting, the electrodograms were visually inspected by two audiologists to determine the maximum contrast between phoneme confusions, which is the maximum difference in electrode activity, illustrated by the red circles in Figure 1. This contrast was manually determined because the onsets and duration of the PRQ stimuli were not aligned, making it difficult to automate this assessment. The use of electrodograms in the fitting procedure was introduced to visually guide the audiologist during fitting sessions. There were no specific guidelines that had to be followed.

Based on the identified contrasts, several manipulations to increase the contrast were proposed. For instance, a targeted intervention to increase the contrast

between HOT-HAT, emphasizing HOT for a default FAT, could be to increase the C-level of electrode 21, and decrease the C-levels of electrode 13 and 12 due to their correspondence to the largest visible differences observed in the electrodograms. Other generic interventions included lowering the pulse rate, increasing the C-level over a broad range of electrodes either all apical or basal, or deactivating electrodes. In case a participant did not have a default FAT (11 out of 23 subjects did not have a default FAT), the audiologist manually looked up the frequency band corresponding to the electrode of the default FAT versus the actual FAT.

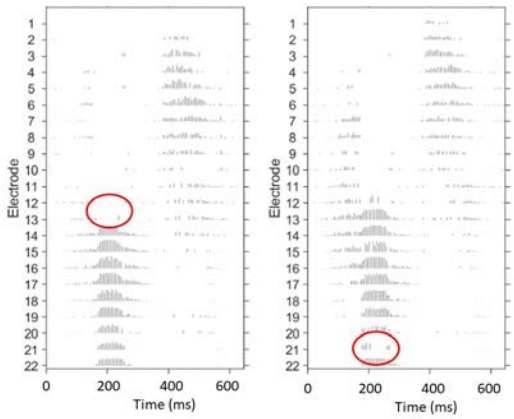

Figure 1. *Electrodograms of the stimuli 'hot' (left) and 'hat' (right). The horizontal lines show the CI-output in terms of pulses per electrode in time. The amplitude of the pulses is modulated by the slow-varying envelope of the speech signal and determined by the signal processing of the CI. The difference in presented versus responded vowels is visible around 200 ms. The red circles indicate the electrodes with the highest contrast between /ɔ/ ('hot') and /ɑ/ ('hat').*

Formant Deviations

Formant analysis was conducted to examine peaks in the vowel frequency spectrum, comparing the formants of presented vowels to the formants of the participants' responses. This comparison was visualized in a formant deviation graph. The formants were depicted from the WAV files of the vowels of the PRQ test using PRAAT (version 6.1.16) (Boersma, 2001). PRAAT extracts the formant frequencies at any given time in an audio file; the first two formants were taken from the vowel part in all the /hVt/ audio files used in testing and plotted on a cartesian graph. The blue lines represent each electrode's upper and lower frequency limit as extracted from each participant's MAP. From the vowel confusions, deviations of the first and second formants were studied, shown for example in Figure 2.

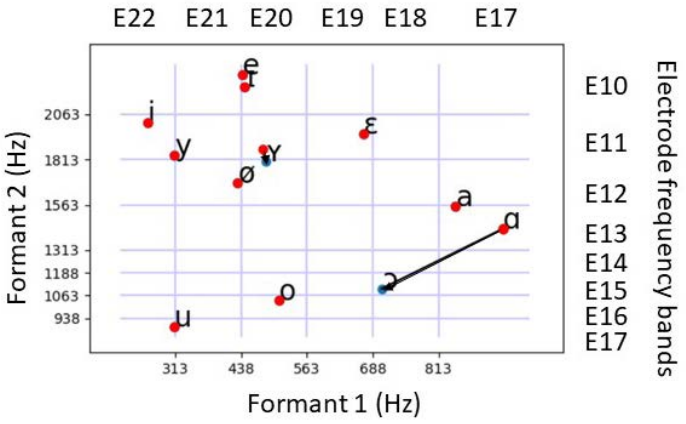

Figure 2. *Formant deviation graph. Example of hot – hat (ɔ-ɑ) confusion. The vertical lines delineate the frequency band between electrodes E22 to E17, which is the part of the spectrum where the first formant is expected, and the horizontal lines delineate the frequency band between E10 and E17. Red dots represent correct formant responses, while blue dots indicate deviant formant responses. The arrows illustrate deviation from the target formant.*

Fitting Procedure

At the second visit, a standard clinical fitting, referred to as a basic fitting check (see supplementary material A for more details), was conducted to ensure that optimal clinical CI fitting care was achieved before advanced fitting (visit 2). Based on the participants' error patterns from the PRQ test obtained at visit 1, the audiologist selected up to three phoneme pairs with the largest systematic errors for each participant. The audiologist aimed to address these errors both during basic fitting and, more explicitly in the advanced fitting.

The basic fitting involved a check of fitting parameters, including impedance, compliance, T and C-levels, and electrode functionality, ensuring sounds were audible and stimulation was comfortable. Speech perception changes were evaluated using a reduced CVC test, guiding the choice between continuing with the new basic fit MAP or reverting to the original MAP.

Consequently, advanced fitting aimed to directly target phoneme confusions was employed starting with either the original MAP or the new basic fit MAP as described in the previous paragraph. Advanced fitting employed generic adjustments (e.g. Pulse rate changes, low- or high-frequency boosts) or specific targeted interventions for the largest systematic errors. These approaches were individually tailored based on clinical experience and consensus discussions between experienced audiologists, utilizing electrodograms and formant

deviations to create intervention proposals. Fitting adjustments were implemented based on participant feedback, focusing on parameters including T- & C-levels, FAT, electrode (in)activation, and Pulse Rate adjustments. Adjustments were guided by live feedback on sound quality and outcome on the reduced CVC and PRQ test.

In summary, the order of activities is:

1) Basic fitting
    i) Perform reduced CVC
    ii) Decide which MAP to continue with

2) Advanced fitting
    i) Repeat per variant reduced CVC + PRQ
    ii) Decide which MAP to take home

The flow diagram in Figure 3 shows the type of intervention participants received, targeted (n=16) or generic (n=7) and the adherence to the new MAPs. After visit 2, all participants were provided with the new take-home MAP but the original MAP was stored in program slot 2, so they retained the option to revert to their original MAP at any time. Two weeks later at visit 3, performance with the new take-home MAP was evaluated.

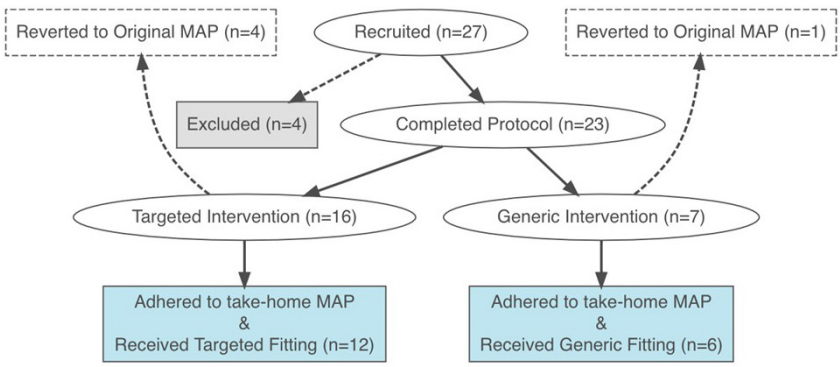

Figure 3. *Flow diagram showing the type of intervention, targeted versus generic, and adherence to the new MAPs.*

From Participant 13 onward, a modified procedure was implemented, utilizing an improved confusion matrix tool (version 2.0) for the immediate assessment of phoneme confusions. In addition, the audiologist now repeatedly presented the

targeted phoneme pairs with live voice to the participants, and based on their subjective feedback—whether the contrast had improved, worsened, or remained unchanged—the audiologist adjusted parameters such as the C-level for targeted electrodes. This fitting-evaluation loop aimed to directly address participant-specific phonemic confusions more effectively. However, a potential weakness of the live fitting-evaluation loop was that the live spoken phonemes differed from the recorded material used in the PRQ test. The most obvious difference was that the audiologist's voice was male, as opposed to the female voice, which was implemented in the PRQ material. Simultaneously fitting the CI and streaming PRQ stimuli was not possible, and programming every single fitting adjustment to the CI processor was too time-consuming. Despite these challenges, the decision regarding which MAP to take home was made based on outcomes from the confusion matrix and participant preferences, rather than on aggregate PRQ scores. The aggregate PRQ scores were monitored to ensure no significant performance decrease occurred.

Statistical methods

To assess the impact of the fitting intervention, PRQ test outcomes from visits 1 and 3 were compared. Following a post-hoc analysis that confirmed the data was within acceptable agreement of a normal distribution (Shapiro-Wilk Normality tests), paired t-tests were applied for statistical comparison, with a Bonferroni correction implemented to mitigate the risk of Type I errors due to multiple comparisons (Nosek & Lakens, 2014). While Bonferroni correction is commonly used to manage the familywise error rate, the exploratory nature of this study also required the incorporation of effect sizes and confidence intervals for a more nuanced understanding of the impact of our interventions (Rubin, 2017).

In addition to aggregate performance metrics, performance on the targeted phoneme pairs was analyzed in depth separately for each participant. For each targeted phoneme pair, which included two distinct stimuli, a sub-score was calculated for these stimuli, referred to as a sub-score of targeted phonemes. In case three phoneme pairs were targeted, the sub-score of the combined six targets was calculated. If a participant had fewer targeted phoneme pairs—two participants had one, four had two, four had three, and one (Participant 25) was an exception who had four targeted pairs due to combining b-p and s-z in a single intervention—the sub-score of targeted phonemes was based on the corresponding number of presentations.

Due to time constraints and to minimize the burden for participants, only the reduced version of the CVC test was performed at visits 2 and 3. The data collection

and subsequent analyses were conducted using Python 3, supported by the SciPy package. Graphical representations were predominantly generated through Matplotlib, except the flow diagram and the estimation plots, which were created with R (RStudio [version 2022.07.1+554]; dabestr version 2023.9.12; Ho et al., 2019)). For statistical tests (t-test, Bonferroni-Holm, Shapiro-Wilk), the rstatix package [version 0.7.2] was employed, and for effect sizes, the dabstr package was utilized.

RESULTS

This study explored CI fitting adjustments based on PRQ test outcomes in experienced CI users. A basic fitting according to clinical practice was performed, followed by advanced fittings. The advanced fittings ranged from generic adjustments to specific adjustments targeting individualized phoneme pair confusions.

Eighteen participants preferred the (new) take-home MAP at the end of the study, while five reverted to their original MAP. Participants reported one or multiple of the following subjective reason(s) for not accepting the take-home MAP: discomfort at higher C-levels in basal electrodes (n = 2), annoyance from environmental sounds or their own voice after changes to apical electrodes (n = 1), or a decrease in perceived sound quality and performance (n = 4). These complaints prevailed even after at least two weeks of trying.

An example of the confusion matrices is shown in Figure 4, displaying the phoneme identification results for Participant 25 during visit 1. This participant consistently confused the phoneme /y/ (as in the Dutch word 'huut') with /Y/ (as in 'hut'). Additional confusions for vowels and consonants were also identified, including but not limited to /Y/ with /ɛ/ and /ɔ/ with /ɑ/, as well as the consonant pairs l-n, b-p, and v-f.

Phoneme contrasts identified in the electrodograms (maximum difference between electrodes) differed from those found in formant deviation graphs (difference in F1/F2). The electrodograms and formant deviations provided distinct insights, guiding the advanced fitting procedure's approach for vowels differently at the electrode level.

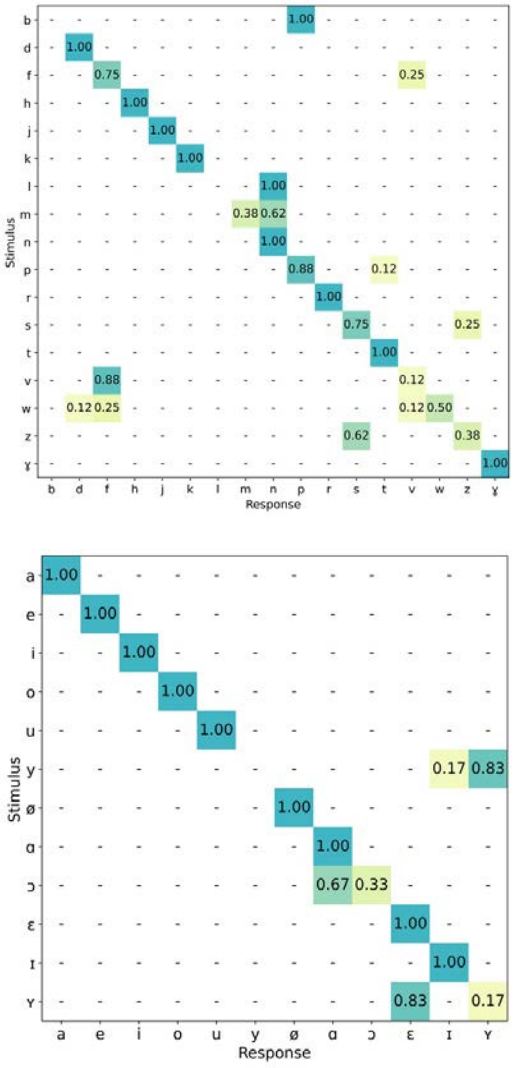

Figure 4. *Example of confusion matrices of vowel confusions (top) and consonant confusions (bottom) of Participant 25 at visit 1. The values in the cells show the ratio of responses to the total instances a particular phoneme was presented. The values on the diagonal indicate the ratio of correct responses for each corresponding phoneme. Vowels were presented six times and consonants eight times and are presented in separate confusion matrices. For example, when the consonant /z/ was presented, the participant replied /s/ in five (0.62) instances and /z/ in three (0.38) instances. The aggregate accuracy rates for vowels and consonants were 80% and 69%, respectively.*

Following data collection, a post-hoc analysis of targeted electrodes was conducted. This analysis, informed by electrodograms rather than the formant deviations, compared the targeted electrodes (using the targeted phoneme

confusions that were implemented in the take-home MAP) and the actual changes (compounded effect of basic fitting and explorative versions of advanced fitting that utilized both electrodograms and formant deviations). The analysis revealed inaccuracies in the phoneme enhancements due to electrodogram creation errors and interpretation issues. An illustrative example of the standardized assessment method (electrodogram) is shown in Figure 1 in the methods section.

Table 1A in Appendix A outlines the proposed C-level changes (increase or decrease) that should be made for the targeted phoneme pairs per participant based on the standardized electrogram analysis for default FATs. It only includes the 12 participants who adhered to the take-home MAP (to ensure that only interventions acceptable in daily life were included) and who received a targeted fitting. For instance, for Participant 20, HOT-HAT was a targeted phoneme pair incorporated in Table 1A, therefore, the recommendation was to increase E21 (+) and decrease E12 (-) and E13 (-)to enhance the contrast (illustrated in Figure 1). Combinations of targeted phoneme pairs could lead to conflicting fitting information (indicated by - +). This occurs when phoneme pair A points at an electrode to an increase in c-level (indicated by the + sign), while phoneme confusion pair B points to a decrease in C-level (indicated by the – sign) at that same electrode. Table 1B in Appendix A shows the difference between the actual adjustments across the full electrode array in the take-home MAPs, which are the combined results of the basic fit and advanced fitting.

Interestingly, participant 12 found the take-home MAP too sharp and uncomfortable upon return (after two weeks of using it), but preferred it after completing the training intervention at the end of the AuDiET study (visit 5), underscoring the need for an extensive adaptation period (at least six weeks) with new MAPs. The generic fitting adjustments in participants 06, 07, 08, 09, 11, 13, and 14 precluded the identification of specifically targeted phoneme pairs included in their take-home MAP and have been left out of the analysis shown in Tables 1A and 1B.

Effect of (Advanced) Fitting on PRQ Outcomes

The difference in PRQ outcomes, between visits 1 using the original MAP and visit 3 using the take-home MAP, for all participants is shown in Figure 5A. This figure displays the aggregate PRQ accuracy results per participant at visit 1, shown as baseline, and the changes measured at visit 3. The average baseline accuracy score of PRQ tests at the group level was 66% at visit 1, and 68% at visit 3. Here, accuracy refers to the percentage of correctly identified phonemes (also known as an error rate). A within-participant comparison of PRQ scores using a Paired t-test, showed

no significant change (t(22) = -1.2, p =.24). The mean effect size was 1.8 percentage points, with a 95% Confidence Interval ranging from -8 to 11 percentage points improvement, as detailed in Figure 7A). Figure 5A shows that changes were not related to baseline performance or affected by a ceiling effect.

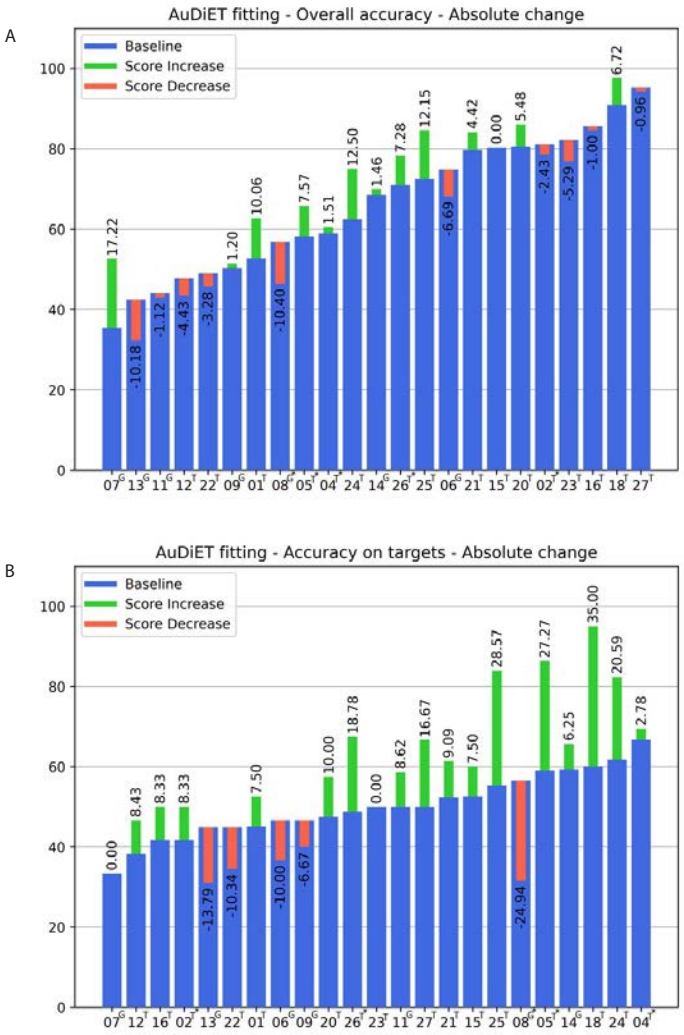

Figure 5. *Difference in PRQ scores between visit 1 and visit 3 for aggregate scores (A) and targeted phonemes sub-scores (B) sorted by baseline performance. Superscript (T) denotes those who received targeted interventions, and (G) denotes those who received generic interventions. Asterisks indicate participants who after testing at visit 3 reverted to their original MAP.*

Effect of (Advanced) Fitting on Targeted Phonemes

The sub-scores of targeted phonemes are shown in Figure 5B. The baseline scores represent performance on the targeted phoneme pairs per individual at visit 1. The average baseline score across the group was 50%, which is substantially lower than in Figure 5A, reflecting participants' difficulty with these phonemes. The average accuracy after the intervention was 57%. The mean effect size was 6.8 percentage points, with a 95% Confidence Interval ranging from -0.8 to 16 percentage points improvement, as detailed in Figure 7B. An examination of performance on the targeted phonemes between visits 1 and 3 revealed a trend that did not reach statistical significance at the 0.05 level after Bonferroni-Holm correction ($t(22) = -2.3$, $p = .03$, adjusted $p = .09$), which hints at potential efficacy in the fitting intervention for these targeted phonemes (illustrated in Figure 5B and 7B). Since the generic fitting adjustments did not specifically target phoneme pairs in participants 06, 07, 08, 09, 11, 13, and 14, the observed effect might be underestimated.

In the subgroup of participants who received a targeted intervention and who adhered to the take-home MAP after the conclusion of the study (labelled 'Adhered to Targeted Fitting, n=12), a larger effect of the fitting adjustment was observed, with average aggregate accuracy changing from 73% at baseline to 76% after the intervention (Figure 6A), which translates to a mean effect size of 3.0 percentage points with a 95% Confidence Interval ranging from -10 to 15 percentage points improvement, as detailed in Figure 7C. Analysis showed non-significant changes in aggregate PRQ scores ($t(11) = -1.64$, $p = .13$, adjusted $p = .25$). The average accuracy on targeted phonemes increased from 50% to 62% (Figure 6B). The mean effect size was 12 percentage points, with a 95% Confidence Interval ranging from 2.4 to 12 percentage points improvement, as detailed in Figure 7D). Paired t-test on the sub-scores of targeted phonemes showed a significant effect ($t(11) = -3.36$, $p = .006$, adjusted $p = .03$), which remained significant after Bonferroni-Holm correction for multiple testing. The standardized effect size expressed as Cohen's d was .88.

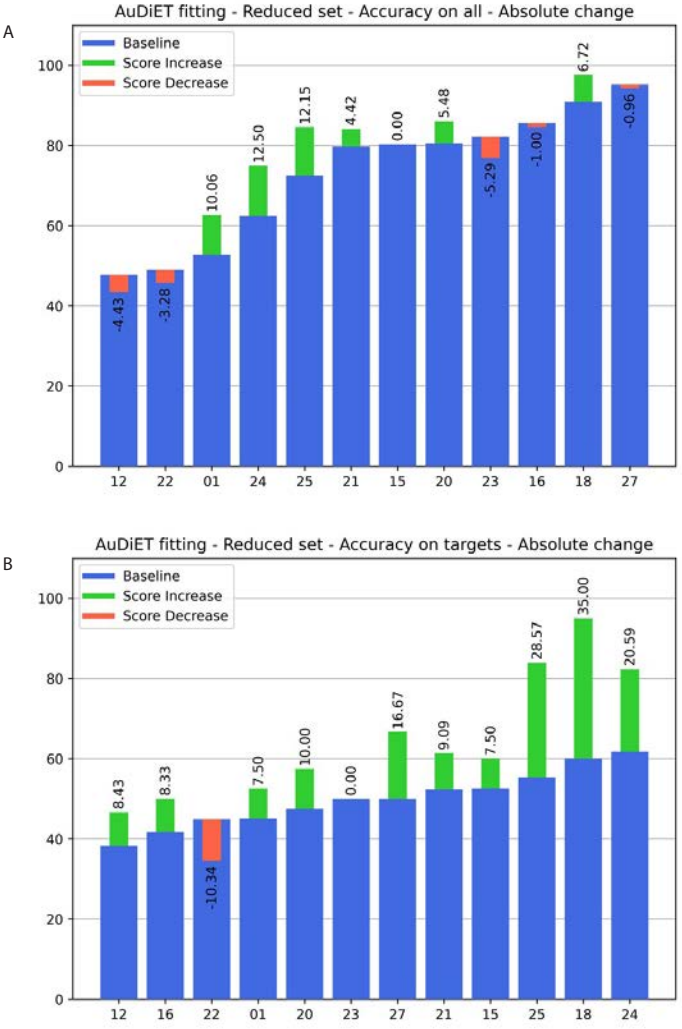

Figure 6. *Difference in PRQ scores between visit 1 and visit 3 of aggregate scores (A) and targeted phonemes sub-scores (B) in the subgroup that adhered to the take-home MAP and received a targeted intervention (participants labelled (T) in Figure 5).*

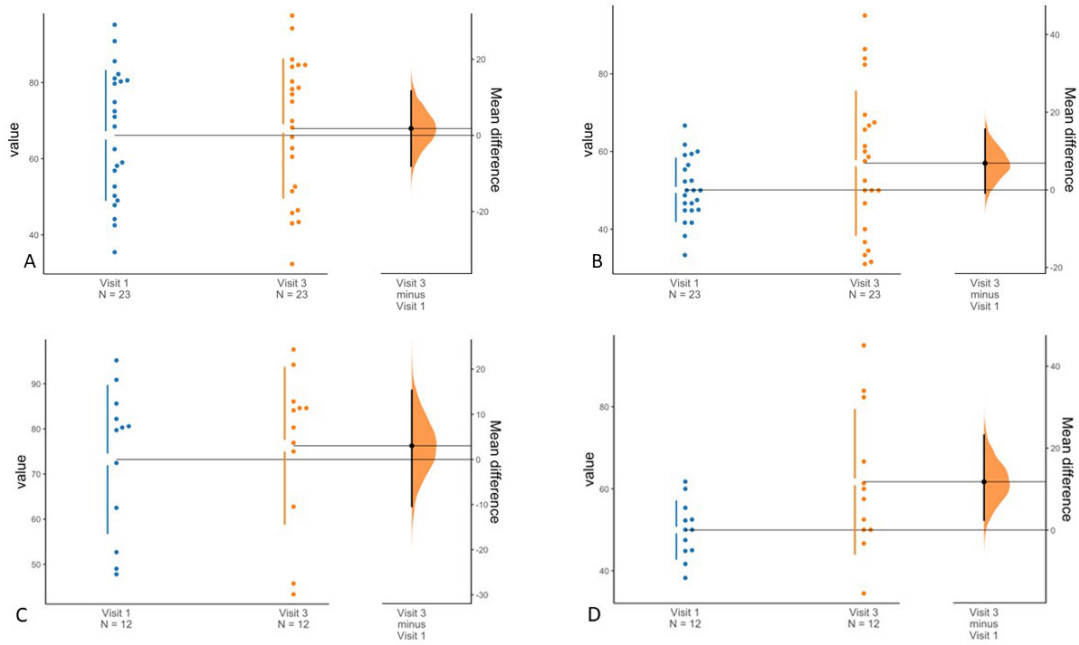

Figure 7. *Paired estimation plots for the full group and subgroup that adhered to the take-home MAP and received a targeted intervention, detailing aggregate scores and targeted phonemes sub-scores between visit 1 and visit 3. Panels: A (full group, aggregate), B (full group, targeted phonemes), C (subgroup, aggregate), D (subgroup, targeted phonemes).*

DISCUSSION

In this feasibility study, phoneme identification performance was measured via a self-testing paradigm (PRQ-test) to inform fitting adjustments in experienced adult CI users. We hypothesized that the largest systematic phoneme confusions can be reduced by adjusting the MAP. First, at visit 1, an initial assessment of performance was done. Then, at visit 2 a basic fitting check took place, directly followed by an advanced fitting procedure. These variants ranged from generic adjustments (e.g. Pulse Rate change) to adjustments targeted to individualized phoneme pair confusions.

Results showed in the full cohort a trend toward improved phoneme identification performance in targeted phonemes when using the new take-home MAP. Although there was variation between participants (Figure 7B), the results suggest that PRQ outcomes can be informative in fitting interventions, even in an experienced cohort of adult CI users. The average targeted phoneme identification performance increased (Cohen's d 0.88) in the subgroup that adhered to the targeted fitting. This

result suggests that targeted interventions could be more impactful than generic interventions, but a preregistered confirmatory study is required to draw more definitive conclusions (Wagenmakers et al., 2012). However, Table 1B shows that the effect is caused by a mix of basic fitting and advanced fitting. This means that there is potential to make the interventions more targeted.

Given that there is no effect on aggregate score, our approach may be more suitable for the "finishing touch," i.e. fine-tuning a few phoneme identification difficulties. In poorer performing CI users who struggle with identifying many phonemes, resulting in numerous systematic and random phoneme errors, those phoneme errors may lead to conflicting suggestions on how to change the MAP. The conflicting directions in Table 1A show limitations in a targeted approach when one phoneme confusion pair results in a suggested increase in C-level at an electrode, while another phoneme confusion pair results in a suggested decrease in C-level at that same electrode or vice versa. In other words, if those suggested interventions are implemented, improving one phoneme confusion error will deteriorate the other, and vice versa. Therefore, for poorer-performing CI users, the outcomes of other test paradigms may be more informative, including tone-decay, a test which has been revisited in recent years and is more sensitive to retrocochlear lesions (F. H. Schmidt et al., 2024; Wasmann et al., 2018).

Interestingly, 18 out of 23 participants preferred to continue using the new fitting (n=12 targeted, n=6 generic fitting) at the end of the study. It remains unclear whether participants based their preference on the effect of the intervention on speech recognition performance or other criteria. For instance, participants 8 and 13 preferred the new fitting despite their lower phoneme identification after the intervention. Browning et al. (2020) did not explicitly study if conflicting information from objective versus subjective evaluations accounts for variability in fitting procedures across centers, but this could be an implicit factor. Interestingly, Vaerenberg et al. (2014) stated that most clinics rely only on subjective feedback to guide the CI fitting procedure.

In this study, we primarily targeted vowel confusions. Out of the 37 targets included in the adhered targeted fittings, 24 were vowel pairs and 13 were consonant pairs. This approach was inspired by vocoder studies from DiNino et al. (2016) and Kasturi et al. (2002), which suggested that vowel confusions might arise from compromised frequency selectivity near the electrode-cochlea interface. In comparison, consonant confusions may instead depend more strongly on the temporal envelope structure. In our fitting approach, we had more options to manipulate spectral

cues (e.g. changing C-levels or the Frequency Allocation Table (FAT)) than options to manipulate temporal cues (e.g. pulse rate or strategy). Therefore, we expected to have a stronger impact on vowel than consonant identification. The distinction could be subtle, as exceptions exist. For example, the contrast between /s/ and /z/ relies on spectral cues, as is generally the case for voiced/unvoiced consonants. Previously, Migliorini et al. (2024) showed that within participants the vowel and consonant performance scores were similar. The effect of our fitting interventions was too small, and the number of targeted vowels versus consonants was too unbalanced to differentiate between effects on vowels or consonants.

Limitations

Even though the outcomes of this feasibility study show that in experienced adult CI users the fine-tuning of the MAP matters, the exploratory nature also requires cautious interpretation (Rubin, 2017). The study has several limitations:

1) The small selection of targeted phoneme pairs per participant provides little room for improvement on the aggregate PRQ score and may be subject to a higher degree of test-retest variability. In addition, there is a risk of priming due to the selection of targeted phoneme pairs since participants may have realized which phonemes were targeted, however, the random presentation order of phonemes in the PRQ test likely minimized this effect.

2) Improving CI outcomes is inherently difficult. Multiple fitting approaches may lead to similar outcomes (Waltzman & Kelsall, 2020), making it difficult to work incrementally towards an 'optimal' fitting, as radically different approaches may lead to similar speech performance outcomes, as currently measured within standard clinical care. There is no gold standard to compare against, and it is uncertain whether there is room for improvement (Wathour et al., 2021). In addition, the inclusion criteria had to be widened to include relatively better performers with less room for improvement since the experienced poorer performers were difficult to recruit and often did not meet our inclusion criteria.

3) There was no strictly defined fitting protocol. The fitting adjustments were made during a single visit without time for acclimatization per fitting variant. The fact that multiple fitting interventions were layered (basic and multiple advanced) led to the inability to disentangle the effects of basic fitting from more targeted advanced fitting interventions. So, while we observe some potential improvements, attributing these directly to the different parts of each fitting intervention is

impossible. A pre-registered confirmatory study with strictly defined actions based on objective outcomes is needed to confirm our results (Wagenmakers et al., 2012).

4) A limitation arises from the phoneme confusions being based on a single speaker's voice, as formant frequencies can vary significantly across different speakers, so optimizing for one speaker may not generalize to better phoneme identification in general. DiNino et al. (2016) showed that vocoders will process some vowels (e.g. HID, HUD, HOOD) differently for male and female voices. Also, the method suffers from regional differences since dialects can lead to different perceptions of vowels (Wright & Souza, 2012). This means that in the limiting case, such an approach may optimize the MAP for a certain speaker and a limited number of utterances without generalization of the benefit. In principle, this could be tackled by introducing tests with a large number of speakers, different prosodic contexts etc. However, this approach would significantly increase (already long) testing time and introduce more conflicting suggestions. In this study, we deliberately choose a single speaker to minimize complexity and first test the feasibility of the method.

5) A more general limitation of using phoneme confusion is evident from Table 1A, showing that changing targeted phoneme contrasts may inadvertently reduce contrast or performance on other phonemes, which may explain the lack of improvement in aggregate scores. Another approach to increase phoneme contrasts without a negative effect on other phonemes could be a new CI strategy that compensates for the spread of excitation leading to more focused electrical stimulation patterns known as SPACE (Bolner et al., 2020).

Recommendations

The goal of this feasibility study was to explore the potential impact of fitting interventions based on individual phoneme identification errors on the performance of experienced adult CI users. Our data seems to suggest that such an approach could be beneficial for experienced CI users. However, given the exploratory nature of our findings, replication in a pre-registered larger clinical study using a strict fitting protocol is necessary to provide stronger evidence. To improve the advanced fitting methodology, we suggest implementing targeted interventions sequentially over multiple visits to facilitate acclimatization. It is recommended to carefully consider the overall changes across the electrode array, as fitting adjustments may involve conflicting targets that require distinct approaches.

In the future, experienced CI users who are motivated to refine their current MAP might use a remotely administered version of the PRQ. An automated, potentially

Machine Learning based approach, that targets the electrodes associated with their most frequent systematic confusions, could then deliver micro-interventions. This approach may entail relatively low risk, similar to the self-fitting that CI users can already perform (Vroegop et al., 2017). The effectiveness of these micro-interventions could be remotely assessed at convenient times for users, following a period of acclimatization and at the time of day they prefer (Wasmann et al., 2024). This iterative cycle of adjustment and evaluation could progressively implement beneficial adjustments. Such an approach would empower CI users to actively participate in their post-implantation care, collecting valuable data on the impact of micro (minor) adjustments on speech recognition, and enabling the use of ecological momentary assessment (EMA) in the day-to-day context of the CI user (Holube et al., 2020). As has been demonstrated in the self-fitting of hearing aids, it is important to empower end users, making them confident in their own capabilities and assuring them of the many tasks they can handle by themselves (Convery et al., 2019). In addition, data-driven (Machine Learning based) fitting approaches may alleviate the clinical workload for audiologists and could be part of a more streamlined clinical protocol where less time is devoted by the clinician to optimize MAPs. If these approaches are also feasible in newly implanted CI users, these new users might start with a first fit based on a population mean and subsequently receive fine-tuning with above-mentioned approaches.

Overall Conclusion

This feasibility study demonstrated the viability of using precise evaluations of phoneme identification performance in fitting procedures. Experienced CI users may benefit from targeted interventions based on specific phoneme identification errors that remain invisible when using only aggregate phoneme identification performance. The exploratory fitting approach resulted in modest improvements in sub-scores of targeted phonemes for the subgroup that received the most targeted interventions. The clinical relevance of these findings in experienced CI users may be modest because participants were potentially already close to an "optimal MAP" in terms of aggregate phoneme and speech perception through their clinical pathway. Now that feasibility has been demonstrated, a more systematic approach, based on the lessons learned from this study, may lead to a stronger effect.

8

General Discussion & Synthesis

This thesis explored how AI and teleaudiology can unlock benefits to improve affordable and accessible hearing health care. When writing our computational audiology perspective paper in 2019, we scarcely anticipated that large language models would soon enable nearly natural chat interactions between humans and machines, approaching the capability to pass the Turing test. Based on these recent advances, one can envision that in the future, individuals conduct hearing tests at home while consulting chatbots for support during the test and afterward for interpreting the results and their implications. In this general discussion, we will reflect on the findings of the studies conducted in this thesis as steps toward achieving such a vision, and we will address some of the remaining challenges we still have to overcome to achieve optimal implementation of AI in audiology.

THE NEED FOR AND POTENTIAL OF COMPUTATIONAL AUDIOLOGY

With the increasing pressure on healthcare in general and audiology in particular, the necessity for the further development and implementation of computational audiology will grow within the field of hearing care in the coming years (Chapter 2). These advancements also give rise to new challenges across various domains, including privacy, data management, and responsibility. To address these challenges, computational audiology will require increased collaboration across disciplines. The intersection of AI and audiology brings together audiologists, engineers, and data scientists to create patient-centered solutions and empower patients to use their own data. Chapter 2 presents a case for improving access, precision, and efficiency of hearing healthcare services through computational audiology. Ethical implications of using AI, such as liability when algorithms make errors and the consequences of biases in large datasets, are examined. We learned that there is still a long way to go. An active role from professional societies is essential to mitigate risks and address challenges for a safe transition to greater adoption of AI within audiology. One advisable approach would involve establishing interoperable systems and implementing shared data policies (FAIR) in computational audiology. This ensures that patients receive support based on high-quality data wherever they are located.

Based on ideas presented in our perspective paper about computational audiology, several initiatives have emerged internationally to promote the adoption and development of computational audiology. For example, the Computational Audiology Network (CAN) and its Virtual Conference on Computational Audiology (VCCA)

conferences bridged the gap between disciplines by encouraging interdisciplinary discussion and knowledge sharing. In the autumn of 2023, CAN was formalized as a professional network. In February 2024, CAN became a member of the World Hearing Forum (Wasmann, 2024). The high interest in the VCCA meetings indicates that the field of computational audiology is being embraced. Today, more than 1,500 people have attended one of the VCCA conferences. In addition, the publication of numerous papers in a dedicated Frontiers Special Issue (Meng et al., 2022) demonstrates the dynamic discourse and progress within this interdisciplinary domain.

AI CHATBOTS IN HEARING HEALTHCARE

Early 2023, we explored the potential of large language model-based AI chatbots for people with hearing loss or tinnitus, outlining opportunities to improve accessibility by counseling people with hearing loss and assisting clinicians (see Chapter 3). After experimenting with prompts to simulate patient conversations and creating a simple AI chatbot using an OpenAI model (Wasmann, 2023), we suggested research priorities to develop AI chatbots responsibly, such as addressing risks related to missing or incorrect information provided by these chatbots (Swanepoel et al., 2023). Within a year, the paper by Swanepoel et al. was followed by new studies showcasing the use of AI chatbots in tinnitus management (Jedrzejczak et al., 2023), counseling before and after CI surgery (Aliyeva et al., 2024), evaluating AI chatbots on professional audiology exams (S. Wang et al., 2023) and their use in audiology training programs (Sooful et al., 2023). There was a quickly increasing amount of literature discussing the opportunities and risks associated with using AI chatbots. For instance, the response variability among AI chatbots was reported depending on the large language model utilized (Jedrzejczak & Kochanek, 2023). Strategies were investigated to improve AI chatbots' responses by forcing the model to primarily use Wikipedia as a resource (Matos et al., 2024), to name a few examples recently appearing as preprints or conference papers.

To ensure the successful implementation of new technology, it is crucial to engage all stakeholders in its development and application (see also Chapter 2). This certainly holds true in the development and deployment of AI chatbots. Now that several applications have been developed, as mentioned above, innovative audiologists will be important for exploring the first steps in integrating AI chatbots into audiological care. Low-risk scenarios for using AI chatbots in hearing health care could include an AI chatbot responding to difficulties patients experience with using their hearing aid or cochlear implant, as illustrated in the examples below.

Low-risk examples of using AI chatbots with patients

Technical Support for Connectivity Issues:

- **Example Question:** "How can I connect my hearing aid via Bluetooth to my phone?"
- **AI Chatbot Response:** The chatbot can guide the patient step-by-step through the process of pairing their hearing aid with their smartphone, including checking if Bluetooth is on, ensuring the hearing aid is in pairing mode, and selecting the device on their phone. This can greatly reduce frustration and improve the patient's experience.

Maintenance Guidance:

- **Example Question:** "How do I replace the microphone filters on my hearing aid?"
- **AI Chatbot Response:** The chatbot can provide simple, clear instructions or videos on how to replace microphone filters, ensuring the patient can maintain their device properly without needing a clinician visit. It can also remind users of necessary maintenance tasks based on time intervals or usage patterns.

Troubleshooting Common Problems:

- **Example Question:** "My hearing aid is not working properly. What should I do?"
- **AI Chatbot Response:** The chatbot can run through a checklist of common issues and solutions, such as checking if the device is on, if the battery needs to be replaced or recharged, or if the device needs to be reset. This helps patients troubleshoot quickly and effectively.

Usage Optimization Tips:

- **Example Question:** "What settings should I use in a noisy environment?"
- **AI Chatbot Response:** The chatbot can suggest optimal settings or adjustments to the hearing aid for different environments, enhancing the user experience without needing a clinician's immediate intervention.

Quickly solving these problems at home can significantly impact successful daily use. Solving connectivity problems is a tedious task that most clinicians will probably be happy to hand over to a chatbot with endless patience. However, innovative clinicians will need clear guidance from other stakeholders to align their efforts with patient-centered care, data privacy, and ethical standards.

This guidance should address the specific challenges of LLMs, such as adapting workflows, ensuring data security, and maintaining meaningful patient interactions. Equipping these pioneers with knowledge and best practices, such as case studies, implementation toolkits, and peer networks, can create the basis for wider AI chatbot adoption and further development. We may also learn from interactions with AI chatbots in the process (see Figure 1). Higher risk applications including interacting with people with acute ear infections or those at risk of suicide due to various factors including tinnitus, will require more involvement from clinicians and appropriate level of oversight. Therefore, collaboration among healthcare professionals, researchers, patient associations, and policymakers is essential. Practical actions include forming interdisciplinary committees and developing shared ethical guidelines to ensure the ethical design and use of AI chatbots.

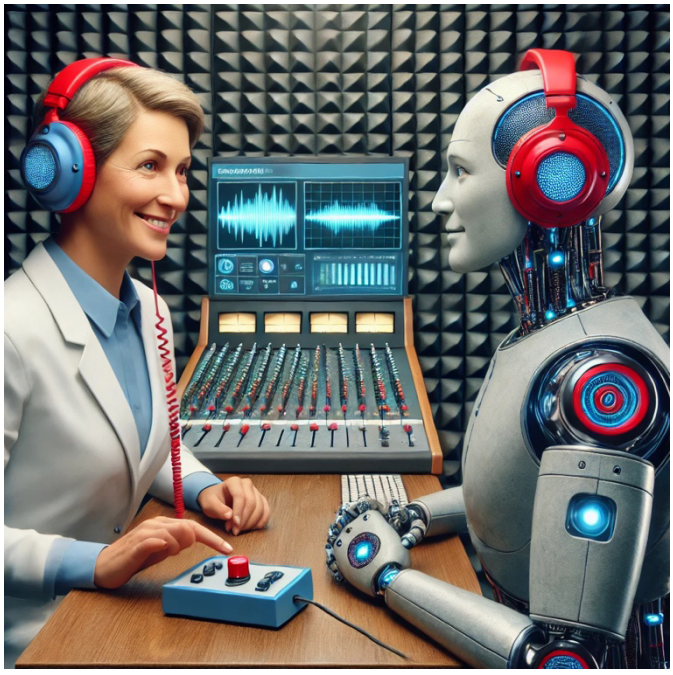

Figure 1. *DALL·E created artwork. Prompt "A surreal scene of a friendly audiologist and a humanoid robot facing each other and testing each other's hearing in a sound booth. The audiologist, a middle-aged person with a warm smile, is seated at a desk, wearing red headphones on the right ear and blue headphones on the left ear, holding a response button in hand. Directly across, a humanoid robot with sleek metallic features and a digital display face is facing the audiologist, also wearing matching headphones and holding a response button, listening intently. The sound booth is detailed with acoustic panels, various audiology equipment, and a computer screen showing sound waveforms, creating a thought-provoking atmosphere of advanced technology and human collaboration."*

TOWARD INTERACTIONS NOT LIMITED BY HEARING STATUS

Speech transcription is a valuable tool for individuals with hearing loss. It can be applied in specific contexts, such as watching television, participating in online meetings, or during appointments with clinicians. ASR technology is quickly adopted around the globe, and the number of use cases is growing (Loizides et al., 2020). For instance, Google's Live Transcribe is a widely used application that instantly converts speech and ambient sounds into text on smartphones, aiding people with hearing difficulties (Slaney et al., 2020). The app supports over 80 languages, identifies various sound types, and includes a sound level indicator for users' spatial awareness. The availability of ASR in meeting platforms like MS Teams and innovations, including Speaksee, which uses multiple microphones to facilitate the participation of people with hearing loss in groups (Hazelebach, 2023), illustrates the technology's advancement. Another use case within audiology is that ASR can potentially automate the scoring of speech recognition performance in standard clinical audiological tests (Venail et al., 2016).

In the assessment of ASR apps, described in Chapter 4, we found that speech recognition performance, as measured by standard audiological tests, was on par with the speech recognition performance of people with moderate hearing loss in low noise conditions. However, in test conditions with competing noise, the ASR apps performance declined compared to the low noise condition. This difficulty in transcribing speech in noise emphasizes the need to position the microphone close to the talker in challenging listening situations, as one would expect from decades of experience with remote microphone systems (Fitzpatrick et al., 2009; Jerger et al., 1996). Furthermore, how and where to display the transcript to the users is a crucial factor for the practical usage of ASR. Several ways to present the text have been researched and tested (Chavez et al., 2024). One example is the use of augmented reality applications in glasses (Jin et al., 2023). This may seem like a good solution, but projecting the text without confusing the user or hindering social interaction is difficult (Rzayev et al., 2020). Another aspect we encountered while evaluating ASR apps is the need for a fast and stable internet connection to avoid missing parts of the conversation or losing lip-sync. For a new evaluation of ASR apps in 2023 (published in a Dutch report), we repeated the transcription of the English dialogue using Live Transcribe. With a more stable internet connection, we measured a WER below 5% instead of the previously estimated 34% (see Figure 5 in Chapter 4) in 2020 using a less stable internet connection (Hazelebach et al., 2023; Pragt et al., 2022). Thus, in an audiology center's consultation room, acoustics and fast, reliable internet bandwidth must be considered to facilitate conversations between clinician and patient.

During my PhD trajectory, I had the opportunity to connect with developers from the National Acoustic Laboratories. They developed a speech-to-text solution using the Apple speech-to-text engine specifically for audiology centers. Adding Dutch as a language turned out to be easy. Within a year, the Dutch version was downloaded more than 500 times in the Apple App Store, presumably for use at hearing aid dispenser shops, audiology centers, and at home. Notably, with the right contacts, specific applications can progress unexpectedly rapidly.

AUTOMATED TELEDIAGNOSTICS ANYWHERE AND ANYTIME

The scoping review of self-administered pure-tone audiometry approaches presented in Chapter 5 is an update and extension of a systematic review by Mahomed et al. (2013). In this review, we identified six approaches for air conduction audiometry that met clinical diagnostic standards, meaning the test outcome was comparable to clinic-based assessments performed by clinicians. Of the 64 reports included in this review, those that used similar approaches were clustered, and the resulting 27 unique approaches were assessed based on clinically relevant criteria. New features found in the literature to increase reliability included monitoring background sound levels and automatically flagging invalid responses. We estimated that the test duration of thirteen approaches was similar to manual assessment; in one approach, the testing time was a potential burden, while ten approaches did not report testing time. In addition, a maximum likelihood-based approach and two Bayesian active learning approaches reported shorter test durations due to more efficient threshold-searching algorithms. Based on our scoping review findings, we identified an opportunity to make these approaches more child-friendly by adding game-design elements and more clinically acceptable by incorporating bone conduction transducers.

Despite the potential of home testing and home care as future solutions using self-administered testing paradigms, the reliability and practicality of home testing remain uncertain. The primary reason is that it is harder to control the home environment and to know under what conditions the test was performed, which may affect its validity. Previous studies of at-home assessments in CI recipients showed that it was feasible to perform but did not consider time-of-day effects and started evaluations in a controlled environment, that is, within the clinic (Cullington et al., 2018; de Graaff et al., 2018; Maruthurkkara et al., 2022). The next step was to look at the possible effect of time-of-day and chronotype on assessments carried out at home in CI users

without prior experience with the test battery. In Chapter 6, we assessed auditory performance at various times of day, including beyond regular office hours. Time of day, fatigue, motivation, or chronotype did not affect the aided thresholds test and outcomes in the digits-in-noise test. Differences of ≥ 4 dB in the digits-in-noise test outcome were rare, which seems to agree with the clinically significant difference of 3.1 dB determined by Maruthurkkara et al. (2022). Although we demonstrated the reliability of at-home cochlear implant performance monitoring, not all participants were able to complete the digits-in-noise test. The speech recognition in quiet score was a good predictor of whether a participant could perform the digits-in-noise test (see Chapter 6, Figure 2). To increase the success rate of the test battery, we recommended including the digits-in-noise test only when a participant's speech recognition in quiet was at least 65% at 65 dB SPL, which is a bit more stringent than the ≥ 40% score recommended by Kaandorp et al. (2015).

It is reasonable to conclude that remote testing can be integrated into a virtual clinic model since at-home assessments are reliable and feasible for many CI users, in line with findings from previous research (Cullington et al., 2018; de Graaff et al., 2018; Maruthurkkara et al., 2022). This may make the CI aftercare for clinics more efficient since routine check-ups in the clinic can be replaced by specific visits based on problems identified at home. However, during the remote assessments, it was important to address questions and issues promptly via easily accessible communication channels, given that our participants had no prior experience with the test battery and that unexpected problems could arise at home. Interacting with our participants via a messenger system helped keep everybody engaged and facilitated quick problem-solving (e.g., in 37% of remote assessments, participants experienced technical difficulties; see Table 3, Chapter 6). However, in practice, synchronous teleaudiology will not provide time-saving benefits to professionals and limits the freedom for CI users to perform the tests at any time. Having AI chatbots to encourage and help CI users complete remote assessments at any time, such as the aided thresholds and digits-in-noise test, as well as administering questionnaires, is the next step one can envision. Such AI chatbots could subsequently summarize test outcomes both for patients and clinicians. AI chatbots would thereby help clinicians reduce their workload by making summaries and highlighting abnormal outcomes. Interacting with AI chatbots may make remote assessment more "enjoyable" while allowing clinicians to dedicate more time to complex cases and less to routine tasks. We should train AI chatbots on audiological data to make their responses and interpretation of test outcomes more reliable.

TOWARD AUTOMATED CI FITTING

Within the field of cochlear implants, computational audiology can also be applied in meaningful ways. Adjusting the parameters of the CI system requires knowledge and expertise. However, the precise justification for adjusting the CI system is still lacking (Wathour et al., 2021). When we gain more insight into the effects of fitting adjustments that can be made to the CI system based on hearing performance errors, the process of fitting the processor will ultimately become more data-driven (or outcome-driven), which in turn provides opportunities to do more efficiently using automation. It was shown in Chapter 7 that phoneme identification performance may provide useful information to the audiologist in refitting experienced adult CI users. Our exploratory fitting method appeared promising regarding effect size on targeted phonemes by enhancing the contrast between those targeted phonemes. However, there was no effect on overall speech performance. One limiting factor was that the subset of targeted phonemes (2 to 6) per participant had only a limited weight in the aggregate score of all phonemes (32 phonemes, 17 consonants, and 15 vowels). Another reason could be that the increased contrast between targeted phonemes by adjusting C-levels may reduce the contrast between non-targeted phonemes, thereby introducing new errors (see Table 1A, Appendix A). New CI strategies to increase phoneme contrasts, such as SPACE instead of ACE (Bolner et al., 2020), might be a better approach with the additional benefit that it does not affect loudness balance across electrodes. Follow-up confirmatory studies are required to provide more substantial evidence of our fitting approach and discern the effects of clinical fitting adjustments from those of more targeted fitting adjustments. In future work, collecting performance data in a teleaudiology setting could further facilitate the development of more outcome-driven fitting approaches. Remotely administered tests may motivate CI users to improve their current MAP, resulting in a more extensive dataset than can be practically collected in laboratory studies. A safe start with AI-powered automated CI programming tools would be low-risk micro-interventions similar to existing self-adjustment tools. Although still speculative, incremental improvements might be achievable by encouraging users to assess their performance with tests after acclimatization to a CI fitting. This process of sequential interventions may gather valuable insights into the effects of micro-interventions on speech recognition. Additionally, it may facilitate evaluations and interventions in real environments that CI users encounter in daily life instead of the less representative setting of the CI clinic. This could be further supported by tools such as ecological momentary assessments (EMAs) to evaluate how changes are experienced in daily usage

situations (Holube et al., 2020). Most importantly, teleaudiology tools enable CI users to play a more active role in their post-implantation care management.

TOWARD INTEGRATED TELEAUDIOLOGY SERVICES

Teleaudiology experienced a peak during COVID-19 due to lockdowns and social distancing, but in-person care has reverted to pre-COVID-19 levels (Chong-White et al., 2023). Despite technological readiness, societal and user acceptance appear to be barriers. In March 2024, the topic of teleaudiology was discussed at a meeting organized by the Dutch Association of Cochlear Implant Users (OPCI in Dutch). It was an opportunity to meet a large group of CI users who receive care from CI Centers across the Netherlands, present them with teleaudiology tools, and administer a survey about their perception of the opportunities and barriers associated with teleaudiology. More than 60 Dutch CI users filled out the survey. The data have not yet been published, but here are several takeaways. The results show that a large group of CI users (around 90% of respondents) is open to teleaudiology; however:

- Many are afraid that their CI system may break down when there is a glitch when receiving teleaudiology services;
- Many worry whether communication with the clinician during a teleaudiology session will be adequate;
- Many worry whether the quality will be on par with in-person care;
- Some do not want teleaudiology (around 10% of respondents) or need help to use it. For instance, they may be unable to perform all tasks at home alone, such as replacing a defective part due to dexterity problems.

Also, surveys among clinicians have identified concerns about teleaudiology provision, including how it may negatively affect the relationship between professional and client (Parmar et al., 2022) and the reliability of remote assessments (Bennett et al., 2023). A pre-pandemic systematic review singled out logistical barriers such as the infrastructure and the billing system (Ravi et al., 2018).

Addressing these worries is vital so that enough end-users feel at ease trying these technologies and so that we may reach the critical mass necessary to make teleaudiology viable as routine clinical care for reasons beyond dealing with lockdowns.

LIMITATIONS OF THIS THESIS

Currently, even the most advanced AI chatbots using LLMs are at risk of hallucinating. It is not clear where they retrieve information from, their reliability is uncertain, and there is no regulatory oversight yet for medical use (Gilbert et al., 2023), which means that we need to be cautious when providing unsupervised chatbots to patients. There have been calls for open-source LLM that can be used for public services (Editorial, 2023). However, even then, with open-source LLMs, regulating these models is complex given the near-infinite range of inputs and outputs these models can generate and the lack of established methods to mitigate all risks (Gilbert et al., 2023). Therefore, we may need to wait until these issues are resolved, and some have urged not to use Chatbots in clinical practice (Au Yeung et al., 2023; Ethics Review Committee (ERC), 2024). Nevertheless, right now, patients are already using search engines to retrieve medical information, and these are increasing using LLMs. Ideally, we should develop best practices as a field to help pioneers move forward with acceptable risks.

Another limitation is that developments in AI are progressing very rapidly. Publications from four years ago (e.g., Chapters 2 and 4) are already outdated. Also, the ASR study was a preliminary study using suboptimal evaluation methods. We primarily used standard audiological tests typically used to assess auditory performance of people with hearing loss. These tests are not explicitly designed to assess speech transcription performance. As a result, neither the potential effect on speechreading nor the readability of the transcript was evaluated. Therefore, we suggested evaluation metrics that reflect a more realistic usage situation than only the word error rate in an ideal listening situation. Better evaluation methods are needed, given the growing adoption of ASR (Szymański et al., 2020). We can assume that the accuracy of currently available speech transcription apps has further improved since developments are progressing quickly. However, even in the Word Error Rate (WER), there is room for improvement, especially regarding speaker accents, background noise, and spontaneous conversations (Ferraro et al., 2023).

In the Remote Check study (Chapter 6), we were restricted to working with the clinically available application, implying that we could not make changes to improve it, even as a scientific tool. Only self-reported measures were used to assess workload or fatigue. Participants may have misjudged their task-related fatigue. In addition, the influence of task-related fatigue may have been overcome by increased attention during the relatively short task, so perhaps it was not measurable in the digits-in-noise anyway. Therefore, although the Remote Check

app could not assess the effect of fatigue or listening effort, the advantage is that this makes it more robust for assessments at any time. Potential improvements to better probe listening effort could include adding more demanding tests, such as the Digit Span Test (S. Wang & Wong, 2024), sentences in noise, or dual-tasks.

Another restriction in our Remote Check study was that, at the time, it could only be carried out with Apple devices. Not being able to recruit CI users with Android smartphones may have led to a bias since people with lower incomes are more likely to own an Android device (Jamalova & Milán, 2019). Furthermore, our sample likely included people more inclined to use teleaudiology, although the age range of the studied group (median age 67) was similar to the adult population of our clinic, and as the survey mentioned above among Dutch CI users revealed, the large majority is willing to use teleaudiology. Digital innovations promise enhanced accessibility. However, the cost and actual availability of these technologies for all patient demographics, especially in under-resourced areas, remain significant barriers. Teleaudiology may reduce the cost of lifelong aftercare for CI users but will not change the cost of surgery, implants, and CI processors (D'Onofrio & Zeng, 2022). Fortunately, all CI manufacturers are investigating remote monitoring and remote fitting options on Android and iOS devices, which means there is increasing potential for teleaudiology to become integrated into clinical pathways without cutting off large groups of people.

The most important limitation of the AuDiET study presented in Chapter 7 was its exploratory nature and broad scope. Within a single project we included new diagnostic tests, new fitting procedures and more individualized training approaches. This lead to interdependencies that made the execution of the project quiet complex. Regarding the fitting part of the project, we investigated many CI fitting adjustments and tested multiple hypotheses, resulting in less-than-desirable statistical rigor. The compounded effect of all implemented fitting interventions was evaluated per CI user, which made it impossible to disentangle the impact of each intervention. Another limitation was that we did not give CI users time for acclimatization to the distinct fitting interventions. A more extensive, standardized preregistered replication focusing only on fitting interventions is needed to confirm and expand our findings.

RECOMMENDATIONS

Teleaudiology holds promise and will see increased adoption in the future. However, the specific implementation in clinical practice needs further refinement. During this implementation phase, there are several recommendations to consider. Clinicians, for example, need to design service delivery models that align with the needs and experiences of individual patients and, thus, meet the human scale. This means that the interests of the end-user are given a central position. For instance, how does the service impact the end-user personally? Can we design simple solutions? Can we adapt the technology to the user instead of vice versa? This calls for a design approach that prioritizes end-user involvement, a strategy successfully adopted in Australia, leading to innovations such as NALscribe and C2Hear (Convery et al., 2020; Ferguson et al., 2018; Young et al., 2022). NALscribe is an ASR system whose interface was developed based on input from potential users and clinicians and was subsequently tested in audiology clinics (Shang, 2023). Additionally, counseling prospective users about these technologies is very important. For instance, people reported fear that teleaudiology might corrupt their CI system. Fortunately, this is not a valid concern, but developers and clinicians should be aware of and address such concerns in their interaction with potential users. These examples underscore the value of placing the end-user behind the steering wheel, empowering them to participate actively in their hearing healthcare pathway.

Another recommendation is that teleaudiology tools should be provided in settings that promote comfort and trust. Our commitment as clinicians to creating a welcoming physical environment in our audiology clinics - complete with clear signage, ample parking, reception desks, and amenities like coffee - should be mirrored in the digital space (see Figure 2 for an impression of the current situation). This ensures patients feel supported online and at ease when accessing teleaudiology services. When patients fail to log in or encounter difficulties later in the process, we need the equivalent guidance available in person to help them get started and feel welcome in a virtual hearing clinic. The high number of technical difficulties encountered during Remote Checks (Chapter 6) shows that there is room to improve such apps and that support from the clinic is needed to make this a success. It takes effort and resources to reach an adequate level of comfort and trust, for which the appropriate reimbursement programs are required in order to give the right incentives to clinicians and end-users.

Figure 2. *DALL·E created artwork. Left-hand side: Prompt "A four-panel cartoon depicting a patient's visit to an audiology clinic. 1. First panel: A patient arrives at an audiology clinic with a friendly exterior. The clinic has ample parking space with clear signage that reads 'Audiology Clinic.' A clinician is outside, smiling and waving to the patient. 2. Second panel: The patient walks into the clinic, greeted by a welcoming reception area. A receptionist behind the desk smiles warmly, and there are clear signs pointing to different areas of the clinic. 3. Third panel: The patient sits in a comfortable waiting area, enjoying a cup of coffee from a small coffee station. There are informational posters about hearing health on the walls, and a sense of comfort and hospitality is evident. 4. Fourth panel: A friendly clinician guides the patient down a well-lit hallway towards the examination rooms. The environment looks professional yet cozy, with artwork and plants enhancing the welcoming atmosphere.". Right-hand side: A cartoon image of a person sitting at home, wearing wireless earbuds, trying to log in for an online test on their smartphone. The person appears frustrated and confused, with a furrowed brow and a slightly open mouth, expressing annoyance. They are holding a smartphone that clearly displays an error message saying 'Login Failed - Please Try Again.' The background features a cozy living room with a comfortable sofa, a small coffee table, and a plant, highlighting a typical home setting. The wireless earbuds are visible in the person's ears, not connected with any wires to the smartphone."*

Teleaudiology should also be accompanied by patient education and professional training. As digital tools become increasingly integral to hearing healthcare, the importance of patient education cannot be overstated. Patients must be well-informed about the capabilities and advantages of these technologies to increase adoption and effective use. Positive reinforcement can lead to better outcomes and might be feasible using teleaudiology. Research shows that patients who have more knowledge and skills relating to their hearing aids obtain better outcomes (Bennett et al., 2018). The same likely holds for AI-driven tools. However, patients will need to be equipped with tools for critical thinking so that they can interact relatively safely with AI chatbots. Clinicians will need to find out what their patients can do by themselves, what their patients cannot do, and how they can best support their patients. One problem is that clinicians must be aware of what happens outside

their clinics in situations beyond their direct observation. Manchaiah et al. (2023) surveyed hearing healthcare professionals about their views on over-the-counter hearing aids. A significant majority of professionals (73%) expressed concern that users would struggle to insert their hearing aids, while Coco et al. (in preparation) recently showed that the majority of first-time hearing aid users are capable of inserting their devices properly without any external help. All professionals in the field will continuously need to update their beliefs about what patients and clinicians can do and ensure that they inform each other.

As explained in Chapter 2, AI expert systems and teleaudiology require standardized protocols in their application to prevent overwhelming end-users and developers with a plethora of user interfaces. In short, standardization helps to achieve interoperability between products and can facilitate the exchange of audiological data among clinics (Vercammen & Buhl, 2024). Such protocols and best practices are also crucial for ensuring the quality and consistency of care. The guidelines for teleaudiology practices proposed by Bennett et al. (2024) address several of these aspects and provide a start. Standardized data formats and tools for data collection may lead to the development of an "AudioHealthNet" database. Such a database could facilitate data exchange and collaboration among clinicians, researchers, and professionals from the industry. By adopting standardized data formats while sharing, for instance, synthetic data versions (or updated priors using federated machine learning) that safeguard privacy, we can create larger datasets for training algorithms, which may propel the field of audiology forward, much as the ImageNet Challenge did for computer vision (Russakovsky et al., 2015; Saak et al., 2022; K. Taylor & Sheikh, 2022).

Integrating AI and remote technologies in audiology offers tremendous potential. However, audiology is built upon and aims for human interaction. Digital communication cannot capture the full spectrum of human interactions, and thus, relying solely on AI-assisted teleaudiology is fundamentally limited. As Einstein famously stated, "Not everything that counts can be counted." Just as not all aspects of music can be fully conveyed through sheet music alone, not all subtleties of patient care and empathy can be captured in algorithms.

The genie is out of the bottle. Given all the potential benefits, we should continue using AI in audiology to increase inclusivity instead of deepening our divides. Working together, we can strive to adapt its use to better meet the needs of both patients and clinicians.

Summaries

SUMMARY

Toward AI-Assisted Teleaudiology

This thesis explored teleaudiology and artificial intelligence (AI)-based methods that can support people with hearing loss and, in the future, could be used to address the growing need for hearing healthcare. Worldwide, 1.5 billion people experience some degree of hearing loss, and according to the World Health Organization, a large portion of them lack access to hearing healthcare or hearing assistive technologies. Due to the global aging population, the capacity challenge in hearing healthcare will further increase. Therefore, innovations are needed.

This thesis focuses on studying new technologies designed to overcome barriers in hearing care. These barriers include limited access to diagnostic tools, the high cost of hearing devices, and challenges in fitting and evaluating those devices. The primary focus of this thesis is on chronic cochlear implant (CI) care. A CI is a device for people with severe hearing loss who do not benefit sufficiently from hearing aids. This thesis does not address other factors that hamper access to hearing healthcare, such as stigma or lack of motivation to seek help. Chapters 1, 2, and 3 examine the potential global impact of AI on hearing healthcare. The subsequent chapters address relatively affordable practical projects and solutions, including automated speech recognition apps for individuals with hearing loss, self-administered hearing tests, and the use of teleaudiology in CI aftercare.

Chapter 1 further explains the capacity challenge in hearing healthcare and proposes solutions based on teleaudiology and AI. Teleaudiology refers to hearing care provided online or from home, for example, via a smartphone. Teleaudiology can lower the threshold for seeking care, especially in areas where hearing care is scarce or absent, and it can free up capacity within the healthcare system. For instance, individuals with hearing loss can test themselves without a professional's assistance or with AI support. AI is an umbrella term for machines and software that mimic human intelligence and can now be used for performing complex tasks. In this thesis, the term "computational audiology" is used for AI-driven applications in hearing healthcare.

Chapter 2 looks at the digital transformation that enables computational audiology and outlines the requirements to improve hearing healthcare. It provides examples of AI applications in hearing diagnosis, hearing rehabilitation, and scientific research from recent literature. Besides practical advice for policymakers and

clinicians, the ethical implications of AI are examined. Issues such as liability and the potential dangers of AI in hearing healthcare, including privacy risks and the risk of incorrect conclusions based on biased data or algorithms, are discussed. The chapter concludes with a call for standards to improve interoperability between audiological equipment, facilitate data exchange between organizations so that it no longer matters where care is provided, and train staff to use AI responsibly in hearing healthcare.

The possibilities of AI chatbots in hearing healthcare are outlined in Chapter 3, based on simulated conversations with an AI chatbot and recent literature. AI chatbots offer opportunities to provide personalized patient care, improve hearing healthcare accessibility, and support researchers. However, there are limitations due to the questionable reliability of AI chatbots' information and the limited accuracy in citing sources. Potential benefits of AI chatbots include their 24/7 availability and their ability to rewrite complex information to meet the specific needs of patients or clinicians. The chapter emphasizes the need to establish guidelines for a responsible and safe implementation.

The effectiveness of automated speech recognition apps for converting speech to text for people with hearing loss was investigated in the study described in Chapter 4. Four apps were tested (Ava, Earfy, Live Transcribe, and Speechy), which all have a free version available in Dutch and run on a smartphone. The apps were tested using standard speech recognition tests in a sound booth. The smartphone was positioned where normally a person with hearing loss is tested. The word error rates in the transcript were determined. Automated speech recognition apps generally need loud speech material, preferably an intensity of 80 dB SPL or higher, to reach sufficient scores, similar to those with a moderate hearing loss. However, performance deteriorated when speech was presented in background noise. Word error rates in the transcript were insufficient to determine the practical usability of the automated speech recognition apps, which also depends on factors such as the readability of the text by the user. Therefore, it was suggested that future evaluations should include criteria related to usage conditions and user-friendliness, in addition to word error rates.

Chapter 5 reviews the accuracy, reliability, and test duration of self-administered hearing tests described in scientific literature. The literature review was conducted according to guidelines for a scoping review and was a follow-up to a 2012 literature study. The review focused on self-administered pure-tone audiometry. In pure-tone audiometry hearing thresholds are determined for frequencies that are important

for speech reception. Fifty-six publications from 2012 to June 2021 were found, and overlapping publications were clustered, resulting in 27 unique approaches that were assessed on clinically relevant criteria. Six approaches had accuracy comparable to standard air-conduction audiometry performed in audiological centers. New techniques to enhance the reliability of self-administered hearing tests were identified, such as measuring background noise levels during testing and automatically flagging invalid measurements. Thirteen approaches had a similar test duration to clinical audiometry, and three reported shorter test durations using more effective algorithms to determine hearing thresholds. Identified gaps in the literature include the limited number of self-administered tests that measure bone conduction and tests designed for children.

The stability and reliability of self-administered tests performed at-home in 50 CI users are described in Chapter 6. Participants determined their aided hearing thresholds and speech reception threshold for digits-in-noise with their CI at ten different sessions. The test outcomes were used to assess auditory functioning in relation to time-of-day, fatigue, motivation, and self-reported chronotype (morning or evening person). The outcomes showed that the time-of-day, fatigue, motivation, or chronotype did not affect test outcomes, implying that tests performed at-home are reliable and can be performed at any time. However, speech recognition with CI must be sufficient to perform the digits-in-noise test. Participants needed to score at least 65% on the Dutch CNC-test at conversational speech intensity (65 dB SPL). Based on the results, there are opportunities to organize chronic CI care from home, providing digitally proficient CI users more control over their follow-up care, such as insight into their auditory capabilities, and this could, if implemented effectively, reduce clinic workload by establishing criteria for in-clinic appointments.

Chapter 7 presents a feasibility study of a CI fitting procedure based on phoneme confusions measured in experienced CI users using self-administered tests. Phonemes are the sounds that make up spoken words. Self-administered tests were used to assess for each participant which phonemes were not identified correctly and with which phonemes they were confused. Participants who received a more targeted fitting intervention based on specific phoneme errors showed a decrease in errors compared to those who received a more generic intervention. Although this procedure shows potential, confirmatory studies are needed to provide stronger evidence and distinguish the effects of the conventional clinical CI fitting procedure from the phoneme-confusion-based fitting procedure.

Chapter 8 summarizes the findings of this thesis and describes the potential of AI-assisted teleaudiology to improve access to and affordability of hearing care worldwide. Self-administered tests performed at-home provide reliable and stable outcomes in CI users. Outcomes of self-administered tests can be used to adjust CI fittings. A literature review found six self-administered approaches that meet clinical standards for determining hearing status. Additionally, automated speech recognition apps for people with hearing loss provide adequate transcription in quiet listening situations. With the further development of teleaudiology and AI chatbots, people with hearing loss can increasingly manage their hearing healthcare at the time and place of their choosing. However, we will need to involve people with hearing loss more in developing these technical solutions to align those solutions with their needs and capabilities. There is still much work to be done before people feel as comfortable in a digital environment as they do in the physical world. It will require interdisciplinary collaboration, as well as standardization of protocols and data collection to integrate these technologies into clinical practice. In this way, computational audiology can be optimally utilized as one of the solutions to the growing capacity challenge in hearing healthcare.

NEDERLANDSE SAMENVATTING

Op Weg Naar AI Ondersteunde Tele-audiologie

In dit proefschrift zijn methoden onderzocht die gebaseerd zijn op tele-audiologie en/of kunstmatige intelligentie (AI) die slechthorenden kunnen ondersteunen en waarmee op termijn het toenemende capaciteitsprobleem in de hoorzorg aangepakt kan worden. Er zijn wereldwijd 1,5 miljard mensen met enige mate van gehoorverlies en een groot deel daarvan heeft volgens de Wereldgezondheidsorganisatie geen toegang tot hoorzorg of hoorhulpmiddelen. Door vergrijzing van de wereldbevolking zal het capaciteitsprobleem in de hoorzorg nog verder toenemen en daarom zijn innovaties nodig.

Er zijn in dit proefschrift nieuwe technologieën bestudeerd om de toegang tot hoorzorg te verbeteren. De huidige beperkingen worden mede veroorzaakt door onvoldoende beschikbaarheid van gehoordiagnostiek, de kosten van hoorhulpmiddelen en de complexiteit van het instellen van deze hoorhulpmiddelen. De primaire focus ligt op langdurige cochleaire implantaat (CI) zorg. Een CI is een hulpmiddel voor mensen met een zeer ernstig gehoorverlies waarvoor hoortoestellen onvoldoende baat bieden. Andere factoren die toegang tot hoorzorg in de weg staan zoals stigma of gebrek aan motivatie om hulp te zoeken, worden in dit proefschrift niet behandeld. Nadat in hoofdstuk 1, 2 en 3 in de volle breedte is gekeken naar de mogelijke impact van AI op de hoorzorg wereldwijd, richten de hoofdstukken daarna zich op relatief betaalbare oplossingen en projecten zoals automatische spraakherkenningapps voor slechthorenden, zelfuitgevoerde hoortesten, en het gebruik van tele-audiologie in CI-nazorg.

In hoofdstuk 1 wordt het capaciteitsprobleem in de hoorzorg toegelicht en worden oplossingen voorgesteld gebaseerd op tele-audiologie en AI. Tele-audiologie is het online of vanuit thuis aanbieden van hoorzorg via bijvoorbeeld een smartphone. Tele-audiologie kan de drempel verlagen tot het verkrijgen van hoorzorg, met name op plekken waar hoorzorg schaars is of ontbreekt, en kan capaciteit in de zorg vrijspelen. Bijvoorbeeld omdat de slechthorende zichzelf kan testen zonder hulp van een professional of juist met ondersteuning van AI. AI is een verzamelterm voor machines en software die menselijke intelligentie nabootsen en kan tegenwoordig gebruikt worden voor het uitvoeren van complexe taken. De term "*computational audiology*" wordt in dit proefschrift gebruikt voor AI-gedreven toepassingen in de hoorzorg.

In hoofdstuk 2 wordt nader ingegaan op de digitale transformatie die *computational audiology* mogelijk maakt en de benodigde randvoorwaarden om hiermee hoorzorg

te verbeteren. Er worden voorbeelden gegeven uit de literatuur van AI-toepassingen op het gebied van gehoordiagnostiek, gehoorrevalidatie en wetenschappelijk onderzoek. Naast praktische adviezen voor beleidsmakers en clinici wordt ook gekeken naar de ethische implicaties van het gebruik van AI. Vraagstukken zoals aansprakelijkheid en de mogelijke gevaren van AI in de hoorzorg komen aan bod. Denk hierbij aan de risico's rondom privacy, of verkeerde conclusies die getrokken worden op basis van vertekende data of bevooroordeelde algoritmen. Het hoofdstuk eindigt met een pleidooi voor standaarden om audiologische apparatuur beter onderling te laten samenwerken (interoperabel), gegevensuitwisseling tussen organisaties beter mogelijk te maken zodat het niet meer uitmaakt waar deze zorg geboden wordt, en om medewerkers te trainen om tot een verantwoord gebruik van AI in de hoorzorg te komen.

De mogelijkheden van AI-chatbots in de hoorzorg worden naar aanleiding van gesimuleerde gesprekken met een AI-chatbot geschetst in hoofdstuk 3 op basis van recente literatuur. AI-chatbots bieden kansen om gepersonaliseerde patiëntenzorg te bieden, de toegankelijkheid van gezondheidszorg te verbeteren, en onderzoekers te ondersteunen. Vooralsnog zijn er beperkingen wegens de wisselende betrouwbaarheid van de informatie die AI-chatbots leveren en het ontbreken van accurate bronverwijzingen. Potentiële voordelen van AI-chatbots zijn dat ze 24 uur per dag inzetbaar zijn en dat ze complexe informatie specifiek kunnen herschrijven zodat dit beter past bij de behoefte van de patiënt of juist die van de clinicus. In het hoofdstuk wordt de noodzaak benadrukt om richtlijnen op te stellen om tot een verantwoorde en veilige implementatie te komen.

Hoe goed spraakherkenningsapps voor slechthorenden spraak omzetten in tekst is onderzocht in de studie beschreven in hoofdstuk 4. We hebben hiertoe vier apps getest (Ava, Earfy, Live Transcribe en Speechy) waarvan een gratis versie in het Nederlands beschikbaar was. Het zijn apps met automatische spraakherkenning die te gebruiken zijn op een smartphone. De apps werden getest met behulp van standaard spraakverstaantesten in de kliniek door de smartphone te plaatsen op de plek van de slechthorende en de fouten in het transcript op woordniveau te tellen. De spraakherkenningsapps hadden een luid spraaksignaal nodig, liefst 80 dB SPL of meer, om een voldoende resultaat te bereiken, dit is vergelijkbaar met mensen met een matig gehoorverlies. Echter, wanneer spraak in achtergrondlawaai werd aangeboden vielen de prestaties van de spraakherkenningsapps tegen. De fouten in het transcript op woordniveau zeggen onvoldoende over de bruikbaarheid van de spraakherkenningsapps in de praktijk, wat onder andere afhangt van de leesbaarheid van de tekst door de gebruiker. Om toekomstige evaluatie van

spraakherkenningsapps representatiever te maken voor de bruikbaarheid in de praktijk wordt daarom de aanbeveling gedaan om naast fouten op woordniveau ook criteria op het gebied van gebruiksomstandigheden en gebruiksgemak mee te nemen.

In hoofdstuk 5 zijn de in de wetenschappelijke literatuur beschreven meetmethoden voor zelfuitgevoerde hoortesten beoordeeld op nauwkeurigheid, betrouwbaarheid en testduur. De literatuurstudie werd uitgevoerd volgens de richtlijnen voor een scoping review en was een vervolg op een literatuurstudie uit 2012. Er werd specifiek gekeken naar zelfuitgevoerde toonaudiometrie. Bij toonaudiometrie worden de gehoordrempels bepaald voor frequenties die belangrijk zijn voor het horen van spraak. Er werden 56 publicaties gevonden uit de periode van 2012 tot juni 2021. Overlappende publicaties werden samengevoegd. Dit leidde tot 27 unieke meetmethoden die werden beoordeeld op klinisch relevante criteria. De studie vond dat de nauwkeurigheid van zes meetmethoden vergelijkbaar was met standaard luchtgeleidingsaudiometrie zoals dat op een audiologisch centrum wordt verricht. Er werden nieuwe technieken gevonden om de betrouwbaarheid van zelfuitgevoerde hoortesten te verhogen zoals het meten van achtergrondgeluid tijdens de meting en het automatisch markeren van ongeldige metingen. Qua tijdsduur waren er dertien meetmethoden die een vergelijkbare testduur hadden als klinische audiometrie. Er werden drie meetmethoden gevonden die een kortere testduur rapporteerden door gebruik te maken van effectievere zoekalgoritmen om de gehoordrempel te bepalen. Verder werden er in de literatuur lacunes gevonden zoals het geringe aantal zelfuitgevoerde hoortesten waarmee beengeleiding kan worden gemeten of hoortesten die aantrekkelijk gemaakt zijn voor kinderen.

De stabiliteit en betrouwbaarheid van zelfuitgevoerde thuistesten bij 50 CI-gebruikers wordt beschreven in hoofdstuk 6. De deelnemers bepaalden op tien testmomenten hun geholpen gehoordrempels en verstaan van cijfers in ruis met CI. De testuitslagen werden gebruikt om het auditief functioneren van CI-gebruikers te bepalen als functie van testmoment, vermoeidheid, motivatie en zelfgerapporteerd chronotype, dat wil zeggen of ze zich als ochtend- of avondmens typeerden. De resultaten toonden dat het moment van uitvoer van de thuistest, de vermoeidheid, motivatie of chronotype geen invloed had op de testuitslagen. Dit impliceert dat thuistesten stabiele resultaten opleveren en op elk moment uitvoerbaar zijn. Wel dient het spraakverstaan met CI voldoende te zijn om het verstaan van de cijfers in ruis goed te kunnen uitvoeren. Deelnemers moesten bij standaard spraakaudiometrie ten minste 65% spraakverstaan scoren op conversatiesterkte (65 dB SPL). Op basis van de studieresultaten wordt gesuggereerd dat er mogelijk-

heden zijn om chronische CI-zorg vanuit thuis te organiseren. Dit biedt CI-gebruikers die digitaal vaardig zijn meer regie over hun nazorg, zoals inzicht in hun auditieve mogelijkheden, en kan, mits doelmatig geïmplementeerd, de werkdruk voor de kliniek te verminderen door middel van criteria voor controles in de kliniek.

In hoofdstuk 7 wordt een haalbaarheidsstudie beschreven van een CI-afregelprocedure gebaseerd op foneemverwisselingen bij ervaren CI-gebruikers met behulp van zelfuitgevoerde testen. Fonemen zijn de klanken waaruit gesproken woorden zijn opgebouwd. Met behulp van zelfuitgevoerde testen werd per CI-gebruiker bepaald welke fonemen niet goed onderscheiden werden en voor welk foneem ze werden verwisseld. Het foutenpercentage van specifiek behandelde foneemverwisselingen nam af bij deelnemers die een specifiek op die foneemverwisselingen-gebaseerde afregelinterventie kregen, in vergelijking met degenen die een meer generieke afregelinterventie kregen. Hoewel deze op foneemverwisselingen-gebaseerde afregelprocedure potentie heeft, zijn vervolgstudies nodig om een sterkere onderbouwing te leveren en de effecten van de gangbare klinische CI-afregelprocedure te onderscheiden van de op foneemverwisselingen-gebaseerde procedure.

Hoofdstuk 8 vat de bevindingen van dit proefschrift samen en beschrijft welke mogelijkheden AI-gedreven tele-audiologie biedt om toegang tot en betaalbaarheid van de gehoorzorg wereldwijd te verbeteren. Zelftesten vanuit thuis bieden betrouwbare en stabiele resultaten bij CI-gebruikers op elk moment van de dag. Resultaten uit zelftesten kunnen worden gebruikt om de CI-instellingen aan te passen. In de literatuur werden zelftestmethoden gevonden voor het bepalen van de gehoorstatus die aan klinische eisen voldoen. Verder geven automatische spraakherkenningsapps voor slechthorenden, met name in rustige situaties, gesprekken goed weer. Met de doorontwikkeling van tele-audiologie en AI-chatbots, zullen mensen met gehoorverlies steeds meer zelf hun hoorzorg vorm kunnen geven, op het moment en vanuit de plek waar zij dat willen. Echter, we zullen mensen met gehoorverlies dan nog meer moeten betrekken bij de ontwikkeling van deze technische oplossingen om deze zo goed mogelijk te laten aansluiten op hun wensen en mogelijkheden. Er is nog veel werk te verzetten voordat mensen zich in een digitale omgeving net zo op hun gemak voelen als in de fysieke wereld. Dit zal interdisciplinaire samenwerking vergen, en standaardisatie van protocollen en gegevensverzameling zal nodig zijn om de nieuwe technieken in de klinische praktijk te kunnen inpassen. Op deze wijze kan *computational audiology* optimaal worden ingezet als één van de oplossingen voor het groeiende capaciteitsprobleem in de hoorzorg.

References

Abdullah, S., Murnane, E. L., Matthews, M., Kay, M., Kientz, J. A., Gay, G., & Choudhury, T. (2016). Cognitive rhythms: Unobtrusive and continuous sensing of alertness using a mobile phone. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 178–189. https://doi.org/10.1145/2971648.2971712

Abegunde, D. O., Mathers, C. D., Adam, T., Ortegon, M., & Strong, K. (2007). The burden and costs of chronic diseases in low-income and middle-income countries. *The Lancet*, *370*(9603), 1929–1938.

Adamopoulou, E., & Moussiades, L. (2020). An Overview of Chatbot Technology. In I. Maglogiannis, L. Iliadis, & E. Pimenidis (Eds.), *Artificial Intelligence Applications and Innovations* (pp. 373–383). Springer International Publishing. https://doi.org/10.1007/978-3-030-49186-4_31

A&L Goodbody. (2016). *The GDPR-A Guide for Businesses*. A&L Goodbody. https://www.algoodbody.com/media/TheGDPR-AGuideforBusinesses1.pdf

Alhanbali, S., Dawes, P., Lloyd, S., & Munro, K. J. (2017). Self-Reported Listening-Related Effort and Fatigue in Hearing-Impaired Adults. *Ear and Hearing*, *38*(1), e39. https://doi.org/10.1097/AUD.0000000000000361

Aliyeva, A., Sari, E., Alaskarov, E., & Nasirov, R. (2024). Enhancing Postoperative Cochlear Implant Care With ChatGPT-4: A Study on Artificial Intelligence (AI)-Assisted Patient Education and Support. *Cureus*. https://doi.org/10.7759/cureus.53897

AlSamhori, J. F., AlSamhori, A. R. F., Amourah, R. M., AlQadi, Y., Koro, Z. W., Haddad, T. R. A., AlSamhori, A. F., Kakish, D., Kawwa, M. J., Zuriekat, M., & Nashwan, A. J. (2024). Artificial intelligence for hearing loss prevention, diagnosis, and management. *Journal of Medicine, Surgery, and Public Health*, *3*, 100133. https://doi.org/10.1016/j.glmedi.2024.100133

American Speech-Language-Hearing Association. (2005). *Guidelines for manual pure-tone threshold audiometry*. https://www.asha.org/policy/gl2005-00014/

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, *8*(1), 19–32. https://doi.org/10.1080/1364557032000119616

Ashique, K. T., Kaliyadan, F., & Aurangabadkar, S. J. (2015). Clinical photography in dermatology using smartphones: An overview. *Indian Dermatology Online Journal*, *6*(3), 158. https://doi.org/10.4103/2229-5178.156381

Au Yeung, J., Kraljevic, Z., Luintel, A., Balston, A., Idowu, E., Dobson, R. J., & Teo, J. T. (2023). AI chatbots not yet ready for clinical use. *Frontiers in Digital Health*, *5*, 1161098. https://doi.org/10.3389/fdgth.2023.1161098

Ausili, S. A. (2019). *Spatial Hearing with Electrical Stimulation Listening with Cochlear Implants* [Radboud University]. https://repository.ubn.ru.nl/handle/2066/203054

Baby, D., Van Den Broucke, A., & Verhulst, S. (2021). A convolutional neural-network model of human cochlear mechanics and filter tuning for real-time applications. *Nature Machine Intelligence*, *3*(2), 134–143. https://doi.org/10.1038/s42256-020-00286-8

Bailey, A. (2020). AirPods Pro Become Hearing Aids in iOS 14. *Hearing Tracker*. https://www.hearingtracker.com/news/airpods-pro-become-hearing-aids-with-ios-14

Baker, J. M., Deng, L., Glass, J., Khudanpur, S., Lee, C., Morgan, N., & O'Shaughnessy, D. (2009). Developments and directions in speech recognition and understanding, Part 1 [DSP Education]. *IEEE Signal Processing Magazine*, *26*(3), 75–80. IEEE Signal Processing Magazine. https://doi.org/10.1109/MSP.2009.932166

Barbour, D. L. (2018). Formal Idiographic Inference in Medicine. *JAMA Otolaryngology–Head & Neck Surgery*, *144*(6), 467–468. https://doi.org/10.1001/jamaoto.2018.0254

Barbour, D. L., DiLorenzo, J. C., Sukesan, K. A., Song, X. D., Chen, J. Y., Degen, E. A., Heisey, K. L., & Garnett, R. (2019). Conjoint psychometric field estimation for bilateral audiometry. *Behavior Research Methods*, *51*(3), 1271–1285. https://doi.org/10.3758/s13428-018-1062-3

Barbour, D. L., Howard, R. T., Song, X. D., Metzger, N., Sukesan, K. A., DiLorenzo, J. C., Snyder, B. R. D., Chen, J. Y., Degen, E. A., Buchbinder, J. M., & Heisey, K. L. (2019). Online Machine Learning Audiometry. *Ear and Hearing*, *40*(4), 918–926. https://doi.org/10.1097/AUD.0000000000000669

Barnett, M., Hixon, B., Okwiri, N., Irungu, C., Ayugi, J., Thompson, R., Shinn, J. B., & Bush, M. L. (2017). Factors involved in access and utilization of adult hearing healthcare: A systematic review. *The Laryngoscope*, *127*(5), 1187–1194. https://doi.org/10.1002/lary.26234

Baskent, D., Eiler, C. L., & Edwards, B. (2007). Using Genetic Algorithms with Subjective Input from Human Subjects: Implications for Fitting Hearing Aids and Cochlear Implants. *Ear and Hearing*, *28*(3), 370. https://doi.org/10.1097/AUD.0b013e318047935e

Bastianelli, M., Mark, A. E., McAfee, A., Schramm, D., Lefrançois, R., & Bromwich, M. (2019). Adult validation of a self-administered tablet audiometer. *Journal of Otolaryngology - Head & Neck Surgery*, *48*(1), 59. https://doi.org/10.1186/s40463-019-0385-0

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Battmer, R.-D., Borel, S., Brendel, M., Buchner, A., Cooper, H., Fielden, C., Gazibegovic, D., Goetze, R., Govaerts, P., Kelleher, K., Lenartz, T., Mosnier, I., Muff, J., Nunn, T., Vaerenberg, B., & Vanat, Z. (2015). Assessment of 'Fitting to Outcomes Expert' FOX™ with new cochlear implant users in a multi-centre study. *Cochlear Implants International*, *16*(2), 100–109. https://doi.org/10.1179/1754762814Y.0000000093

Bean, B. N., Roberts, R. A., Picou, E. M., Angley, G. P., & Edwards, A. J. (2021). Automated Audiometry in Quiet and Simulated Exam Room Noise for Listeners with Normal Hearing and Impaired Hearing. *Journal of the American Academy of Audiology*. https://doi.org/10.1055/s-0041-1728778

Békésy, G. v. (1947). A new audiometer. *Acta Oto-Laryngologica*, *35*(5–6), 411–422.

Bennett, B., Kelsall-Foreman, I., Barr, C., Campbell, E., Coles, T., Paton, M., & Vitkovic, J. (2023). Barriers and facilitators to tele-audiology service delivery in Australia during the COVID-19 pandemic: Perspectives of hearing healthcare clinicians. *International Journal of Audiology*, *62*(12), 1145–1154. https://doi.org/10.1080/14992027.2022.2128446

Bennett, B., Laird, E., Timmer, B., & Campbel, E. (2024). Development and Implementation of National Teleaudiology Guidelines. *The Hearing Journal*, *77*(2), 23–24.

Bennett, B., Meyer, C. J., Eikelboom, R. H., & Atlas, M. D. (2018). Evaluating Hearing Aid Management: Development of the Hearing Aid Skills and Knowledge Inventory (HASKI). *American Journal of Audiology*, *27*(3), 333–348. https://doi.org/10.1044/2018_AJA-18-0050

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2012). GenBank. *Nucleic Acids Research*, *41*(D1), D36–D42. https://doi.org/10.1093/nar/gks1195

Berenger, M. (2021). *Hearing Australia*. New App from National Acoustic Laboratories Improves Communication at Hearing Health Clinics. https://www.hearing.com.au/About-Hearing-Australia/Hearing-news/New-app-from-National-Acoustic-Laboratories-improv

Berke, L. (2017). Displaying confidence from imperfect automatic speech recognition for captioning. *ACM SIGACCESS Accessibility and Computing*, *117*, 14–18. https://doi.org/10.1145/3051519.3051522

Bernstein, L. E., Tucker, P. E., & Demorest, M. E. (2000). Speech perception without hearing. *Perception & Psychophysics*, *62*(2), 233–252.

Bharathi, C. U., Ragavi, G., & Karthika, K. (2021). Signtalk: Sign Language to Text and Speech Conversion. *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*, 1–4. https://doi.org/10.1109/ICAECA52838.2021.9675751

Biadsy, F., Weiss, R. J., Moreno, P. J., Kanevsky, D., & Jia, Y. (2019). *Parrotron: An End-to-End Speech-to-Speech Conversion Model and its Applications to Hearing-Impaired Speech and Speech Separation* (arXiv:1904.04169). arXiv. https://doi.org/10.48550/arXiv.1904.04169

Bis, J. C., DeCarli, C., Smith, A. V., van der Lijn, F., Crivello, F., Fornage, M., Debette, S., Shulman, J. M., Schmidt, H., Srikanth, V., Schuur, M., Yu, L., Choi, S.-H., Sigurdsson, S., Verhaaren, B. F. J., DeStefano, A. L., Lambert, J.-C., Jack, C. R., Struchalin, M., … the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium. (2012). Common variants at 12q14 and 12q24 are associated with hippocampal volume. *Nature Genetics*, *44*(5), 545–551. https://doi.org/10.1038/ng.2237

Bizios, D., Heijl, A., & Bengtsson, B. (2011). Integration and fusion of standard automated perimetry and optical coherence tomography data for improved automated glaucoma diagnostics. *BMC Ophthalmology*, *11*(1), 1–11. https://doi.org/10.1186/1471-2415-11-20

Blamey, P., Artieres, F., Başkent, D., Bergeron, F., Beynon, A., Burke, E., Dillier, N., Dowell, R., Fraysse, B., Gallégo, S., Govaerts, P. J., Green, K., Huber, A. M., Kleine-Punte, A., Maat, B., Marx, M., Mawman, D., Mosnier, I., O'Connor, A. F., … Lazard, D. S. (2013). Factors Affecting Auditory Performance of Postlinguistically Deaf Adults Using Cochlear Implants: An Update with 2251 Patients. *Audiology and Neurotology*, *18*(1), 36–47. https://doi.org/10.1159/000343189

Bloem, B. R., Marks, W. J., Silva de Lima, A. L., Kuijf, M. L., van Laar, T., Jacobs, B. P. F., Verbeek, M. M., Helmich, R. C., van de Warrenburg, B. P., Evers, L. J. W., intHout, J., van de Zande, T., Snyder, T. M., Kapur, R., & Meinders, M. J. (2019). The Personalized Parkinson Project: Examining disease progression through broad biomarkers in early Parkinson's disease. *BMC Neurology*, *19*(1), 160. https://doi.org/10.1186/s12883-019-1394-3

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot. Int.*, *5*(9), 341–345.

Bolner, F., Magits, S., van Dijk, B., & Wouters, J. (2020). Precompensating for spread of excitation in a cochlear implant coding strategy. *Hearing Research*, *395*, 107977. https://doi.org/10.1016/j.heares.2020.107977

Bornman, M., Swanepoel, D. W., De Jager, L. B., & Eikelboom, R. H. (2019). Extended high-frequency smartphone audiometry: Validity and reliability. *Journal of the American Academy of Audiology*, *30*(3), 217–226.

Bosman, A. J., & Smoorenburg, G. F. (1992). Woordenlijst voor spraakaudiometrie. *Nederlandse Vereniging Voor Audiologie, Utrecht*.

Bosman, A. J., & Smoorenburg, G. F. (1995). Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. *Audiology*, *34*(5), 260–284. https://doi.org/10.3109/00206099509071918

Botros, A., & Psarros, C. (2010). Neural response telemetry reconsidered: I. The relevance of ECAP threshold profiles and scaled profiles to cochlear implant fitting. *Ear and Hearing*, *31*(3), 367–379. https://doi.org/10.1097/AUD.0b013e3181c9fd86

Bramsløw, L., Naithani, G., Hafez, A., Barker, T., Pontoppidan, N. H., & Virtanen, T. (2018). Improving competing voices segregation for hearing impaired listeners using a low-latency deep neural network algorithm. *The Journal of the Acoustical Society of America*, *144*(1), 172–185. https://doi.org/10.1121/1.5045322

Brendel, M., Buechner, A., Krueger, B., Frohne-Buechner, C., & Lenarz, T. (2008). Evaluation of the Harmony soundprocessor in combination with the speech coding strategy HiRes 120. *Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, *29*(2), 199–202. https://doi.org/10.1097/mao.0b013e31816335c6

Brennan-Jones, C. G., Eikelboom, R. H., Bennett, R. J., Tao, K. F., & Swanepoel, D. W. (2018). Asynchronous interpretation of manual and automated audiometry: Agreement and reliability. *Journal of Telemedicine and Telecare*, *24*(1), 37–43.

Brennan-Jones, C. G., Eikelboom, R. H., Swanepoel de, W., Friedl, P. L., & Atlas, M. D. (2016). Clinical validation of automated audiometry with continuous noise-monitoring in a clinically heterogeneous population outside a sound-treated environment. *Int J Audiol*, *55*(9), 507–513.

Brittz, M., Heinze, B., Mahomed-Asmail, F., Swanepoel, D. W., & Stoltz, A. (2019). Monitoring Hearing in an Infectious Disease Clinic with mHealth Technologies. *Journal of the American Academy of Audiology*, *30*(6), 482–492. https://doi.org/10.3766/jaaa.17120

Bronkhorst, A. W., Bosman, A. J., & Smoorenburg, G. F. (1993). A model for context effects in speech recognition. *The Journal of the Acoustical Society of America*, *93*(1), 499–509.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., … Amodei, D. (2020). Language Models are Few-Shot Learners. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), *Advances in Neural Information Processing Systems* (Vol. 33, pp. 1877–1901). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

Browning, L. M., Nie, Y., Rout, A., & Heiner, M. (2020). Audiologists' preferences in programming cochlear implants: A preliminary report. *Cochlear Implants International*, *21*(4), 179–191. https://doi.org/10.1080/14670100.2019.1708553

Buchman, C. A., Gifford, R. H., Haynes, D. S., Lenarz, T., O'Donoghue, G., Adunka, O., Biever, A., Briggs, R. J., Carlson, M. L., Dai, P., Driscoll, C. L., Francis, H. W., Gantz, B. J., Gurgel, R. K., Hansen, M. R., Holcomb, M., Karltorp, E., Kirtane, M., Larky, J., … Zwolan, T. (2020). Unilateral Cochlear Implants for Severe, Profound, or Moderate Sloping to Profound Bilateral Sensorineural Hearing Loss: A Systematic Review and Consensus Statements. *JAMA Otolaryngology–Head & Neck Surgery*, *146*(10), 942–953. https://doi.org/10.1001/jamaoto.2020.0998

Buolamwini, J. A. (2017). *Gender shades: Intersectional phenotypic and demographic evaluation of face datasets and gender classifiers* [Thesis, Massachusetts Institute of Technology]. http://dspace.mit.edu/handle/1721.1/114068

Bush, M. L., Burton, M., Loan, A., & Shinn, J. B. (2013). Timing discrepancies of early intervention hearing services in urban and rural cochlear implant recipients. *Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, *34*(9), 1630–1635. https://doi.org/10.1097/MAO.0b013e31829e83ad

Byrne, D., Dillon, H., Ching, T., Katsch, R., & Keidser, G. (2001). NAL-NL1 procedure for fitting nonlinear hearing aids: Characteristics and comparisons with other procedures. *Journal of the American Academy of Audiology*, *12*(1).

Caposecco, A., Hickson, L., Meyer, C., & Khan, A. (2016). Evaluation of a Modified User Guide for Hearing Aid Management. *Ear and Hearing*, *37*(1), 27–37. https://doi.org/10.1097/AUD.0000000000000221

Carroll, M. W. (2015). Sharing Research Data and Intellectual Property Law: A Primer. *PLOS Biology*, *13*(8), e1002235. https://doi.org/10.1371/journal.pbio.1002235

Castor EDC. (2022). *Castor Electronic Data Capture (EDC)*. Castor. https://www.castoredc.com

Cha, D., Pae, C., Seong, S.-B., Choi, J. Y., & Park, H.-J. (2019). Automated diagnosis of ear disease using ensemble deep learning with a big otoendoscopy image database. *EBioMedicine*, *45*, 606–614. https://doi.org/10.1016/j.ebiom.2019.06.050

Chan, J., Raju, S., Nandakumar, R., Bly, R., & Gollakota, S. (2019). Detecting middle ear fluid using smartphones. *Science Translational Medicine*, *11*(492), eaav1102. https://doi.org/10.1126/scitranslmed.aav1102

Charih, F., Bromwich, M., Mark, A. E., Lefrançois, R., & Green, J. R. (2020). Data-Driven Audiogram Classification for Mobile Audiometry. *Scientific Reports*, *10*(1), 3962. https://doi.org/10.1038/s41598-020-60898-3

Chavez, M. A., Feanny, M., Seita, M., Thompson, B., Delk, K., Officer, S., Glasser, A., Kushalnagar, R., & Vogler, C. (2024). *How Users Experience Closed Captions on Live Television: Quality Metrics Remain a Challenge* (arXiv:2404.10153). arXiv. https://doi.org/10.48550/arXiv.2404.10153

Chen, F., Wang, S., Li, J., Tan, H., Jia, W., & Wang, Z. (2019). Smartphone-Based Hearing Self-Assessment System Using Hearing Aids With Fast Audiometry Method. *IEEE Trans Biomed Circuits Syst*, *13*(1), 170–179.

Chong-White, N., Incerti, P., Poulos, M., & Tagudin, J. (2023). Exploring teleaudiology adoption, perceptions and challenges among audiologists before and during the COVID-19 pandemic. *BMC Digital Health*, *1*(1), 24. https://doi.org/10.1186/s44247-023-00024-1

Christensen, J. H., Saunders, G. H., Havtorn, L., & Pontoppidan, N. H. (2021). Real-World Hearing Aid Usage Patterns and Smartphone Connectivity. *Frontiers in Digital Health*, *3*, 722186. https://doi.org/10.3389/fdgth.2021.722186

Christensen, J. H., Saunders, G. H., Porsbo, M., & Pontoppidan, N. H. (2021). The everyday acoustic environment and its association with human heart rate: Evidence from real-world data logging with hearing aids and wearables. *Royal Society Open Science*, *8*(2), 201345. https://doi.org/10.1098/rsos.201345

Cieri, C., Miller, D., & Walker, K. (2004). The Fisher Corpus: A Resource for the Next Generations of Speech-to-Text. *LREC*, *Vol. 4*, 69–71.

Coco, L., Champlin, C. A., & Eikelboom, R. H. (2016). Community-Based Intervention Determines Tele-Audiology Site Candidacy. *American Journal of Audiology*, *25*(3S), 264–267. https://doi.org/10.1044/2016_AJA-16-0002

Coldewey, D. (2020, December 10). *Ava expands its AI captioning to desktop and web apps, and raises $4.5M to scale*. TechCrunch. https://social.techcrunch.com/2020/12/10/ava-expands-its-ai-captioning-to-desktop-and-web-apps-and-raises-4-5m-to-scale/

Colsman, A., Supp, G. G., Neumann, J., & Schneider, T. R. (2020). Evaluation of Accuracy and Reliability of a Mobile Screening Audiometer in Normal Hearing Adults. *Frontiers in Psychology*, *11*, 744.

Convery, E., Heeris, J., Ferguson, M., & Edwards, B. (2020). Human–Technology Interaction Considerations in Hearing Health Care: An Introduction for Audiologists. *American Journal of Audiology*, *29*(3S), 538–545. https://doi.org/10.1044/2020_AJA-19-00068

Convery, E., Keidser, G., Hickson, L., & Meyer, C. (2019). Factors Associated With Successful Setup of a Self-Fitting Hearing Aid and the Need for Personalized Support. *Ear and Hearing*, *40*(4), 794. https://doi.org/10.1097/AUD.0000000000000663

Corona, A. P., Ferrite, S., Bright, T., & Polack, S. (2020). Validity of hearing screening using hearTest smartphone-based audiometry: Performance evaluation of different response modes. *International Journal of Audiology*, *59*(9), 666–673. https://doi.org/10.1080/14992027.2020.1731767

Corry, M., Sanders, M., & Searchfield, G. D. (2017). The accuracy and reliability of an app-based audiometer using consumer headphones: Pure tone audiometry in a normal hearing group. *International Journal of Audiology*, *56*(9), 706–710.

Crowson, M. G., Lee, J. W., Hamour, A., Mahmood, R., Babier, A., Lin, V., Tucci, D. L., & Chan, T. C. Y. (2020). AutoAudio: Deep Learning for Automatic Audiogram Interpretation. *Journal of Medical Systems*, *44*(9), 163. https://doi.org/10.1007/s10916-020-01627-1

Crum, P. (2019). Hearables: Here come the: Technology tucked inside your ears will augment your daily life. *IEEE Spectrum*, *56*(5), 38–43. https://doi.org/10.1109/MSPEC.2019.8701198

Cullington, H., Kitterick, P., Weal, M., & Margol-Gromada, M. (2018). Feasibility of personalised remote long-term follow-up of people with cochlear implants: A randomised controlled trial. *BMJ Open*, *8*(4), e019640. https://doi.org/10.1136/bmjopen-2017-019640

Davies-Venn, E., & Glista, D. (2019). *Connected hearing healthcare: The realisation of benefit relies on successful clinical implementation*. *28*(5), 2.

de Graaff, F., Huysmans, E., Merkus, P., Theo Goverts, S., & Smits, C. (2018). Assessment of speech recognition abilities in quiet and in noise: A comparison between self-administered home testing and testing in the clinic for adult cochlear implant users. *International Journal of Audiology*, *57*(11), 872–880. https://doi.org/10.1080/14992027.2018.1506168

de Graaff, F., Huysmans, E., Vanpoucke, F. J., Merkus, P., Goverts, S. T., & Smits, C. (2016). The development of remote speech recognition tests for adult cochlear implant users: The effect of presentation mode of the noise and a reliable method to deliver sound in home environments. *Audiology and Neurotology*, *21*(Suppl. 1), 48–54.

de Graaff, F., Lissenberg-Witte, B. I., Kaandorp, M. W., Merkus, P., Goverts, S. T., Kramer, S. E., & Smits, C. (2020). Relationship between speech recognition in quiet and noise and fitting parameters, impedances and ECAP thresholds in adult cochlear implant users. *Ear and Hearing*, *41*(4), 935–947.

De Sousa, K. C., Smits, C., Moore, D. R., Myburgh, H. C., & Swanepoel, D. W. (2020). Pure-tone audiometry without bone-conduction thresholds: Using the digits-in-noise test to detect conductive hearing loss. *International Journal of Audiology*, 1–8. https://doi.org/10.1080/14992027.2020.1783585

De Sousa, K. C., Swanepoel, D. W., Moore, D. R., Myburgh, H. C., & Smits, C. (2020). Improving Sensitivity of the Digits-In-Noise Test Using Antiphasic Stimuli. *Ear and Hearing*, *41*(2), 442–450. https://doi.org/10.1097/AUD.0000000000000775

Deng, L. (2016). Deep learning: From speech recognition to language and multimodal processing. *APSIPA Transactions on Signal and Information Processing*, *5*.

Deng, L., & Li, X. (2013). Machine Learning Paradigms for Speech Recognition: An Overview. *IEEE Transactions on Audio, Speech, and Language Processing*, *21*(5), 1060–1089. IEEE Transactions on Audio, Speech, and Language Processing. https://doi.org/10.1109/TASL.2013.2244083

Desarnaulds, V., Carvalho, A. P., & Monay, G. (2002). Church acoustics and the influence of occupancy. *Building Acoustics*, *9*(1), 29–47.

Dewyer, N. A., Jiradejvong, P., Lee, D. S., Kemmer, J. D., Henderson Sabes, J., & Limb, C. J. (2019). Automated Smartphone Audiometry: A Preliminary Validation of a Bone-Conduction Threshold Test App. *Ann Otol Rhinol Laryngol*, *128*(6), 508–515.

Dille, M. F., Jacobs, P. G., Gordon, S. Y., Helt, W. J., & McMillan, G. P. (2013). OtoID: New extended frequency, portable audiometer for ototoxicity monitoring. *Journal of Rehabilitation Research and Development*, *50*(7), 997–1006. https://doi.org/10.1682/JRRD.2012.09.0176

Dingemanse, J. G., houben, R., & Soede, W. (2023). *Kwaliteitsnorm NVKF voor meetruimten en spreekkamers Audiologische Centra*. https://nvkf.nl/resources/media/Kwaliteitsnorm%20NVKF%20voor%20meetruimten%20en%20spreekkamers%20ACs.pdf

DiNino, M., Winn, M. B., & Bierer, J. A. (2016). Cochlear implant listener vowel identification performance and confusion patterns with selective channel activation programs. *The Journal of the Acoustical Society of America*, *140*(4_Supplement), 3438. https://doi.org/10.1121/1.4971082

D'Onofrio, K. L., & Zeng, F.-G. (2022). Tele-Audiology: Current State and Future Directions. *Frontiers in Digital Health*, *3*. https://doi.org/10.3389/fdgth.2021.788103

Dubno, J. R., Eckert, M. A., Lee, F.-S., Matthews, L. J., & Schmiedt, R. A. (2013). Classifying human audiometric phenotypes of age-related hearing loss from animal models. *Journal of the Association for Research in Otolaryngology: JARO*, *14*(5), 687–701. https://doi.org/10.1007/s10162-013-0396-x

Dwyer, R. T., Gifford, R. H., Bess, F. H., Dorman, M., Spahr, A., & Hornsby, B. W. Y. (2019). Diurnal Cortisol Levels and Subjective Ratings of Effort and Fatigue in Adult Cochlear Implant Users: A Pilot Study. *American Journal of Audiology*, *28*(3), 686–696. https://doi.org/10.1044/2019_AJA-19-0009

Earfy. (2017, September 11). *New Earfy functions based on user feedback!* -. Earfy. https://www.earfy.net/earfy/new-functions-for-earfy-based-on-user-feedback/

Editorial. (2023). Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature*, *613*(7945), 612–612. https://doi.org/10.1038/d41586-023-00191-1

Eijpe. (2014). *Overview of the national laws on electronic health records in the EU Member States National Report for the Netherlands* (p. 40). Milieu Ltd and Time.lex. https://ec.europa.eu/health/sites/health/files/ehealth/docs/laws_netherlands_en.pdf

Eikelboom, R. H., Swanepoel de, W., Motakef, S., & Upson, G. S. (2013). Clinical validation of the AMTAS automated audiometer. *Int J Audiol*, *52*(5), 342–349.

Eksteen, S., Launer, S., Kuper, H., Eikelboom, R. H., Bastawrous, A., & Swanepoel, D. W. (2019). Hearing and vision screening for preschool children using mobile technology, South Africa. *Bulletin of the World Health Organization*, *97*(10), 672–680. https://doi.org/10.2471/BLT.18.227876

Ethics Review Committee (ERC). (2024, February 9). *Using AI to write clinical notes and reports*. Ethics Review Committee (ERC). https://auderc.org.au/guidance-for-practitioners/using-ai-to-write-clinical-notes-and-reports/

Ezzatian, P., Pichora-Fuller, M. K., & Schneider, B. A. (2010). Do Circadian Rhythms Affect Adult Age-Related Differences in Auditory Performance? *Canadian Journal on Aging / La Revue Canadienne Du Vieillissement*, *29*(2), 215–221. https://doi.org/10.1017/S0714980810000139

Faber, B. M. (2017). Acoustical measurements with smartphones: Possibilities and limitations. *Acoustics Today*, *13*(2), 10–16.

Ferguson, M., Leighton, P., Brandreth, M., & Wharrad, H. (2018). Development of a multimedia educational programme for first-time hearing aid users: A participatory design. *International Journal of Audiology*, *57*(8), 600–609. https://doi.org/10.1080/14992027.2018.1457803

Ferraro, A., Galli, A., La Gatta, V., & Postiglione, M. (2023). Benchmarking open source and paid services for speech to text: An analysis of quality and input variety. *Frontiers in Big Data*, *6*. https://doi.org/10.3389/fdata.2023.1210559

Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America*, *88*(4), 1725–1736.

Fitzpatrick, E. M., Séguin, C., Schramm, D. R., Armstrong, S., & Chénier, J. (2009). The Benefits of Remote Microphone Technology for Adults with Cochlear Implants. *Ear and Hearing*, *30*(5), 590. https://doi.org/10.1097/AUD.0b013e3181acfb70

Flynn, M. C., Dowell, R. C., & Clark, G. M. (1998). Aided Speech Recognition Abilities of Adults With a Severe or Severe-to-Profound Hearing Loss. *Journal of Speech, Language, and Hearing Research*, *41*(2), 285–299. https://doi.org/10.1044/jslhr.4102.285

Forrest, C. (2018, April 24). How to request your personal data under GDPR. *TechRepublic*. https://www.techrepublic.com/article/how-to-request-your-personal-data-under-gdpr/

Foulad, A., Bui, P., & Djalilian, H. (2013). Automated audiometry using apple iOS-based application technology. *Otolaryngol Head Neck Surg*, *149*(5), 700–706.

Francart, T., Van Wieringen, A., & Wouters, J. (2011). Comparison of fluctuating maskers for speech recognition tests. *International Journal of Audiology*, *50*(1), 2–13.

Gardner, J. R., Malkomes, G., Garnett, R., Weinberger, K. Q., Barbour, D. L., & Cunningham, J. (2015). Bayesian Active Model Selection with an Application to Automated Audiometry. *NIPS*.

Garland, A. F., Jenveja, A. K., & Patterson, J. E. (2021). Psyberguide: A useful resource for mental health apps in primary care and beyond. *Families, Systems, & Health*, *39*(1), 155–157. https://doi.org/10.1037/fsh0000587

Gatehouse, S., Naylor, G., & Elberling, C. (2003). Benefits from hearing aids in relation to the interaction between the user and the environment. *International Journal of Audiology*, *42*(sup1), 77–85. https://doi.org/10.3109/14992020309074627

Gatsios, D., Antonini, A., Gentile, G., Marcante, A., Pellicano, C., Macchiusi, L., Assogna, F., Spalletta, G., Gage, H., Touray, M., Timotijevic, L., Hodgkins, C., Chondrogiorgi, M., Rigas, G., Fotiadis, D. I., & Konitsiotis, S. (2020). Mhealth for remote monitoring and management of Parkinson's disease: Determinants of compliance and validation of a tremor evaluation method. *JMIR mHealth and uHealth*. https://epubs.surrey.ac.uk/856376/

Gaur, Y., Metze, F., Miao, Y., & Bigham, J. P. (2015). *Using keyword spotting to help humans correct captioning faster.*

Gescheider, G. A. (2013). *Psychophysics: The Fundamentals*. Psychology Press.

Gilbert, S., Harvey, H., Melvin, T., Vollebregt, E., & Wicks, P. (2023). Large language model AI chatbots require approval as medical devices. *Nature Medicine*, *29*(10), 2396–2398. https://doi.org/10.1038/s41591-023-02412-6

Glasser, A., Kushalnagar, K., & Kushalnagar, R. (2017). Deaf, Hard of Hearing, and Hearing Perspectives on Using Automatic Speech Recognition in Conversation. *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, 427–432. https://doi.org/10.1145/3132525.3134781

Godfrey, J. J., Holliman, E. C., & McDaniel, J. (1992). SWITCHBOARD: Telephone speech corpus for research and development. *Acoustics, Speech, and Signal Processing, IEEE International Conference On*, *1*, 517–520.

Goehring, T., Bolner, F., Monaghan, J. J. M., van Dijk, B., Zarowski, A., & Bleeck, S. (2017). Speech enhancement based on neural networks improves speech intelligibility in noise for cochlear implant users. *Hearing Research*, *344*, 183–194. https://doi.org/10.1016/j.heares.2016.11.012

Goehring, T., Keshavarzi, M., Carlyon, R. P., & Moore, B. C. (2019). Using recurrent neural networks to improve the perception of speech in non-stationary noise by people with cochlear implants. *The Journal of the Acoustical Society of America*, *146*(1), 705–718. https://doi.org/10.1121/1.5119226

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation." *AI Magazine*, *38*(3), 50. https://doi.org/10.1609/aimag.v38i3.2741

Google. (2021). *How Google technology is improving accessibility for deaf people—Google*. About Google. https://about.google/intl/ALL_us/stories/making-conversation-more-accessible-with-live-transcribe

Gorin, A. L., Riccardi, G., & Wright, J. H. (1997). How may I help you? *Speech Communication*, *23*(1–2), 113–127. https://doi.org/10.1016/S0167-6393(97)00040-X

Govender, S. M., & Mars, M. (2018). Validity of automated threshold audiometry in school aged children. *International Journal of Pediatric Otorhinolaryngology*, *105*, 97–102. https://doi.org/10.1016/j.ijporl.2017.12.008

Güçlü, U., & van Gerven, M. A. J. (2017). Modeling the Dynamics of Human Brain Activity with Recurrent Neural Networks. *Frontiers in Computational Neuroscience*, *11*. https://doi.org/10.3389/fncom.2017.00007

Haile, L. M., Kamenov, K., Briant, P. S., Orji, A. U., Steinmetz, J. D., Abdoli, A., Abdollahi, M., Abu-Gharbieh, E., Afshin, A., Ahmed, H., Ahmed Rashid, T., Akalu, Y., Alahdab, F., Alanezi, F. M., Alanzi, T. M., Al Hamad, H., Ali, L., Alipour, V., Al-Raddadi, R. M., … Chadha, S. (2021). Hearing loss prevalence and years lived with disability, 1990–2019: Findings from the Global Burden of Disease Study 2019. *The Lancet*, *397*(10278), 996–1009. https://doi.org/10.1016/S0140-6736(21)00516-X

Handzel, O., Ben-Ari, O., Damian, D., Priel, M. M., Cohen, J., & Himmelfarb, M. (2013). Smartphone-Based Hearing Test as an Aid in the Initial Evaluation of Unilateral Sudden Sensorineural Hearing Loss. *Audiology and Neurotology*, *18*(4), 201–207. https://doi.org/10.1159/000349913

Hazan, A., Rivilla, J., Méndez, N., Wack, N., Paytuvi, O., Zarowski, A., Offeciers, E., & Kinsbergen, J. (2020, June 4). *Test-retest analysis of aggregated audiometry testing data using Jacoti Hearing Center self-testing application*. Proceedings of the VCCA2020 conference. https://computationalaudiology.com/test-retest-analysis-of-aggregated-audiometry-testing-data-using-jacoti-hearing-center-self-testing-application/

Hazelebach, J. (2023). *Hoe werkt Speaksee in vergelijking met andere apps?* Speaksee. https://speak-see.com/nl/blogs/introducing-speaksee/speaksee-outperforms-other-apps-in-terms-of-accuracy-and-in-background-noise

Hazelebach, J., Pragt, L., & Wasmann, J.-W. A. (2023). *Rapportage technische validatie Speaksee*.

Héder, M. (2017). From NASA to EU: the evolution of the TRL scale in Public Sector Innovation. *THE INNOVATION JOURNAL*, *22*(2), 1–23. http://eprints.sztaki.hu/9204/

Heisey, K. L., Buchbinder, J. M., & Barbour, D. L. (2018). Concurrent Bilateral Audiometric Inference. *Acta Acustica United with Acustica*, *104*(5), 762–765. https://doi.org/10.3813/AAA.919218

Heisey, K. L., Walker, A. M., Xie, K., Abrams, J. M., & Barbour, D. L. (2020). Dynamically Masked Audiograms With Machine Learning Audiometry. *Ear and Hearing*, *41(6)*, 1692–1702. https://doi.org/10.1097/AUD.0000000000000891

Helfer, K. S. (1997). Auditory and auditory-visual perception of clear and conversational speech. *Journal of Speech, Language, and Hearing Research: JSLHR*, *40*(2), 432–443. https://doi.org/10.1044/jslhr.4002.432

Hendrikse, M. M. E., Dingemanse, G., Grimm, G., Hohmann, V., & Goedegebure, A. (2024). Development of Virtual Reality scenes for clinical use with hearing device fine-tuning. *Proceedings of the 10th Convention of the European Acoustics Association Forum Acusticum 2023*, 1401–1408. https://doi.org/10.61782/fa.2023.0300

Henriksen, V., Kvaløy, O., & Svensson, U. P. (2014). Development and Calibration of a New Automated Method to Measure Air Conduction Auditory Thresholds Using an Active Earplug. *Acta Acustica United with Acustica*, *100*(1), 113–117.

Heutink, F., Koch, V., Verbist, B., van der Woude, W. J., Mylanus, E., Huinck, W., Sechopoulos, I., & Caballo, M. (2020). Multi-Scale deep learning framework for cochlea localization, segmentation and analysis on clinical ultra-high-resolution CT images. *Computer Methods and Programs in Biomedicine*, *191*, 105387. https://doi.org/10.1016/j.cmpb.2020.105387

Hildebrand, M. S., DeLuca, A. P., Taylor, K. R., Hoskinson, D. P., Hur, I. A., Tack, D., McMordie, S. J., Huygen, P. L. M., Casavant, T. L., & Smith, R. J. H. (2009). AudioGene Audioprofiling: A Machine-based Candidate Gene Prediction Tool for Autosomal Dominant Non-syndromic Hearing Loss. *The Laryngoscope*, *119*(11), 2211–2215. https://doi.org/10.1002/lary.20664

Ho, J., Tumkaya, T., Aryal, S., Choi, H., & Claridge-Chang, A. (2019). Moving beyond P values: Data analysis with estimation graphics. *Nature Methods*, *16*(7), 565–566. https://doi.org/10.1038/s41592-019-0470-3

Holden, L. K., Finley, C. C., Firszt, J. B., Holden, T. A., Brenner, C., Potts, L. G., Gotter, B. D., Vanderhoof, S. S., Mispagel, K., Heydebrand, G., & Skinner, M. W. (2013). Factors affecting open-set word recognition in adults with cochlear implants. *Ear and Hearing*, *34*(3), 342–360. https://doi.org/10.1097/AUD.0b013e3182741aa7

Holmes, A. E., Shrivastav, R., Krause, L., Siburt, H. W., & Schwartz, E. (2012). Speech based optimization of cochlear implants. *Int J Audiol*, *51*(11), 806–816. https://doi.org/10.3109/14992027.2012.705899

Holube, I., von Gablenz, P., & Bitzer, J. (2020). Ecological Momentary Assessment in Hearing Research: Current State, Challenges, and Future Directions. *Ear and Hearing*, *41 Suppl 1*, 79S-90S. https://doi.org/10.1097/AUD.0000000000000934

Hornsby, B. W. Y., Naylor, G., & Bess, F. H. (2016). A Taxonomy of Fatigue Concepts and Their Relation to Hearing Loss. *Ear and Hearing*, *37*, 136S. https://doi.org/10.1097/AUD.0000000000000289

Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. (2018). Artificial intelligence in radiology. *Nature Reviews Cancer*, *18*(8), 500–510. https://doi.org/10.1038/s41568-018-0016-5

Huang, N., Slaney, M., & Elhilali, M. (2018). Connecting Deep Neural Networks to Physical, Perceptual, and Electrophysiological Auditory Signals. *Frontiers in Neuroscience*, *12*. https://doi.org/10.3389/fnins.2018.00532

Huang, W.-C., Hayashi, T., Wu, Y.-C., Kameoka, H., & Toda, T. (2019). Voice Transformer Network: Sequence-to-Sequence Voice Conversion Using Transformer with Text-to-Speech Pretraining. *arXiv:1912.06813 [Cs, Eess]*. http://arxiv.org/abs/1912.06813

Individuals' Right under HIPAA to Access Their Health Information (2016). https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/access/index.html

Initiative (WAI), W. W. A. (2021). *Home*. Web Accessibility Initiative (WAI). https://www.w3.org/WAI/

International Organization for Standardization. (2010). *ISO 8253-1: 2010. Acoustics: Audiometric test methods. Part 1: Pure-tone air and bone conduction audiometry*. International Organization for Standardization Geneva.

Jacobs, P. G., Silaski, G., Wilmington, D., Gordon, S., Helt, W., McMillan, G., Fausti, S. A., & Dille, M. (2012). Development and Evaluation of a Portable Audiometer for High-Frequency Screening of Hearing Loss From Ototoxicity in Homes/Clinics. *IEEE Transactions on Biomedical Engineering*, *59*(11), 3097–3103. IEEE Transactions on Biomedical Engineering. https://doi.org/10.1109/TBME.2012.2204881

Jakovljević, N., Janev, M., Pekar, D., & Mišković, D. (2008). Energy normalization in automatic speech recognition. *International Conference on Text, Speech and Dialogue*, 341–347.

Jamalova, M., & Milán, C. (2019). The Comparative Study of the Relationship Between Smartphone Choice and Socio-Economic Indicators. *International Journal of Marketing Studies*, *11*(3), Article 3. https://doi.org/10.5539/ijms.v11n3p11

Jedrzejczak, W. W., & Kochanek, K. (2023). *Comparison of the audiological knowledge of three chatbots – ChatGPT, Bing Chat, and Bard* (p. 2023.11.22.23298893). medRxiv. https://doi.org/10.1101/2023.11.22.23298893

Jedrzejczak, W. W., Skarzynski, P. H., Raj-Koziak, D., Sanfins, M. D., Hatzopoulos, S., & Kochanek, K. (2023). *ChatGPT for tinnitus information and support: Response accuracy and retest after three months* (p. 2023.12.19.23300189). medRxiv. https://doi.org/10.1101/2023.12.19.23300189

Jerger, J., Chmiel, R., Florin, E., Pirozzolo, F., & Wilson, N. (1996). Comparison of Conventional Amplification and an Assistive Listening Device in Elderly Persons. *Ear and Hearing*, *17*(6), 490. https://journals.lww.com/ear-hearing/fulltext/1996/12000/comparison_of_conventional_amplification_and_an.5.aspx?casa_token=TBURrYRoAjYAAAAA:dudTpZDZ7xn72NDGDwAsM9Dx19TaqfMhgehx7r1XIjlGKCxwvSUTP0-mhV-aa5svX2x1715v-hWqZvvTzGY5

Jin, Y., Choi, S., Gao, Y., Li, J., Li, Z., & Jin, Z. (2023). TransASL: A Smart Glass based Comprehensive ASL Recognizer in Daily Life. *Proceedings of the 28th International Conference on Intelligent User Interfaces*, 802–818. https://doi.org/10.1145/3581641.3584071

Johansen, B., Petersen, M. K., Korzepa, M. J., Larsen, J., Pontoppidan, N. H., & Larsen, J. E. (2018). Personalizing the fitting of hearing aids by learning contextual preferences from internet of things data. *Computers*, *7*(1), 1. https://doi.org/10.3390/computers7010001

Jonsson, P., Carson, S., Blennerud, G., Shim, J., Arendse, B., Husseini, A., Lindberg, P., & Öhman, K. (2019). *Ericsson Mobility Report November 2019* (Mobility Reports, p. 36). Ericsson. https://www.ericsson.com/en/mobility-report/reports/november-2019

Juang, B.-H., & Rabiner, L. R. (2005). Automatic speech recognition–a brief history of the technology development. *Georgia Institute of Technology. Atlanta Rutgers University and the University of California. Santa Barbara*, *1*(67), 1.

Jurafsky, D., & Martin, J. H. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (2nd ed.). Pearson Prentice Hall.

Kaandorp, M. W., De Groot, A. M., Festen, J. M., Smits, C., & Goverts, S. T. (2016). The influence of lexical-access ability and vocabulary knowledge on measures of speech recognition in noise. *International Journal of Audiology*, *55*(3), 157–167.

Kaandorp, M. W., Smits, C., Merkus, P., Goverts, S. T., & Festen, J. M. (2015). Assessing speech recognition abilities with digits in noise in cochlear implant and hearing aid users. *International Journal of Audiology*, *54*(1), 48–57. https://doi.org/10.3109/14992027.2014.945623

Kader, S. E., Eckert, A. M., & Gural-Toth, V. (2021). Voice-to-Text Technology for Patients with Hearing Loss. *The Hearing Journal*, *74*(2), 11–14.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kamenov, K., Martinez, R., Kunjumen, T., & Chadha, S. (2021). Ear and Hearing Care Workforce: Current Status and its Implications. *Ear and Hearing*, *42*(2), 249. https://doi.org/10.1097/AUD.0000000000001007

Kasturi, K., Loizou, P. C., Dorman, M., & Spahr, T. (2002). The intelligibility of speech with "holes" in the spectrum. *The Journal of the Acoustical Society of America*, *112*(3), 1102–1111. https://doi.org/10.1121/1.1498855

Kell, A. J. E., Yamins, D. L. K., Shook, E. N., Norman-Haignere, S. V., & McDermott, J. H. (2018). A Task-Optimized Neural Network Replicates Human Auditory Behavior, Predicts Brain Responses, and Reveals a Cortical Processing Hierarchy. *Neuron*, *98*(3), 630-644.e16. https://doi.org/10.1016/j.neuron.2018.03.044

Kell, A. J., & McDermott, J. H. (2019). Deep neural network models of sensory systems: Windows onto the role of task constraints. *Current Opinion in Neurobiology*, *55*, 121–132. https://doi.org/10.1016/j.conb.2019.02.003

Kelly, E. A., Stadler, M. E., Nelson, S., Runge, C. L., Friedl, & D. R. (2018). Tablet-based Screening for Hearing Loss: Feasibility of Testing in Nonspecialty Locations. *Otol Neurotol*, *39*(4), 410–416.

Khoza-Shangase, K., & Kassner, L. (2013). Automated screening audiometry in the digital age: Exploring uhear™ and its use in a resource-stricken developing country. *International Journal of Technology Assessment in Health Care*, *29*(1), 42–47. https://doi.org/10.1017/S0266462312000761

Kim, S., Arora, A., Le, D., Yeh, C.-F., Fuegen, C., Kalinli, O., & Seltzer, M. L. (2021). Semantic Distance: A New Metric for ASR Performance Analysis Towards Spoken Language Understanding. *arXiv Preprint arXiv:2104.02138*.

Kincaid, J. (2018, September 9). Which Automatic Transcription Service is the Most Accurate? — 2018. *Descript*. https://medium.com/descript/which-automatic-transcription-service-is-the-most-accurate-2018-2e859b23ed19

Knecht, H. A., Nelson, P. B., Whitelaw, G. M., & Feth, L. L. (2002). *Background noise levels and reverberation times in unoccupied classrooms*.

Kocielnik, R., Agapie, E., Argyle, A., Hsieh, D. T., Yadav, K., Taira, B., & Hsieh, G. (2020). HarborBot: A Chatbot for Social Needs Screening. *AMIA Annual Symposium Proceedings*, *2019*, 552–561. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7153089/

Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., Toups, C., Rickford, J. R., Jurafsky, D., & Goel, S. (2020). Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(14), 7684–7689. https://doi.org/10.1073/pnas.1915768117

Kollmeier, B., & Kiessling, J. (2018). Functionality of hearing aids: State-of-the-art and future model-based solutions. *International Journal of Audiology*, *57*(sup3), S3–S28. https://doi.org/10.1080/14992027.2016.1256504

Konečný, J., McMahan, H. B., Ramage, D., & Richtárik, P. (2016). Federated Optimization: Distributed Machine Learning for On-Device Intelligence. *arXiv:1610.02527 [Cs]*. http://arxiv.org/abs/1610.02527

Kramer, S. E., Kapteyn, T. S., & Houtgast, T. (2006). Occupational performance: Comparing normally-hearing and hearing-impaired employees using the Amsterdam Checklist for Hearing and Work. *International Journal of Audiology*, *45*(9), 503–512. https://doi.org/10.1080/14992020600754583

Kropp, M. H., Hocke, T., Agha-Mir-Salim, P., & Müller, A. (2021). Evaluation of a synthetic version of the digits-in-noise test and its characteristics in CI recipients. *International Journal of Audiology*, *60*(7), 507–513. https://doi.org/10.1080/14992027.2020.1839678

Kruk, M. E., Gage, A. D., Joseph, N. T., Danaei, G., García-Saisó, S., & Salomon, J. A. (2018). Mortality due to low-quality health systems in the universal health coverage era: A systematic analysis of amenable deaths in 137 countries. *The Lancet*, *392*(10160), 2203–2212. https://doi.org/10.1016/S0140-6736(18)31668-4

Kung, B., Kunda, L., Groff, S., Miele, E., Loyd, M., & Carpenter, D. M. (2021). Validation Study of Kids Hearing Game: A Self-Administered Pediatric Audiology Application. *The Permanente Journal*, *25*. https://doi.org/10.7812/TPP/20.157

Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., & Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digital Health*, *2*(2), 1–12. https://doi.org/10.1371/journal.pdig.0000198

Lai, W. K., Dillier, N., Weber, B. P., Lenarz, T., Battmer, R., Gantz, B., Brown, C., Cohen, N., Waltzman, S., Skinner, M., Holden, L., Cowan, R., Busby, P., & Killian, M. (2009). TNRT profiles with the nucleus research platform 8 system. *International Journal of Audiology*, *48*(9), 645–654. https://doi.org/10.1080/14992020902962413

Lai, Y.-H., Tsao, Y., Lu, X., Chen, F., Su, Y.-T., Chen, K.-C., Chen, Y.-H., Chen, L.-C., Po-Hung Li, L., & Lee, C.-H. (2018). Deep Learning–Based Noise Reduction Approach to Improve Speech Intelligibility for Cochlear Implant Recipients. *Ear and Hearing*, *39*(4), 795. https://doi.org/10.1097/AUD.0000000000000537

Leese, M. (2014). The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union. *Security Dialogue*, *45*(5), 494–511. https://doi.org/10.1177/0967010614544204

Lehne, M., Sass, J., Essenwanger, A., Schepers, J., & Thun, S. (2019). Why digital medicine depends on interoperability. *Npj Digital Medicine*, *2*(1), 1–5. https://doi.org/10.1038/s41746-019-0158-1

Lesica, N. A., Mehta, N., Manjaly, J. G., Deng, L., Wilson, B. S., & Zeng, F.-G. (2021). Harnessing the power of artificial intelligence to transform hearing healthcare and research. *Nature Machine Intelligence*, *3*(10), 840–849. https://doi.org/10.1038/s42256-021-00394-z

Lieu, J. E. C., Kenna, M., Anne, S., & Davidson, L. (2020). Hearing Loss in Children: A Review. *JAMA*, *324*(21), 2195–2205. https://doi.org/10.1001/jama.2020.17647

Liu, Y., Yang, D., Xiong, F., Yu, L., Ji, F., & Wang, Q. J. (2015). Development and Validation of a Portable Hearing Self-Testing System Based on a Notebook Personal Computer. *J Am Acad Audiol*, *26*(8), 716–723.

Livingston, G., Sommerlad, A., Orgeta, V., Costafreda, S. G., Huntley, J., Ames, D., Ballard, C., Banerjee, S., Burns, A., & Cohen-Mansfield, J. (2017). Dementia prevention, intervention, and care. *The Lancet*, *390*(10113), 2673–2734. https://doi.org/10.1016/S0140-6736(17)31363-6

Loizides, F., Basson, S., Kanevsky, D., Prilepova, O., Savla, S., & Zaraysky, S. (2020). Breaking Boundaries with Live Transcribe: Expanding Use Cases Beyond Standard Captioning Scenarios. *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, 1–6. https://doi.org/10.1145/3373625.3417300

Lüdecke, D. (2018, May 18). *Plotting Estimates (Fixed Effects) of Regression Models*. Plotting Estimates (Fixed Effects) of Regression Models. http://strengejacke.de/sjPlot/articles/plot_model_estimates.html

Lyon, R. F. (2017). *Human and Machine Hearing*. Cambridge University Press.

Maclennan-Smith, F., Swanepoel, D. W., & Hall, J. W. (2013). Validity of diagnostic pure-tone audiometry without a sound-treated environment in older adults. *International Journal of Audiology*, *52*(2), 66–73. https://doi.org/10.3109/14992027.2012.736692

Maddox, T. M., Rumsfeld, J. S., & Payne, P. R. O. (2019). Questions for Artificial Intelligence in Health Care. *JAMA*, *321*(1), 31. https://doi.org/10.1001/jama.2018.18932

Magro, I., Clavier, O., Mojica, K., Rieke, C., Eisen, E., Fried, D., Stein-Meyers, A., Fellows, A., Buckey, J., & Saunders, J. (2020). Reliability of Tablet-based Hearing Testing in Nicaraguan Schoolchildren: A Detailed Analysis. *Otology & Neurotology*, *41*(3), 299–307. https://doi.org/10.1097/MAO.0000000000002534

Maharani, A., Dawes, P., Nazroo, J., Tampubolon, G., & Pendleton, N. (2018). Longitudinal Relationship Between Hearing Aid Use and Cognitive Function in Older Americans. *Journal of the American Geriatrics Society*, *66*(6), 1130–1136. https://doi.org/10.1111/jgs.15363

Mahomed, F., Swanepoel, D. W., Eikelboom, R. H., & Soer, M. (2013). Validity of Automated Threshold Audiometry: A Systematic Review and Meta-Analysis. *Ear and Hearing*, *34*(6), 745–752. https://doi.org/10.1097/01.aud.0000436255.53747.a4

Manchaiah, V., Eikelboom, R. H., Bennett, R. J., & Swanepoel, D. W. (2021). International survey of audiologists during the COVID-19 pandemic: Effects on the workplace. *International Journal of Audiology*, 1–8.

Manchaiah, V., Sharma, A., Rodrigo, H., Bailey, A., De Sousa, K. C., & Swanepoel, D. W. (2023). Hearing Healthcare Professionals' Views about Over-The-Counter (OTC) Hearing Aids: Analysis of Retrospective Survey Data. *Audiology Research*, *13*(2), Article 2. https://doi.org/10.3390/audiolres13020018

Manganella, J. L., Stiles, D. J., Kawai, K., Barrett, D. L., O'Brien, L. B., & Kenna, M. A. (2018). Validation of a portable hearing assessment tool: Agilis Health Mobile Audiogram. *Int J Pediatr Otorhinolaryngol*, *113*, 94–98.

Margolis, R. H., Bratt, G., Feeney, M. P., Killion, M. C., & Saly, G. L. (2018). Home Hearing Test: Within-Subjects Threshold Variability. *Ear and Hearing*, *39*(5), 906–909. https://doi.org/10.1097/AUD.0000000000000551

Margolis, R. H., Frisina, R., & Walton, J. P. (2011). AMTAS®: Automated method for testing auditory sensitivity: II. Air conduction audiograms in children and adults. *International Journal of Audiology*, *50*(7), 434–439.

Margolis, R. H., Glasberg, B. R., Creeke, S., & Moore, B. C. J. (2010). AMTAS: Automated method for testing auditory sensitivity: validation studies. *International Journal of Audiology*, *49*(3), 185–194. https://doi.org/10.3109/14992020903092608

Margolis, R. H., Killion, M. C., Bratt, G. W., & Saly, G. L. (2016). Validation of the Home Hearing Test™. *J Am Acad Audiol*, *27*(5), 416–420.

Margolis, R. H., & Moore, B. C. (2011). AMTAS®: Automated method for testing auditory sensitivity: III. Sensorineural hearing loss and air-bone gaps. *International Journal of Audiology*, *50*(7), 440–447.

Margolis, R. H., & Saly, G. L. (2008). Asymmetric hearing loss: Definition, validation, and prevalence. *Otology & Neurotology*, *29*(4), 422–431.

Margolis, R. H., Saly, G. L., Le, C., & Laurence, J. (2007). Qualind: A method for assessing the accuracy of automated tests. *Journal of the American Academy of Audiology*, *18*(1), 78–89. https://doi.org/10.3766/jaaa.18.1.7

Maruthurkkara, S., Allen, A., Cullington, H., Muff, J., Arora, K., & Johnson, S. (2021). Remote check test battery for cochlear implant recipients: Proof of concept study. *International Journal of Audiology*, *0*(0), 1–10. https://doi.org/10.1080/14992027.2021.1922767

Maruthurkkara, S., Case, S., & Rottier, R. (2022). Evaluation of Remote Check: A Clinical Tool for Asynchronous Monitoring and Triage of Cochlear Implant Recipients. Ear and Hearing, 43(2), 495–506. https://doi.org/10.1097/AUD.0000000000001106

Masalski, M., Grysiński, T., & Kręcicki, T. (2018). Hearing tests based on biologically calibrated mobile devices: Comparison with pure-tone audiometry. *JMIR mHealth and uHealth*, *6*(1), e10.

Masalski, M., Kipiński, L., Grysiński, T., & Kręcicki, T. (2016). Hearing tests on mobile devices: Evaluation of the reference sound level by means of biological calibration. *Journal of Medical Internet Research*, *18*(5), e130.

Masalski, M., & Kręcicki, T. (2013). Self-test web-based pure-tone audiometry: Validity evaluation and measurement error analysis. *Journal of Medical Internet Research*, *15*(4), e71. https://doi.org/10.2196/jmir.2222

Mashmous, M. H. A. B. (2022). Efficacy of Remote Hearing Aids Programming Using Teleaudiology: A Systematic Review. *E-Health Telecommunication Systems and Networks*, *11*(1), Article 1. https://doi.org/10.4236/etsn.2022.111002

Masterson, E. A., Tak, S., Themann, C. L., Wall, D. K., Groenewold, M. R., Deddens, J. A., & Calvert, G. M. (2013). Prevalence of hearing loss in the United States by industry. *American Journal of Industrial Medicine*, *56*(6), 670–681. https://doi.org/10.1002/ajim.22082

Matos, H., Araujo, E., Agostinho, R., Wasmann, J.-W. A., Morata, T., Mondelli, M., Alvarenga, K. de F., & Jacob, L. (2024). *Improving the delivery of hearing care information through GPT-4 enhanced by Wikipedia*. 36th World Congress of Audiology., Paris.

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*(7–8), 953–978.

McDaid, D., Park, A.-L., & Chadha, S. (2021). Estimating the global costs of hearing loss. *International Journal of Audiology*, *60*(3), 162–170. https://doi.org/10.1080/14992027.2021.1883197

McRackan, T. R., Bauschard, M., Hatch, J. L., Franko-Tobin, E., Droghini, H. R., Nguyen, S. A., & Dubno, J. R. (2018). Meta-analysis of quality-of-life improvement after cochlear implantation and associations with speech recognition abilities. *The Laryngoscope*, *128*(4), 982–990. https://doi.org/10.1002/lary.26738

Meeuws, M., Pascoal, D., Bermejo, I., Artaso, M., Ceulaer, G. D., & Govaerts, P. J. (2017). Computer-assisted CI fitting: Is the learning capacity of the intelligent agent FOX beneficial for speech understanding? *Cochlear Implants International*, *18*(4), 198–206. https://doi.org/10.1080/14670100.2017.1325093

Meinke, D. K., Norris, J. A., Flynn, B. P., & Clavier, O. H. (2017). Going wireless and booth-less for hearing testing in industry. *Int J Audiol*, *56*, 41–51.

Mellor, J., Stone, M. A., & Keane, J. (2018). Application of Data Mining to a Large Hearing-Aid Manufacturer's Dataset to Identify Possible Benefits for Clinicians, Manufacturers, and Users. *Trends in Hearing*, *22*, 2331216518773632. https://doi.org/10.1177/2331216518773632

Meng, Q., Chen, J., Zhang, C., Wasmann, J.-W. A., Barbour, D. L., & Zeng, F.-G. (2022). Editorial: Digital hearing healthcare. *Frontiers in Digital Health*, *4*. https://www.frontiersin.org/articles/10.3389/fdgth.2022.959761

Merikanto, I., Kronholm, E., Peltonen, M., Laatikainen, T., Lahti, T., & Partonen, T. (2012). Relation of Chronotype to Sleep Complaints in the General Finnish Population. *Chronobiology International*, *29*(3), 311–317. https://doi.org/10.3109/07420528.2012.655870

Metz, C. (2018, November 26). Efforts to Acknowledge the Risks of New A.I. Technology. *The New York Times*. https://www.nytimes.com/2018/10/22/business/efforts-to-acknowledge-the-risks-of-new-ai-technology.html

Meyer, L., Rachman, L., Araiza-Illan, G., Gaudrain, E., & Başkent, D. (2023). Use of a humanoid robot for auditory psychophysical testing. *PLOS ONE*, *18*(12), e0294328. https://doi.org/10.1371/journal.pone.0294328

Migliorini, E., Wasmann, J.-W. A., Philpott, N., Dijk, B. van, Philips, B., & Huinck, W. (2024). *Comparing Phoneme and Word Recognition Test Outcomes in Adult CI users: Data Analysis from the AuDieT Study* (p. 2024.03.28.24304843). medRxiv. https://doi.org/10.1101/2024.03.28.24304843

Miller, G. A., & Nicely, P. E. (1955). An Analysis of Perceptual Confusions Among Some English Consonants. *The Journal of the Acoustical Society of America*, *27*(2), 338–352. https://doi.org/10.1121/1.1907526

Miner, A. S., Haque, A., Fries, J. A., Fleming, S. L., Wilfley, D. E., Wilson, G. T., Milstein, A., Jurafsky, D., Arnow, B. A., & Agras, W. S. (2020). Assessing the accuracy of automatic speech recognition for psychotherapy. *NPJ Digital Medicine*, *3*(1), 1–8.

Monk, T., Buysse, D., Reynolds Iii, C., Berga, S., Jarrett, D., Begley, A., & Kupfer, D. (1997). Circadian rhythms in human performance and mood under constant conditions. *Journal of Sleep Research*, *6*(1), 9–18. https://doi.org/10.1046/j.1365-2869.1997.00023.x

Moore, D. R., Zobay, O., & Ferguson, M. A. (2020). Minimal and Mild Hearing Loss in Children: Association with Auditory Perception, Cognition, and Communication Problems. *Ear and Hearing*, *Publish Ahead of Print*. https://doi.org/10.1097/AUD.0000000000000802

Mościcki, E. K., Elkins, E. F., Baum, H. M., & McNamara, P. M. (1985). Hearing loss in the elderly: An epidemiologic study of the Framingham Heart Study Cohort. *Ear and Hearing*, *6*(4), 184–190.

Mosley, C. L., Langley, L. M., Davis, A., McMahon, C. M., & Tremblay, K. L. (2019). Reliability of the Home Hearing Test: Implications for Public Health. *Journal of the American Academy of Audiology*, *30*(3), 208–216. https://doi.org/10.3766/jaaa.17092

Motlagh Zadeh, L., Silbert, N. H., Sternasty, K., Swanepoel, D. W., Hunter, L. L., & Moore, D. R. (2019). Extended high-frequency hearing enhances speech perception in noise. *Proceedings of the National Academy of Sciences*, *116*(47), 23753–23759. https://doi.org/10.1073/pnas.1903315116

Myburgh, H. C., Jose, S., Swanepoel, D. W., & Laurent, C. (2018). Towards low cost automated smartphone- and cloud-based otitis media diagnosis. *Biomedical Signal Processing and Control*, *39*, 34–52. https://doi.org/10.1016/j.bspc.2017.07.015

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Nielsen, J. B. B., Nielsen, J., & Larsen, J. (2014). Perception-based personalization of hearing aids using Gaussian processes and active learning. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *23*(1), 162–173. https://doi.org/10.1109/TASLP.2014.2377581

Nosek, B. A., & Lakens, D. (2014). Registered Reports. *Social Psychology*, *45*(3), 137–141. https://doi.org/10.1027/1864-9335/a000192

Oba, S. I., Fu, Q.-J., & Galvin, J. J. I. (2011). Digit Training in Noise Can Improve Cochlear Implant Users' Speech Understanding in Noise. In *Ear and Hearing* (Vol. 32, Issue 5, pp. 573–581). https://journals.lww.com/ear-hearing/Fulltext/2011/09000/Digit_Training_in_Noise_Can_Improve_Cochlear.4.aspx

O'Brien, E. (2020, May 14). The critical role of computing infrastructure in computational audiology. *Computational Audiology*. https://computationalaudiology.com/the-critical-role-of-computing-infrastructure-in-computational-audiology/

OECD. (2024). *Old-age dependency ratio (indicator)* [Dataset]. OECD. https://doi.org/10.1787/e0255c98-en

Office of the Director, Defense Research and Engineering (DDR&E). (2009). *Appendix E. Biomedical Technology Readiness Levels (TRLs). Technology Readiness Assessment (TRA) Deskbook* (Appendix E., p. 130). https://apps.dtic.mil/sti/pdfs/ADA524200.pdf

Olusanya, B. O., Neumann, K. J., & Saunders, J. E. (2014). The global burden of disabling hearing impairment: A call to action. *Bulletin of the World Health Organization*, *92*(5), 367–373. https://doi.org/10.2471/BLT.13.128728

O'Neill, E. R., Parke, M. N., Kreft, H. A., & Oxenham, A. J. (2020). Development and validation of sentences without semantic context to complement the basic English lexicon sentences. *Journal of Speech, Language, and Hearing Research*, *63*(11), 3847–3854.

Opstal, A. J., & Noordanus, E. (2023). Towards personalized and optimized fitting of cochlear implants. *Front Neurosci*, *17*. https://doi.org/10.3389/FNINS.2023.1183126

Oremule, B., Saunders, G. H., Kulk, K., d'Elia, A., & Bruce, I. A. (2024). Understanding, experience, and attitudes towards artificial intelligence technologies for clinical decision support in hearing health: A mixed-methods survey of healthcare professionals in the UK. *The Journal of Laryngology & Otology*, 1–31. https://doi.org/10.1017/S0022215124000550

Orji, A., Kamenov, K., Dirac, M., Davis, A., Chadha, S., & Vos, T. (2020). Global and regional needs, unmet needs and access to hearing aids. *International Journal of Audiology*, *59*(3), 166–172. https://doi.org/10.1080/14992027.2020.1721577

Palacios, G., Noreña, A., & Londero, A. (2020). Assessing the Heterogeneity of Complaints Related to Tinnitus and Hyperacusis from an Unsupervised Machine Learning Approach: An Exploratory Study. *Audiology and Neurotology*, *25*(4), 173–188. https://doi.org/10.1159/000504741

Panayotov, V., Chen, G., Povey, D., & Khudanpur, S. (2015). Librispeech: An ASR corpus based on public domain audio books. *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5206–5210. https://doi.org/10.1109/ICASSP.2015.7178964

Parmar, B., Beukes, E., & Rajasingam, S. (2022). The impact of COVID-19 on provision of UK audiology services & on attitudes towards delivery of telehealth services. *International Journal of Audiology*, *61*(3), 228–238. https://doi.org/10.1080/14992027.2021.1921292

Patel, K., Thibodeau, L., McCullough, D., Freeman, E., & Panahi, I. (2021). Development and Pilot Testing of Smartphone-Based Hearing Test Application. *International Journal of Environmental Research and Public Health*, *18*(11). https://doi.org/10.3390/ijerph18115529

Philipsen, R. H. H. M., Sánchez, C. I., Melendez, J., Lew, W. J., & van Ginneken, B. (2019). Automated chest X-ray reading for tuberculosis in the Philippines to improve case detection: A cohort study. *The International Journal of Tuberculosis and Lung Disease: The Official Journal of the International Union Against Tuberculosis and Lung Disease*, *23*(7), 805–810. https://doi.org/10.5588/ijtld.18.0004

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). *Ear and Hearing*, *37 Suppl 1*, 5S-27S. https://doi.org/10.1097/AUD.0000000000000312

Plomp, R., & Mimpen, A. M. (1979a). Improving the reliability of testing the speech reception threshold for sentences. *Audiology*, *18*(1), 43–52.

Plomp, R., & Mimpen, A. M. (1979b). Speech-reception threshold for sentences as a function of age and noise level. *The Journal of the Acoustical Society of America*, *66*(5), 1333–1342. https://doi.org/10.1121/1.383554

Poling, G. L., Kunnel, T. J., & Dhar, S. (2016). Comparing the Accuracy and Speed of Manual and Tracking Methods of Measuring Hearing Thresholds. *Ear Hear*, *37*(5), e336-40.

Potgieter, J.-M., Swanepoel, D. W., & Smits, C. (2018). Evaluating a smartphone digits-in-noise test as part of the audiometric test battery. *The South African Journal of Communication Disorders = Die Suid-Afrikaanse Tydskrif Vir Kommunikasieafwykings*, *65*(1), e1–e6. https://doi.org/10.4102/sajcd.v65i1.574

Pragt, L., van Hengel, P., Grob, D., & Wasmann, J.W. (2020). *Speech recognition apps for the hearing impaired and deaf*. Proceedings of the VCCA2020 conference. https://computationalaudiology.com/ai-speech-recognition-apps-for-hearing-impaired-and-deaf/

Pragt, L., van Hengel, P., Grob, D., & Wasmann, J.-W. A. (2022). Preliminary Evaluation of Automated Speech Recognition Apps for the Hearing Impaired and Deaf. *Frontiers in Digital Health*, *4*, 806076. https://doi.org/10.3389/fdgth.2022.806076

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257–286. Proceedings of the IEEE. https://doi.org/10.1109/5.18626

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine Learning in Medicine. *New England Journal of Medicine*, *380*(14), 1347–1358. https://doi.org/10.1056/NEJMra1814259

Ratanjee-Vanmali, H., Swanepoel, D. W., & Laplante-Lévesque, A. (2020). Patient uptake, experience, and satisfaction using web-based and face-to-face hearing health services: Process evaluation study. *Journal of Medical Internet Research*, *22*(3), e15875.

Ratnanather, J., Wang, L., Bae, S.-H., O'Neill, E., Sagi, E., & Tward, D. (2021). Visualization of Speech Perception Analysis via Phoneme Alignment: A Pilot Study. *Front. Neurol.*, *12:724800*. https://doi.org/10.3389/fneur.2021.724800

Ravi, R., Gunjawate, D. R., Yerraguntla, K., & Driscoll, C. (2018). Knowledge and perceptions of teleaudiology among audiologists: A systematic review. *Journal of Audiology & Otology*, *22*(3), 120.

Regulation (EU) 2016/679, 88 (2016). https://eur-lex.europa.eu/eli/reg/2016/679/oj?eliuri=eli%3Areg%3A2016%3A679%3Aoj

Remus, J. J., Throckmorton, C. S., & Collins, L. M. (2007). Expediting the Identification of Impaired Channels in Cochlear Implants Via Analysis of Speech-Based Confusion Matrices. *IEEE Transactions on Biomedical Engineering*, *54*(12), 2193–2204. IEEE Transactions on Biomedical Engineering. https://doi.org/10.1109/TBME.2007.908336

Rocher, L., Hendrickx, J. M., & Montjoye, Y.-A. de. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications*, *10*(1), 1–9. https://doi.org/10.1038/s41467-019-10933-3

Rodrigues, L. C., Ferrite, S., & Corona, A. P. (2020). Validity of hearTest Smartphone-Based Audiometry for Hearing Screening in Workers Exposed to Noise. *Journal of the American Academy of Audiology*. https://doi.org/10.1055/s-0040-1718931

Rubin, M. (2017). Do p Values Lose Their Meaning in Exploratory Analyses? It Depends How You Define the Familywise Error Rate. *Review of General Psychology*, *21*, 269–275. https://doi.org/10.1037/gpr0000123

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, *115*(3), 211–252. https://doi.org/10.1007/s11263-015-0816-y

Russell, L. B., Ibuka, Y., & Carr, D. (2008). How Much Time Do Patients Spend on Outpatient Visits? *The Patient: Patient-Centered Outcomes Research*, *1*(3), 211–222. https://doi.org/10.2165/1312067-200801030-00008

Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: A modern approach*. Pearson.

Rzayev, R., Korbely, S., Maul, M., Schark, A., Schwind, V., & Henze, N. (2020). Effects of Position and Alignment of Notifications on AR Glasses during Social Interaction. *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, 1–11. https://doi.org/10.1145/3419249.3420095

Saak, S., Huelsmeier, D., Kollmeier, B., & Buhl, M. (2022). A flexible data-driven audiological patient stratification method for deriving auditory profiles. *Frontiers in Neurology*, *13*. https://doi.org/10.3389/fneur.2022.959582

Saliba, J., Al-Reefi, M., Carriere, J. S., Verma, N., Provencal, C., & Rappaport, J. M. (2017). Accuracy of Mobile-Based Audiometry in the Evaluation of Hearing Loss in Quiet and Noisy Environments. *Otolaryngology–Head and Neck Surgery*, *156*(4), 706–711. https://doi.org/10.1177/0194599816683663

Sanchez Lopez, R., Bianchi, F., Fereczkowski, M., Santurette, S., & Dau, T. (2018). Data-driven approach for auditory profiling and characterization of individual hearing loss. *Trends in Hearing*, *22*, 2331216518807400. https://doi.org/10.1177/2331216518807400

Sandström, J., Swanepoel, D. W., Carel Myburgh, H., & Laurent, C. (2016). Smartphone threshold audiometry in underserved primary health-care contexts. *International Journal of Audiology*, *55*(4), 232–238.

Sandström, J., Swanepoel, D. W., Laurent, C., Umefjord, G., & Lundberg, T. (2020). Accuracy and Reliability of Smartphone Self-Test Audiometry in Community Clinics in Low Income Settings: A Comparative Study. *Annals of Otology, Rhinology & Laryngology*, *129*(6), 578–584. https://doi.org/10.1177/0003489420902162

Saon, G., Kurata, G., Sercu, T., Audhkhasi, K., Thomas, S., Dimitriadis, D., Cui, X., Ramabhadran, B., Picheny, M., Lim, L.-L., Roomi, B., & Hall, P. (2017). English Conversational Telephone Speech Recognition by Humans and Machines. *arXiv:1703.02136 [Cs]*. http://arxiv.org/abs/1703.02136

Saunders, G. H., Bott, A., & Tietz, L. H. (2020). Hearing care providers' perspectives on the utility of datalogging information. *American Journal of Audiology*, *29*(3S), 610–622. https://doi.org/10.1044/2020_AJA-19-00089

Saunders, G. H., & Roughley, A. (2021). Audiology in the time of COVID-19: Practices and opinions of audiologists in the UK. *International Journal of Audiology*, *60*(4), 255–262. https://doi.org/10.1080/14992027.2020.1814432

Schepers, K., Steinhoff, H.-J., Ebenhoch, H., Böck, K., Bauer, K., Rupprecht, L., Möltner, A., Morettini, S., & Hagen, R. (2019). Remote programming of cochlear implants in users of all ages. *Acta Oto-Laryngologica*, *139*(3), 251–257. https://doi.org/10.1080/00016489.2018.1554264

Schlittenlacher, J., Turner, R. E., & Moore, B. C. (2018a). A hearing-model-based active-learning test for the determination of dead regions. *Trends in Hearing*, *22*, 2331216518788215. https://doi.org/10.1177/2331216518788215

Schlittenlacher, J., Turner, R. E., & Moore, B. C. J. (2018b). Audiogram estimation using Bayesian active learning. *The Journal of the Acoustical Society of America*, *144*(1), 421–430. https://doi.org/10.1121/1.5047436

Schmidt, C., Collette, F., Cajochen, C., & Peigneux, P. (2007). A time to think: Circadian rhythms in human cognition. *Cognitive Neuropsychology*, *24*(7), 755–789. https://doi.org/10.1080/02643290701754158

Schmidt, F. H., Hocke, T., Zhang, L., Großmann, W., & Mlynski, R. (2024). Tone Decay Reconsidered: Preliminary Results of a Prospective Study in Hearing-Aid Users with Moderate to Severe Hearing Loss. *Journal of Clinical Medicine*, *13*(2), Article 2. https://doi.org/10.3390/jcm13020500

Schmidt, J. H., Brandt, C., Pedersen, E. R., Christensen-Dalsgaard, J., Andersen, T., Poulsen, T., & Bælum, J. (2014). A user-operated audiometry method based on the maximum likelihood principle and the two-alternative forced-choice paradigm. *International Journal of Audiology*, *53*(6), 383–391.

Schwab, K. (2016). *The Fourth Industrial Revolution: What it means and how to respond*. World Economic Forum. https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/

Sekhri, N., Feachem, R., & Ni, A. (2011). Public-private integrated partnerships demonstrate the potential to improve health care access, quality, and efficiency. *Health Affairs*, *30*(8), 1498–1507. https://doi.org/10.1377/hlthaff.2010.0461

Shang, X. (2023, May 11). *NALscribe: Live captions*. NAL. https://www.nal.gov.au/nalscribeapp/

Shield, B. (2019). *Hearing Loss –Numbers and Costs. Evaluation of the social and economic costs of hearing impairment* (pp. 1–249). Hear it AISBL. https://m.hear-it.org/sites/default/files/BS%20-%20report%20files/HearitReportHearingLossNumbersandCosts.pdf

Shillingford, B., Assael, Y., Hoffman, M. W., Paine, T., Hughes, C., Prabhu, U., Liao, H., Sak, H., Rao, K., Bennett, L., Mulville, M., Coppin, B., Laurie, B., Senior, A., & de Freitas, N. (2018). Large-Scale Visual Speech Recognition. *arXiv:1807.05162 [Cs]*. http://arxiv.org/abs/1807.05162

Shuren, J., & Califf, R. M. (2016). Need for a National Evaluation System for Health Technology. *JAMA*, *316*(11), 1153. https://doi.org/10.1001/jama.2016.8708

Siggaard, L. D., Jacobsen, H., Hougaard, D. D., & Hoegsbro, M. (2023). Effects of remote ear-nose-and-throat specialist assessment screening on self-reported hearing aid benefit and satisfaction. *International Journal of Audiology*, *0*(0), 1–10. https://doi.org/10.1080/14992027.2023.2298506

Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., Scales, N., Tanwani, A., Cole-Lewis, H., & Pfohl, S. (2023). Large language models encode clinical knowledge. *Nature*, *620*(7972), 172–180. https://doi.org/10.1038/s41586-023-06291-2

Sininger, Y. S., Hunter, L. L., Hayes, D., Roush, P. A., & Uhler, K. M. (2018). Evaluation of Speed and Accuracy of Next-Generation Auditory Steady State Response and Auditory Brainstem Response Audiometry in Children With Normal Hearing and Hearing Loss. *Ear and Hearing*, *39*(6), 1207. https://doi.org/10.1097/AUD.0000000000000580

Skinner, M. W., Holden, L. K., Holden, T. A., & Demorest, M. E. (1995). Comparison of Procedures for Obtaining Thresholds and Maximum Acceptable Loudness Levels With the Nucleus Cochlear Implant System. *Journal of Speech, Language, and Hearing Research*, *38*(3), 677–689. https://doi.org/10.1044/jshr.3803.677

Skinner, M. W., Holden, L. K., Whitford, L. A., Plant, K. L., Psarros, C., & Holden, T. A. (2002). Speech Recognition with the Nucleus 24 SPEAK, ACE, and CIS Speech Coding Strategies in Newly Implanted Adults. *Ear and Hearing*, *23*(3), 207. https://journals.lww.com/ear-hearing/fulltext/2002/06000/Speech_Perception_as_a_Function_of_Electrical.00005.aspx?casa_token=lDB1IJ7B1BAAAAAA:ZL3L2C1D_lL-7KTwnmElP-l_SHfLhAG6LTSu8lpDcgN5wQlCaD1wopArD7NXR0JA2YGAypdZU0SVa63iPArE

Slaney, M., Lyon, R. F., Garcia, R., Kemler, B., Gnegy, C., Wilson, K., Kanevsky, D., Savla, S., & Cerf, V. G. (2020). Auditory Measures for the Next Billion Users. *Ear and Hearing*, *41*, 131S. https://doi.org/10.1097/AUD.0000000000000955

Smith, M., Cunningham, K. T., & Haley, K. L. (2019). Automating error frequency analysis via the phonemic edit distance ratio. *Journal of Speech, Language, and Hearing Research*, *62*(6), 1719–1723.

Smits, C., Theo Goverts, S., & Festen, J. M. (2013a). The digits-in-noise test: Assessing auditory speech recognition abilities in noise. *The Journal of the Acoustical Society of America*, *133*(3), 1693–1706. https://doi.org/10.1121/1.4789933

Søgaard Jensen, N., Hau, O., Bagger Nielsen, J. B., Bundgaard Nielsen, T., & Vase Legarth, S. (2019). Perceptual effects of adjusting hearing-aid gain by means of a machine-learning approach based on individual user preference. *Trends in Hearing*, *23*, 2331216519847413. https://doi.org/10.1177/2331216519847413

Song, X. D., Wallace, B. M., Gardner, J. R., Ledbetter, N. M., Weinberger, K. Q., & Barbour, D. L. (2015). Fast, Continuous Audiogram Estimation Using Machine Learning. *Ear Hear*, *36*(6), e326-35.

Sooful, P., Simpson, A., Thornton, M., & Šarkić, B. (2023). The AI Revolution: Rethinking Assessment in Audiology Training Programs. *The Hearing Journal*, *76*(11), 000. https://doi.org/10.1097/01.HJ.0000995264.80206.87

Sorkin, D. L., & Buchman, C. A. (2023). Cochlear Implants Now More Accessible to Older Adults. *The Hearing Journal*, *76*(01), 21–22.

Stepanov, I. (2020). Introducing a property right over data in the EU: The data producer's right – an evaluation. *International Review of Law, Computers & Technology*, *34*(1), 65–86. https://doi.org/10.1080/13600869.2019.1631621

Stickney, G. S., Loizou, P. C., Mishra, L. N., Assmann, P. F., Shannon, R. V., & Opie, J. M. (2006). Effects of electrode design and configuration on channel interactions. *Hearing Research*, *211*(1–2), 33–45. https://doi.org/10.1016/j.heares.2005.08.008

Stokel-Walker, C. (2023). ChatGPT listed as author on research papers: Many scientists disapprove. *Nature*, *613*(7945), 620–621. https://doi.org/10.1038/d41586-023-00107-z

Storey, K. K., Muñoz, K., Nelson, L., Larsen, J., & White, K. (2014). Ambient noise impact on accuracy of automated hearing assessment. *International Journal of Audiology*, *53*(10), 730–736. https://doi.org/10.3109/14992027.2014.920110

Sun, C., Liu, Y., & Wang, X. (2019). *An Automated Hearing Test Equipment Based on Active Noise Control Technology* (rayyan-91470105). 1–5.

Swanepoel, D. W., & Biagio, L. (2011). Validity of Diagnostic Computer-Based Air and Forehead Bone Conduction Audiometry. *Journal of Occupational and Environmental Hygiene*, *8*(4), 210–214. https://doi.org/10.1080/15459624.2011.559417

Swanepoel, D. W., & Clark, J. L. (2019). Hearing healthcare in remote or resource-constrained environments. *The Journal of Laryngology & Otology*, *133*(1), 11–17. https://doi.org/10.1017/S0022215118001159

Swanepoel, D. W., De Sousa, K. C., Smits, C., & Moore, D. R. (2019). Mobile applications to detect hearing impairment: Opportunities and challenges. *Bulletin of the World Health Organization*, *97*(10), 717–718. https://doi.org/10.2471/BLT.18.227728

Swanepoel, D. W., & Hall, J. W. (2020). Making Audiology Work During COVID-19 and Beyond. *The Hearing Journal*, *73*(6), 20–22. https://doi.org/10.1097/01.HJ.0000669852.90548.75

Swanepoel, D. W., Manchaiah, V., & Wasmann, J.-W. A. (2023). The Rise of AI Chatbots in Hearing Health Care. *The Hearing Journal*, *76*(04), 26. https://doi.org/10.1097/01.HJ.0000927336.03567.3e

Swanepoel, D. W., Mngemane, S., Molemong, S., Mkwanazi, H., & Tutshini, S. (2010). Hearing assessment-reliability, accuracy, and efficiency of automated audiometry. *Telemedicine Journal and E-Health: The Official Journal of the American Telemedicine Association*, *16*(5), 557–563. https://doi.org/10.1089/tmj.2009.0143

Swanepoel, D. W., Myburgh, H. C., Howe, D. M., Mahomed, F., & Eikelboom, R. H. (2014). Smartphone hearing screening with integrated quality control and data management. *International Journal of Audiology*, *53*(12), 841–849. https://doi.org/10.3109/14992027.2014.920965

Swanepoel de, W., Matthysen, C., Eikelboom, R. H., Clark, J. L., & Hall, J. W., 3rd. (2015). Pure-tone audiometry outside a sound booth using earphone attentuation, integrated noise monitoring, and automation. *Int J Audiol*, *54*(11), 777–785.

Swanson, B., & Mauch, H. (2006). Nucleus Matlab Toolbox 4.20 software user manual. *Cochlear Ltd, Lane Cove NSW, Australia*.

Szatmari, T.-I., Petersen, M. K., Korzepa, M. J., & Giannetsos, T. (2020). Modelling Audiological Preferences using Federated Learning. *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, 187–190. https://doi.org/10.1145/3386392.3399560

Szudek, J., Ostevik, A., Dziegielewski, P., Robinson-Anagor, J., Gomaa, N., Hodgetts, B., & Ho, A. (2012). Can Uhear me now? Validation of an iPod-based hearing loss screening test. *Journal of Otolaryngology - Head & Neck Surgery = Le Journal D'oto-Rhino-Laryngologie Et De Chirurgie Cervico-Faciale*, *41 Suppl 1*, S78-84.

Szymański, P., Żelasko, P., Morzy, M., Szymczak, A., Żyła-Hoppe, M., Banaszczak, J., Augustyniak, L., Mizgajski, J., & Carmiel, Y. (2020). WER we are and WER we think we are. In T. Cohn, Y. He, & Y. Liu (Eds.), *Findings of the Association for Computational Linguistics: EMNLP 2020* (pp. 3290–3295). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.findings-emnlp.295

Taylor, K. I., Staunton, H., Lipsmeier, F., Nobbs, D., & Lindemann, M. (2020). Outcome measures based on digital health technology sensor data: Data- and patient-centric approaches. *Npj Digital Medicine*, *3*(1), Article 1. https://doi.org/10.1038/s41746-020-0305-8

Taylor, K., & Sheikh, W. (2022). Automated Hearing Impairment Diagnosis Using Machine Learning. *2022 Intermountain Engineering, Technology and Computing (IETC)*, 1–6. https://doi.org/10.1109/IETC54973.2022.9796707

Thabit, H., & Hovorka, R. (2016). Coming of age: The artificial pancreas for type 1 diabetes. *Diabetologia*, *59*(9), 1795–1805. https://doi.org/10.1007/s00125-016-4022-4

Thompson, G. P., Sladen, D. P., Borst, B. J. H., & Still, O. L. (2015). Accuracy of a Tablet Audiometer for Measuring Behavioral Hearing Thresholds in a Clinical Population. *Otolaryngology--Head and Neck Surgery: Official Journal of American Academy of Otolaryngology-Head and Neck Surgery*, *153*(5), 838–842. https://doi.org/10.1177/0194599815593737

Tsagris, M., & Pandis, N. (2021). Multicollinearity. *American Journal of Orthodontics and Dentofacial Orthopedics*, *159*(5), 695–696. https://doi.org/10.1016/j.ajodo.2021.02.005

Tsimpida, D., Kontopantelis, E., Ashcroft, D. M., & Panagioti, M. (2021). Conceptual Model of Hearing Health Inequalities (HHI Model): A Critical Interpretive Synthesis. *Trends in Hearing*, *25*, 23312165211002963. https://doi.org/10.1177/23312165211002963

United Nations. (2023). *GLOBAL SUSTAINABLE DEVELOPMENT REPORT*. https://sdgs.un.org/goals

Vaerenberg, B., Smits, C., De Ceulaer, G., Zir, E., Harman, S., Jaspers, N., Tam, Y., Dillon, M., Wesarg, T., Martin-Bonniot, D., Gärtner, L., Cozma, S., Kosaner, J., Prentiss, S., Sasidharan, P., Briaire, J. J., Bradley, J., Debruyne, J., Hollow, R., … Govaerts, P. J. (2014). Cochlear Implant Programming: A Global Survey on the State of the Art. *The Scientific World Journal*, *2014*, e501738. https://doi.org/10.1155/2014/501738

Valentino-DeVries, J. (2019, April 13). Tracking Phones, Google Is a Dragnet for the Police. *The New York Times*. https://www.nytimes.com/interactive/2019/04/13/us/google-location-tracking-police.html

Van Dis, E. A., Bollen, J., Zuidema, W., Van Rooij, R., & Bockting, C. L. (2023). ChatGPT: five priorities for research. *Nature*, *614*(7947), 224–226. https://doi.org/10.1038/d41586-023-00288-7

Van Tasell, D. J., & Folkeard, P. (2013). Reliability and accuracy of a method of adjustment for self-measurement of auditory thresholds. *Otol Neurotol*, *34*(1), 9–15.

van Tonder, J., Swanepoel, W., Mahomed-Asmail, F., Myburgh, H., & Eikelboom, R. H. (2017). Automated Smartphone Threshold Audiometry: Validity and Time Efficiency. *J Am Acad Audiol*, *28*(3), 200–208.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. ukasz, & Polosukhin, I. (2017). Attention is All you Need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Venail, F., Legris, E., Vaerenberg, B., Puel, J.-L., Govaerts, P. J., & Ceccato, J. C. (2016). Validation of the French-language version of the OTOSPEECH automated scoring software package for speech audiometry. *European Annals of Otorhinolaryngology, Head and Neck Diseases*, *133*(2), 101–106. https://doi.org/10.1016/j.anorl.2016.01.001

Veneman, C. E., Gordon-Salant, S., Matthews, L. J., & Dubno, J. R. (2013). Age and measurement time-of-day effects on speech recognition in noise. *Ear and Hearing*, *34*(3), 288–299. https://doi.org/10.1097/AUD.0b013e31826d0b81

Vercammen, C., & Buhl, M. (2024). Special session: Big data and data standards in audiology. *Virtual Conference on Computational Audiology (VCCA 2024), Date: 2024/06/20-2024/06/21, Location: Virtual*. VCCA2024. https://computationalaudiology.com/big-data-and-data-standards-in-audiology/

Vercoulen, J. H., Swanink, C. M., Fennis, J. F., Galama, J. M., van der Meer, J. W., & Bleijenberg, G. (1994). Dimensional assessment of chronic fatigue syndrome. *Journal of Psychosomatic Research*, *38*(5), 383–392. https://doi.org/10.1016/0022-3999(94)90099-x

Verhulst, S., Altoè, A., & Vasilkov, V. (2018). Computational modeling of the human auditory periphery: Auditory-nerve responses, evoked potentials and hearing loss. *Hearing Research*, *360*, 55–75. https://doi.org/10.1016/j.heares.2017.12.018

Vijayasingam, A., Frost, E., Wilkins, J., Gillen, L., Premachandra, P., Mclaren, K., Gilmartin, D., Picinali, L., Vidal-Diez, A., Borsci, S., Ni, M. Z., Tang, W. Y., Morris-Rosendahl, D., Harcourt, J., Elston, C., Simmonds, N. J., & Shah, A. (2020). Tablet and web-based audiometry to screen for hearing loss in adults with cystic fibrosis. *Thorax*, *75*(8), 632–639. https://doi.org/10.1136/thoraxjnl-2019-214177

Vinay, Svensson, U. P., Kvaløy, O., & Berg, T. (2015). A comparison of test–retest variability and time efficiency of auditory thresholds measured with pure tone audiometry and new early warning test. *Applied Acoustics*, *90*, 153–159. https://doi.org/10.1016/j.apacoust.2014.11.002

Visagie, A., Swanepoel, D. W., & Eikelboom, R. H. (2015). Accuracy of Remote Hearing Assessment in a Rural Community. *Telemedicine and E-Health*, *21*(11), 930–937. https://doi.org/10.1089/tmj.2014.0243

Vos, T., Allen, C., Arora, M., Barber, R. M., Bhutta, Z. A., Brown, A., Carter, A., Casey, D. C., Charlson, F. J., Chen, A. Z., Coggeshall, M., Cornaby, L., Dandona, L., Dicker, D. J., Dilegge, T., Erskine, H. E., Ferrari, A. J., Fitzmaurice, C., Fleming, T., … Murray, C. J. L. (2016). Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: A systematic analysis for the Global Burden of Disease Study 2015. *The Lancet*, *388*(10053), 1545–1602. https://doi.org/10.1016/S0140-6736(16)31678-6

Vroegop, J. L., Dingemanse, J. G., van der Schroeff, M. P., Metselaar, R. M., & Goedegebure, A. (2017). Self-Adjustment of Upper Electrical Stimulation Levels in CI Programming and the Effect on Auditory Functioning. *Ear and Hearing*, *38*(4), e232–e240. https://doi.org/10.1097/AUD.0000000000000404

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An Agenda for Purely Confirmatory Research. *Perspectives on Psychological Science*, *7*(6), 632–638. https://doi.org/10.1177/1745691612463078

Wallace, R. S. (2009). The Anatomy of A.L.I.C.E. In R. Epstein, G. Roberts, & G. Beber (Eds.), *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer* (pp. 181–210). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6710-5_13

Waltzman, S. B., & Kelsall, D. C. (2020). The Use of Artificial Intelligence to Program Cochlear Implants. *Otology & Neurotology*, *41*(4), 452. https://doi.org/10.1097/MAO.0000000000002566

Wang, D. (2017). Deep learning reinvents the hearing aid. *IEEE Spectrum*, *54*(3), 32–37. https://doi.org/10.1109/MSPEC.2017.7864754

Wang, S., Mo, C., Chen, Y., Dai, X., Wang, H., & Shen, X. (2023). *Exploring the Performance of ChatGPT-4 in Taiwan Audiologist Examination: Indicating the Potential of AI Chatbots in Hearing Care (Preprint)*. https://doi.org/10.2196/preprints.55595

Wang, S., & Wong, L. L. N. (2024). An Exploration of the Memory Performance in Older Adult Hearing Aid Users on the Integrated Digit-in-Noise Test. *Trends in Hearing*, *28*, 23312165241253653. https://doi.org/10.1177/23312165241253653

Wang, Y., Naylor, G., Kramer, S. E., Zekveld, A. A., Wendt, D., Ohlenforst, B., & Lunner, T. (2018). Relations between self-reported daily-life fatigue, hearing status, and pupil dilation during a speech perception in noise task. *Ear and Hearing*, *39*(3), 573.

Wasmann, J.-W. A. (2023, April 13). AI's Latest Frontier part 3: An AI chatbot for audiology. *Computational Audiology*. https://computationalaudiology.com/ais-latest-frontier-part-3-an-ai-chatbot-for-audiology/

Wasmann, J.-W. A. (2024, April 1). *About*. Computational Audiology. https://computationalaudiology.com/about/

Wasmann, J.-W. A., Huinck, W. J., & Lanting, C. P. (2024). Remote Cochlear Implant Assessments: Validity and Stability in Self-Administered Smartphone-Based Testing. *Ear and Hearing*, *45*(1), 239–249. https://doi.org/10.1097/AUD.0000000000001422

Wasmann, J.-W. A., Lanting, C. P., Huinck, W. J., Mylanus, E. A. M., van der Laak, J. W. M., Govaerts, P. J., Swanepoel, D. W., Moore, D. R., & Barbour, D. L. (2021). Computational Audiology: New Approaches to Advance Hearing Health Care in the Digital Age. *Ear and Hearing*, *42*(6), 1499–1507. https://doi.org/10.1097/AUD.0000000000001041

Wasmann, J.-W. A., Pragt, L., Eikelboom, R., & Swanepoel, D. W. (2022). Digital Approaches to Automated and Machine Learning Assessments of Hearing: Scoping Review. *Journal of Medical Internet Research*, *24*(2), e32581. https://doi.org/10.2196/32581

Wasmann, J.-W. A., & Swanepoel, D. W. (2023, February 14). AI's Latest Frontier part 2: AI chatbots and clinical audiology. *Computational Audiology*. https://computationalaudiology.com/ais-latest-frontier-part-2-ai-chatbots-and-clinical-audiology/

Wasmann, J.-W. A., van Eijl, R. H., Versnel, H., & van Zanten, G. A. (2018). Assessing auditory nerve condition by tone decay in deaf subjects with a cochlear implant. *International Journal of Audiology*, *57*(11), 864–871. https://doi.org/10.1080/14992027.2018.1498598

Wasmann, J.-W., & Laat, J. de. (2022). *The Carbon footprint of hearing healthcare and how to reduce it*. OSF Preprints. https://doi.org/10.31219/osf.io/3sj5u

Wathour, J., Govaerts, P. J., & Deggouj, N. (2021). Variability of fitting parameters across cochlear implant centres. *European Archives of Oto-Rhino-Laryngology*, *278*(12), 4671–4679. https://doi.org/10.1007/s00405-020-06572-w

Wathour, J., Govaerts, P. J., Lacroix, E., & Naïma, D. (2023). Effect of a CI Programming Fitting Tool with Artificial Intelligence in Experienced Cochlear Implant Patients. *Otology & Neurotology*, *44*(3), 209–215. https://doi.org/10.1097/MAO.0000000000003810

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45. https://doi.org/10.1145/365153.365168

Wells, H. R. R., Freidin, M. B., Zainul Abidin, F. N., Payton, A., Dawes, P., Munro, K. J., Morton, C. C., Moore, D. R., Dawson, S. J., & Williams, F. M. K. (2019). GWAS Identifies 44 Independent Associated Genomic Loci for Self-Reported Adult Hearing Difficulty in UK Biobank. *The American Journal of Human Genetics*, *105*(4), 788–802. https://doi.org/10.1016/j.ajhg.2019.09.008

Whitton, J. P., Hancock, K. E., Shannon, J. M., & Polley, D. B. (2016). Validation of a Self-Administered Audiometry Application: An Equivalence Study. *Laryngoscope*, *126*(10), 2382–2388.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., & Zerbi, M. (1993). *Design and evaluation of a continuous interleaved sampling (CIS) processing strategy for multichannel cochlear implants*.

Wilson, B. S., Tucci, D. L., Merson, M. H., & O'Donoghue, G. M. (2017). Global hearing health care: New findings and perspectives. *The Lancet*, *390*(10111), 2503–2515. https://doi.org/10.1016/S0140-6736(17)31073-5

Wilson, B. S., Tucci, D. L., O'Donoghue, G. M., Merson, M. H., & Frankish, H. (2019). A Lancet Commission to address the global burden of hearing loss. *The Lancet*. https://doi.org/10.1016/S0140-6736(19)30484-2

Wilson, H. J., & Daugherty, P. R. (2018). Collaborative Intelligence: Humans and AI Are Joining Forces. *Harvard Business Review*, *July–August 2018*. https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces

Wolfgang, K. (2019). Artificial Intelligence and Machine Learning: Pushing New Boundaries in Hearing Technology. *The Hearing Journal*, *72*(3), 26. https://doi.org/10.1097/01.HJ.0000554346.30951.8d

World Economic Forum. (2021). *Diagnostics for Better Health: Considerations for Global Implementation* (p. 30). World Economic Forum. http://www3.weforum.org/docs/WEF_Diagnostics_for_Better_Health_Considerations_for_Globa_%20Implementation_2021.pdf

World Health Organization. (2013). *Multi-country assessment of national capacity to provide hearing care*. https://www.who.int/pbd/publications/WHOReportHearingCare_Englishweb.pdf

World Health Organization. (2017). *Global costs of unaddressed hearing loss and cost-effectiveness of interventions*. http://apps.who.int/iris/bitstream/10665/254659/1/9789241512046-eng.pdf

World Health Organization. (2019, March 31). *Global estimates on prevalence of hearing loss*. Deafness Prevention. http://www.who.int/deafness/estimates/en/

World Health Organization. (2021). *World report on hearing*. World Health Organization. https://www.who.int/publications-detail-redirect/world-report-on-hearing

Wright, R., & Souza, P. (2012). Comparing Identification of Standardized and Regionally Valid Vowels. *Journal of Speech, Language, and Hearing Research*, *55*(1), 182–193. https://doi.org/10.1044/1092-4388(2011/10-0278)

Wu, Y.-H., Stangl, E., Zhang, X., & Bentler, R. A. (2015). Construct validity of the ecological momentary assessment in audiology research. *Journal of the American Academy of Audiology*, *26*(10), 872–884. https://doi.org/10.3766/jaaa.15034

Xiong, W., Droppo, J., Huang, X., Seide, F., Seltzer, M. L., Stolcke, A., Yu, D., & Zweig, G. (2017). Toward Human Parity in Conversational Speech Recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *25*(12), 2410–2423. https://doi.org/10.1109/TASLP.2017.2756440

Yalamanchali, S., Albert, R. R., Staecker, H., Nallani, R., Naina, P., & J Sykes, K. (2020). Evaluation of Portable Tablet-Based Audiometry in a South Indian Population. *Indian Journal of Otolaryngology and Head & Neck Surgery*. https://doi.org/10.1007/s12070-020-02094-3

Yeung, J. C., Heley, S., Beauregard, Y., Champagne, S., & Bromwich, M. A. (2015). Self-administered hearing loss screening using an interactive, tablet play audiometer with ear bud headphones. *Int J Pediatr Otorhinolaryngol*, *79*(8), 1248–1252.

Yeung, J., Javidnia, H., Heley, S., Beauregard, Y., Champagne, S., & Bromwich, M. (2013). The new age of play audiometry: Prospective validation testing of an iPad-based play audiometer. *J Otolaryngol Head Neck Surg*, *42*(1), 21.

Yeung, W. K., Dawes, P., Pye, A., Charalambous, A.-P., Neil, M., Aslam, T., Dickinson, C., & Leroi, I. (2019). eHealth tools for the self-testing of visual acuity: A scoping review. *Npj Digital Medicine*, *2*(1), 1–7. https://doi.org/10.1038/s41746-019-0154-5

Yi, H., Pingsterhaus, A., & Song, W. (2021). Effects of Wearing Face Masks While Using Different Speaking Styles in Noise on Speech Intelligibility During the COVID-19 Pandemic. *Frontiers in Psychology*, *12*, 682677. https://doi.org/10.3389/fpsyg.2021.682677

Young, T., Pang, J., & Ferguson, M. (2022). Hearing From You: Design Thinking in Audiological Research. *American Journal of Audiology*, *31*(3S), 1003–1012. https://doi.org/10.1044/2022_AJA-21-00222

Appendices

Appendix A

Research Data Management

Acknowledgments

About the Author

PhD portfolio

List of Publications

Donders Graduate School

APPENDIX A

ID	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11
001	-	-	-	-	- +	- + +	+		+	+	+
012					+	+				+	
015							-	-			++
016											+
018								-		-	
020					+	+				+	
021					+	+	+				
022					+						
023	-	-	-	-	-			+	+	+	+
024						-		+	+	+	
025	+							-	+	- +	+
027											

Table 1A

ID	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11
001	NA	NA	11	4	-4	0	0	-3	3	6	6
012	-18	-12	-14	-12	12	4	0	0	0	0	0
015	-6	-4	0	0	-5	-4	3	4	-8	-9	0
016	-28	-23	-17	-7	0	0	0	-4	0	-7	0
018	-6	-8	-5	-7	0	0	0	-6	-10	-16	-9
020	-3	-4	-7	-6	-6	0	0	-6	-6	-6	-6
021	-9	-12	0	0	-8	0	0	-7	-7	-9	-15
022	0	-4	0	-6	20	-5	0	0	0	-6	-7
023	NA	7	5	6	9	0	-3	-5	-3	3	3
024	0	-5	-6	-5	-13	4	-11	0	6	-4	0
025	NA	-5	-4	0	0	0	0	NA	0	0	NA
027	-9	-17	-6	-6	-7	-5	-5	-3	0	0	-5

Table 1B

E12	E13	E14	E15	E16	E17	E18	E19	E20	E21	E22
	+	++	++							
						-				
- + +	-		+			-	-		+	
- +	-		++	+	+	-	-			
	+			-	-	-	-	+	+	
-	-					-			+	
						-				
- -	- -								++	+
-	- -	-	-						+	
	-	- +	+	+	+	+	+		-	-
-	-			-	-			- + +	- + + +	- +
-	-								+	

E12	E13	E14	E15	E16	E17	E18	E19	E20	E21	E22
-5	7	10	14	-12	-16	-13	-12	-11	-7	-7
-4	0	0	0	0	0	3	0	0	0	0
-7	-5	0	0	-3	0	3	3	0	0	-6
0	0	-3	-3	-3	-3	0	-3	-8	-3	-3
-4	0	-4	0	0	0	6	0	0	0	-7
-11	-6	-6	-6	-6	0	-20	-21	-6	NA	0
0	-7	-7	-7	-13	-9	6	-6	-4	-9	-7
4	0	0	0	0	-9	0	-4	0	4	0
0	0	0	6	-9	-4	-19	-15	-14	12	-16
-4	0	8	4	0	0	6	9	-4	10	12
-3	0	0	0	0	0	0	-3	0	0	5
8	0	0	0	0	-5	-8	-7	0	0	-3

RESEARCH DATA MANAGEMENT

Ethics and privacy

The study in Chapter 6 was set up following a single-center repeated measures cohort study design. A statement that the study in Chapter 6 was not subject to the Dutch Medical Research Involving Human Subjects Act (WMO), was obtained from the institutional ethical review committee CMO Radboudumc, Nijmegen, the Netherlands (Filenumber: 2020-7203). The study adhered to good clinical practice. Informed consent was obtained from participants to collect and process their data for this study.

This feasibility study detailed in Chapter 7 was conducted using a single-center pre-post interventional design performed at the Radboud university medical center's outpatient clinic. The institutional ethical review committee CMO Radboudumc, Nijmegen, the Netherlands, has given approval to conduct this study (CMO Radboudumc dossier number: 2022-13495). The study was part of the Auditory Diagnostics and Error-based Treatment (AuDiET) trial, which is pre-registered. Additional information on the AuDiET trial can be found at https://clinicaltrials.gov/study/NCT05307952.

To protect participant privacy:

- The privacy of the participants in these studies in Chapter 6 and 7 was warranted by the use of pseudonymization. The pseudonymization key was stored on a secured network drive that was only accessible to members of the project who needed access to it because of their role within the project. The pseudonymization key was for both studies stored separately from the research data. Participant data of Chapter 6 and 7 contains sensitive and personally identifiable data. Only the performance data are shared to safeguard the privacy of the participants

Funding

Cochlear Ltd funded the study presented in Chapter 6 and provided the loaner CI processors used for the study presented in Chapter 7. The study in Chapter 7 was part of the MOSAICS project. It was made possible by a European Industrial Doctorate project funded by the European Union's Horizon 2020 framework program for research and innovation under the Marie Sklodowska-Curie Grant Agreement No. 860718.

Data collection and storage

The raw Data from Chapter 4, 5, 6 and 7 were stored and analyzed on the department server, with access restricted to project members working at Radboudumc. The data for the metareview from Chapter 5 was retrieved from the electronic databases of PubMed, IEEE, and Web of Science was conducted to identify relevant reports from the peer-reviewed literature. The complete set of terms and the applied search strategy are provided in Multimedia Appendix 1 in the published paper. The Chronotype, CIS, and MM questionnaires from Chapter 6 were filled out online by the participants using Castor EDC for secured online questionnaires. Data from the online questionnaires were stored in Castor EDC. The Remote Check data from Chapter 6 were securely stored within the Cochlear clinician's portal. Paper (hardcopy) data containing questionnaires from Chapter 7 are stored in cabinets on the department. The analyzed auditory performance data from Chapter 6 and 7 are stored in secure repositories that comply with GDPR and other relevant privacy laws.

Data sharing according to the FAIR principles

Findable and Accessible

The auditory performance data from Chapter 6 and 7 are published in Data Sharing Collections (DSC's) in the Radboud Data Repository, with access regulated under the RUMC-RA-DUA-1.0 Data Use Agreement. Requests for access will be checked by the PI against the conditions for sharing the data as described in the signed Informed Consent. The data not suitable for reuse will be archived for 15 years after termination of the study.

Reusability and Interoperability

To promote data reusability and interoperability, the following measures have been implemented:

- **File Formats**: The raw data are stored in XYZ in their original form. Where relevant, data has also been stored pseudonymized in tidy data format in.xlsx and.csv files, ensuring that data remains usable in the future.
- **Reproducibility**: Detailed specifications of the experimental setups and analysis scripts are provided. The R code used in Chapters 7 is available in the repository. R code used in Chapter 6 and the Python code employed in Chapter 7, may be made available upon reasonable request. Version numbers for all software used are documented within the provided scripts.

The table below details where the data and research documentation for Chapter 6 and 7 can be found. All data archived as a Data Sharing Collection remain available for at least 15 years after termination of the studies.

Chapter	Repository	*Data Sharing Collection* (DSC)	DSC License
6	Radboud Data Repository	DOI: https://doi.org/10.34973/3ykx-5g32	RUMC-RA-DUA-1.0
7	Radboud Data Repository	DOI: https://doi.org/10.34973/qr4j-y723	RUMC-RA-DUA-1.0

ACKNOWLEDGMENTS

I would like to use this section to thank the people who inspired, helped, and motivated me, and to share a few lessons I learned along the way.

Academic Supervisors, Mentors, and Collaborators

First and foremost, I wish to thank my primary supervisors:
Cris, you are a strong scientist. Your ability to seamlessly switch between high-level overviews and detailed scrutiny significantly improved our proposals and analyses. Throughout my PhD, you consistently provided clear feedback on both the weaknesses in the storyline (kapstok, "framework") and the style of my written manuscripts.

Wendy, you have a keen eye for the human scale. Whenever I bit off more than I could chew, you helped me prioritize and prevented me from stalling. In all our writing, you ensured that the message was clear and considered how it impacted the receiver.

Emmanuel, you put people and human relationships first. This quality is very important for collaboration and creating a positive work environment. Your idealism provided a great backdrop, reminding me not to forget the bigger picture.

I am also deeply grateful to my informal advisors, and I have been fortunate to collaborate with many talented individuals.

Dave, I'm glad I could express my thanks in person in Manchester, and your inspiration to consider strangers as unmet friends has greatly influenced me both professionally and personally. Dennis, I can consider you literally as an unmet friend. Our first in-person meeting is still in the making. I immensely enjoyed our online chats and virtual meetings about basically anything that stirs our interest. We are like brothers-in-arms in trying to get more Bayesian approaches into the clinic.

Lucas, when I told you I was ready for "Hora Est," you responded "Hoera Est" (Hurray Est). It shows your sense of humor, your elegant way of providing feedback, and your talent for making every letter count. You have acted as a protector, advisor, and critical reviewer. You read texts completely different from me, and your feedback was invaluable in helping me to avoid many mistakes. Without your help, I would have remained too vague about the concepts I wanted to convey. Marc, you always

put scientific integrity and your students' interests first, which you certainly did in your role as independent advisor.

De Wet, you came to our rescue when our computational audiology paper was rejected for the umpteenth time. Our ongoing collaboration has taught me many valuable lessons. I'm very grateful that you and Rob E. were willing to meet with Leontien and me every two weeks for almost a full year when we conducted our scoping review. Rob E., you have been a very thoughtful reader of some of my texts, helping me to express myself more precisely.

While being surrounded by idealists, it was good to talk with you, Paul, every now and then to set my feet back on the ground. Leontien, Dagmar, and Peter, it was fun working with you. I had not expected that my playful proposal of testing ASR with audiological assessments would, with your help, grow into a genuine project. Bas and Birgit, our collaboration within AuDiET provided many new insights. Not only will cochlear care be further personalized in the future, but you already personalized research mentorship to a high degree by taking into account the individual needs of Nikki, Enrico, and me. Nikki, it was a pleasure working with you. You have an eye for timing and a talent for making things work just in time. Enrico, although we differ on many things except our interest in Sci-Fi, this often complemented our collaboration. Vinay, I'm glad we finally met this summer and grateful for the tips you provided on how to plan my next steps in research.

VCCA and CAN Community

Creating the Virtual Conference of Computational Audiology (VCCA) series as a platform to share knowledge has been a fantastic experience. I would like to acknowledge:

Volker, you reached out to me to protect me against any liability for the information on our website. It was great working with you on the VCCA conferences; you are one of my go-to persons on ethical issues and open science. Stefan, your lessons about the importance of trust resonated very well and continue to play an important role. Brent, your leadership in our field has been an inspiring example. Tobias, you are one of those other compasses when discussing the right (ethical) direction. Your decision to organize the second VCCA helped to turn it into a series of events that continues to grow and develop. I'm glad you inspired many others to step in. In addition, it was a lot of fun! Waldo, Anna, Jess, Karina, and Kaye, thanks a lot for the nice collaboration in subsequent VCCAs. Simone, Trevor, and Jon, it was great working with you. You basically wrote the template for future events. Seba, I'm glad that the next VCCA is in good hands.

There is a wider community I have enjoyed being part of with a shared goal of advancing hearing healthcare through the application of data science, computational methods, and artificial intelligence (AI) in hearing loss research and technology. Tilak, your frank feedback about diversity and openness about your experience with hearing loss helped me better appreciate how we could use new tools. If Dennis is my brother-in-arms, then Deniz is my sister-in-arms, trying to advance more equality and diversity into our ranks and approaches. Deniz, your help was important in establishing CAN as a formal network. Fan-Gang, it was great to get your support when we just started reaching out to the community. It certainly helped us get further traction, and I want to thank you for your unwavering support. Blake, your kind words gave me wings. Similarly, I would like to extend my thanks to Astrid for all your advice over the years, which led to a good collaboration with ISA. George, you are our ideal problem solver. Thais, thank you for moving CAN towards the World Hearing Forum and, together with Deborah, introducing me to Brazilian researchers and activities. Elle, your suggestion to start a Slack network was a further step in building a community, and suggesting the acronym Computational Audiology Network (CAN) was excellent from a brand-building perspective. Hector, Mareike, and Charlotte V., your many initiatives and enthusiasm for further expanding the CAN group have been very rewarding.

Jan, whenever I had a new idea, I could count on your support and enthusiasm. Without you, our sustainability agenda would not have progressed as it did. When you retired, I thought you would have more free time. However, it is impressive to see how your agenda is still more packed than mine, but even then, you are always willing to join an activity and make valuable notes and observations. Inga, your enthusiasm for training new generations of audiologists and scientists is contagious. Thanks for your support and the opportunities you provided. Alessia, you bring in very useful experience and skills that I hope to learn more about. Valeriy, our discussions have always brought me new insights and better expressions. I'm confident we will continue to be part of a larger community. Mel, as we agree, what's in a name is vital. I very much enjoyed the idea of using pop art to disseminate knowledge better. Helen H., thanks again for the coffee and nice talk in Reno. Bec, your talent to get innovation really into clinical practice is unparalleled. I'm looking forward to carrying out future projects. Gaby, I'm sure we can make those innovations more intuitive. How else will it ever get adopted?

Professional Support

I am deeply grateful to my colleagues at the Radboud University Medical Center:
It was my former boss, Henri, who encouraged me to start a PhD. I would like to thank you for entrusting me with the freedom to chart my own course and for providing timely feedback that helped me avoid overpromising and underdelivering. You also gave me the opportunity to explore activities that were a bit off course but which I enjoyed a lot. Ronald, you smoothly took over Henri's guidance, and I'm grateful you helped to create an elegant "PhD exit strategy." At a certain point, one needs to finish things, but it also helped to have positive motivation for what would be in store next. Another big thanks goes to Saskia. It would have been impossible to combine clinical and research duties without your support.

Hilde, as you know from firsthand experience, my planning was often imperfect. Thanks for coming to the rescue in time and finding new ways of organizing clinical activities. Meanwhile, the training of audiologists was, and is, in good hands. Charlotte S. and Arno, you made sure that all clinical care was provided, even when we were short of hands. Martijn, thanks for always having a plan B for my PhD. Andy, your support helped me to chart my own course.

I like to thank all my colleagues at our Audiology Center. Dorien B., your skills in organizing and clear communication were vital for instructing participants on how to do tests at home. I cannot thank you enough for all the evenings/nights you stood on duty to help participants with urgent problem-solving. The feedback from the speech therapists was important to make official communications to participants more clear. I would like to thank all of my colleagues in recruiting participants. Over these years, I relied much on the autonomy and expertise of all my "VAC colleagues," who ensured that clinical care was never interrupted. Herman, sometimes it comes in handy that you never say no. Anne, your flexibility was also unparalleled. Teja and Mieki, I'm glad you supported my move towards more research. Helma, Henriette, and Donneke, your eye for detail has given me a lot of new insights. Esther, Anja, and José, you made sure things got done.

John, your feedback was a great way to sharpen my thoughts. Yagmur and Marcel, it was nice to learn about more AI activities across campus, and I hope we can continue collaborating. Frank and Klaasjan, you made our hospital a pleasant environment for training medical physicists, and you were always available when our group needed support. Marinette, I'm happy you took the initiative to start helping children with hearing loss in Indonesia in collaboration with Kolewa. I'm sure this will further grow.

Industry Partners

I appreciate the collaboration and support from industry partners:
Annes, Feike, Rob B., Dorien vdZ, Niels vD., thanks for helping with the Remote Check project. Jari and Karlien, thanks for your openness and help in further evaluating new ASR technologies. Geert, thank you for sharing vital contacts. Tarien, thank you for your support towards a sustainable path. Niels P., thanks for your industry-wide insights. Steve, thanks for showing the ropes regarding podcast recording, and Abram, thank you for your hearing industry insights. A special thanks is deserved by the people involved in the ESG working group, in particular Chaojun, Michael, Maurits, and Nicola, for their open collaboration toward more sustainable hearing healthcare. The last "industry entities" I need to acknowledge are ChatGPT 3.5 and GPT-4o. A human servant could not have equaled your obedience when rephrasing, grammar checking, and image creating.

Personal Reflections and Experiences

There were several events and conversations that motivated me to pursue research. In 2013, at the Eargroup in Antwerp, I witnessed AI-driven fitting methods, which opened my eyes to new possibilities. Participating in the leadership course at the Radboud University Medical Center played an important role in my decision to delve into research. The examples, discussions, and feedback from Jolt, Alexander, Frank Erik, Hedi, Bregje, Rick, Lisenka, Erwin, Helen B., and Judith were instrumental in helping me determine what I wanted to do and how. A conversation with Jeroen over coffee made a significant difference. When I shared ideas about potential projects using AI in audiology, he remarked, "What you like to do sounds like computational audiology to me."

At the first VCCA in 2020, I met a young man from Bolivia with profound hearing loss. We communicated using Google Live Transcribe, and his resourcefulness in using assistive technology inspired me. This encounter also reminded me that providing clinical care in Nijmegen does not necessarily expose one to all new developments. Sometimes, being in a small and wealthy country can inhibit us from seeing the pressing needs and arising opportunities elsewhere. The COVID-19 pandemic shifted my perspective on remote care from a "nice-to-have" to a necessity. While there was initial reluctance among clinicians to adopt AI and remote care, I became motivated to break through these limitations and contribute to globally available solutions.

Peers, Friends, and Family

I am grateful to my peers, and friends for their support and positive distractions: When I needed coffee, I could always kidnap you from your desk, Hugo. Thanks for lending an ear. Ietske and Arjan, it was nice to share PhD struggles and successes. Ignacio, your gentle and thoughtful suggestions provided many insights. Whenever I arrived early in the office, it was nice to chat and start the day with you, Robert T.. Coby, thank you for helping me stay within protocol and preventing me from bending the rules too much. Joke, I very much enjoyed our talks about music and hearing. Marloes, thanks for the fun distractions. Thijs and Niels vH., your projects beyond the clinic are inspiring, and I hope my research can be a stepping stone.

Pim and Sal, your opinions matter. How else could we save the planet? Skander, it's always nice to talk to you; your kindness is overwhelming. Michiel, thanks for the outdoor adventures. Martin, it is always nice to let old times revive with you. Dear Catshouse members, you thrive in family distractions. Jaap and Robert W., thanks for guiding me towards the hills and mountains whenever possible. In addition, Robert W., your suggestion to read Cialdini was very influential.

I am grateful to my family for their unwavering support:
Iris, thanks for all your support in website design. Jasmijn, thank you for your help finding a video editor. Mom and Dad, thanks for your unconditional support. Pieter, it's invaluable to always have a brother I can count on. Emily, thanks for your proofreading. Lastly, Donna, Daniel, and Hannah, you made sure to show me what things really matter.

ABOUT THE AUTHOR

Jan-Willem Wasmann was born on November 21, 1984, in Utrecht, the Netherlands. He completed his secondary education at the Twents Carmel Lyceum in Oldenzaal (2003). He pursued a Bachelor of Physics at Radboud University Nijmegen (2003–2006). He continued his studies at Radboud University, completing a Master of Physics in 2009. His academic journey included an Erasmus exchange at Universidad Autónoma in Madrid (2006–2007) and research internships at the Hahn-Meitner Institute in Berlin (2007) and the National Aerospace Laboratory in Amsterdam (2009).

After completing his studies, Jan-Willem began his career as a Technology Advisor at the Netherlands Space Office (2009–2011). He later transitioned from aerospace to audiology, undertaking his training as a Medical Physicist Audiologist at the University Medical Center Utrecht (2011–2015). Upon completion, he briefly continued as a Medical Physicist Audiologist in Utrecht before joining Radboudumc Nijmegen in 2015. Currently, Jan-Willem combines clinical care for adults with hearing loss and ongoing research. His research interests focus on modernizing hearing healthcare through the development of remote testing, outcome-guided cochlear implant fitting, and applying new AI tools in audiology.

Jan-Willem contributes to the international audiology community. He launched computationalaudiology.com, a forum for AI and audiology, and organized the first Virtual Conference on Computational Audiology (VCCA) in 2020. He has also been actively involved in its subsequent editions. He founded the Computational Audiology Network (CAN) with an international group of enthusiasts.

PHD PORTFOLIO

Name PhD candidate: Jan-Willem Wasmann
Graduate School: Donders Graduate School
PhD period: 17-08-2021 – 18-03-2024
Supervisor: Prof. E.A.M. Mylanus

Courses & Workshops	Organizer	Hours
Graduate School Introduction Day (2022)	Donders Graduate School	7
Basic course for clinical investigators (2023)	NFU BROK Academie	42
Scientific Integrity Course (2024)	Donders Graduate School	7
Writing Scientific Articles (2024)	Radboud University	96

Conferences	Location
VCCA (2021)	Online
VCCA (2022)	Online
VCCA (2023)	Online
VCCA (2024)	Online
Hearing across the lifespan (HEAL, 2022)	Cernobbio, Italy
European Symposium on Pediatric Cochlear Implantation (ESPCI, 2023)	Rotterdam, The Netherlands
International Hearing-Aid Research Conference (IHCON, 2024)	Lake Tahoe, United States

LIST OF PUBLICATIONS

Wasmann, J.-W. A., Lanting, C. P., Huinck, W. J., Mylanus, E. A. M., van der Laak, J. W. M., Govaerts, P. J., Swanepoel, D. W., Moore, D. R., & Barbour, D. L. (2021). Computational Audiology: New Approaches to Advance Hearing Health Care in the Digital Age. Ear and Hearing, 42(6), 1499–1507. https://doi.org/10.1097/AUD.0000000000001041

Swanepoel, D. W., Manchaiah, V., & Wasmann, J.-W. A. (2023). The Rise of AI Chatbots in Hearing Health Care. The Hearing Journal, 76(04), 26. https://doi.org/10.1097/01.HJ.0000927336.03567.3e

Pragt, L., van Hengel, P., Grob, D., & Wasmann, J.-W. A. (2022). Preliminary Evaluation of Automated Speech Recognition Apps for the Hearing Impaired and Deaf. Frontiers in Digital Health, 4, 806076. https://doi.org/10.3389/fdgth.2022.806076

Wasmann, J.-W. A, Pragt, L., Eikelboom, R., & Swanepoel, D. W. (2021). Digital approaches to automated and machine learning assessments of hearing: A scoping review. Journal of Medical Internet Research. https://doi.org/10.2196/32581

Wasmann, J.-W. A., Huinck, W. J., & Lanting, C. P. (n.d.). Remote Cochlear Implant Assessments: Validity and Stability in Self-Administered Smartphone-Based Testing. Ear and Hearing,10.1097/AUD.0000000000001422. https://doi.org/10.1097/AUD.0000000000001422

Wasmann, J.-W. A., Migliorini, E., Philpott, N., Philips, B., van Dijk, B., & Huinck, W. (2024). *Feasibility of a Cochlear Implant Fitting Approach Based on Phoneme Confusions: Lessons Learned from the AuDiET Study*. https://doi.org/10.31219/osf.io/k8f2y

Other publications

Barbour, D. L., & Wasmann, J.-W. A. (2021). Performance and Potential of Machine Learning Audiometry. The Hearing Journal, 74(3), 40–43. https://doi.org/10.1097/01.HJ.0000737592.24476.88

Meng, Q., Chen, J., Zhang, C., Wasmann, J.-W. A., Barbour, D. L., & Zeng, F.-G. (2022). Editorial: Digital hearing healthcare. Frontiers in Digital Health, 4. https://doi.org/10.3389/fdgth.2022.959761

Vijverberg, M. A., Caspers, C. J., Kruyt, I. J., Wasmann, J.-W., Bosman, A. J., Mylanus, E. A., & Hol, M. K. (2019). Comment on "Baha Skin Complications in the Pediatric Population: Systematic Review with Meta-Analysis." Otology & Neurotology, 40(5), 689–691. https://doi.org/10.1097/MAO.0000000000002223

Vogt, K., Wasmann, J.-W., Van Opstal, A. J., Snik, A. F., & Agterberg, M. J. (2020). Contribution of spectral pinna cues for sound localization in children with congenital unilateral conductive hearing loss after hearing rehabilitation. Hearing Research, 385, 107847. https://doi.org/10.1016/j.heares.2019.107847

Wasmann, J. A., Janssen, A. M., & Agterberg, M. J. H. (2020). A mobile sound localization setup. Methods X, 7. https://doi.org/10.1016/j.mex.2020.101131

Wasmann, J.-W. A., & Barbour, D. L. (2021). Emerging Hearing Assessment Technologies for Patient Care. The Hearing Journal, 74(3), 44–45. https://doi.org/10.1097/01.HJ.0000737596.12888.22

Wasmann, J.-W. A., & de Laat, J. (2022). *The Carbon Footprint of Hearing Healthcare and How to Reduce it*. https://doi.org/10.31219/osf.io/3sj5u

Wasmann, J.-W. A., van Eijl, R. H., Versnel, H., & van Zanten, G. A. (2018). Assessing auditory nerve condition by tone decay in deaf subjects with a cochlear implant. International Journal of Audiology, 57(11), 864–871. https://doi.org/10.1080/14992027.2018.1498598

DONDERS GRADUATE SCHOOL

For a successful research Institute, it is vital to train the next generation of scientists. To achieve this goal, the Donders Institute for Brain, Cognition and Behaviour established the Donders Graduate School in 2009. The mission of the Donders Graduate School is to guide our graduates to become skilled academics who are equipped for a wide range of professions. To achieve this, we do our utmost to ensure that our PhD candidates receive support and supervision of the highest quality.

Since 2009, the Donders Graduate School has grown into a vibrant community of highly talented national and international PhD candidates, with over 500 PhD candidates enrolled. Their backgrounds cover a wide range of disciplines, from physics to psychology, medicine to psycholinguistics, and biology to artificial intelligence. Similarly, their interdisciplinary research covers genetic, molecular, and cellular processes at one end and computational, system-level neuroscience with cognitive and behavioural analysis at the other end. We ask all PhD candidates within the Donders Graduate School to publish their PhD thesis in de Donders Thesis Series. This series currently includes over 600 PhD theses from our PhD graduates and thereby provides a comprehensive overview of the diverse types of research performed at the Donders Institute. A complete overview of the Donders Thesis Series can be found on our website: https://www.ru.nl/donders/donders-series

The Donders Graduate School tracks the careers of our PhD graduates carefully. In general, the PhD graduates end up at high-quality positions in different sectors, for a complete overview see https://www.ru.nl/donders/destination-our-former-phd. A large proportion of our PhD alumni continue in academia (>50%). Most of them first work as a postdoc before growing into more senior research positions. They work at top institutes worldwide, such as University of Oxford, University of Cambridge, Stanford University, Princeton University, UCL London, MPI Leipzig, Karolinska Institute, UC Berkeley, EPFL Lausanne, and many others. In addition, a large group of PhD graduates continue in clinical positions, sometimes combining it with academic research. Clinical positions can be divided into medical doctors, for instance, in genetics, geriatrics, psychiatry, or neurology, and in psychologists, for instance as healthcare psychologist, clinical neuropsychologist, or clinical psychologist. Furthermore, there are PhD graduates who continue to work as researchers outside academia, for instance at non-profit or government organizations, or in pharmaceutical companies. There are also PhD graduates who work in education, such as teachers in high school, or as lecturers in higher

education. Others continue in a wide range of positions, such as policy advisors, project managers, consultants, data scientists, web- or software developers, business owners, regulatory affairs specialists, engineers, managers, or IT architects. As such, the career paths of Donders PhD graduates span a broad range of sectors and professions, but the common factor is that they almost all have become successful professionals.

For more information on the Donders Graduate School, as well as past and upcoming defences please visit:

http://www.ru.nl/donders/graduate-school/phd/

DONDERS INSTITUTE

Radboud University

Radboudumc