Contextually Adapting Hearing Aids by Learning User Preferences from Data

Pasta, Alessandro

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

Contextually Adapting Hearing Aids by Learning User Preferences from Data

Alessandro Pasta

Kongens Lyngby 2021

Technical University of Denmark
Department of Applied Mathematics and Computer Science
Richard Petersens Plads, building 324,
2800 Kongens Lyngby, Denmark
Phone +45 4525 3031
compute@compute.dtu.dk
www.compute.dtu.dk

Summary (English)

Although hearing aid users perceive sound in individual ways, current approaches do not fully exploit the potential for personalization. Providing a comprehensively personalized hearing aid solution is a complex and multidimensional challenge and requires a deep understanding of patients' preferences and behavior. This thesis leverages real-world data collected through smartphone-connected hearing aids to address two main themes.

First, personalizing hearing aid settings requires learning the audiological preferences of users. We adopted a smartphone-based method to make users explore three audiological parameters (Noise Reduction, Brightness, and Soft Gain) and to gather their audiological preferences in real-world listening environments. The collected data was modeled to investigate the feasibility of a context-aware system for providing users with a number of relevant hearing aid settings to choose from. We found that having access to different intervention levels of two audiological parameters (Brightness and Soft Gain) affected listening satisfaction. Moreover, context significantly impacted the perceived usefulness of having access to different intervention levels, as well as the intervention level preferences.

Second, offering a comprehensively personalized solution, as well as transferring the learned audiological preferences to new or inactive users, requires learning users' behavior. Large scale data logged by commercially available products were analyzed to explore patterns of hearing aid use, as well as the provision and context of use of listening programs. We found that, on average, users used the hearing aids 10 hours/day and we identified three clusters of users, each characterized by a predominant daily pattern of hearing aid use. Moreover, we

identified a default listening program, a primary additional program, and two secondary additional programs. We also found that users used the additional listening programs in sound environments different than the sound environment measured when using the default program.

This thesis contributes to the progress towards a data-driven approach to real-time hearing aid personalization by learning users' preferences and behavior from data. It also demonstrates that smartphone-connected hearing aids can be useful to both perform experimental studies aimed at exploring novel ways of personalizing the device, and observational studies aimed at investigating how users naturally use commercially available devices in real-world settings.

Summary (Danish)

Selvom høreapparatsbrugere opfatter lyde individuelt, har nuværende løsninger ikke formået at udnytte potentialet for at personalisere lytte oplevelsen. En udførlig personliggjort høreapparatsløsning er en kompleks og flerdimensionel udfordring, som kræver en dyb forståelse for patienternes præferencer og adfærd. Denne afhandling benytter data, indsamlet i den virkelig verden igennem høreapparater tilkoblet smartphones, med det formål at adressere to tematikker.

For det første; Personliggjorte høreapparats indstillinger kræver en forståelse af høreapparatsbrugernes audiologiske præferencer. Vi har brugt en smartphonebaseret metodik til at få høreapparatsbrugere til at udforske tre audiologiske præferencer (støjreduktion, klarhed, svage lyde). Denne metodik er ligeledes brugt for at indsamle disse præferencer i lydmiljøer i den virkelige verden. Den indsamlede data blev modelleret med det formål at undersøge mulighederne for et kontekst-bevidst system som kunne foreslå høreapparatsbrugerne et antal relevante indstillinger at vælge ud fra. Vi fandt at have adgang til interventionsniveauer på to audiologiske parameter (klarhed, svage lyde) har indflydelse på tilfredsheden med lytteoplevelsen. Ydermere, fandt vi at den kontekst høreapparatsbrugeren befinder sig i har en signifikant indflydelse på den opfattede nytte ved at have adgang til forskellige interventionsniveauer. Ligeledes, har den kontekst høreapparatsbrugeren befinder sig i indflydelse på dennes præferencer for disse interventionsniveauer.

Dernæst; For at tilbyde en udførlig personliggjort løsning, samt at oversætte disse læringer om audiologiske præferencer til nye eller inaktive høreapparatsbrugere, kræves en læring af disse brugers adfærd. For at analysere og udforske mønstre hos høreapparatsbrugere, samt disses anvendelse af høreapparatspro-

grammer og i hvilken kontekst disse blev brugt, blev data brugt i stor skala, tilgængelige via datalogging i kommercielt tilgængelige programmer. Vi fandt at høreapparatsbrugere, i gennemsnit, bruger deres høreapparater 10 timer om dagen. Ligeledes, identificerede vi tre grupper af høreapparatsbrugere, som hver især er karakteriseret ved forskelle i typisk brug af høreapparat i løbet af en dag. Endvidere identificerede vi at høreapparatsbrugere benytter deres supplerende høreapparatsprogrammer i andre lydmiljøer end de lydmiljøer målt ved brugen af deres standard høreapparatsprogram.

Denne afhandling yder et bidrag til udviklingen imod en mere data-dreven tilgangsvinkel i personliggørelsen af høreapparater i realtid, ved at tilbyde en forståelse af høreapparatsbrugernes preferencer og adfærd igennem brugen af data. Ligeledes, demonstrerer afhandlingen at høreapparater tilkoblet smartphones kan være brugbare i udformningen af eksperimenter med det formål at udforske nye måder at personliggøre høreapparater på. Afhandlingen demonstrerer også hvordan høreapparater tilkoblet smartphones kan bruges i observationsstudier med det formål at undersøge hvordan høreapparatsbrugere bruger kommercielt tilgængelige høreapparater i den virkelige verden.

Preface

This thesis was prepared at DTU Compute in fulfilment of the requirements for acquiring a Ph.D. in Engineering, and was prepared at the Cognitive Systems section of DTU Compute, under the supervision of Associate Professor Jakob Eg Larsen, Professor Ole Winther, Senior Manager Kasper Juul Jensen, and Research Manager Niels Henrik Pontoppidan. The thesis investigates the audiological preferences and real-world behavior of hearing aid users to improve the personalization of hearing aids. The thesis includes three published papers, and two manuscripts under review.

Lyngby, 31-December-2021

Alessandro Pasta

Scientific Contributions

Peer-reviewed publications

- **Alessandro Pasta**, Michael Kai Petersen, Kasper Juul Jensen, and Jakob Eg Larsen (2019). Rethinking Hearing Aids as Recommender Systems. *CEUR Workshop Proceedings*, 2439, 11-17, http://ceur-ws.org/Vol-2439/4-paginated.pdf. (Appendix A)

- **Alessandro Pasta**, Michael Kai Petersen, Kasper Juul Jensen, and Jakob Eg Larsen (2020) Designing Audiologist Bots Fusing Soundscapes and User Feedback. *CHI2020, Workshop on Conversational Agents for Health and Wellbeing*. (Appendix C).

- **Alessandro Pasta**, Tiberiu-Ioan Szatmari, Jeppe Høy Christensen, Kasper Juul Jensen, Niels Henrik Pontoppidan, Kang Sun, and Jakob Eg Larsen (2021). Clustering Users Based on Hearing Aid Use: An Exploratory Analysis of Real-World Data. *Frontiers in digital health*, 3, 725130, https://doi.org/10.3389/fdgth.2021.725130. (Appendix D).

Manuscripts under review

- **Alessandro Pasta**, Michael Kai Petersen, Kasper Juul Jensen, Niels Henrik Pontoppidan, Jakob Eg Larsen, and Jeppe Høy Christensen. Measuring and Modeling Context-Dependent Preferences for Hearing Aid Settings. (Appendix B).

- **Alessandro Pasta**, Tiberiu-Ioan Szatmari, Jeppe Høy Christensen, Kasper Juul Jensen, Niels Henrik Pontoppidan, Kang Sun, and Jakob Eg Larsen. Investigating the Provision and Context of Use of Hearing Aid Listening Programs from Real-World Data. (Appendix E).

The above contributions are part of this thesis and are included in the appendices A to E.

Acknowledgements

I am very grateful to Michael Kai Petersen, who has made it possible for me to pursue this PhD and has been my external supervisor during the first half of my PhD. Thank you for your visionary guidance, for your infinite kindness, and your openness to new ideas.

A heartfelt thanks to Kasper Juul Jensen, my second external supervisor, for always believing in me since I joined Oticon as an intern, for embracing my ambitions, and for always ensuring I felt part of the team. I would also like to thank Jonas Garder Kristiansen and the other former and present colleagues in the Customer Insights team for their interest in my research, and for making my journey very pleasant. I feel grateful to be in such a great team.

Thank you to my university supervisor, Associate Professor Jakob Eg Larsen, for his generous guidance and constant availability throughout the PhD journey and during the writing of this thesis. I would like to thank Niels Henrik Pontoppidan, who has been my supervisor during the second half of my PhD, for his precious advice and constant support. Thank you to Professor Ole Winther for his insightful Deep Learning lessons.

I am grateful to Jeppe Høy Christensen, who has helped me shaping my PhD journey by encouraging me to formulate relevant and clear research questions. Thank you for your helpfulness and your invaluable feedback. I would like to thank Tiberiu-Ioan Szatmari, with whom I intensively cooperated during the second half of my PhD. I enjoyed working with you and learning from you.

I would like to thank Ole Asboe Jørgensen, President of the Oticon Brand, and

Finn Möhring, Senior Vice President, R&D, Demant for believing in me and agreeing to fund this PhD project. I would also like to thank my colleagues in Oticon and Eriksholm Research Centre for our collaboration on several projects.

A sincere thank you to my mother, Patrizia, and my father, Andrea, for their care, interest and trust. Their unconditioned support and their constant encouragement have been extremely important for my personal growth and academic achievements. I am grateful to my sister, Elisa, who, with her sweetness and determination, has always been an example to follow.

My special thanks go to my girlfriend, Giulia, for her lovely presence and for being constantly by my side throughout these years.

I would like to express my gratitude to my professors and my fellow students, who made my studies interesting and stimulating. Finally, thank you to everyone else with whom I shared little but significant moments of my life. They all contributed, in their unique and invaluable way, to make me the person I am today.

Acronyms

BR brightness. 20, 21

NF noise floor. 37–39, 47

NR noise reduction. 20, 21

SG soft gain. 20, 21

SML sound modulation level. 37–39, 47

SNR signal-to-noise ratio. 20, 23, 42–44

SPL sound pressure level. 20, 24, 37, 38, 43, 44, 47

Contents

Chapter 1

Introduction

This chapter provides an introduction to the hearing loss condition (Section 1.1) and to hearing aids as a way to compensate for hearing loss (Section 1.2). Moreover, it discusses how a paradigm shift in hearing healthcare might enable addressing the main challenges with hearing aids (Section 1.3). Finally, Section 1.4 states the research objectives and Section 1.5 summarizes the thesis structure.

1.1 Hearing loss

This section defines the hearing loss condition, the different types of hearing loss, and its causes. Furthermore, it provides an overview of the prevalence and the consequences of hearing loss, and it summarizes the available treatments.

1.1.1 Definition, Classification and Etiology

Hearing is the sense that enables us to perceive the sounds around us, to engage with our environment and to connect to the world [92]. Humans can commonly hear sounds having frequencies from 20 to 20,000 Hz [112], and intensity from

0 to 120 dB HL. A person is said to have normal hearing if she is able to hear sounds softer than 20 dB in both ears (computed as an average of hearing thresholds at 500, 1000, 2000 and 4000 Hz) [97]. Hearing loss is defined as a partial or total inability to hear [23]. A person is said to have hearing loss if she is not able to hear sounds louder than 20 dB in any of the two ears [97].

Depending on the affected part of the ear, three types of hearing loss can be identified [9]. Conductive hearing loss is caused by a damage or a blockage in the outer and/or middle ear [5]. Sensorineural hearing loss is the most common type of hearing loss and occurs when there is a damage or malfunction in the inner ear or auditory nerve [8]. This type of hearing loss is usually associated to a damage of the hair cells, which are positioned in the inner ear and are responsible for transmitting sound information to the auditory nerve [36]. Mixed hearing loss includes both a conductive and a sensorineural hearing loss [7]. The degree of hearing loss can widely vary and be mild (20-34 dB HL), moderate (35-49 dB HL), moderately severe (50-64 dB HL), severe (65-79 dB HL), profound (80-94 dB HL), or complete (>94 dB HL) [97]. While a person with mild hearing loss has difficulties in hearing soft sounds, a person with profound hearing loss cannot hear any speech in a noisy environment [97].

Hearing loss can be caused by several factors that may occur at different stages in life. During the prenatal phase, hearing loss can be caused by genetic factors or intrauterine infections [133]. In the perinatal period, common causes are birth asphyxia, hyperbilirubinemia and low-birth weight [133]. During childhood and adolescence, chronic ear infections, meningitis and collection of fluid in the ear can lead to hearing loss [133]. During adulthood and older age, chronic diseases, otosclerosis, age-related sensorineural degeneration can cause hearing loss [133]. Finally, some factors, such as trauma, loud sounds, nutritional deficiencies, and viral infections, can affect hearing across the life span [133].

1.1.2 Epidemiology

It is estimated that 1.57 billion people (20% of the world population) globally have some degree of hearing loss [47]. Among those, 430 million people (5% of the population) suffer from a moderate or higher level of hearing loss [47].

The prevalence of hearing impairment is higher in middle- and low-income countries than in high-income countries [120]. The largest number of people with a moderate or higher level of hearing loss resides in the Western Pacific region (127 million), the South-East Asia region (103 million), and the region of the Americas (58 million) [47]. The greater prevalence of moderate or higher levels of hearing loss in middle- and low-income countries are partially due to preventable conditions, such as impacted earwax, otitis media, use of ototoxic drugs and pre-natal or post-natal childhood infections [120, 72, 40]. Furthermore, the prevalence of hearing loss is higher in older adults. Indeed, 62% of people with hearing loss are older than 50 years old [47]. In 2019, age-related and other hearing loss was the fourth largest cause of global years lived with disability (YLDs), and the leading cause of global YLDs for individuals older than 70 years [53].

Partially due to the aging of the population, the prevalence of hearing loss is increasing over time. The percentage of people with hearing loss increased from 15.9% in 1990 to 20.3% in 2019 [47]. By 2050, it is estimated that 2.45 billion people (25% of the world population) will have hearing loss and that 689 million people (7% of the population) will have a moderate or higher level of hearing loss [47]. The greatest relative increase of people with a moderate or higher level of hearing loss is projected in the African region (154.9% increase from 2019), and in the Eastern Mediterranean region (138.4% increase from 2019) [47].

1.1.3 Consequences of Hearing Loss

Not only does hearing loss have high and increasing prevalence, but it also has profound consequences on a broad range of dimensions. Hearing loss has several repercussions at an individual level. First, people with hearing loss were found to possess lower physical abilities. Hearing loss is associated with lower balance scores [15], greater risk for falls due to poorer postural control [126, 69], incident mobility disability [12], greater risk to develop impaired lower extremity function and frailty syndrome [136], and greater difficulties in performing activities of daily living (ADL), which include basic tasks necessary for independent living [129, 12, 30]. Second, people with hearing loss were found to have a poorer psychological status. Hearing loss is associated with lower self-reported mental health and major depressive episodes [129], accelerated cognitive decline [43], and incident cognitive impairment [70]. Moreover, hearing loss is associated with incident dementia [71, 76] and is identified as the biggest modifiable dementia risk factor in midlife. If hearing loss is eliminated, a 8% reduction in dementia prevalence is estimated [73]. Third, people with hearing loss encounter greater difficulties in performing instrumental activities of daily living (IADL), which require complex thinking, as well as communication and organizational skills [136, 30]. Consequently, hearing loss is associated with difficulties in gaining education, higher unemployment rate, and lower grade of employment [133]. The severity of hearing loss is associated with lower household income [62] and reduced quality of life [30].

Additionally, hearing loss has repercussions on society. Indeed, people with hearing loss have a greater risk of incident hospitalization and greater annual rate of hospitalization [44]. This is reflected in higher total medical expenditures [38]. However, the cost of untreated hearing loss goes beyond health sector costs and also includes costs of educational support, loss of productivity, and societal costs [133]. The annual global cost of untreated hearing loss is estimated around US$ 980 billion [133]. The majority of this amount (57%) burdens low- and middle-income countries [133].

1.1.4 Diagnosis and Treatment of Hearing Loss

Throughout the life course, many of the causes that lead to hearing loss can be avoided through public health strategies and clinical interventions, such as immunization, ototoxicity prevention, and noise control [133]. When prevention is not possible, early identification and diagnosis are crucial. Hearing screenings are performed by healthcare providers with a variety of tools, depending on the condition and age of the patient. For instance, transient evoked otoacoustic emissions (TEOAEs) and automated auditory brainstem response (AABR) are most commonly used for newborns [56], while audiometric evaluation is usually performed on school-age children and adults [135].

Once a hearing impairment has been diagnosed, it is important to provide the patient with the appropriate treatment. While conductive hearing loss can often be treated with medicine or surgery, the effects of sensorineural hearing loss can be reduced through hearing technologies [5]. Different types of hearing technologies can be adopted, such as hearing aids and cochlear implants. Hearing aids are electronic devices that selectively amplify sound and are useful in treating hearing losses resulting from damage to the small sensory cells in the inner ear [90]. Cochlear implants are surgically implanted neuroprostheses that directly stimulate the auditory nerve and are useful when the middle- and inner-ear structures are damaged [89].

1.2 Hearing Aids

This section focuses on hearing aids, one of the ways to compensate for hearing loss. It provides a definition of hearing aids and an overview of the types of hearing aids available, the benefits of hearing aids, and the prevalence of hearing aid use. Furthermore, the main challenges posed by hearing aids are discussed.

1.2.1 Definition and Classification

Hearing aids are electronic devices that selectively amplify sound to compensate for a sensorineural hearing loss. Hearing aids are composed of three basic parts: a microphone that converts the sound into electrical signals, an amplifier that modifies and selectively amplifies the signals, and a speaker that reproduces the amplified signal [90]. By magnifying the acoustic sounds, hearing aids help the surviving hair cells sense the sound and convert it into electric impulses that are transmitted, through the hearing nerve, to the brain [36].

Different types of hearing aids exist, which differ on the size and placement of the main components. A behind-the-ear (BTE) device is worn behind the ear and transfers the sound to the ear canal via a plastic tube. A mini BTE, also known as receiver-in-the-ear (RITE) or receiver-in-the-canal (RIC), has a smaller size and a speaker placed in the ear canal. This hearing aid style offers reduced occlusion and feedback [39], is produced by all main manufacturers, and is usually equipped with advanced technology. An in-the-ear (ITE) device is custom made and fits completely into the outer ear. An in-the-canal (ITC) device is smaller than the ITE and fits into a smaller portion of the outer ear. A completely-in-the-canal (CIC) device is even smaller and is placed deeper into the ear canal. While being less visible, ITC and CIC devices have limited features and are only suitable for milder hearing losses [90].

1.2.2 Hearing Aid Features

Hearing loss affects several dimensions of hearing. Modern hearing aids do not simply amplify sound, but are equipped with several features to partly account for the changes occurred in the environment and in the sound perception of hearing impaired people. First, hearing impaired people have lower sensitivity to sounds [36]. To compensate for this, sounds are amplified according to the hearing threshold measured at different frequencies. Second, hearing impaired people have a reduced dynamic range, that is a smaller difference between the softest sound they can hear and the loudest sound they deem comfortable to listen to [36]. For this reason, non-linear amplification is provided, by amplifying soft sounds more than loud sounds. Third, hearing impaired people have reduced ability to locate sounds, since they cannot rely on normal binaural hearing [36]. Using two hearing aids improves hearing by reconstructing the full sound information [36].

Additionally, hearing aids have features aimed to adapt the amplification to the environment and to reduce the disturbance of background noise. These include advanced noise reduction algorithms and directional microphones, which suppress background noise and sounds coming from the back [90]. Modern hearing aids also adapt the noise reduction and the amplification to the environment, by providing more support in the complex situations. Finally, hearing aids have feedback management features aimed to reduce whistling due to acoustic feedback, a sound loop which occurs between the microphone and the speaker.

1.2.3 Benefits of Hearing Aids

Hearing aids have proven to be effective in reducing some of the negative consequences of hearing loss and in improving the quality of life of their users on several dimensions. First, hearing aid use is associated with better cognition [87] and lower deterioration of cognitive functions [78], either by improving audibility [32] or by reducing social isolation [108]. Second, using hearing aids appears to reduce the excess risk for dementia deriving from hearing loss [73]. A 25-year prospective study found increased dementia incidence in people with untreated hearing loss, but not in people using hearing aids [10]. Third, hearing aid use is associated with lower depression scores [87, 10], as well as higher social, emotional, and communication outcomes [87], greater listening ability and health-related quality of life [37]. Finally, treated hearing loss is associated with better performance at work. The difference in income between treated and untreated hearing impaired people increases at the rate of approximately $1,000 for every 10% increase in hearing loss severity [62].

1.2.4 Prevalence of Hearing Aid Use

Although 20% of the global population has hearing impairment, the prevalence of hearing aid use is lower and varies extensively from developing to developed countries. Globally, among the individuals who could benefit from hearing aids, only 17% use them [135].

In high-income countries, 40 million adults use hearing aids [120]. However, hearing aids are far from being adopted by all individuals who could benefit from them. In such countries, hearing aid use prevalence is low (6%) among individuals with mild hearing loss, and increases with the severity of the hearing impairment [120]. Among adults with a moderate or higher level of hearing loss, 43% use a hearing aid [120]. Among individuals with profound hearing loss, 89% use a hearing aid [120]. Moreover, among adults with moderate or higher level of hearing loss, the prevalence of hearing aid use increases with age [26]. In the United States, among hearing impaired adults aged 70 and older, 30% have ever used hearing aids; among adults aged 20 to 69, 16% have ever used hearing aids [91].

In developing countries, the prevalence of hearing aid use is even lower [120]. Although two-thirds of the hearing impaired people live in developing countries, only one in every eight hearing aids is sent to such developing countries [134].

1.2.5 Challenges with Hearing Aids

Barriers to Hearing Aid Use and Satisfaction

Different factors contribute to such a limited prevalence of hearing aid use. A major barrier to hearing aid uptake is the fear of stigma, which causes hearing aid owners to feel disrespected [113]. Indeed, hearing aids are often connected with old age and reduced cognitive abilities [31]. Another factor associated to hearing aid uptake is the real or perceived severity of hearing loss [54]. Individuals with milder hearing losses are less likely to own hearing aids, due to a lower sense of need and urgency. A third barrier is represented by the cost of hearing aids, although the worthiness of hearing aids might play a role in evaluating their cost [54].

Among the hearing aid owners, a substantial percentage does not wear the devices. Such percentage ranges from 5% in Germany, 6% in France, 7% in United Kingdom, to 12% in United States [51], and 24% in Australia [48]. One of the main reasons for not wearing the hearing aids is related to problems in perceiving the hearing aid value. Hearing aid owners often mention difficulties in coping with noisy situations, as well as poor benefit from the hearing aids [81].

Among the hearing aid owners who use their hearing aids, overall satisfaction is generally high [107]. Satisfaction with hearing aids is positively correlated with variables related to the user such as time of use and previous hearing aid experience, and, to a lower extent, it depends on expectations and personality [132]. Additionally, satisfaction is correlated with some variables related to the device. Hearing aids that have better sound quality are associated to higher satisfaction, while problems in hearing aid use have negative effects on satisfaction [132]. Moreover, satisfaction significantly depends on the listening situation. Hearing aid users tend to be less satisfied in noisy situations, while hearing aids that can perform in complex situations yield higher satisfaction [132].

All in all, enhancing the value provided to hearing aid users and increasing the hearing aid availability worldwide could improve the prevalence of hearing aid use, as well as the proportion of hearing aid owners who use their device, and the hearing aid satisfaction in certain situations. Addressing a major challenge - the lack of personalization in hearing healthcare - is crucial for progressing towards such direction.

One Hearing Aid Configuration Does Not Fit All Users

Sensorineural hearing loss is often caused by a damage of the hair cells, causing a degraded sound information to be transmitted to the auditory nerve [105]. Additionally, in age-related hearing loss this coincides with a natural decline in cognitive abilities [105]. Therefore, hearing impaired older adults have to process a degraded signal by using declined cognitive skills [105]. This implicates that speech processing regions are complemented by additional neural networks, and that even a mild hearing loss has profound effects on neural activities [105].

Since the auditory and cognitive domains are intertwined, hearing aid users with similar hearing pure-tone thresholds (PTTs) have been shown to perceive sounds in highly individual ways. First, individuals with similar hearing thresholds have different performances in noise. That is, the signal-to-noise ratio required to understand speech is poorly predicted by the audiometric loss (i.e., audiogram) [61]. Indeed, while the outer cells are responsible for the sensitivity to sounds, the inner hair cells convey information to the brain [60]. Depending on whether the loss interests inner or outer cells, the repercussions on speech understanding might be different [60]. Second, individuals with similar hearing thresholds have different loudness perception [96]. That is, the interindividual loudness perception of binaural broadband signals is poorly predicted by the audiometric loss [96]. Indeed, the audiometric loss is inferred from monaural narrowband signals, while real-world signals such as speech or environmental sounds are typically broadband and binaural [96]. Third, hearing impaired people widely vary in their perception of sounds close to their hearing threshold [80]. While some perceive such sounds to be soft, others perceive them to be loud [80]. Additionally, hearing thresholds fail to predict hearing handicap as they cannot account for several factors, such as motivation, linguistic backgrounds, and age of hearing loss onset [27].

Lack of Audiological Personalization

In summary, performance in noise, loudness perception, perception of soft sounds and hearing handicap cannot be predicted by the patient's pure-tone thresholds.

Despite that, hearing aids address the lack of sensitivity to sounds by prescribing amplification at different frequencies based on pure-tone threshold audiometry, a test that measures the hearing thresholds for pure tones at different frequencies [128]. The results of such test are usually recorded in graphic form by presenting the audible thresholds in an audiogram [6]. Pure-tone threshold audiometry does not provide specific information on the status of the central auditory nervous, nor the auditory processing of real-world signals [88, 97]. Therefore, it does not represent real-world hearing abilities [61, 14, 97].

Traditionally, based on the pure-tone audiometric loss of the patient [96], the amplification is prescribed using a fitting formula, also known as fitting rationale. Different fitting rationales exist, either developed by hearing aid manufacturers (e.g., VAC+ [98]) or by independent organizations (e.g., NAL-NL2 [58]). Some fitting rationales additionally take into account perceptual dimensions such as the narrowband loudness perception, by considering a measured or predicted Uncomfortable Loudness Level [114], i.e., the level at which a sound becomes uncomfortable for the listener. However, such measurements are not always effective as they rely on narrowband monaural sounds and do not control for loudness perception of real-world binaural broadband sounds [96]. Some fitting rationales also take into account the profile of the hearing aid user (e.g., age, gender, experience) [58]. In general, when a rationale considers dimensions other than the audiometric loss, it assumes that all patients with the same audiometric profile share the same perceptual characteristics [114]. Relying on average corrections does not account for the large individual perceptual variations [96, 114].

The major implication is that the initial prescription should be regarded as a good starting point, because it is based on average expectations [114, 2]. However, it should not be seen as the optimal solution, as it is not based on individual needs or preferences [114, 2]. Assuming that the manufacturer defaults are correct for each patient is regarded as one the key mistakes made by hearing care professionals [63]. A fine-tuning of the hearing aid is recommended, during which the hearing care professional modifies the hearing aid settings based on patient's input [63]. A successful fine-tuning, as well as patient's benefit and satisfaction, depends on the hearing care professional's ability to use the programming software, and to interpret and translate users' recollections of past listening experiences [35, 11]. This is a time-consuming procedure, requiring multiple visits to obtain a satisfactory configuration [2]. Yet, it is not an optimal procedure, since fine-tuning and additional hearing tests performed in the clinic do not guarantee a significant advantage over a default prescription [29, 115].

One Hearing Aid Configuration Does Not Fit All Contexts

Furthermore, the satisfaction of hearing aid users significantly depends on context. Users cope with several real-world situations and often report difficulties in noisy environments [132]. Previous studies have shown that an important factor for the lack of adoption of hearing aids is their unsatisfactory performance in noisy environments [20, 48, 1], in conversations [121], and multi-talker scenarios [42]. When provided with different hearing aid settings, users have been found to select contrasting settings, suggesting that context has a crucial impact on the preference toward specific hearing aid settings [55]. Typically, to account for context, hearing users can be provided with predefined settings aimed at improving the listening experience in specific listening environments. However, these settings are determined by the average user and do not reflect the individual audiological preferences.

Lack of Comprehensive Personalization

In addition to personalized audiological settings, hearing aid manufacturers face the challenge of providing a comprehensively personalized solution to patients. The surge in online interactions has exposed consumers to the personalization practices of e-commerce companies and has raised expectations for any product and service [13]. Nowadays, the great majority of consumers (71%) expects companies to deliver personalized interactions, and an even greater percentage of consumers (76%) gets frustrated when that does not happen [13]. Hearing aid patients can potentially benefit from more personalized interactions in different steps of their journey: recommendations during the hearing aid selection, fitting, and use; maintenance advice; counseling during the rehabilitation phase. Developing a deep understanding of the needs of hearing aid users and offering a personalized solution throughout these phases would create value for hearing aid users.

1.3 Paradigm Shift in Hearing Healthcare

This section discusses how promoting user-driven personalization and leveraging real-world data might constitute a paradigm shift in hearing healthcare and enable addressing the current challenges with hearing aids.

1.3.1 User-driven Personalization

Traditionally, the hearing care professional has been the main point of contact for hearing aid users, being responsible for evaluating the patient's hearing loss, recommending the appropriate device, personalizing the hearing aid settings, counselling the patient, and ultimately ensuring a good experience with the device. While the role of the hearing care professional is crucial for providing high-quality hearing healthcare, uniquely relying on such role for personalizing the experience of hearing aid users has several limitations.

First, the health system capacity suffers from a lack of human resources. The vast majority of low-income countries (93%) and lower-middle-income countries (76%) have fewer than one audiologist per million inhabitants [135]. While the availability of audiologists increases in higher income countries, 86% of upper-middle-income countries and 35% of high-income countries still have fewer than ten audiologists per million inhabitants [135]. This restricts the access to hearing health care and poses challenges for hearing impaired people [135]. Second, the lack of human resources places unreasonable demands on the existing hearing care professionals providing these services [135]. As a consequence, hearing care professionals might streamline their workflow and resort to non-personalized approaches, such as default audiological settings. Third, even when they have access to a hearing care professional, hearing aid users are often hesitant to seek help, so the need of a patient might never be reported nor addressed [17]. Fourth, while the hearing care professionals can potentially establish a fruitful dialogue with the patient and obtain information about her needs and preferences, it is not always possible to leverage on data obtained in the clinic. Indeed, the hearing aid user might not be able to provide accurate data. For instance, reporting real-world listening experiences, audiological preferences and past behavior might suffer from recall bias [67]. Moreover, the information reported by the patient might require some interpretation and translation into an audiological language. However, such information might lack a faithful contextualization, making an effective translation difficult.

For these reasons, in contrast to the current clinical practices, patient empowerment and self-management should be promoted [17]. Empowering hearing aid users would enable gathering more accurate and contextualized data on their preferences and behavior, thereby developing a deeper understanding of their needs and ultimately offering them a truly personalized solution.

1.3.2 Learning User Preferences and Behavior from Data

mHEalth, i.e., the use of mobile devices for health practice, can potentially enhance the access to hearing aids and improve their use. This can be achieved by empowering hearing aid users throughout their journey, while also more efficiently learning their preferences. Indeed, the introduction of smartphone-connected hearing aids enables, on one hand, enhancing the experience of hearing aid users by providing them with targeted information and with the opportunity to adjust their device settings. On the other hand, it enables a better understanding of users' preferences and behavior, by gathering accurate real-world data about their hearing aid use, their interactions with it, and the context. Such data can either be subjective or objective. Subjective data consists of the patient's perceived experiences or subjective preferences, shared either spontaneously or in response to questions. Objective data refers to observable and measurable data, gauged by the smartphone or by the hearing aids.

Fundamentally, user preferences and behavior can be learned in two ways. First, experimental studies can be conducted by establishing an intervention, enrolling some test users, measuring specific variables and analyzing the effects of the intervention [124]. Such studies enable exploring new trajectories and evaluating novel ways of interaction with the hearing aids. When aiming to learn personalized settings, these studies allow to investigate new device configurations and to gather specific data about the listening experience while the patient is trying the settings. However, such studies can be time-consuming and place a burden on the user. Indeed, for instance, test users might be asked to use new hearing models, experiment with new features or settings, and repeatedly report their experience. Second, observational studies can be conducted based on the commercially available products (e.g., apps and hearing aid models). The widespread adoption of smartphones among older adults [94] and the introduction of smartphone-connected hearing aids enables logging data while hearing aid users autonomously use their devices. Participants in these types of studies are simply observed in a more natural setting, and independent variables are allowed to vary naturally, without interfering with the standard clinical practice [125]. Such studies enable inferring the behavior and preferences of hearing aid users by observing how they naturally use their hearing aids or interact with them. By analyzing data coming from commercially available products, these studies can leverage a larger and more diverse sample of patients. However, they can only rely on data commonly gathered in the standard clinical practice.

1.4 Research Objectives

Providing a comprehensively personalized hearing aid solution is a complex and multidimensional challenge and requires a deep understanding of patients' preferences and behavior. This thesis leverages real-world data collected through smartphone-connected hearing aids to address two main themes.

First, personalizing hearing aid settings requires learning the audiological preferences of the users and poses several challenges. To address this theme, three specific research objectives were formulated:

- *Investigate the feasibility of a smartphone-based method to gather the audiological preferences of hearing aid users in real-world listening environments.* Indeed, users' perceptual preferences and environmental context need to be measured. This requires making users try and evaluate different settings by navigating a complex audiological space.
- *Investigate the feasibility of a context-aware system for providing hearing aid users with a number of relevant hearing aid settings to choose from.* Indeed, the gathered preferences should be meaningful and the offered settings should improve the listening experience of the patient. By leveraging contextual data and experience assessments, it might be possible to make more informed decisions on which settings to provide the user with.
- *Explore how a conversational agent could combine real-world user feedback and context to recommend personalized settings.* Indeed, considering the lack of audiological resources, alternative ways of gathering user preferences should be evaluated.

Second, offering a comprehensively personalized solution, as well as transferring the learned audiological preferences to new or inactive users, requires learning users' behavior. To address this theme, two specific research objectives were formulated:

- *Explore patterns of hearing aid use throughout the day and assess whether clusters of users with similar use patterns can be identified.* Indeed, exploring how patients use their hearing aids can help shed light on their needs and on the similarity and variability among them.
- *Investigate the provision and context of use of listening programs currently available in the market.* Indeed, investigating whether users contextually adapt the device settings in specific listening situations can pave the way for more personalized solutions.

1.5 Thesis Structure

The remainder of this thesis is structured as follows. Chapter 2 summarizes the studies aimed to understand the audiological preferences of hearing aid users. Chapter 3 summarizes the studies aimed to understand the real-world behavior of hearing aid users. The research objectives addressed in Chapter 2 and 3, as well as the respective methods and main findings, are detailed in Table 1.1. Chapter 4 discusses the main findings, their implications and the perspectives for future research. Chapter 5 concludes the thesis.

Table 1.1: Detailed structure of Chapter 2 and Chapter 3

Section	Objective	Methods	Main Findings
	Chapter 2: *Understanding Users' Audiological Preferences*		
2.1 Measuring Audiological Preferences	To investigate the feasibility of a smartphone-based method to gather the audiological preferences of hearing aid users in real-world listening environments.	During their everyday life, seven participants were asked to optimize three audiological parameters which were subsequently combined into a personalized device configuration. This configuration was blindly compared against a configuration personalized in a standard clinical workflow.	Six out of seven participants preferred the device configuration learned in real-world listening environments.
2.2 Modeling Audiological Preferences	To investigate the feasibility of a context-aware system for providing hearing aid users with a number of relevant hearing aid settings to choose from.	During their everyday life, seven participants were asked to optimize three audiological parameters, by evaluating four intervention levels for each parameter. The listening experience, audiological preferences, and contextual data were collected as both self-reports and objective data logging.	1. Having access to different intervention levels of two audiological parameters affected listening satisfaction. 2. The perceived usefulness of having access to different intervention levels was significantly modulated by context. 3. Contextual data improved the prediction of intervention level preferences.
2.3 A Future Use Case	To explore how a conversational agent could combine real-world user feedback and context to recommend personalized settings.	A conversational agent model was outlined and two use cases were proposed: troubleshooting and contextual personalization.	Implementing a conversational agent potentially allows to automatically gather user feedback in real-world environments, while monitoring the context, in order to recommend personalized settings.
	Chapter 3: *Understanding Users' Real-World Behavior*		
3.1 How Do Users Use the Hearing Aids?	To explore patterns of hearing aid use throughout the day and assess whether clusters of users with similar use patterns can be identified.	We analyzed 453,612 logged days of objective hearing aid use logged from 15,905 real-world users. We explored the daily amount of hearing aid use, identified typical days of hearing aid use based on hourly patterns, and then clustered users based on use patterns. Finally, we validated the user clustering by training a supervised ensemble to predict user clusters.	1. On average, users used the hearing aids for 10.01 h/day, exhibiting a substantial between-user and within-user variability. 2. Three typical days of hearing aid use were identified. 3. Three distinct user groups were found, each characterized by a predominant typical day of hearing aid use. The supervised ensemble achieved an 86% accuracy.
3.2 How Do Users Personalize the Hearing Aids?	To investigate the provision and context of use of listening programs currently available in the market.	We explored the provision of listening programs among 32,336 hearing aid users. We analyzed 396,723 program selections from 1,312 users to investigate the sound environments in which specific programs are used.	1. 57% of users had additional listening programs for specific listening situations. We identified a default program, an additional primary program, two additional secondary programs, and two programs related to the use of external accessories. 2. Users used the listening programs in sound environments different than the sound environment measured when using the default program.

Chapter 2

Understanding Users' Audiological Preferences

In this chapter, the audiological preferences of hearing aid users are gathered in real-world environments via a smartphone-based method (Section 2.1). Such preferences are then modeled to investigate the feasibility of a context-aware system for providing hearing aid users with a number of relevant hearing aid settings to choose from (Section 2.2). Finally, the chapter explores how a conversational agent could combine user feedback and context to recommend personalized settings (Section 2.3).

This chapter presents the contribution of the articles entitled "Rethinking Hearing Aids as Recommender Systems" [104] (Appendix A), "Measuring and Modeling Context-Dependent Preferences for Hearing Aid Settings" [103] (Appendix B), and "Designing Audiologist Bots Fusing Soundscapes and User Feedback" [101] (Appendix C). All figures reported in this chapter (except Figure 2.1) are extracted from the aforementioned three articles.

2.1 Measuring Audiological Preferences

This section discusses the challenges posed by measuring audiological preferences and investigates the feasibility of a smartphone-based method to gather the audiological preferences of hearing aid users in real-world environments.

2.1.1 Related Challenges

Previous research conducted under controlled laboratory conditions has shown that user-driven adjustments of hearing aid settings are feasible and potentially beneficial [137, 22, 93, 116]. Some studies have also focused on personalizing the hearing aid settings based on user feedback and contextual information gathered in the real-world [4, 59]. However, results were mixed and some open challenges remain.

Since audiological preferences depend on the human perception of sounds and are highly individual, it is important to repetitively gather subjective perceptual preferences by enabling users to evaluate alternative hearing aid settings. This means that users need to be able to effectively explore a complex audiological design space, defined by the possible combinations of hearing aid settings. Moreover, users need to be able to effectively report on their preferences by submitting timely and relevant feedback.

When gathering users' audiological preferences, real-world context should be taken into consideration. Indeed, users' audiological preferences might vary depending on the listening environment [55]. Moreover, empowering users to self-adjust their hearing aid settings in real-world contexts ensures that they evaluate them in a realistic context, which is highly representative of their present and future listening challenges. This entails that users might be more engaged in interacting with the hearing aids, and that the gathered listening preferences might be more relevant for their listening experience. However, gathering audiological preferences in real-world contexts increases the complexity of the task, as the user has to personalize the device settings in a sound environment which is constantly changing, while performing other activities (e.g., conversing), and without the assistance of a hearing care professional.

Reducing the complexity of the exploration is therefore crucial. Assuming that the audiological design space is defined by several audiological parameters, different methods could be employed to enable users exploring it. Via *parameter tweaking*, users are asked to tune the hearing aid settings by acting on a set of either continuous or discrete parameters. While this method provides users with a predictable set of handles, acting on several parameters at a time might be overwhelming. Via *pairwise comparison*, users can choose between two alternative configurations. Although this method provides users with a low and manageable number of configurations at a time, the alternative configurations are not predictable and several iterations are required to obtain the preferred configuration. Via a *one-dimensional parameter tweaking*, users can act on one discrete or continuous parameter at a time. This method allows to reduce the complexity of a multi-parameter configuration, while ensuring that participants can consciously track the effects of their actions on their listening experience.

2.1.2 An Experimental Study

A study was conducted to investigate the feasibility of a smartphone-based method to gather the audiological preferences of hearing aid users in real-world listening environments. Seven experienced hearing aid users were enrolled. They had a mild to moderately severe hearing loss. The participants received a pair of Oticon Opn S1 MiniRITE, and installed a custom app on their smartphone (Figure 2.1).

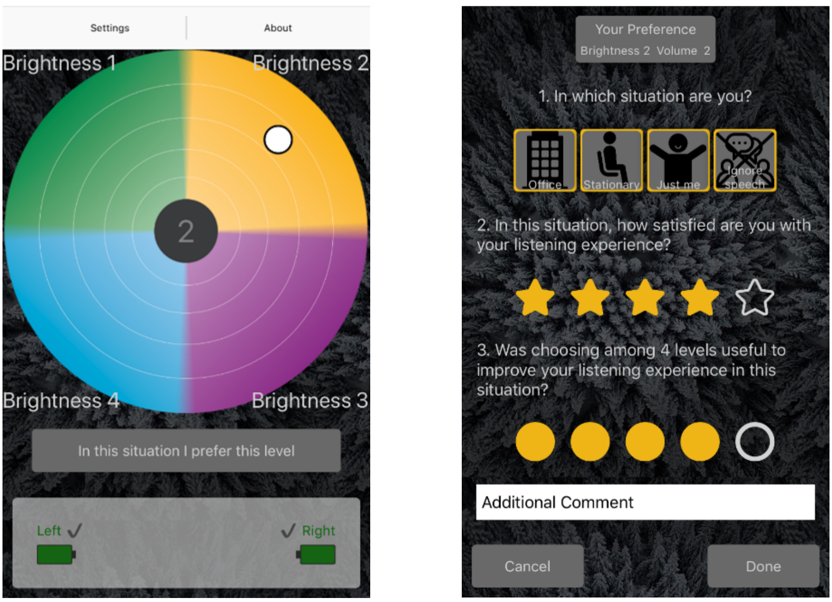


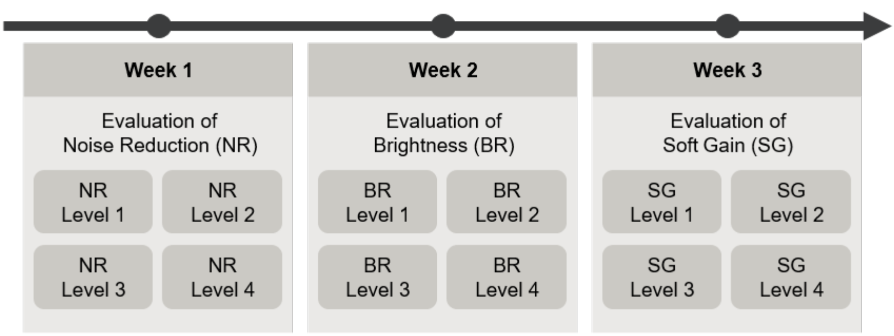
Figure 2.1: Interface of the app used by the study participants. The app enabled them to explore different hearing aid settings (left), while also gathering their feedback and data about the context (right).

The app was connected to the hearing aids via Bluetooth and had a threefold function:

- to enable the participants to explore different hearing aid settings. One audiological parameter was evaluated each week, and the participants were provided with four intervention levels of the specific parameter of the week.
- to gather data about the listening experience, be it self-reported (listening satisfaction, from 1 to 5; usefulness of choosing among the four intervention levels, from 1 to 5; preferred intervention level) or automatically logged (level selections and usage).
- to gather data about the context, be it self-reported (listening environment, listening intent, motion state - all from predefined categories) or automatically logged (sound pressure level (SPL), signal-to-noise ratio (SNR) - both in dB). The SPL is the most commonly used indicator of the acoustic wave strength and correlates well with human perception of loudness [74]. The SNR is the difference between the energy of a signal and the energy of any present noise and it is key to speech intelligibility [75].

The study lasted four weeks. Each of the first three weeks was devoted to the evaluation of an audiological parameter, and the participants were provided with four intervention levels of that parameter, in ascending order of intensity, from level 1 to 4 (Figure 2.2).

Figure 2.2: Data collection timeline. Each parameter (NR, BR, SG) was evaluated for the duration of one week. Each week, the participants were provided with four intervention levels of the parameter of the week (Appendix B).



The three parameters were: Noise Reduction (NR), providing four levels of noise reduction and directionality; Brightness (BR), adjusting the amplification of high frequency sounds; Soft Gain (SG), adjusting the amplification of soft sounds. These parameters have all been shown to be important for the listening experience of hearing aid users [95, 55, 131] and to be perceived differently by individuals [61, 80]. Each week, the participants were asked to evaluate the four intervention levels of the parameter of the week during their everyday life and to submit their preferred level, together with an evaluation of their listening experience and an assessment of the context.

In order to evaluate the feasibility of the employed method, during the fourth week, participants compared two configurations in a blind test: a configuration individually personalized based on data gathered in real-world environments, by combining the preferred levels of the three audiological parameters gathered during the previous three weeks; a configuration personalized in a standard clinical workflow based on questions and on pairwise comparisons of prerecorded sound samples.

2.1.3 Could We Gather Audiological Preferences?

Overall, when looking at the preferences for the different intervention levels of the three audiological parameters, a substantial between- and within-participant variability was observed. Indeed, the participants had different audiological preferences among each other. Moreover, when evaluating the four intervention levels in different situations, the participants were not found to strive for a single optimum, but rather selected different levels within each parameter. Consistently with this, after the fourth week, participants wished to keep more than one configuration after the end of the study. However, when asked to choose between the configuration based on the feedback gathered in the real-world and the configuration based on the standard clinical workflow, six out of seven participants preferred the former. The smartphone-based method gathered more faithful audiological preferences compared to the standard in-clinic personalization procedure based on asking users to answer questions and evaluate prerecorded sounds. These results suggest that, by using the proposed smartphone-based method, the participants efficiently explored the audiological design space in real-world environments and successfully communicated their preferences.

2.2 Modeling Audiological Preferences

In this section, the gathered audiological preferences are modeled to investigate the feasibility of a context-aware system for providing hearing aid users with a number of relevant hearing aid settings to choose from.

2.2.1 The Relevance of Audiological Preferences

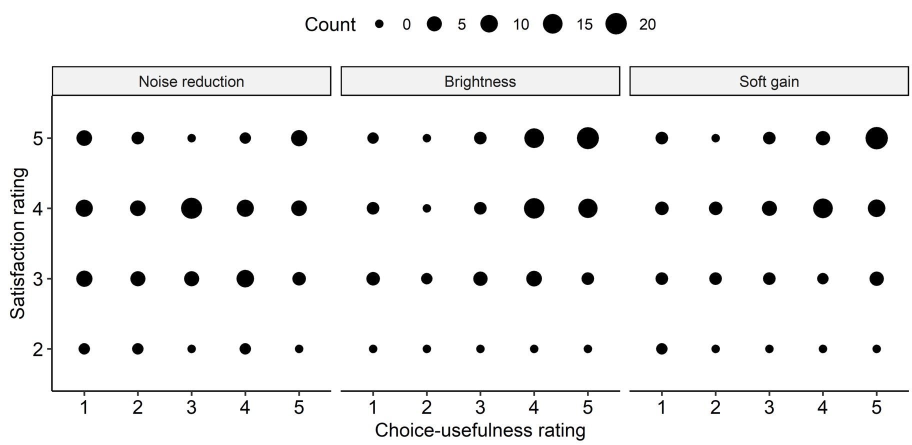
As discussed above, audiological preferences need to be gathered in several real-world contexts by making users explore a vast space of possible hearing aid settings. For this reason, not only should the complexity of the exploration be reduced, but it is also important to focus on settings and contexts that cause a tangible improvement in the listening experience. Indeed, previous work has assumed that users have one and only one preference in each context. However, when choosing between two alternative hearing aid settings in a specific sound environment, some users are not consistent in their reported preferences [130]. This inconsistency might be due to the fact that a selected setting does not yield a significant improvement in the listening experience and therefore leads to noisy preference assessments. Thus, understanding when a preferred setting is perceived to truly improve the listening experience would help focusing on the relevant audiological parameters and contexts.

In view of the above, in the following sections, the collected subjective evaluations of the listening experience, as well as the audiological preferences, were analyzed. In doing so, the context could explain part of the variation among the perceptual evaluations, being informative of whether a parameter affects the listening satisfaction or of which intervention level of an audiological parameter is preferred. Although users might have individual perceptions and preferences, such contextual effect on their individual preferences might be similar across users.

2.2.2 Relationship Between Choice-Usefulness and Listening Satisfaction

A user-adaptive system should be able to provide a choice among settings that, when relevant, leads to higher user satisfaction. When evaluating the four levels of an audiological parameter, participants rated the usefulness of having four levels to choose from (i.e., choice-usefulness), as well as their satisfaction with the listening experience. Thus, we analyzed if listening satisfaction is related to the perceived usefulness of an audiological parameter (Figure 2.3). Having access to useful levels of Noise Reduction did not affect listening satisfaction ($r = 0.064$, $t = 0.637$, $df = 96$, $p = 0.526$), while having access to useful levels of Brightness ($r = 0.383$, $t = 3.908$, $df = 89$, $p < 0.001$) and Soft Gain ($r = 0.400$, $t = 3.830$, $df = 77$, $p < 0.001$) significantly affected listening satisfaction. Furthermore, the effect of context was analyzed by computing the correlation between choice-usefulness and satisfaction separately for sound environments with high and low quality (i.e., SNR levels respectively above and below the median). For the Brightness parameter, a significant correlation was only found in low quality sound environments ($r = 0.443$, $t = 3.129$, $df = 40$, $p = 0.003$), while for the Soft Gain parameter a higher correlation was found in high quality sound environments ($r = 0.572$, $t = 3.886$, $df = 31$, $p < 0.001$) compared to low quality sound environments ($r = 0.366$, $t = 2.080$, $df = 28$, $p = 0.047$).

Figure 2.3: Contingency tables for ratings of choice-usefulness (x-axis) and listening satisfaction (y-axis), for each audiological parameter. Having access to useful levels of Noise Reduction was not associated with higher listening satisfaction, while having access to useful levels of Brightness and Soft Gain was associated with higher listening satisfaction (Appendix B).



2.2.3 Contextual Impact on Choice-Usefulness

A user-adaptive system might leverage contextual information to provide the hearing aid user with more relevant parameters. Thus, we tested the hypothesis that context has a distinct impact on the usefulness of choosing among different intervention levels of the three parameters. This would make it possible to reduce the space of possible hearing aid settings by assigning context-aware usefulness to the audiological parameters.

We applied a model for the cumulative probability of the i^{th} choice-usefulness rating falling in the j^{th} category or below:

$$logit(P(Y_i \leq j)) = \theta_j - \beta_1(env_i) - \beta_2(intent_i) - \beta_3(SPL_i) - \beta_4(SNR_i) - u(ID_i),$$
$$i = 1, ..., n, j = 1, ..., J - 1$$

where i indexes all observations, and $j = 1, ..., J$ indexes the choice-usefulness categories ($J = 5$). The continuous SPL and SNR variables were converted into categorical variables (respectively "Low Intensity", "High Intensity", and "Low Quality", "High Quality"). The participant effect (ID) was included as random effect. The self-reported listening environment (env) and intention ($intent$), and the objective contextual predictors (SPL, SNR) were included as fixed effects.

Across the three audiological parameters, the contextual predictors significantly increased the prediction of choice-usefulness ratings (likelihood ratio test, $\chi^2(6) = 21.71$, $p = 0.002$). Subsequently, three separate models were fitted, one per audiological parameter. The model coefficients (Figure 2.4a) indicate that the Noise Reduction parameter is perceived to be more useful in noisy environments (both indoor and outdoor); the Brightness parameter is perceived to be more useful in quieter environments (lower SPL, i.e., lower intensity) and when having a "Focus" listening intention; the Soft Gain parameter is perceived to be more useful when having a "Social" listening intention.

2.2.4 Contextual Impact on Audiological Preferences

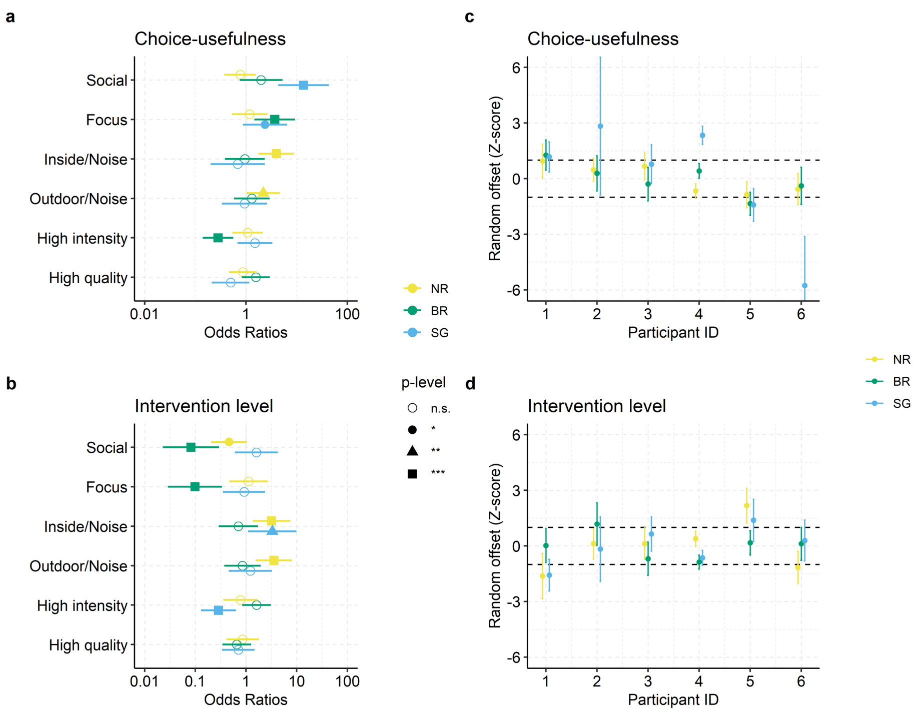
A user-adaptive system might also benefit from contextual information when predicting the preferred level of a parameter in a specific situation. Thus, we tested the hypothesis that context has a distinct impact on the preferred level of the parameters.

First, we applied a model for the cumulative probability of the i^{th} explicit self-reported preference falling in the j^{th} level or below:

$$logit(P(Y_i \leq j)) = \theta_j - \beta_1(env_i) - \beta_2(intent_i) - \beta_3(SPL_i) - \beta_4(SNR_i) - u(ID_i),$$
$$i = 1, ..., n, j = 1, ..., J - 1$$

where i indexes all observations, and $j = 1, ..., J$ indexes the parameter levels ($J = 4$). As in the previous model, the participant effect (ID) was included

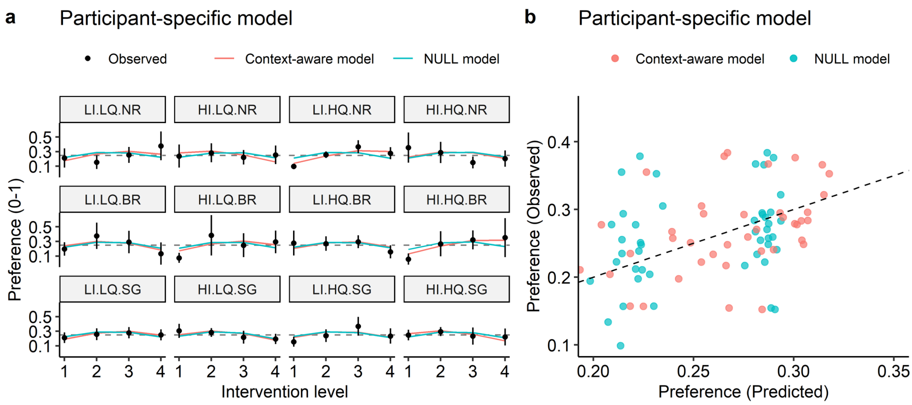
Figure 2.4: In (A) and (B), coefficients and 95% confidence intervals for predicting choice-usefulness and explicit level preference. In (C) and (D), the corresponding random offsets due to participant effects. The baseline conditions for the contextual predictors were: "Only me" (for listening intention), "Quiet/Indoor" (for listening environment), "Low intensity" (for SPL), "Low quality" (for SNR) (Appendix B).



as random effect, while the self-reported listening environment (env), listening intention ($intent$), and the objective contextual predictors (SPL, SNR, previously converted into categorical variables) were included as fixed effects.

Across the three audiological parameters, the contextual predictors significantly increased the prediction of explicit level preferences (likelihood ratio test, $\chi^2(6) = 14.418$, $p = 0.025$). Subsequently, three separate models were fitted, one per audiological parameter. The model coefficients (Figure 2.4b) indicate that higher levels of Noise Reduction were preferred in noisy environments (both indoor and outdoor), while lower levels were preferred when having "Social" listening intentions; higher levels of Brightness were preferred when having an "Only me" listening intention; higher levels of Soft Gain were preferred in quieter environments and in noisy environments indoor.

Figure 2.5: Observed and predicted implicit preference for intervention level for the most active participant. In (A), model predictions are shown together with the observed relative preference for each intervention level, grouped by all combinations of SPL and SNR (columns) and separated by audiological parameter (rows). LI=Low Intensity; HI=High Intensity; LQ=Low Quality; HQ=High Quality. In (B), the difference between predicted and observed preference is shown as a scatter plot with the dashed line indicating a $y = x$ relationship (Appendix B).



In addition to the explicitly self-reported level preferences, implicit preferences can be inferred from the selections of the intervention levels. Indeed, participants made ~ 8 times more active level selections (i.e., level selections that are set for at least three minutes) than explicit preference submissions. If a system could rely on implicit preferences, the training phase would be less burdensome for the user. The same model was fitted for implicit level preferences, with the only difference that context was only represented by the objective contextual predictors (SPL, SNR). The model was first fitted to data from all participants. As for the explicit preferences, the contextual predictors significantly increased the prediction of implicit level preferences ($\chi^2(8) = 17.43$, $p = 0.026$). Subsequently, the model was fitted to data from the most active participant. Also in this case, the contextual predictors significantly increased the prediction of implicit level preferences ($\chi^2(8) = 19.97$, $p = 0.010$). Figure 2.5 shows the predictions of both a null model and a context-aware model for the most active participant. The latter produced predictions closer to the observed preferences

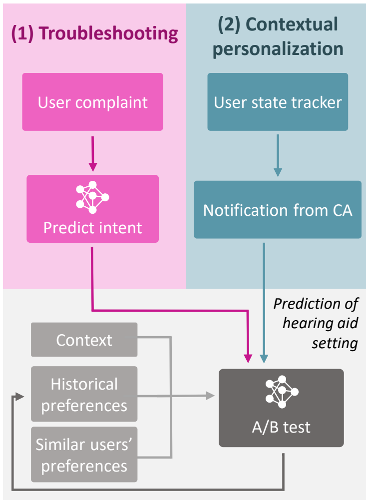
(Person's correlation; null model: $r = 0.45$, 95%, $CI = [0.19, 0.65]$, $df = 46$, $p = 0.001$; context-aware model: $r = 0.67$, 95%, $CI = [0.48, 0.80]$, $df = 46$, $p < 0.001$).

2.3 A Future Use Case

The results presented above suggest that offering hearing aid users to personalize the device settings in real-world environments can potentially lead to higher listening satisfaction. Moreover, context plays an important role in determining the usefulness of the audiological parameters that should be adjusted, as well as the preferences for specific levels of the parameters. However, gathering feedback in real-world environments is time consuming and requires multiple interactions with the device, which potentially can improve the long term overall experience, but at the cost of short term user experience. A solution that might accelerate this phase is the implementation of a conversational agent that autonomously gathers user feedback in real-world environments and collects information about the context to learn users' audiological preferences. Such a conversational agent could combine natural language understanding with sequential patterns of contextual features to predict the most likely preferred hearing aid setting. A conversational agent could address different user needs by interacting with the hearing aid user in two ways (Figure 2.6).

First, a conversational agent might help troubleshooting users' problems and fine-tune newly acquired hearing aids during the trial phase. In this use case, the interaction is initiated by the user who expresses a complaint. The complaint is translated into an audiological intent, while contextual information is collected. Based on the user's complaint and on contextual features, the agent proposes a setting adjustments. Second, a conversational agent might monitor both the current device settings and the context and, in some situations, proactively suggest a comparison between two alternative settings. In both use cases, the dialogue with the user enables gathering immediate user feedback on the proposed adjustment and therefore progressively learning the best hearing aid settings to recommend. The track outlined in this section was not prioritized internally over other activities and was therefore put on indefinite hold.

Figure 2.6: Overview of two different conversational agent use cases with different objectives: Troubleshooting (1) and Contextual personalization (2) [101].



2.4 Chapter Summary

In this chapter, seven participants optimized, via a smartphone-based method, three audiological parameters in real-world environments and their audiological preferences, as well as their context, were gathered. When combining such preferences into a personalized device configuration and comparing it against a configuration personalized in a standard clinical workflow, six out of seven participants preferred the former configuration. The collected preferences were then modeled to investigate the feasibility of a context-aware system for providing hearing aid users with a number of relevant hearing aid settings to choose from. We found that having access to different intervention levels of two audiological parameters affected listening satisfaction. Moreover, context significantly modulated the perceived usefulness of having access to different intervention levels, as well as the intervention level preferences. Finally, the chapter explored how implementing a conversational agent potentially allows to automatically gather user feedback in real-world environments, while monitoring the context, in order to recommend personalized settings.

Chapter 3

Understanding Users' Real-World Behavior

In this chapter, the real-world behavior of hearing aid users is investigated based on objective data logged by commercially available hearing aids. In particular, Section 3.1 investigates how users use their hearing aids, by exploring patterns of hearing aid use throughout the day and assessing whether clusters of users with similar use patterns can be identified. Section 3.2 investigates how users personalize their hearing aids, by analyzing the provision and context of use of listening programs currently available in the market.

This chapter presents the contribution of the articles entitled "Clustering Users Based on Hearing Aid Use: An Exploratory Analysis of Real-World Data" [100] (Appendix D) and "Investigating the Provision and Context of Use of Hearing Aid Listening Programs from Real-World Data" [102] (Appendix E). All figures reported in this chapter are extracted from the aforementioned two articles.

3.1 How Do Users Use the Hearing Aids?

This section investigates how users use their hearing aids in the real-world, by analyzing the amount of hearing aid use, patterns of hearing aid use throughout the day, and assessing whether clusters of users with similar use patterns can be identified.

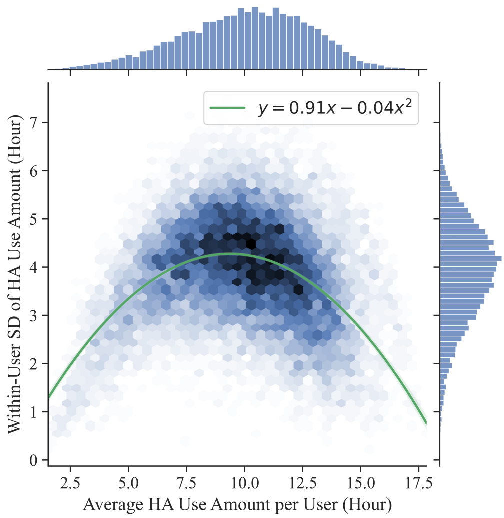
3.1.1 Investigating Hearing Aid Use from Real-World Data

The success of hearing aids in mitigating the effects of hearing loss depends on the intervention provided to the patient, but also on the patient compliance with the intervention program [106]. Investigating hearing aid use can provide insight into patient compliance, but also deepen the understanding of patients' needs and behavior. Most of the existing studies assessing hearing aid use are based on self-reported measures, such as interviews and questionnaires, while only a few studies are based on objective measures, such as data logging and battery consumption [106]. However, self-reported measures have been found to cause overreport of hearing aid use [41, 67, 117, 79]. Smartphone-connected hearing aids make it possible to log objective data about hearing aid use. In addition to avoiding overreport, objective measures enable assessing the hearing aid use of a large number of users and with a greater temporal resolution [34].

3.1.2 Amount of Hearing Aid Use

When evaluating hearing aid use, the amount of hearing aid use is traditionally regarded as an indicator of treatment success [66] and often analyzed [122, 20, 79, 117, 123]. We investigated the daily amount of hearing aid use and its within-user and between-user variability, by analyzing 453,612 days logged from 15,905 real-world users over a 4-month period. The average user had an average amount of hearing aid use of 10.01 hours, however the between-user standard deviation was substantial (2.76 hours) and 50% of the users had an average hearing aid use either below 8.18 (light users) or above 12.04 hours (heavy users) (Figure 3.1). Furthermore, users had a within-user standard deviation of 3.88 hours, indicating that they did not use the hearing aids uniformly across the logged days. Compared to the middle 50% of users, the light users (Two-sample t-test: $t = 23.06$, $p < 0.001$; Effect size: $d = 0.44$) and the heavy users (Two-sample t-test: $t = 41.85$, $p < 0.001$; Effect size: $d = 0.81$) exhibited a significantly larger within-user standard deviation, suggesting that they were more consistent in their low or high hearing aid use.

Figure 3.1: Distribution of users by their average amount of hearing aid use and their within-user standard deviation. A second order linear regression model (line ± 99% confidence interval) was fitted to the data to model the relationship between average hearing aid use (x) and within-user standard deviation (y) [100].



3.1.3 Patterns of Hearing Aid Use

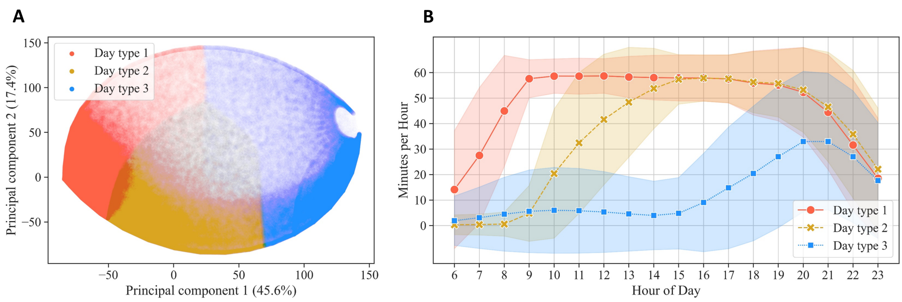
Despite being a popular metric, the amount of hearing aid use does not necessarily equate with benefit [52]. Indeed, depending on the degree of hearing loss, users might be more or less dependent on the hearing aids. Hearing aid users that exhibit a low amount of hearing aid use are not necessarily dissatisfied [65] but could only need them in the most challenging situations. Moreover, the amount of hearing aid use is not informative of how the hearing aids are used during the day. Objective data logging enables analyzing hourly patterns of hearing aid use. We investigated how users use the hearing aids during the day by clustering the 453,612 logged days. The input data consisted of a 453,612 × 18 matrix

$$A_{r\times c} = A_{453612\times 18} = \begin{bmatrix} a_{11} & \cdots & a_{1c} \\ \vdots & \ddots & \vdots \\ a_{r1} & \cdots & a_{rc} \end{bmatrix} = (a_{ij}) \in [0, 60];\ i = 1, ..., r;\ j = 1, ..., c$$

where each row i represents a day of hearing aid use, each column j represents an hour of the day (from 6 to 23) and a_{ij} is the amount of hearing aid use (from

0 to 60 minutes) in the day i and hour j. The choice to only analyze usage from 6:00 to 23:59 was justified by the limited data during night time and by the risk to overrepresent night usage due to users forgetting to turn off the hearing aids when going to bed. The k-means clustering technique was applied and the

Figure 3.2: In (A), the days of hearing aid use are displayed as scatterplot against the two main principal components and colored by the three clusters (i.e., three day types). In (B), the mean (±SD) of hourly hearing aid use for each cluster is displayed [100].



3-cluster solution was selected, which accounted for almost 50% of the variation among the days. Based on the hourly mean of hearing aid use for each of the three clusters (Figure 3.2B), the three types of days of hearing aid use could be characterized: a full day of hearing aid use (day type 1, containing 44% of days), a day of afternoon use (day type 2, containing 27% of days), and a day of sporadic evening use (day type 3, containing 26% of days).

3.1.4 Clustering Users Based on Hearing Aid Use

According to both previous research [79, 118, 123] and the results presented in Section 3.1.2, the amount of hearing aid use varies among users. Additionally, previous research suggests that the patterns of hearing aid use might also vary among users [67]. Consistently, Section 3.1.3 has shown that the days of hearing aid use pooled across all users exhibit widely different patterns. These differences might be driven both by users behaving differently among each other (between-user variation) and by the same user behaving differently throughout the logged days (within-user variation). When comparing users based on their hearing aid use, the within-user variation is usually disregarded and the average use per user is usually analyzed [118, 67, 123]. However, Section 3.1.2 suggests that the same user does not always use the hearing aids in the same way throughout the logged days.

In view of the above, we investigated the behavior of the 15,905 users by analyzing their patterns of hearing aid use throughout the logged days and we assessed whether clusters of users with similar use patterns can be identified. In doing so, we adopted a metric that accounts for the within-user variability. We clustered the 15,905 users based on the proportion of days belonging to each of the three typical days of hearing aid use. The input data consisted of a 15,905 $\times$ 3 matrix

$$B_{r\times c} = B_{15905\times 3} = \begin{bmatrix} b_{11} & \cdots & b_{1c} \\ \vdots & \ddots & \vdots \\ b_{r1} & \cdots & b_{rc} \end{bmatrix} = (b_{ij}) \in [0,1];\ i = 1,...,r;\ j = 1,...,c$$

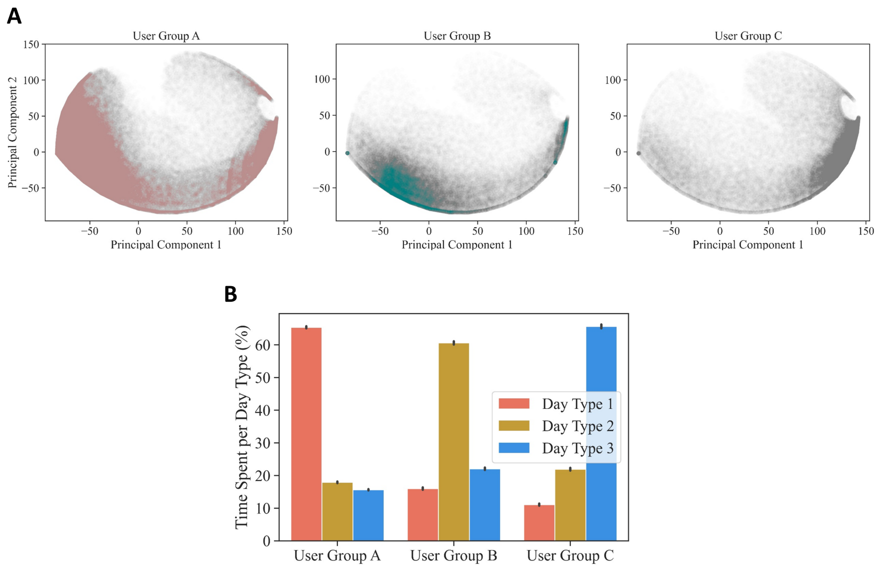
where each row i represents a hearing aid user, each column j represents one of the three types of days of hearing aid use (from 1 to 3) and b_{ij} is the proportion of days belonging to day type j for user i. Four clustering techniques were evaluated: k-means, Hierarchical Agglomerative Clustering (HAC) with Euclidean distance and Ward's method, HAC with Pearson correlation and average linkage method, and HDBSCAN. Based on three internal validation metrics (Silhouette, Davies-Bouldin, and Caliñski-Harabasz), HDBSCAN was selected. HDBSCAN identified three user clusters (Figure 3.3), in addition to labelling some users as noise. The three user groups exhibited widely different hearing aid use patterns: Group A (49% of users) predominantly had full days of hearing aid use, group B (15% of users) predominantly had days of afternoon use, group C (20% of users) predominantly had days of sporadic evening use. However, the users belonging to all three groups also exhibited a substantial day-to-day variability, with the predominant type of day only accounting for $\sim$ 60% of the days.

A good clustering is defined by compact and well separated clusters and enables a supervised classifier to accurately predict the cluster to which an unseen user belongs [111]. Thus, we validated the user clustering by training an ensemble of classifiers to predict the group of each user (group A, B, C, or noisy user) based on her average day of hearing aid use. The input data for classification consisted of a 15,905 $\times$ 18 matrix

$$D_{r\times c} = D_{15905\times 18} = \begin{bmatrix} d_{11} & \cdots & d_{1c} \\ \vdots & \ddots & \vdots \\ d_{r1} & \cdots & d_{rc} \end{bmatrix} = (d_{ij}) \in [0,60];\ i = 1,...,r;\ j = 1,...,c$$

where each row i represents the average day of a hearing aid user, each column j represents an hour of the day (from 6 to 23) and a_{ij} is the average amount of hearing aid use (from 0 to 60 minutes) for user i in the hour j. Three classifiers were defined (multiclass logistic regression, XGBoost and fully connected neural network). When evaluated individually on two metrics, XGBoost resulted to be the best performing one (accuracy = 87%, AUC-ROC = 0.98). In order to

Figure 3.3: In (A), for each of the three user clusters, the days of hearing aid use are displayed as scatter plot against the two main principal components. The distinct densities indicate that the three user groups experienced substantially different days of hearing aid use. In (B), the average proportion of time spent (±95% confidence interval) in each day type is displayed for each user cluster [100].



reduce the bias caused by an individual classifier [111], an ensemble was defined based on the three classifiers and predicted the group of each user by majority voting between the three classifiers. When no majority could be determined, the prediction was based on the best performing classifier (XGBoost). The ensemble reached an accuracy of 86% and an AUC-ROC score of 0.98.

3.2 How Do Users Personalize the Hearing Aids?

This section introduces listening programs as a way to personalize the hearing aids in specific listening situations, investigates the provision of listening programs, and analyzes the real-world context in which listening programs are used.

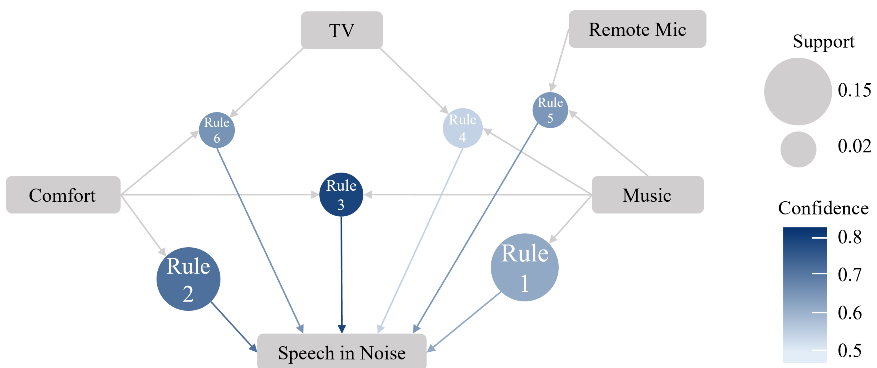
3.2.1 Investigating Listening Programs from Real-World Data

Since hearing aid users cope with a wide range of real-world situations, the devices need to ensure a satisfactory listening experience in different contexts. Thus, multimemory hearing aids have been introduced, that is, hearing aids that provide the user with different programs for specific listening environments. Listening programs set pre-defined rules for contextually adapting different audiological parameters such as overall gain, frequency shaping of the gain, noise reduction, and directionality. In 2019, 41% of hearing aid owners were found to have a program button or switch to change the hearing aid settings for different listening situations [107]. Previous research has found that programs for specific listening environments can be beneficial for users, by improving speech understanding [49, 109]. Moreover, experienced hearing aid users were found to be able to select the same program in the same situation at a rate exceeding pure guess [19], suggesting that programs can have a discernible impact on the listening experience. Nowadays, programs are the most common way for users to contextually adapt the hearing aid settings and thereby personalize their listening experience. Despite that, only a few studies investigated, mainly based on self-reported data, the use of multiple programs for various listening environments. Previous research suggests that some but not all hearing aid users value and use the option to switch between listening programs [119, 57, 16, 28, 127, 99, 110]. Smartphone-connected hearing aids enable leveraging objective data to deepen our understanding of the real-world use of listening programs. Indeed, objective data logging allows to analyze the program use of a large number of users in their everyday life contexts. Moreover, it avoids any bias inherent in self-reports and it allows complementing real-world program use data with objective data about the context. Compared to questionnaires or diaries, objective data logging also allows gathering data possessing higher temporal resolution. Finally, the observational nature of the study allows to observe long-term user behavior compared to an experimental study. Indeed, users enrolled in an experimental study might either use the programs to be compliant with the received instructions or might fail to realize the benefit of listening programs due to the short study duration [64].

3.2.2 Provision of Listening Programs

By leveraging objective data logging, we can deepen our understanding of listening programs and explore which situations motivate users to seek personalization by adjusting their device settings. We investigated the provision of listening programs by analyzing the programs provided to 32,336 hearing aid users. First, we computed how many and which programs were provided to the users. The majority (57%) of the users was found to have at least a program for a specific listening situation in addition to the default program, "General". After some coding of the program names, the most common additional programs were "Speech in Noise" (provided to 26% of users), "TV" (18%), "Music" (13%), "Remote Mic" (12%), and "Comfort" (10%). Second, we explored the relationships among programs by determining association rules. Given a set of n programs $P = p_1, p_2, \ldots, p_n$, and a set of m users $U = u_1, u_2, \ldots, u_m$, where each user is provided with a subset of the programs in P, a rule is defined as an implication of the form: $X \Rightarrow Y$ where X is the antecedent, Y is the consequent, $X, Y \subseteq P$, and $X, Y \cap \emptyset$ [46]. The rules were evaluated based on several metrics: support, coverage, confidence, and lift. The association rules with support≥0.02, confidence>0.5, and lift>1 are presented in Figure 3.4. "Speech in Noise", the most common additional listening program, is also the consequent of all rules, indicating that users provided with other additional programs also receive "Speech in Noise". Conversely, "Music" and/or "Comfort" are always in the antecedent set. The confidence values indicate that 62%, 71%, 79% of users who have "Music" (rule 1), "Comfort" (rule 2), or both of them (rule 3) also have "Speech in Noise". Instead, "TV" and "Remote Mic" are never in a single-program antecedent set, suggesting that getting those programs does not increase the likelihood of receiving other additional programs.

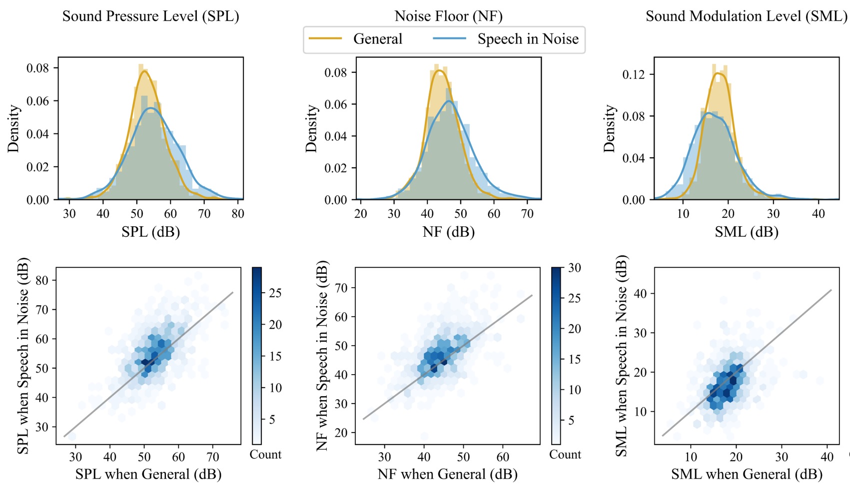
Figure 3.4: Association rules with support≥0.02, confidence>0.5 and lift>1. The support of each rule is indicated by the area of the circle, while the confidence is conveyed by the color intensity. "Speech in Noise" is the consequent of all rules, suggesting that it is frequently provided to the users that are also provided with other programs such us "Comfort" and "Music" (Appendix E).



3.2.3 Context of Use of Listening Programs

Establishing the need and interest for listening programs does not necessarily imply that users will benefit from them. In order to benefit from listening programs, users need to be able to correctly classify the environment and select the appropriate program [64]. Whether users can match the programs to the sound environments for which they are intended deserves further study [45]. Thus, we investigated the context of use of three listening programs ("Speech in Noise", "Comfort", "Music"), by analyzing the sound environment (sound pressure level (SPL), noise floor(NF), and sound modulation level (SML)) occurring during 396,723 program selections from 1,312 users. The NF is the level of background noise in a signal, while the SML describes how much the modulated variable of the signal varies around its unmodulated level. For each logged selection of

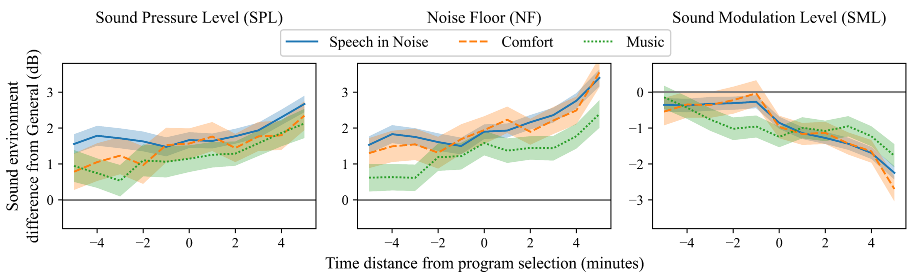
Figure 3.5: Analysis of the sound environment (SPL, NF, SML) in which "Speech in Noise" and "General" were selected. In the upper figures, distribution of users (using histograms and kernel density estimation) by their average sound environment when selecting "General" and "Speech in Noise". In the lower figures, 2d histograms displaying, for each user, the sound environment when selecting "Speech in Noise" (y-axis) and "General" (x-axis). The color of the hexagon is determined by the number of users in the hexagon. The identity line ($y = x$) is drawn in grey (Appendix E).



a specific listening program, the sound environment measured in a 10-minute time-window centered on the program selection was considered. Firstly, we analyzed the sound environment in which "Speech in Noise" was selected. Figure 3.5 displays the distribution of users by their average sound environment when selecting "General" and "Speech in Noise". On average, users selected "Speech in Noise" in significantly louder, noisier and less modulated environments compared to "General", as confirmed by a Wilcoxon Signed-Rank Test (P<.001 for all three parameters). As shown by the lower graphs in Figure 3.5, the majority of the users selected "Speech in Noise" in louder (64% of users), noisier (66%), and less modulated (62%) environments. Subsequently, we explored the sound environment in which "Comfort" and "Music" were selected. Users selected both programs in louder, noisier, and less modulated (Wilcoxon Signed-Rank test, all P<.001) environments compared to "General". No sound environment difference was found between "Comfort" and "Speech in Noise", as confirmed by a Mann-Whitney U test (SPL, P=.405; NF, P=.595; SML, P=.344). Instead, "Music" was selected in less loud (Mann-Whitney U test, P=.009) and less noisy (Mann-Whitney U test, P<.001) sound environments compared to "Speech in Noise".

In summary, by analyzing a 10-minute time window centered on program selection, we found that the three investigated programs were used in sound environments different than the default program, "General". Additionally, exploring whether the sound environment difference changes before and after program selection enables better understanding users' behavior and interaction with the hearing aids. For each of the three programs ("Speech in Noise", "Comfort", "Music") we computed a 5-minutes rolling average of the sound environment difference from "General" (Figure 3.6). The difference deviates from zero through-

Figure 3.6: 5-minutes running average (± SE) of the sound environment difference from "General", computed in a time-window near the program selection. The difference deviates from zero throughout the whole time-window. However, especially for NF and SML, the difference increased after program selection (Appendix E).



out the whole time-window. However, especially for NF and SML, the difference seems to increase after program selection. We quantified such increase by comparing the sound environment difference from "General" in the 5 minutes before program selection with the sound environment difference from "General" in the 5 minutes after program selection. A Wilcoxon Signed-Rank test revealed that, for all three programs and sound environment features, the difference from "General" significantly increases after program selection (all $P<.05$).

3.3 Chapter Summary

In this chapter, large scale data logged by commercially available products were analyzed to deepen the understanding of users' behavior. First, we investigated how users use the hearing aids and we found that, on average, they used the hearing aids 10 hours/day. We identified three typical days of hearing aid use and three distinct user groups, each characterized by a predominant typical day of hearing aid use. Moreover, a supervised ensemble was trained to predict the group to which an unseen user belongs. Second, we explored how users personalize their hearing aids in specific listening situations. We found that 57% of the sampled users had listening programs beyond the default one, "General". We identified a primary additional program ("Speech in Noise") and two secondary additional programs ("Comfort", "Music"). Finally, we analyzed the sound environment in which such additional programs are used and we found that users selected them in louder, noisier, and less modulated environments compared with the environment in which they selected the default program, "General".

Chapter 4

Discussion and Future Work

This chapter discusses the findings presented in the previous chapters and highlights opportunities for future work. Section 4.1 discusses the findings and future work related to the studies aimed at understanding users' audiological preferences. Section 4.2 discusses the findings and future work related to the studies aimed at understanding users' real-world behavior.

4.1 Understanding Users' Audiological Preferences

This section discusses the findings presented in the articles entitled "Rethinking Hearing Aids as Recommender Systems" [104] (Appendix A), "Measuring and Modeling Context-Dependent Preferences for Hearing Aid Settings" [103] (Appendix B), and "Designing Audiologist Bots Fusing Soundscapes and User Feedback" [101] (Appendix C).

Although hearing aid users perceive sound in individual ways, current approaches do not fully exploit the potential for personalization. However, smartphone-connected hearing aids potentially enable learning audiological preferences in real-world contexts by directly involving the user. Learning audiological preferences requires, first of all, to effectively gather user preferences. A smartphone-based method was adopted to make users explore a complex audiological space and communicate their preferences. Such exploration was simplified by decoupling three audiological parameters (Noise Reduction, Brightness, and Soft Gain), and allowing users to adjust them one at a time. Meanwhile, data about their experience, preferences and context were gathered.

The three parameters are traditionally deemed important for the listening experience. In order to investigate whether it is equally advisable to provide users with the opportunity to adjust such parameters, we analyzed whether higher perceived usefulness of the parameters was associated with a better listening experience. We found that the usefulness of the Noise Reduction parameter was not associated with the listening experience. Previous research shows that noise reduction in hearing aids enables a more natural speech [24], reduces listening effort [18, 33] and allows users to tolerate a higher noise level [86, 77]. However, if the Noise Reduction adjustments are not discernible, the feedback might be unreliable. Moreover, if the adjustments are not meaningful, the impact on the perceived listening experience might be negligible, thus enabling users to adjust the Noise Reduction parameter might not be worth it. Previous research has shown that, regardless of hearing loss, a change in SNR of 3 dB is necessary for a reliably discernible difference [83]. Moreover, a change in SNR of 6 to 8 dB is required for the difference to be meaningful, i.e., not only discernible but also large enough to induce a behavior change [82]. This might explain why a higher usefulness of the parameter does not correspond to higher listening satisfaction. Conversely, we found that, for both the Brightness and Soft Gain parameters, a higher usefulness was associated with a better listening experience. The significant role of Brightness confirms the idea that high-frequency information has an important role in speech localization [21, 84, 25], speech understanding [84], and perceived naturalness [85]. Previous studies have shown that specifically hearing-impaired listeners benefit from high-frequency amplification for speech understanding [50, 3, 84]. Moreover, the extent to which a hearing aid user benefits from high-frequency amplification is highly individual [84] and, among listeners with similar high-frequency thresholds, seems to be higher for the ones with flat hearing losses compared to the ones with sloping losses [50]. Similarly, hearing impaired listeners widely differ in the loudness perception of sounds close to the hearing thresholds (i.e., soft sounds), and exhibit loudness-growth functions ranging from rapid growth to softness imperception [80]. The significant correlation between choice-usefulness and listening satisfaction is consistent with the fact that the benefit of both high-frequency and soft sounds amplification are highly individual, and suggests that giving users the opportunity to personalize such parameters can significantly enhance their listening experience.

Subsequently, the impact of context was analyzed by applying mixed-effects modeling for predicting choice-usefulness and intervention level preferences. While gathering user feedback in several real-world contexts is necessary to truly understand audiological preferences, a user-adaptive system might also leverage contextual information to provide the hearing aid user with more relevant parameters. We found that both logged and self-reported context had a significant impact on the usefulness of choosing among different intervention levels of the three parameters. For instance, choosing among four levels of the Brightness parameter was evaluated to be more useful when having a "Focus" listening intention (i.e., watching TV, having a meal, listening to music) and in low intensity sound environments. In this situations, hearing aid users might personalize the high-frequency amplification to improve sound localization [21, 84, 25] and speech understanding [50, 3, 84]. The significant role of context suggests that, by assigning context-aware usefulness to the audiological parameters, it might be possible to provide more relevant parameters. This would ensure an improved short-term listening experience, but also a more meaningful interaction and the collection of more relevant audiological preferences.

Additionally, context had a significant impact on the explicitely preferred intervention levels of the audiological parameters. Context-aware predictions of the preferred intervention level can help decide which levels of the parameter should be offered to the user, depending on the situation. For instance, participants selected higher Brightness levels (i.e., high-frequency amplification) when being alone (i.e., "Only me" listening intention compared to "Focus" and "Social"). Previous research have focused on the benefit derived from high-frequency amplification in noisy [3] or complex sound environments, and in speech situations where the target and the masking sounds are spatially separated [68, 84, 3]. The results of our study suggest that users like to select even higher high-frequency amplification in situations that do not involve a conversation. Users, when being alone and in control of the situation, might enjoy increasing high-frequency amplification to improve sound localization, without fearing of being disturbed by sudden noises. Overall, the relevance of both logged and self-reported context in predicting choice-usefulness suggests that it is important to model context both by directly observable features (e.g., SPL and SNR), as well as by hidden features reflecting the user's subjective listening intentions (e.g., enhancing speech or ignoring voices) and environment.

Gathering explicit preferences requires users to actively evaluate the alternative settings and submit their preference. On the one hand, this makes it easier to gather preferences that are perceived as meaningful and more likely to impact the listening experience. On the other hand, this is a time consuming process that places a burden on the users. Indeed, some users might not be willing to actively report their preferences. Motivated users might report their preferences but still miss some challenging situations. Indeed, this task requires engaging in the submission of audiological preferences in a situation which might already be effortful for a hearing impaired listener (e.g., conversation in noise). In addition to explicitly preferred intervention levels, implicit preferences inferred from the active level selections (i.e. level selections that are set for at least three minutes) were analyzed. The impact of logged context (i.e., SPL and SNR) was analyzed by applying mixed-effects modeling for predicting implicit preferences. Context had a significant impact on the implicitly preferred intervention levels of the audiological parameters. A system aimed at learning audiological preferences could rely on active level selections, combined with contextual data. While implicit preferences are potentially less reliable, they are less demanding for the user and might enable preference modeling of less engaged users.

The proposed mixed effect model treated context as fixed effect and assumed that the preferences of users are impacted by context in the same way across users. By gathering more repeated measures, future work could expand on such model to account for the individual sensitivity towards context. This could be done by introducing participant-specific random slopes. Moreover, the proposed model treated the individual-level effects as random effects, estimated with partial pooling. The model can make predictions for new users assuming that the new users exhibit similar traits (e.g., hearing loss, age, perceptual characteristics) as the users used for training the model. By recruiting more participants, future research could investigate what characteristics cause two users to have similar audiological preferences. Better understanding which user features explain the variation in preferences among individuals would enable transferring the learning from active users to new or inactive users.

Additionally, a conversational agent model was outlined, which could accelerate the collection of user feedback by autonomously gathering user feedback in real-world environments and collecting information about the context to learn users' audiological preferences. The data collected by such conversational agent could, in the short term, be beneficial to hearing care professionals by facilitating patient-centered and data-driven decisions in the clinical practice. Indeed, the process does not rely on further in-clinic tests, so it does not place an additional burden on the hearing aid professionals, which are currently a scarce resource in the hearing healthcare field. In the longer term, such data could enable the autonomous learning of user preferences and thereby alleviate the lack of audiological resources.

4.2 Understanding Users' Real-World Behavior

This section discusses the findings presented in the articles entitled "Clustering Users Based on Hearing Aid Use: An Exploratory Analysis of Real-World Data" [100] (Appendix D) and "Investigating the Provision and Context of Use of Hearing Aid Listening Programs from Real-World Data" [102] (Appendix E).

In addition to learning audiological preferences, smartphone-connected hearing aids enable learning about the real-world behavior of users. Indeed, observing the behavior of users in their natural environment offers a new perspective on how they use the hearing aids. While lacking the control of an experimental study, it allows to gather data from numerous and diverse users around the world, with high temporal resolution and for extended periods of time.

To investigate hearing aid use in the real-world, the amount of hearing aid use was explored by analyzing 453,612 days of HA use logged by 15,905 users. On average, users used the hearing aids for 10.01 hours per day. This value is similar [67, 122] or slightly larger [41, 123] than previous studies objectively measuring hearing aid use. Moreover, among the 15,905 users, 25% used the hearing aids, on average, for less than 8.18 hours. Previous studies analyzing hearing aid use based on objective or subjective measures found a similar [67] or larger [117, 20, 123] percentage. The higher amount of hearing aid use and the lower percentage of light users in our study might be partly explained by choices made during the data manipulation process (e.g., including only days with at least an hour of hearing aid use, selecting the larger value between the right and left ear). Additionally, the selected participants all used a self-tracking feature via a smartphone app. When relying on objective data logged via smartphone-connected hearing aids, it is important to remember that, while the sampled is large and international, it also has peculiar characteristics in terms of tech-savyness, openness to experimenting innovative solutions, and engagement.

Not only does data logging potentially improve the accuracy in assessing the amount of hearing aid use, but it also makes it easier to evaluate how hearing aids are used by analyzing patterns of hearing aid use. The 453,612 days of hearing aid use were clustered into three typical days. The most frequent day (44% of days) was denoted by full hearing aid use. This suggests that, in the most typical day, hearing aid users use their device uninterruptedly from the morning (around 7) to the evening (around 22). The second most frequent day (27%) was characterized by afternoon use, from 11 to 22. The third typical day (26%) was denoted by sporadic evening use. Such wide difference among the types of hearing aid use might be driven either by between-user variability or by within-user variability. Based on the proportion of time spent in each of the days of hearing aid use, the 15,905 users were clustered in three user groups. The most common group (49% of users) predominantly had full days of hearing aid use. Due to the inclusion criteria of the study, this group of heavy users might be overrepresented. Moreover, 15% of users mainly had days of afternoon use, and 20% predominantly had days of sporadic evening use. The latter user group might either be composed by unsatisfied users or by hearing impaired people that are not dependent on their hearing aids and only use the device in the most challenging situations. The remaining 15% of users exhibited uncommon behaviors and did not belong to any cluster. The users belonging to the three clusters (85%) tended to have predominant daily patterns of hearing aid use. However, this does not mean that they used the hearing aids uniformly from day to day. Indeed, on average only 60% of days were characterized by the predominant day type. Such day-to-day variation might not emerge in studies based on self-reports, which suffer from recall bias. Indeed, a previous study found that 77% of the participants reported in a questionnaire that they use the hearing aids in the same way every day, while only 23% said that they use the hearing aids differently from day to day [67]. Finally, a supervised classifier was successfully trained (accuracy 86%) to predict the cluster a user belongs to. Training a classifier not only validates the quality of the partition, but it also potentially enables leveraging the knowledge of existing users to predict the user group of a new user.

Investigating patterns of hearing aid use enables a deeper understanding of the needs and behavior of hearing aid users. Future research could complement the objective data about hearing aid use with subjective evaluations of the listening experience. For instance, 20% of users predominantly used the hearing aids in isolated occasions, mainly during the evening. Knowing whether these users are satisfied with their listening experience would help to verify whether the sporadic use is due to a malfunctioning hearing aid or a suboptimal fitting. Furthermore, collecting data describing the hearing aid user (e.g., hearing loss, age) would enable better characterizing the hearing aid users belonging to each cluster, as well as transferring the existing knowledge to similar new users. Ultimately, better understanding the users would help hearing care professionals make data-driven decisions and thereby enhance the provided hearing treatment.

In addition to investigating the use of hearing aids, smartphone-connected hearing aids enable exploring how users personalize them in specific listening situations. The provision of listening programs to 32,336 hearing aid users was analyzed. Fifty-seven percent of hearing aid users had listening programs in addition to the default one. Such percentage is higher than the 41% of users which reported, in a previous study, to have a switch to change the hearing aid settings for different environments [107]. Similarly to previous cases, the inclusion criteria might explain the higher percentage found in our study. The association mining analysis revealed that "Speech in Noise" is a primary program, provided as first option to users that are interested in personalizing their listening experience in specific environments. This is consistent with the difficulty encountered by hearing impaired people in background noise [20, 48, 1], and multi-talker scenarios [42]. However, whether the provision of listening programs is mostly influenced by hearing aid professionals or by hearing aid users cannot be determined based on such objective data and would deserve further study.

The analysis of the context in which the listening programs were selected shew that users selected "Speech in Noise", "Comfort", and "Music" in a louder, noisier and less modulated sound environments compared to "General". The sound environment difference is consistent with a context characterized by a conversation in noise, or by music. This indicates that users are, to some extent, able to correctly characterize the listening environment and select the most appropriate program. Interestingly, there was no difference between the sound environments in which users selected "Comfort" and "Speech in Noise". This indicates that two apparently similar sound environments might in fact require different hearing aid settings, depending on the listening intent. The analyzed high-level sound environment features (i.e., SPL, NF, and SML) might not be sufficient to distinguish among different listening intents. Moreover, we found that the sound environment difference from "General" increased in the five minutes following program selection compared to the five minutes preceding it. This suggests that some users might select additional listening programs in anticipation, rather than as a reaction, to a change in the sound environment. This would confirm that users are aware of what the contextually most appropriate program is and proactively select it before entering a specific listening situation. Overall, the significant difference in the context of use of specific additional programs indicates that hearing aid users could benefit from being empowered to adjust the device settings in different contexts. However, given that hearing aid users differ in behavior and preferences, the benefit from such solution might depend on several factors, such as the hearing loss, age, and experience with hearing aids. Further research is needed to investigate the role played by these factors in benefiting from listening programs.

Chapter 5

Conclusion

Despite the fact that hearing aid users perceive the sounds in individual ways and can potentially benefit from more personalized interactions in different steps of their journey, current approaches do not fully exploit the potential for personalization. Indeed, hearing aid manufacturers face the challenge of providing personalized hearing aid settings, as well as a comprehensively personalized solution to patients. This thesis contributes to the progress towards a data-driven approach to hearing aid personalization by learning users' preferences and behavior from real-world data.

First, we adopted a smartphone-based method to empower hearing aid users to adjust three audiological parameters (Noise Reduction, Brightness, and Soft Gain) in real-world environments and gather contextualized data on their audiological preferences. By analyzing the collected data, we found that having access to different intervention levels of two audiological parameters (Brightness and Soft Gain) affected listening satisfaction. Moreover, the contextual information had a significant impact on the perceived usefulness of having access to different intervention levels, as well as the intervention level preferences. Therefore, the gathered contextual data can help provide the users with more relevant settings to choose from, thereby improving the listening experience while gathering more meaningful data.

Second, we observed the behavior of users by analyzing data logged by commercially available hearing aids. We investigated hearing aid use and identified three typical daily patterns of hearing aid use. Users were clustered based on hearing aid use and three groups were identified, each characterized by a predominant daily pattern of hearing aid use. Moreover, we explored the provision and context of listening programs. We identified a default program, a primary additional program, and two secondary additional programs. We also found that users use the additional listening programs in sound environments different than the default program.

Our results show that smartphone-connected hearing aids can be useful to both perform experimental studies aimed at exploring novel ways of personalizing the device, and observational studies aimed at investigating how users naturally use commercially available devices. Gathering data in real-world environments and developing a deeper understanding of the preferences and behavior of hearing aid users can help hearing care professionals offer better services, manufacturers develop truly personalized solutions, and researchers progress in the quest for a more comprehensive understanding of the diverse needs of hearing impaired people.

Appendix A

Rethinking Hearing Aids as Recommender Systems

Alessandro Pasta, Michael Kai Petersen, Kasper Juul Jensen, and Jakob Eg Larsen (2019). Rethinking Hearing Aids as Recommender Systems. *CEUR Workshop Proceedings*, 2439, 11-17, http://ceur-ws.org/Vol-2439/4-paginated.pdf.

Rethinking Hearing Aids as Recommender Systems

Alessandro Pasta
Technical University of Denmark
Kongens Lyngby, Denmark
alpas@dtu.dk

Michael Kai Petersen
Eriksholm Research Centre
Snekkersten, Denmark
mkpe@eriksholm.com

Kasper Juul Jensen
Oticon A/S
Smørum, Denmark
kjen@oticon.com

Jakob Eg Larsen
Technical University of Denmark
Kongens Lyngby, Denmark
jaeg@dtu.dk

ABSTRACT

The introduction of internet-connected hearing aids constitutes a paradigm shift in hearing healthcare, as the device can now potentially be complemented with smartphone apps that model the surrounding environment in order to recommend the optimal settings in a given context and situation. However, rethinking hearing aids as context-aware recommender systems poses some challenges. In this paper, we address them by gathering the preferences of seven participants in real-world listening environments. Exploring an audiological design space, the participants sequentially optimize three audiological parameters which are subsequently combined into a personalized device configuration. We blindly compare this configuration against settings personalized in a standard clinical workflow based on questions and pre-recorded sound samples, and we find that six out of seven participants prefer the device settings learned in real-world listening environments.

CCS CONCEPTS

• **Information systems** → **Personalization**; **Recommender systems**; • **Human-centered computing** → *Ambient intelligence*; *User centered design*.

KEYWORDS

Personalization, recommender systems, hearing healthcare, hearing aids

ACM Reference Format:
Alessandro Pasta, Michael Kai Petersen, Kasper Juul Jensen, and Jakob Eg Larsen. 2019. Rethinking Hearing Aids as Recommender Systems. In *Proceedings of the 4th International Workshop on Health Recommender Systems co-located with 13th ACM Conference on Recommender Systems (HealthRecSys'19), Copenhagen, Denmark, September 20, 2019,* 7 pages.

1 INTRODUCTION

Despite decades of research and development, hearing aids still fail to restore normal auditory perception as they mainly address the lack of amplification due to loss of hair cells in the cochlea [16], rather than compensating for the resulting distortion of neural activity patterns in the brain [22]. However, the full potential of hearing aids is rarely utilized as devices are frequently dispensed with a "one size fits all" medium setting, which does not reflect the varying needs of users in real-world listening scenarios. The recent introduction of internet-connected hearing aids represents a paradigm shift in hearing healthcare, as the device might now be complemented with smartphone apps that model the surrounding environment in order to recommend the optimal settings in a given context.

Whereas a traditional recommender system is built based on data records of the form < *user,item,rating* > and may apply collaborative filtering to suggest, for instance, new items based on items previously purchased and their features, recommending the optimal hearing aid settings in a given context remains highly complex. Rethinking hearing aids as recommender systems, different device configurations could be interpreted as items to be recommended to the user based on previously expressed preferences as well as preferences expressed by similar users in similar contexts. In this framework, information about the sound environment and user intents in different soundscapes could be treated as contextual information to be incorporated in the recommendation, building a context-aware recommender system based on data records of the form < *user,item,context,rating* > [1]. However, addressing some challenges related to the four aforementioned data types is essential to make it possible to build an effective context-aware recommender system in the near future. In this paper, we discuss the main challenges posed when rethinking hearing aids as recommender systems and we address them in an experiment conducted with seven hearing aid users.

1.1 Rating

In order to be able to precisely and accurately recommend optimal device settings in every situation, gathering relevant user preferences (expressed as ratings) is essential. However, learning user preferences poses some challenges. Firstly, the device settings reflect a highly complex audiological design space involving multiple interacting parameters, such as beamforming, noise reduction, compression and frequency shaping of gain. It is important to explore the different parameters, in order not to disregard some parameters that might have relevant implications for the user listening experience, and to identify which parameters in an audiological design space [10] define user preferences in a given context. Secondly, the preferred device settings depend on the human perception of the

HealthRecSys'19, September 20, 2019, Copenhagen, Denmark

listening experience and it is therefore difficult to represent the perceptual objective using an equation solely calculated by computers [21]. Having to rely on user feedback, it is important to limit the complexity of the interface, to make the interaction as effective as possible. Thirdly, capturing user preferences in multiple real-world situations not only guarantees that the situations are relevant and representative of what the user will experience in the future, but it also allows the user to test the settings with a precise and real intent in mind. However, this increases the complexity of the task, since the real-world environment is constantly changing and a user might explore the design space while performing other actions (e.g. conversing).

A traditional approach to find the best parameter combination (i.e. the best device configuration) is *parameter tweaking*, which consists in acting on a set of (either continuous or discrete) parameters to optimize them. Similarly to enhancing a photograph by manipulating sliders defining brightness, saturation and contrast [21], the hearing aid user could control her listening experience by tweaking the parameters that define the design space and find the optimal settings in different listening scenarios. However, this method can be tedious when the user is moving in a complex design space defined by parameters that interact among each other [13]. One frequently used method to simplify the task of gathering preferences is *pairwise comparison*, which consists in making users select between two contrasting examples. A limitation of this approach is efficiency, given that a single choice between two examples provides limited information and many iterations are required to obtain the preferred configuration. Based on pairwise comparisons, an active learning algorithm may apply Bayesian optimization [2] to automatically reduce the number of examples needed to capture the preferences [3], assuming that the samples selected for comparison capture all parameters across the domain. Alternatively, one might decompose the entire problem into a sequence of unique *one-dimensional slider manipulation* tasks. As exemplified by Koyama et al. [13], the color of photographs can be enhanced by proposing users a sequence of tasks. At every step, the method determines the one-dimensional slider that can most efficiently lead to the best parameter set in a multi-dimensional design space defined by brightness, contrast and saturation. Compared to pairwise comparison tasks, the single-slider method makes it possible to obtain richer information at every iteration and accelerates the convergence of the optimization.

Inspired by the latter approach we likewise formulate the learning of audiological preferences in a given listening scenario as an optimization problem:

$$z = \arg\max_{x \in \mathcal{X}} f(x)$$

where x defines parameters related to beamforming, attenuation, noise reduction, compression, and frequency shaping of gain in an audiological design space $\mathcal{X}$ [10] and the global optimum of the function $f : \mathcal{X} \rightarrow \mathfrak{R}$ returns values defining the preferred hearing aid settings in a given listening scenario.

However, while it remains sensible to assume that individual adjustments would converge when crowdsourcing (i.e. asking crowd workers to complete the tasks independently) the task of enhancing an image [13], it is less likely that hearing impaired users would have similar preferences due to individual differences in their sensorineural processing [16, 22]. Therefore, at least in the first phase, we need to ask the same user many times about her preferences, until her optimal configuration is found. Furthermore, in order to optimize the device in different listening scenarios, we need to ask the same user to move in the same design space multiple times. Altering the one-dimensional slider at every step of the evaluation procedure might make the task difficult, since the user would not know the trajectory defined by the new slider. We believe that decoupling the parameters and allowing users to manipulate one parameter at a time, moving in a one-dimensional space that is clearly understood, would allow them to better predict the effects of their actions and hence more effectively assess their preferences.

1.2 Item

In order to enhance the hearing aid user experience, it is important to appropriately select the parameters that define the hearing aid configurations evaluated by users. Indeed, not only should the parameters have a relevant impact on the user listening experience, but the different levels of the parameters should also be discernible by untrained users. Three parameters have been demonstrated to be particularly important for the experience of hearing impaired users:

(1) *Noise reduction and directionality.* Noise reduction reduces the effort associated with speech recognition, as indicated by pupil dilation measurements, an index of processing effort [23]. By allowing speedier word identification, noise reduction also facilitates cognitive processing and thereby frees up working memory capacity in the brain [18]. Moreover, fast-acting noise reduction proved to increase recognition performances and reduce peak pupil dilation compared to slow-acting noise reduction [23]. Given that the ability of users to understand speech in noisy environments may vary by up to 15 dB [4], it is essential to be able to individualize the threshold levels for the activation of noise reduction.
(2) *Brightness.* While a lot of research has been focused on adapting the frequency-specific amplification which compensates for a hearing loss based on optimized rationales like VAC+ [5], rationales still reflect average preferences across a population rather than individual ones. Several studies indicate that some users may benefit from increasing high-frequency gain in order to enhance speech intelligibility [11, 12].
(3) *Soft gain.* The perception of soft sounds varies largely among individuals. Hearing aid users with similar hearing losses can perceive sounds close to the hearing threshold as being soft or relatively loud. Thus, proposing a medium setting for amplification of soft sounds may seem right when averaging across a population, but would not be representative of the large differences in loudness perception found among individual users [17]. For this reason, modern hearing aids provide the opportunity to fine-tune the soft gain by acting on a compression threshold trimmer [14].

Taking a naive approach, treating each parameter independently, the preferences could subsequently be summed up in a general hearing aid setting, by simply applying the most frequently preferred values along each audiological parameter.

1.3 User

Hearing aids are often fitted based on a pure tone audiometry, a test used to identify the hearing threshold of users. However, as mentioned above, users perceive the sounds differently and might benefit from a fully personalized hearing aid configuration. For this reason, it is essential to fully understand what drives user preferences and which is the relative importance of users' characteristics and context. It is interesting to analyse whether users exhibit similar preferences when optimizing the hearing aids in several real-world environments and whether they result into similar configurations.

1.4 Context

Users often prefer to switch between highly contrasting settings depending on the context [11]. It has been shown that a context-aware hearing aid needs to combine different contextual parameters, such as location, motion, and soundscape information inferred by auditory measures (e.g. sound pressure level, noise floor, modulation envelope, modulation index, signal-to-noise ratio) [12]. However, these contextual parameters might fail to capture the audiological intent of the user, which depends not only on the characteristics of the sound environment but also on the situation the user is in. For this reason, in addition to retrieving the characteristics of the sound environment and the preferred device settings, it is also important to capture the contextual intents of users in the varying listening scenarios. Contextual information, in this exploratory phase, can be explicitly obtained by directly asking the user to define the situation she is in. However, in the future, to enable an automatic adaptation to the needs of users in real-world environments, relevant contextual information will need to be inferred using a predictive model that classifies the surrounding environment.

2 METHOD

2.1 Participants

Seven participants (6 men and 1 woman), from a screened population provided by Eriksholm Research Centre, participated in the study. Their average age was 58.3 years (std. 12 years). Five of them were working, while two were retired. They were suffering from a binaural hearing loss ranging from mild to moderately severe, as classified by the American Speech-Language-Hearing Association [6]. The average hearing threshold levels are shown in Figure 1. They were all experienced hearing aid users, ranging from 5 to 20 years of experience with hearing aids. All test subjects received information about the study and signed an informed consent before the beginning of the experiment.

2.2 Apparatus

The participants were fitted according to their individual hearing loss with a pair of Oticon Opn S 1 miniRITE [8]. All had iPhones with iOS 12 installed and additionally downloaded a custom smartphone app connected to the hearing aids via Bluetooth. The app enabled collecting data about the audiological preferences and the corresponding context.

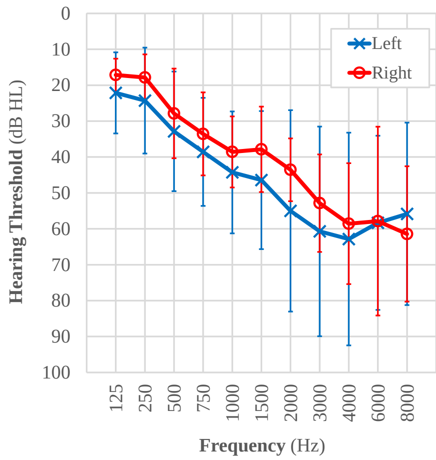


Figure 1: Average hearing threshold (i.e. the sound level below which a person's ear is unable to detect any sound [7]) levels for the 7 participants. The participants had a hearing loss ranging from mild to moderately severe. Error bars indicate ±1 standard deviation of the hearing thresholds.

2.3 Procedure

The experiment was divided into four weeks. As shown in Table 1, the first three weeks were devoted to optimizing the three audiological parameters, one at a time. Each of the first three weeks, the participants were fitted with four levels of the respective parameter, while the other two parameters were kept neutral at a default level. For instance, in week 1, each participant could select between four levels of noise reduction and directionality. The participants were instructed to compare, using a smartphone app, the four levels of the parameter in different situations during their daily life and to report their *preference*. To ensure that the participants would evaluate the different levels in relevant listening situations and when motivated to optimize their device, they were instructed to perform the task on a voluntary basis. Moreover, every time they reported their preference, the participants were asked to specify:

- The *environment* they were in (e.g. office, restaurant, public space outdoor). Different environments are characterised by different soundscapes and pose disparate challenges for hearing aid users.
- Their *motion* state (e.g. stationary, walking, driving). Motion tells more about the activity conducted by the person, but may also mark the transition to a different activity or environment [9].
- Their audiological *intent* (e.g. conversation, work meeting, watching TV, listening to music, ignoring speech). Complementing the contextual information by gathering the intent of the participants in the specific situation might provide a

Table 1: Study timeline

Week	Activity
W. 1	Optimization of *noise reduction and directionality*
W. 2	Optimization of *brightness* (amplification of high-frequency sounds)
W. 3	Optimization of *soft gain* (amplification of soft sounds)
W. 4	Final *test of preference*

deeper insight into how the different audiological parameters help them in coping with different sounds.

- The *usefulness of the parameter* in the specific situation (on a scale ranging from 1 to 5). This evaluation is important not only to understand the relative importance of each preference, but also to assess the perceived benefit of the parameter in diverse situations.

The fourth week each participant compared two different device configurations in a blind test:

- An individually personalized configuration combining the most frequently selected preferences of the three audiological parameters gathered in *real-world listening environments* during the previous three weeks.
- A configuration personalized in a *standard clinical workflow* based on questions and on pairwise comparisons of pre-recorded sound samples capturing different listening scenarios including, for instance, speech with varying levels of background noise.

The participants were instructed to compare the two personalized configurations in different listening situations throughout the day and report their preference, while also labeling the context. At the end of the week, the participants were asked to select the configuration they preferred.

3 RESULTS

During the four weeks of test, the participants actively interacted with their devices, changing the hearing aid settings, overall, 4328 times (i.e. the level of the parameter during the first three weeks or the final configuration during the last week) and submitting 406 preferences. On average, the participants tried the different hearing aid settings 11 times before submitting a preference. Although one parameter affects the perception of the others, isolating them allows to analyse their perceived impact on the listening experience. As illustrated in Figure 2, the brightness parameter was on average rated higher in perceived usefulness. This result is consistent among the seven participants. Conversely, the noise reduction and directionality parameter resulted to have the lowest perceived usefulness for five participants out of seven. The soft gain parameter resulted to have an average perceived usefulness between those of the other two parameters.

Recording, together with each preference, the perceived usefulness of the parameter in the specific situation also allows to understand how much each parameter contributes to the overall setting of the hearing aid. Figures 3, 4, 5 display the preferences of test participants for different levels of noise reduction and directionality, brightness, and soft gain, respectively. Only the preferences

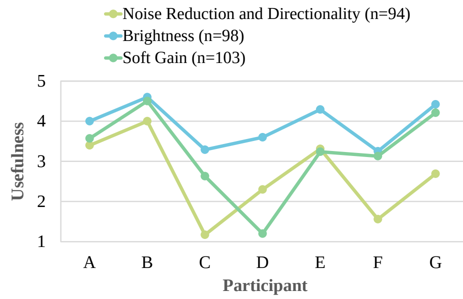


Figure 2: Average perceived usefulness of three parameters (noise reduction and directionality, brightness, soft sounds). Brightness is perceived to be the most useful parameter. Noise reduction and directionality tends to be perceived as the least useful parameter.

recorded in situations where the usefulness of the parameter is rated higher than two out of five are considered.

Firstly, the results indicate that the participants have widely different audiological preferences, rather than converging towards a shared optimal value. As the participants are ordered by age (*A* being the youngest), there seem, nevertheless, to be some common tendencies among younger or older participants across all parameters.

Secondly, most participants are not searching for a single optimum but select different values within each parameter. When adjusting the perceived brightness (Figure 4), six participants out of seven prefer, most of the time, the two highest levels along this parameter. Thirdly, the participants frequently prefer highly contrasting values within each parameter, depending on the context.

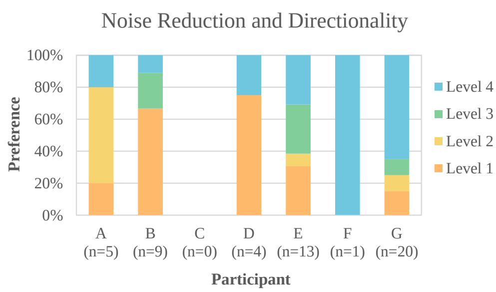


Figure 3: Preferences for the 4 levels of noise reduction and directionality, which correspond (from level 1 to level 4) to increasing directionality settings, increasing levels of noise reduction in simple and complex environments and earlier activation of noise reduction [15]. The participants exhibited different noise reduction and directionality preferences and five of them preferred more than one level in different situations.

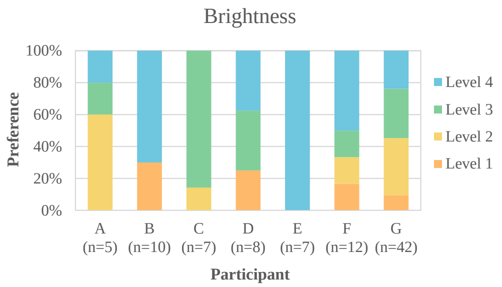


Figure 4: Preferences for the 4 levels of brightness, which correspond (from level 1 to level 4) to increasing amplification of high-frequency sounds. The participants exhibited different brightness preferences and six of them preferred more than one level in different situations.

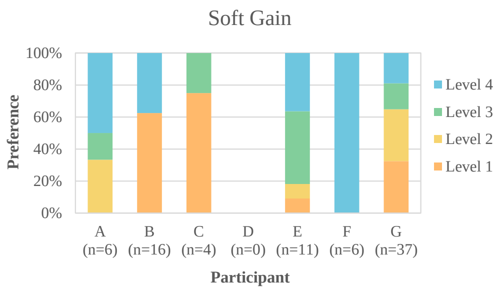


Figure 5: Preferences for the 4 levels of soft gain, which correspond (from level 1 to level 4) to increasing amplification of soft sounds, thus increasing dynamic range compression [14]. The participants exhibited different soft gain preferences and five of them preferred more than one level in different situations.

In order to combine the sequentially learned preferences, we summed up the most frequently chosen values along each parameter into a single hearing aid configuration. For each participant, we subsequently compared it against individually personalized settings configured in a standard clinical workflow based on questions and pre-recorded sound samples. After the fourth week, six out of seven participants responded they appreciated having more than one general hearing aid setting, as they used both configurations in different situations. They also wished to keep both personalized configurations after the end of the test. However, in a blind comparison of the two configurations, six out of seven participants preferred the hearing aid settings personalized by sequentially optimizing parameters in real-world listening scenarios.

4 DISCUSSION

Due to the aging population, the number of people affected by hearing loss will double by 2050 [20] and this will have large implications for hearing healthcare. Rethinking hearing aids as recommender systems might enable the implementation of devices that automatically learn the preferred settings by actively involving hearing impaired users in the loop. Not only would this enhance the experience of current hearing aid users, but it could also help overcome the growing lack of clinical resources. Personalizing hearing aids by integrating audiological domain-specific recommendations might even make it feasible to provide scalable solutions for the 80% of hearing impaired users who currently have no access to hearing healthcare worldwide [19]. The accuracy of the recommendation primarily depends on the ability of the system to gather user preferences, while the user explores a highly complex design space. In this study, we proposed an approach to effectively optimize the device settings by decoupling three audiological parameters and allowing the participants to manipulate one parameter at a time, comparing four discrete levels. The fact that the participants preferred the hearing aid configuration personalized in real-world environments suggests that the proposed optimization approach manages to capture the main individual parameter preferences.

Looking into the individual preferences learned when sequentially adjusting the three parameters, several aspects stand out. The results suggest that the *brightness* parameter has the highest perceived usefulness. This could be due to the fact that enhancing the gain of high frequencies may increase the contrasts between consonants and as a result improve speech intelligibility. Likewise, it may amplify spatial cues reflected from the walls and ceiling, improving the localization of sounds and thereby facilitating the separation of voices. The participants seemed to appreciate a brighter sound when listening to speech or when paying attention to specific sources in a quiet environment. Despite the advances in technology that reduce the risk of audio feedback and allow the new instruments to be fitted to target and deliver the optimal gain [8], in some situations most of the participants seemed to benefit from even more brightness. Conversely, users might prefer a more round sound in noisy situations or when they want to detach themselves.

Adjusting the *noise reduction and directionality* parameter is perceived as having the lowest usefulness. Essentially, this parameter defines how ambient sounds coming from the sides and from behind are attenuated, while still amplifying signals with speech characteristics. Although the benefits of directionality and noise reduction are proven, our results indicate that users find it more difficult to differentiate the levels of this parameter if the ambient noise level is not sufficiently challenging. The four levels of the parameter mainly affect the threshold for when the device should begin to attenuate ambient sounds. However, these elements of signal processing are partly triggered automatically based on how noisy the environment is. Therefore, in some situations, changing the attenuation thresholds (i.e. the parameter levels) might not make a difference. Thus, users may feel less empowered to adjust this parameter. On the other hand, the data also shows that participants actively select the lowest level of the parameter (level 1), which provides an immersive omnidirectional experience without attenuation of ambient sounds in simple listening scenarios. This suggests that, in some contexts, users express a need for personalizing the directionality settings and the activation thresholds of noise reduction. Furthermore, previous studies have shown that the perception of *soft sounds* varies largely among individuals. Our results not only confirm that users have widely different audiological preferences, but also suggest they would benefit from a personalized dynamic adaptation of soft gain dependent on the context.

Focusing on the optimization problem in the audiological design space, some indications can be inferred. The large differences among the participants suggest that, in a first phase, users' interaction is essential to gather individual preferences and thereby reach the optimum configuration for each single user. Simplifying the optimization task and offering a clear explanation of the one-dimensional slider made the process more transparent and increased users' empowerment. Once a recommender system is in place, this component might also prove useful in enhancing users' trust in the recommendations provided. Moreover, performing the optimization task in real-world environments ensured an accurate assessment and communication of users' preferences. In the short term, user preferences collected with this approach could flow into the standard clinical workflow and help hearing care professionals to fine-tune the hearing aids. However, a single static configuration, although personalized, might not fully satisfy the user. Our results indicate that such recommender systems should not simply model users as a sole set of optimized audiological parameters, because the preferred configuration varies depending on the context. It is therefore essential for these models to likewise classify the sound environment and motion state in order to infer the intents of the user. Being fully aware of the intent, by automatically labeling it, would add further value to the collected preferences and would allow to ask for user feedback in specific situations. That would make it feasible to verify hypotheses based on previous data, and progressively optimize several device configurations for different real-world listening scenarios. Once some configurations are learned, the hearing aids could automatically recommend them in specific situations and, by monitoring users' behavior, continuously calibrate to the preference of the user.

5 CONCLUSION

Internet-connected hearing aids open the opportunity for truly personalized hearing aids, which adapt to the needs of users in real-world listening scenarios. This study addressed the main challenges posed when rethinking hearing aids as recommender systems. It investigated how to effectively optimize the device settings by gathering user preferences in real-world environments. A complex audiological space was simplified by decoupling three audiological parameters and allowing the participants to manipulate one parameter at a time, comparing four discrete levels. The participants sequentially optimized the three audiological parameters, which were subsequently combined into a personalized device configuration. This configuration was blindly compared against a configuration personalized in a standard clinical workflow based on questions and pre-recorded sound samples, and six out of seven participants preferred the device settings learned in real-world listening environments. Thus, the approach seemed to effectively gather the main individual audiological preferences. The parameters resulted to have a different perceived usefulness, differently contributing to the listening experience of hearing aid users. The seven participants exhibited widely different audiological preferences. Furthermore, our results indicate that hearing aid users do not simply explore the audiological design space in search of a global optimum. Instead, most of them select multiple highly contrasting values along each parameter, depending on the context.

ACKNOWLEDGMENTS

We would like to thank Oticon A/S, Eriksholm Research Centre, and Research Clinician Rikke Rossing for providing hardware, access to test subjects, clinical approval and clinical resources.

REFERENCES

[1] Gediminas Adomavicius and Alexander Tuzhilin. 2011. Context-Aware Recommender Systems. In *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. Kantor (Eds.). Springer, Boston, MA, USA, 217–253. https://link.springer.com/chapter/10.1007/978-0-387-85820-3_7

[2] Eric Brochu, Vlad M. Cora, and Nando de Freitas. 2010. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. *CoRR* abs/1012.2599 (2010).

[3] Eric Brochu, Nando de Freitas, and Abhijeet Ghosh. 2007. Active Preference Learning with Discrete Choice Data. In *Proceedings of the 20th International Conference on Neural Information Processing Systems (NIPS '07)*. 409–416.

[4] Mead C. Killion. 2002. New Thinking on Hearing in Noise: A Generalized Articulation Index. *Seminars in Hearing* 23 (January 2002), 057–076. https://doi.org/10.1055/s-2002-24976

[5] Susanna L. Callaway and Andreea Micula. 2016. *Client Target and Real-ear Measurements.* Technical Report. Oticon A/S, Smørum, Denmark.

[6] John Clark. 1981. Uses and Abuses of Hearing Loss Classification. *ASHA: a journal of the American Speech-Language-Hearing Association* 23 (August 1981), 493–500.

[7] European Commission. 2019. Glossary: Hearing threshold. Retrieved August 16, 2019 from https://ec.europa.eu/health/scientific_committees/opinions_layman/en/hearing-loss-personal-music-player-mp3/glossary/ghi/hearing-threshold.htm

[8] Josefine J. Jensen. 2019. *Oticon Opn S Clinical Evidence.* Technical Report. Oticon A/S, Smørum, Denmark.

[9] Benjamin Johansen, Maciej J. Korzepa, Michael K. Petersen, Niels H. Pontoppidan, and Jakob E. Larsen. 2018. Inferring User Intents from Motion in Hearing Healthcare. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium.* https://doi.org/10.1145/3267305.3267683

[10] Benjamin Johansen, Michael K. Petersen, Niels H. Pontoppidan, and Jakob E. Larsen. 2019. Modelling User Utterances as Intents in an Audiological Design Space. In *Workshop on Computational Modeling in Human-Computer Interaction (CHI 2019).*

[11] Benjamin Johansen, Michael K. Petersen, Maciej J. Korzepa, Jan Larsen, Niels H. Pontoppidan, and Jakob E. Larsen. 2018. Personalizing the Fitting of Hearing Aids by Learning Contextual Preferences From Internet of Things Data. *Computers* 7, 1 (2018). https://doi.org/10.3390/computers7010001

[12] Maciej J. Korzepa, Benjamin Johansen, Michael K. Petersen, Jan Larsen, Niels H. Pontoppidan, and Jakob E. Larsen. 2018. Learning Preferences and Soundscapes for Augmented Hearing. In *IUI Workshops.*

[13] Yuki Koyama, Issei Sato, Daisuke Sakamoto, and Takeo Igarashi. 2017. Sequential Line Search for Efficient Visual Design Optimization by Crowds. *ACM Trans. Graph.* 36, 4, Article 48 (July 2017), 11 pages. https://doi.org/10.1145/3072959.3073598

[14] Nicolas Le Goff. 2015. *Amplifying Soft Sounds - a Personal Matter.* Technical Report. Oticon A/S, Smørum, Denmark.

[15] Nicolas Le Goff, Jesper Jensen, Michael S. Pedersen, and Susanna L. Callaway. 2016. *An Introduction to OpenSound Navigator^{TM}*. Technical Report. Oticon A/S, Smørum, Denmark.

[16] Nicholas A. Lesica. 2018. Why Do Hearing Aids Fail to Restore Normal Auditory Perception? *Trends in Neurosciences* 41, 4 (April 2018), 174–185. https://doi.org/10.1016/j.tins.2018.01.008

[17] Jeremy Marozeau and Mary Florentine. 2007. Loudness Growth in Individual Listeners with Hearing Losses: A Review. *The Journal of the Acoustical Society of America* 122, 3 (2007), EL81–EL87. https://doi.org/10.1121/1.2761924

[18] Elaine H. N. Ng, Mary Rudner, Thomas Lunner, Michael Syskind Pedersen, and Jerker Rönnberg. 2013. Effects of Noise and Working Memory Capacity on Memory Processing of Speech for Hearing-aid Users. *International Journal of Audiology* 52, 7 (2013), 433–441. https://doi.org/10.3109/14992027.2013.776181

[19] World Health Organization. 2013. *Multi-Country Assessment of National Capacity to Provide Hearing Care.* Geneva, Switzerland.
[20] World Health Organization. 2019. Deafness and Hearing Loss. Retrieved June 30, 2019 from https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss
[21] Antti Oulasvirta, Per O. Kristensson, Xiaojun Bi, and Andrew Howes. 2018. *Computational Interaction.* Oxford University Press.
[22] Jonathan E. Peele and Arthur Wingfield. 2016. The Neural Consequences of Age Related Hearing Loss. *Trends in Neurosciences* 39, 7 (2016), 486–497. https://doi.org/10.1016/j.tins.2016.05.001
[23] Dorothea Wendt, Renskje K. Hietkamp, and Thomas Lunner. 2017. Impact of Noise and Noise Reduction on Processing Effort. *Ear and Hearing* 38, 6 (2017), 690–700. https://doi.org/10.1097/aud.0000000000000454

Appendix B

Measuring and Modeling Context-Dependent Preferences for Hearing Aid Settings

Alessandro Pasta, Michael Kai Petersen, Kasper Juul Jensen, Niels Henrik Pontoppidan, Jakob Eg Larsen, Jeppe Høy Christensen. Measuring and Modeling Context-Dependent Preferences for Hearing Aid Settings.

Measuring and Modeling Context-dependent Preferences for Hearing Aid Settings

Alessandro Pasta* • Michael Kai Petersen • Kasper Juul Jensen • Niels Henrik Pontoppidan • Jakob Eg Larsen • Jeppe Høy Christensen

Abstract Despite having individual perceptual preferences toward sounds, hearing aid users often end up with default hearing aid settings that have no contextual awareness. However, the introduction of smartphone-connected hearing aids has enabled a rethinking of hearing aids as user-adaptive systems considering both individual and contextual differences.

In this study, we aimed to investigate the feasibility of such context-aware system for providing hearing aid users with a number of relevant hearing aid settings to choose from. During normal real-world hearing aid usage, we applied a smartphone-based method for capturing participants' listening experience and audiological preference for different intervention levels of three audiological parameters (Noise Reduction, Brightness, Soft Gain). Concurrently, we collected contextual data as both self-reports (listening environment and listening intention) and continuous data logging of the acoustic environment (sound pressure level, signal-to-noise ratio).

First, we found that having access to different intervention levels of the Brightness and Soft Gain parameters affected listening satisfaction. Second, for all three audiological parameters, the perceived usefulness of having access to different intervention levels was significantly modulated by context. Third, contextual data improved the prediction of both explicit and implicit intervention level preferences. Our findings highlight that context has a significant impact on hearing aid preferences across participants and that contextual data logging can help reduce the space of potential interventions in a user-adaptive system so that the most useful and preferred settings can be offered. Moreover, the proposed mixed-effects model is suitable for capturing predictions on an individual level and could also be expanded to predictions on a group level by including relevant user features.

Keywords contextual awareness · user-adaptive system · user preferences · user satisfaction · hearing healthcare · hearing aids

*Corresponding author, alpas@dtu.dk

Alessandro Pasta · Jakob Eg Larsen
Technical University of Denmark, Kongens Lyngby, Denmark

Jeppe Høy Christensen · Michael Kai Petersen

Eriksholm Research Centre, Snekkersten, Denmark

Kasper Juul Jensen

Oticon A/S, Smørum, Denmark

This paper or a similar version is not currently under review by a journal or conference. This paper is void of plagiarism or self-plagiarism as defined by the Committee on Publication Ethics and Springer Guidelines.

Introduction

Background

Hearing aid users with similar hearing loss perceive sounds in highly individual ways, exhibiting differences in the ability to understand speech in noisy environments (Killion 2002), in the loudness perception (Oetting et al. 2018), and in the perception of sounds close to their hearing threshold (Marozeau and Florentine 2007). Despite that, the prescription of hearing aid amplification is primarily based on pure-tone audiometry, a test that measures the hearing thresholds for tonal stimuli at typically eight different frequencies (Walker et al. 2013). Pure-tone audiometry is a threshold test of signal detection but does not adequately represent real-world hearing abilities (Killion 2002, Baguley et al. 2016), because it does not convey information about central auditory processing, nor the auditory processing of real-world signals (Musiek et al. 2017). For these reasons, the initial prescription is considered a starting point rather than the optimal solution to treat a hearing loss (Abrams et al. 2011). A subsequent fine-tuning of the hearing aid might be performed in follow-up visits, during which the hearing care professional modifies the hearing aid settings based on users' recollections of past listening experiences (Kochkin et al. 2010). However, the success of fine-tuning depends on the hearing care professional's ability to interpret and translate users' recollections (Elberling and Hansen 1999; Arlinger et al. 2017). Moreover, this is a time-consuming procedure, often requiring multiple visits to obtain a satisfactory configuration (Abrams et al. 2011), and it does not guarantee a significant advantage over a default initial prescription (Cunningham et al. 2001; Shi et al. 2007). Crucially, hearing aid users are often hesitant to seek help from their hearing care professional, which highlights the importance of promoting user empowerment and self-management through new technology (Bennett et al. 2019). All in all, alternative user-driven ways of personalizing hearing aids are warranted.

Furthermore, hearing aid users report listening difficulties in different real-life situations ranging from face-to-face conversations to social interactions (Galvez et al.

2012). To cope with the different situations, hearing aid users seem to prefer switching between highly contrasting hearing aid settings depending on contextual variables such as sound environment and listening intention (Johansen et al. 2017, Korzepa et al. 2018). This emphasizes the importance of everyday context on users' listening experience and on their preferences toward specific hearing aid settings. To address a need for contextual adaptation, hearing aid users can currently be provided with different pre-configured programs. Such programs are aimed at improving the listening experience in specific contexts (e.g., speech in noise, music (Hockley et al. 2010)) by setting predefined levels for different audiological parameters. Other hearing aid programs dynamically change their level of intervention according to the sensed environment - e.g., the level of noise reduction can be adjusted based on the ambient sound intensity levels (Schum 2003) within some predefined ranges. However, such programs are based on the average hearing aid user and disregard the fact that listening preferences are highly individual (Brons et al. 2013). As a consequence, users tend not to use other programs than the default one (Nelson et al. 2006). Ultimately, this indicates that it is crucial to account both for individual preferences and for the impact of real-world context45 on users' listening experience and preferences when prescribing or fine-tuning hearing aid settings.

Related Work

Several studies have documented that user-driven adjustments of hearing aids are feasible and potentially beneficial when investigated under controlled laboratory conditions (Yoon et al. 2017; Boothroyd and Mackersie 2017; Nelson et al. 2018; Jensen et al. 2019). Some studies have also reported on the benefits of training a personalized hearing aid program with a combination of sensor data and subjective preference feedback obtained from users in different real-life situations. However, results are mixed. Keidser and Alamudi (2013) trained the hearing aid settings using the SoundLearning algorithm (Chalupper et al. 2009), which learns and adjusts the amplification gain independently in four frequency bands and according to six different sound environments classified by the hearing aids (i.e., speech, speech and noise, quiet, noise, music, and car noise). They reported that 8 out of 18 participants preferred their trained hearing aid prescription in an evaluation phase (8 showed no preference and 2 preferred the un-trained prescription) and they concluded that training was efficient in those participants who initially wanted a change in their prescription. In another study, Aldaz et al. (2016) used smartphone-connected hearing aids to train settings with more "contextual awareness" by having participants perform A/B comparisons between the general program and context-specific programs alternating the microphone directionality or the noise reduction on and off. In the evaluation phase, the authors concluded that 7 out of 15 participants preferred the trained setting, 1 preferred the untrained setting, and 7 showed no preference. Notably, the learned preferences for microphone directionality and noise reduction were found to be nearly uniform across the different sound environments, which suggests that training did not effectively account for context. Overall, for roughly half of the participants in the two studies above (Keidser and Alamudi 2013; Aldaz et al. 2016), training was not efficient as the trained settings were not preferred outside of the training phase.

While training a personalized hearing aid configuration shows promise for specific individuals, it poses some challenges. Firstly, it assumes that users have one and only one preference in each context. However, when choosing between two alternative hearing aid settings in a specific sound environment, some users are consistent in their reported preferences, while others are not (Walravens et al. 2020). This inconsistency might be due to the fact that a selected setting does not yield a significant improvement in user experience and therefore leads to noisy preference assessments. Thus, understanding when a preferred setting is perceived to truly improve the listening experience would help focusing on the relevant audiological parameters and contexts. Alternatively, the inconsistency might be due to an incomplete notion of context, with two situations classified under the same context resulting in two different preferences. For example, different listening intentions might require different hearing aid settings even though the sound environment does not change. Previous studies attempting to adjust hearing aid programs to contextualized and individualized preferences did not consider the listening intention.

Furthermore, learning context-dependent listening preferences requires gathering preferences on multiple audiological parameters and in several real-world contexts. Typically, listening experiences in hearing aid users are measured with experience sampling – that is, having users explicitly provide in-situ ratings of their listening experience as often as possible e.g., via smartphone apps (Shiffman et al. 2008). This method has proven successful in terms of documenting real-world benefits of hearing aid settings (Andersson et al. 2021). Despite being more reliable (Amatriain et al. 2009), explicit feedback is scarce and places a burden on the user (Jawaheer et al. 2010). Moreover, since the hearing aid signal processing acts on several parameters of the sound (e.g., frequency compression, gain amplification) with varying strengths (e.g., levels of amplification), the space of possible hearing aid settings is vast (Pasta et al. 2019).

Therefore, when gathering user preferences in such a vast space and in several contexts, it is important to focus on settings and situations that cause a tangible improvement in user experience. Incidentally and importantly, data logging from modern hearing aids can provide information about implicit preferences toward hearing aid settings as well as information about the environmental context without imposing a burden on the user (Christensen et al. 2021). Previous research across different domains has shown that implicit and explicit feedback possess different characteristics and can complement each other (Jawaheer et al. 2010, 2014; Akehurst et al. 2012).

Research Objective

The above-stated challenges with hearing aid prescription and personalization are addressed by investigating the feasibility of a context-aware user-adaptive system, which aims to offer a choice among relevant hearing aid settings based on collected preferences and contexts of many other users (Pasta et al. 2019). Specifically, we apply a method for capturing users' experiences and (explicit and implicit) audiological preferences for different intervention levels of three audiological parameters. The data are collected by smartphone-connected hearing aids, which enabled users to evaluate different settings during their everyday life. Concurrently, contextual data are acquired both through self-reporting and through continuous data logging. Importantly, all data collection is performed using the typical daily-life setup (i.e., a smartphone and a pair of hearing aids) of a hearing aid user.

First, we analyze if listening satisfaction is related to the perceived usefulness of an audiological parameter. A user-adaptive system should be able to offer a choice among settings that, when relevant, leads to higher user satisfaction. Thus, we gather and compare in-situ ratings of listening satisfaction and of usefulness of choosing among different intervention levels (henceforth called “choice-usefulness”) of the three audiological parameters.

Second, we analyze whether everyday contexts influence the choice-usefulness. Indeed, we hypothesized that context has a measurable and distinct impact on the explicitly reported usefulness of choosing among different intervention levels of the parameters. This would entail that a user-adaptive system can reduce the space of possible hearing aid settings by assigning context-aware usefulness to the audiological parameters.

Third, we apply statistical modeling of user preferences for different intervention levels and hypothesize that contextual predictors enable a better account of the observed preferences. If so, a user-adaptive system would benefit from contextual information when predicting the preferred levels of intervention for a specific audiological parameter. The statistical modeling is performed for both explicitly reported level preferences and for implicit preferences derived from user interactions. Indeed, if a system could rely only on implicit preferences, the training phase would be less burdensome for the user.

Methods

Participants

We recruited experienced hearing aid users having a hearing loss compatible with the Oticon Opn™ S1 MiniRITE hearing aids and being iOS users. Seven participants (6 men and 1 woman) with mean age 58 years (SD = 12 years) were recruited. Five of them were working, while two were retired. The participants all had more than five years of experience with hearing aid usage. All participants had a binaural hearing loss ranging from mild to moderately severe, as classified by the American Speech-Language-Hearing Association (Clark 1981). The study was approved by the Research Ethics Committees of the Capital Region of Denmark. Before the study began, all participants received written information about the study and gave their informed consent. One participant did not allow for contextual data collection and was therefore excluded from the analysis.

Apparatus

The participants were prescribed a pair of Oticon Opn™ S1 miniRITE (Oticon A/S, Smoerum, Denmark) hearing aids and a frequency-specific amplification according to their hearing loss profile. All had iPhones with iOS 12 installed and additionally downloaded a custom smartphone app connected to the hearing aids via low-energy Bluetooth. Via the app, participants could control their hearing aid settings and submit in-situ reports (see details in the "Procedure" section). Furthermore, the app enabled continuous data logging of the active hearing aid settings and of the sound environment. The latter consisted of timestamped minute-based logs of the ambient acoustic environment sensed by the hearing aid microphones (see section "Contextual data"). The app interface also included an open-ended response form for optional additional comments.

Audiological parameters

During the study, three audiological parameters were evaluated: Noise Reduction (NR), Brightness (BR), and Soft Gain (SG). Each parameter targeted a specific dimension of the sound with four levels of intervention. The Noise Reduction parameter provides varying strength of noise reduction and directionality depending on the selected level (in ascending order of intensity, from level 1 to level 4). Thus, level 1 provides the lowest level of noise reduction and directionality, which amplifies most sound sources coming from all directions. In contrast, level 4 suppresses all sounds classified as non-speech and only amplifies sounds coming from the frontal direction. The Brightness parameter adjusts the amplification gain for high frequencies (i.e., frequencies above 1.5 kHz), while the Soft Gain parameter adjusts the amplification gain for soft sounds (i.e., sounds below 50 dB SPL). Common to all parameters, levels 1 and 2 provide a lower intervention compared to the default prescription (i.e., the level that would be automatically prescribed by the fitting software), while levels 3 and 4 provide increased intervention compared to the default prescription. The three targeted audiological parameters have been shown to be particularly important for the listening experience of hearing aid users (Ng et al. 2013; Johansen et al. 2017; Wendt et al. 2017) and to be perceived differently by individuals (Killion 2002; Marozeau and Florentine 2007).

Procedure

The participants were instructed to use their hearing aids "as usual" in their everyday lives for three consecutive weeks and to regularly select and compare, via the supplied smartphone app, the four contrasting intervention levels. Only one audiological parameter was active in each week, while the others were temporarily set at default prescription levels. This was a deliberate design choice aimed at simplifying participant interactions (i.e., less settings to navigate) and ensuring that participants could consciously track the effects of their actions on their listening experience (Pasta et al. 2019). The order by which the parameters were evaluated was fixed (week 1: NR; week 2: BR; week 3: SG). A visualization of the study timeline that each participant went through is given in Figure 1.

Each time the participants changed level, they had the option to submit an in-situ report of their explicit preference (i.e., preferred intervention level from 1 to 4); their current listening satisfaction (Likert rating scale from 1 to 5); the usefulness of having a choice among the four contrasting levels (Likert rating scale from 1 to 5, henceforth referred to as "choice-usefulness"); and the listening intention, listening environment and the state of motion (e.g., stationary, walking) from predefined categories. Moreover, the level selections during normal hearing aid usage were logged and used to define implicit preferences (see section "Statistical Modeling").

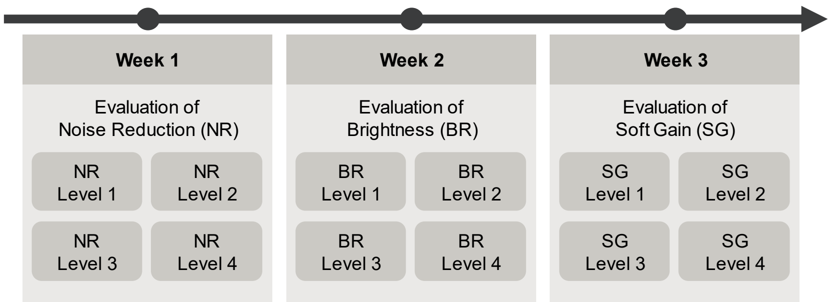


Figure 1. Study Timeline. Each parameter (Noise Reduction, Brightness, Soft Gain) was evaluated for the duration of one week. Each week, the participants were provided with 4 intervention levels of the parameter of the week.

Contextual data

Self-reported context is represented by in-situ reports of listening intention, listening environment, and motion state selected from drop-down lists with predefined categories. Note that for simplicity and due to sparse data, the motion state is not included in further analysis. In addition, due to the fairly low number of assessments received for some contexts (e.g., n = 8 for 'Meeting'), categories were collapsed across similar contexts. Table 1 shows the labels for all possible listening intentions (Table 1a) and listening environments (Table 1b) before ('Original label') and after ('New label') collapsing.

Besides the self-reported context, timestamped acoustic data logged from the hearing aids measured the ambient sound pressure levels (SPLs) and signal-to-noise ratios (SNRs) in decibels across a broad frequency band (0.1 – 10kHz)(Christensen et al. 2019). The SPL is the most used indicator of the sound wave strength and correlates well with human perception of loudness (Long 2014a). The SNR is the difference between the energy of a signal and the energy of any present noise and it is the key to speech intelligibility (Long 2014b). Each in-situ report and level selection was associated with acoustic data averaged across a 3-minute preceding time-window.

Table 1. Self-reported context (i.e., listening intention and listening environment) labels. Original labels that the participants could select among, and new labels after collapsing across similar contexts.

(a) Listening intention

Original label	n	New label	n
Ignore speech	12	Only me	128
Just me	116		
Meal	42	Focus	170
Music	42		
TV	86		
Speech	84	Social	238
Talk	108		
Meeting	8		

(b) Listening environment

Original label	n	New label	n
Home	240	Inside/Quiet	282
Kitchen	12		
Lecture	4		
Office	26		
Meeting	26	Inside/Noise	114
Party	8		
Public Indoor	54		
Restaurant	22		
Restroom	4		
Heavy duty	4	Outside/Noise	140
Outdoor	16		
Public outdoor	44		
Traffic	4		
Transport	72		

Statistical Modeling

Predictions of choice-usefulness and of explicit and implicit intervention level preferences were made using cumulative link proportional-odds mixed models. These models are ideal for multilevel modeling of longitudinal ordinal data (Hedeker 2008) and they are a class of the generalized mixed-effects modeling framework, which is popular among recommender systems (Condliff et al. 1999; Hedeker 2005; Chen et al. 2020b). Since the number of observations (in-situ reports and level selections) for each participant varied, we included data from all participants into global models. The individual-level effects were modeled as random effects and estimated with partial pooling. Such random effects allow model predictions to differ among participants, while partial pooling entails that, if a participant has fewer observations, her effect estimate will be partially based on the more abundant data from other participants. This is a good compromise between estimating an effect by completely pooling all users, which masks participant-level variation, and estimating an effect for all participants completely separately, which could give poor estimates for low-sample participants (Gelman and Hill 2006).

Prior to modeling, the continuous predictor SPL was converted into “Low intensity” and “High intensity”, while the continuous predictor SNR was converted into “Low quality” and “High quality”. This was done by using the median values for each participant as the cut-off between low and high. General recommendations for mixed-effects modeling were followed (Harrison et al. 2018). Fitting and supplementary statistics were performed in *R* using base functions and the ’ordinal’ package (RDocumentation 2019).

For in-situ reports, two separate models were applied for predicting the choice-usefulness rating and the explicitly preferred intervention level. The models were specified with both subjective and objective contextual predictors on the form:

$$logit\big(P(Y_i \leq j)\big) = \theta_j - \beta_1(env_i) - \beta_2(intent_i) - \beta_3(SPL_i) - \beta_4(SNR_i) - u(ID_i), \qquad (1)$$
$$i = 1, \dots, n, j = 1, \dots, J - 1$$

This is a model for the cumulative probability of the i^{th} choice-usefulness rating (or preferred intervention level) falling in the j^{th} category or below, where *i* indexes all observations and $j = 1, \dots, J$ indexes the response categories. In the model for choice-usefulness, $J = 5$. In the model for preferred intervention level, $J = 4$. θ_j are threshold parameters (or cut-points), which are assumed to be equidistant between the response categories. We take the participant effects (ID) to be random and assume that the effects are IID and normal: $u(ID_i) \sim N(0, \sigma_u^2)$. The self-reported listening environment (*env*) and listening intention (*intent*) are added as fixed effects predictors together with the categorical SPL and SNR.

The same model, but without the subjective contextual predictors, was applied to predict the implicit preferences (i.e., level selections) during normal hearing aid usage from user interaction event-logs. Note that only level selections that were kept for minimum three minutes were included as observations to the latter model. This was to ensure that random level selections (i.e., playing around) did not confound the outcome.

Besides inspection of coefficient magnitude and confidence intervals, likelihood ratio-tests based on the χ^2 test statistic were employed to test the significance of contextual predictors.

Results

Descriptive statistics: hearing aid usage and auditory ecology

Prior to assessing the main hypotheses of the study, we describe the main features of the collected data.

The number of level selections and the number of submitted assessments varied across participants (see Table 1). Overall, 8.8% (SD=3.0%) of all level selections led to an in-situ report and a preference submission. This percentage did not differ markedly among the three audiological programs (Noise reduction: M = 9.8%; Brightness: M = 10.3%; Soft gain: M = 6.8%), indicating a fair comparison of the programs.

Table 2. Participants' characteristics and data logs. PTA refers to the average of hearing threshold levels at four specified frequencies (0.5, 1, 2, 4 kHz). The median SNR and SPL are computed from all data-logs for each level selection.

Participant	Hearing loss (PTA) (dB)	In-situ reports (count)	Level selections (count)	Percentage of selections leading to an in-situ report (%)	Median SPL (dB)	Median SNR (dB)
1	67	36	858	4.2	56.40	19.17
2	47	21	227	9.3	49.54	14.76
3	44	19	248	7.7	56.28	10.11
4	36	131	1072	12.2	59.33	17.71
5	54	35	293	12.0	59.03	13.41
6	35	26	343	7.6	66.68	23.97
Mean	47	45	507	8.8	57.88	20.50
SD	11	43	363	3.0	5.57	6.35

The logged acoustic data documented that the participants had different exposure to different sound environments (see Table 2 and Figure 2). However, importantly, there was agreement between the sound exposure they experienced during their normal device usage (i.e., changing levels throughout the day) and when submitting in-situ self-reports (of their listening experience and level preferences). The scatter plots in Figure 2 show the distribution of SPL (Figure 2a) and SNR (Figure 2b) as deciles measured either when performing in-situ ratings (y-axis) or when changing levels throughout the day (x-axis). Notably, despite participant-specific offsets (e.g., participant n. 1 consistently experiences higher SPL during preference submission than during everyday device usage), the relationship between SPL deciles for preference submissions and level selections is linear with slope $\beta = 0.958$ ($F = 121.57$, $p < 0.001$).

On a group level, the relationship between SNR deciles for preference submission and level selections is also linear with slope $\beta = 0.729$ ($F = 37.00$, $p < 0.001$). However, participant n. 6 experiences, on average, much higher SNRs during everyday level selections than when performing in-situ ratings, which indicates that most of the participant's ratings were performed under noisy or quiet conditions (i.e., low quality of the signal). Please note that the discrepancy might also be driven by the participant experiencing very high SNRs for some of the logged level selections.

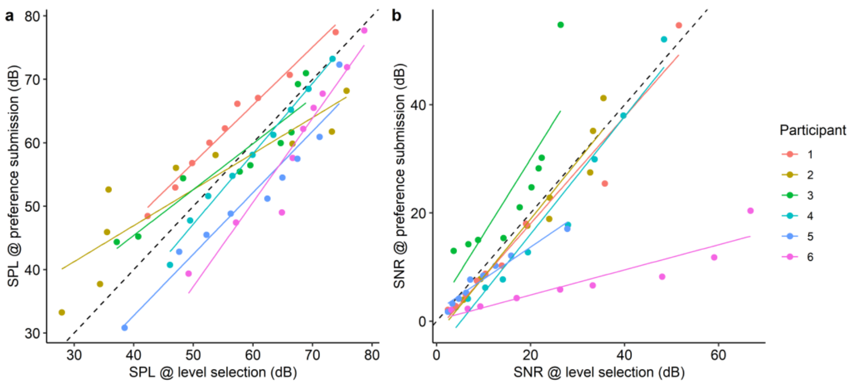


Figure 2: Relationship between the distribution of acoustic characteristics for in-situ preference submissions (y-axis) and while selecting intervention levels during normal device usage (x-axis). Each dot represents the acoustic value at a decile (1st to 9th) for one participant (colors). The dashed line indicates a slope of y = x.

We also assessed whether the self-reported contexts possessed different acoustic characteristics. If so, subjective self-reports conveyed more than the individual perception of auditory scenes. Figure 3 shows boxplots of each reported listening environment (Figure 3a) and listening intention (Figure 3b) against either SPL (top panels) or SNR (bottom panels). Please note that the boxplots are based on pooled data among all participants.

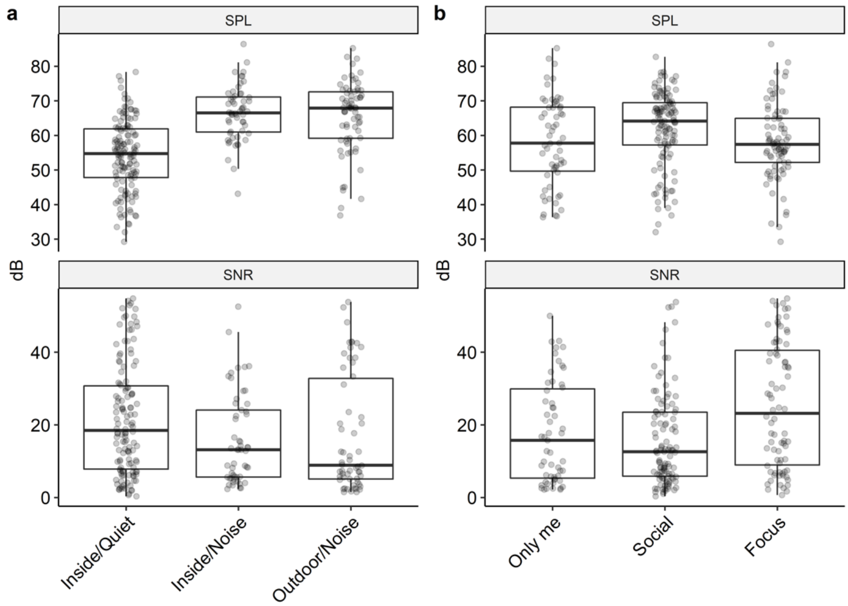


Figure 3: SPL and SNR for self-reported listening environments (a) and listening intentions (b).

The moderating effects of the self-reported contexts on SNR and SPL were evaluated by applying linear mixed-effects models. These models predict either SPL or SNR while controlling for time-of-day with random-effects offsets (e.g., SPL might simply be higher mid-day compared to end-of-day due to daily life activities). We observed main effects of listening environment on SPL ($F(2,221) = 42.844$, $p < 0.001$) and of listening intention on SNR ($F(2,229) = 3.450$, $p = 0.033$). For SPL, the largest effect was between listening environments “Inside/Quiet” and “Outdoor/Noise” ($\beta = 12.457$, $SE = 2.881$, $t = 4.324$, $p < 0.001$). For SNR, the largest effect was between listening intentions “Only me” and “Focus” ($\beta = 7.176$, $SE = 3.288$, $t = 2.182$, $p = 0.030$).

Relationship between choice-usefulness and listening satisfaction

Participants were asked to rate the usefulness of having four intervention levels available to choose from (i.e., choice-usefulness) and then to rate the current listening satisfaction. Thus, high ratings of choice-usefulness followed by high ratings of listening satisfaction are assumed to represent situations where audiological needs are met. To investigate how strongly a useful choice of intervention levels impacts satisfaction, we computed the correlation between the two types of ratings across all participants. Figure 4 shows contingency tables for each audiological parameter, which indicate a stronger correlation for the Brightness and Soft gain parameters than for the Noise reduction parameter. Indeed, Pearson's correlation tests revealed that satisfaction and choice-usefulness were not related when using the Noise Reduction parameter ($r = 0.064$, $t = 0.637$, $df = 96$, $p = 0.526$), but they were for the Brightness ($r = 0.383$, $t = 3.908$, $df = 89$, $p < 0.001$) and Soft Gain ($r = 0.400$, $t = 3.830$, $df = 77$, $p < 0.001$) parameters.

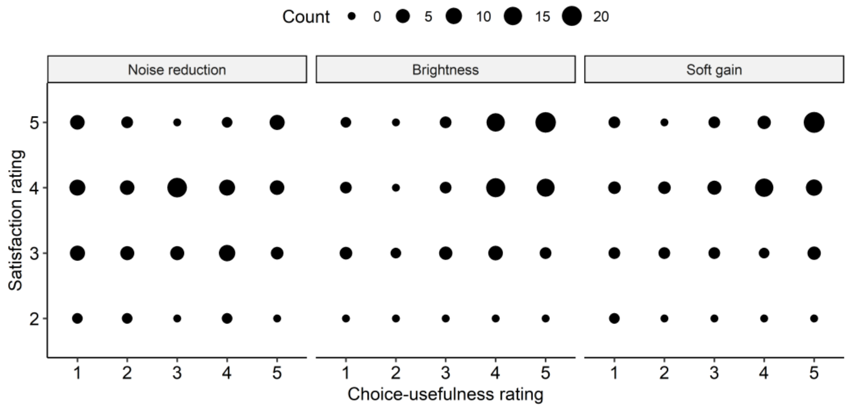


Figure 4: Contingency tables for rating the listening satisfaction (y-axis) and the choice-usefulness (x-axis) of each audiological program. Data are pooled among all in-situ assessments from all participants.

Separating the rating data by the contextual SNR revealed a higher correlation between satisfaction and choice-usefulness for Brightness in lower quality environments ($r = 0.443$, $t = 3.129$, $df = 40$, $p = 0.003$) compared to in higher quality environments ($r = 0.071$, $t = 0.423$, $df = 35$, $p = 0.675$), suggesting that having access to different levels of the Brightness parameter is more strongly associated to listening satisfaction when the quality of the listening environment is below the median. In contrast, Soft Gain exhibited higher correlation in high quality listening environments ($r = 0.572$, $t = 3.886$, $df = 31$, $p < 0.001$) compared to lower quality listening environments ($r = 0.366$, $t = 2.080$, $df = 28$, $p = 0.047$). The correlation between satisfaction and choice-usefulness for the Noise reduction parameter was again not significant after splitting the data by SNR. In summary, the relationship between choice-usefulness and listening satisfaction is distinct among the audiological parameters and varies with the context (here, SNR).

Contextual impact on choice-usefulness and explicit level preferences

One of the main aims of the study is to investigate, for the three audiological parameters, the impact of context on the perceived choice-usefulness and on explicit level preferences.

Ideally, this can lead to context-aware recommendations of audiological programs combining the most relevant parameters and intervention levels for each situation.

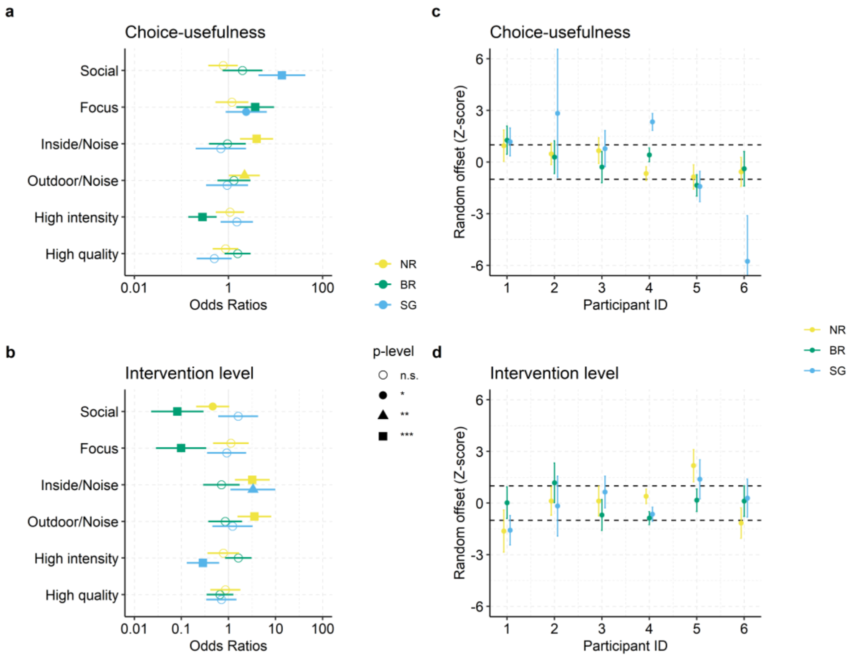


*Figure 5: In **a** and **b**, coefficients (as log odds ratios) and 95% confidence intervals for predicting choice-usefulness and explicit level preference from in-situ ratings. In **c** and **d**, the corresponding random offsets due to participant effects. Note that the models were fitted separately for each audiological parameter (NR, BR, and SG). The baseline conditions for the contextual predictors were: "Only me" (for listening intention), "Quiet/Indoor" (for listening environment), "Low intensity" (for SPL), "Low quality" (for SNR).*

In this section, we investigate the contextual impact by applying mixed-modeling of the in-situ ratings using both subjective and objective contextual predictors. Across the three audiological parameters, the contextual predictors (self-reported listening environment and intention, SPL, SNR) were found to significantly increase the prediction of choice-usefulness ratings (likelihood ratio test, $\chi^2(6) = 21.71$, $p = 0.002$) and intervention level preferences (likelihood ratio test, $\chi^2(6) = 14.418$, $p = 0.025$). Figure 5a-b shows the estimated coefficients when modeling data from each parameter separately with random effects offsets for participants (i.e., Eq 1). Notably, listening intention, listening environment, and SPL modulated both the choice-usefulness and level preference.

The random effects offsets (Figure 5c-d) indicate that participants had comparable ratings and level preferences (i.e., most falling within ±1 SD). Nevertheless, a few outliers were observed. For instance, participant n. 4 consistently rated the Soft Gain choice-usefulness higher than 1 SD from the group mean and participant n. 6 rated it much lower than the group mean, albeit in the latter case, the large error bars indicate that the estimated random offset is based on few observations. For Noise Reduction, participant n. 5 preferred significantly higher levels than the group mean.

Preference prediction from real-world usage patterns

The level preferences modeled in Figure 5 represent explicitly preferred levels. That is, levels that the participants purposefully reported as preferences. However, during normal real-world usage, participants made ~11 times more level selections than preference submissions (see Table 2), with automatically logged SPL and SNR associated to them. While some of these level selections were made to perform momentary comparisons, other level selections were made and used for longer periods of time. Participants made on average 377 active level selections (i.e., level selections that are set for at least three minutes), which is ~8 times more than preference submissions. Thus, a user-adaptive system could potentially leverage on these in case explicit feedback is not available.

We first assessed the contextual modulation of the implicit preferences by applying the statistical model in Eq. 1 to data from all participants. As was the case with the explicit preference data (Figure 5), SPL and SNR significantly improved the model's ability to predict intervention level ($\chi^2(8) = 17.43$, $p = 0.026$). The context-aware model prediction is shown in Figure 6a as a red solid line together with both the observed preferences (dots with error bars) and the prediction from a NULL model – i.e., an intercept only model (blue solid line in Figure 6a). Visually, differences in predictions between the two models are subtle. However, the contextual aware model does capture more variation in the observed preferences (Person's correlation - NULL model: $r = 0.48$, 95% CI = [0.23 to 0.67], df = 46, $p < 0.001$; Context-aware model: $r = 0.57$, 95% CI = [0.34 to 0.73] df = 46, $p < 0.001$), which is evident in Figure 6b with the context-aware model being able to predict a wider range of preference. For example, for the "High intensity" / "High quality" condition with the Brightness parameter the context aware model is able to better fit the observed data.

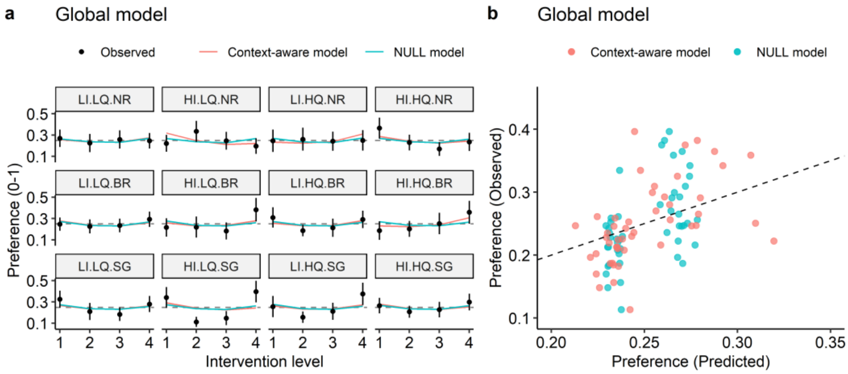


*Figure 6: Observed and predicted implicit preference for intervention level using data from all participants. In **a**, model predictions are shown together with the observed relative preference for each intervention grouped by all combinations of SPL and SNR (columns) and separated by audiological parameter (rows). LI = "Low intensity"; HI = "High intensity"; LQ = "Low quality"; HQ = "High quality". In **b**, the difference between predicted and observed preference is shown as a scatter plot with the dashed line indicating a y=x relationship.*

In addition, we assessed how well the model performed on a single-user level by fitting it only on data gathered from participant n. 4 (the participant with most data logged, see Table 2). The predictions are shown in Figure 7, and again SPL and SNR significantly improved the model fit ($\chi^2(8) = 19.97$, $p = 0.010$), and produced closer fitting predictions (Person's correlation - NULL model: r = 0.45, 95% CI = [0.19 to 0.65], df = 46, p =0.001; Context-aware model: r = 0.67, 95% CI = [0.48 to 0.80] df = 46, p < 0.001).

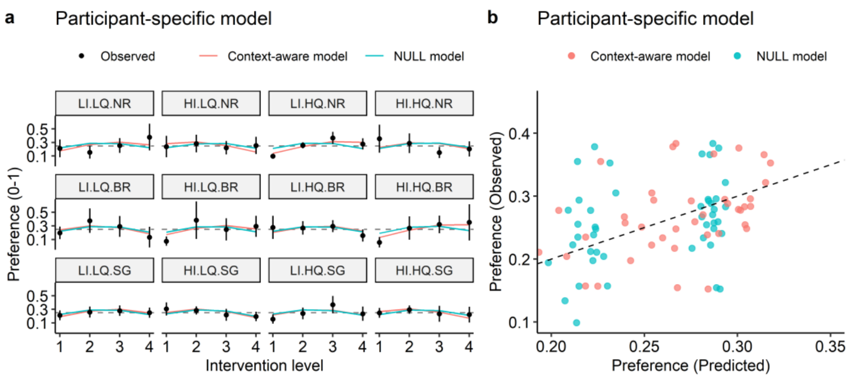


*Figure 7: Observed and predicted implicit preference for intervention level for participant n. 4. In **a,** model predictions are shown together with the observed relative preference for each intervention grouped by all combinations of SPL and SNR (columns) and separated by audiological parameter*

*(rows). LI = "Low intensity"; HI = "High intensity"; LQ = "Low quality"; HQ = "High quality". In **b**, the difference between predicted and observed preference is shown as a scatter plot with the dashed line indicating a y=x relationship.*

Discussion

This study applied a novel smartphone-based method for capturing real-world in-situ experiences and preferences, combined with data of environmental sound logged from the hearing aids. By investigating the impact of everyday context on users' listening experience and preferences, this study aimed to shed light on the feasibility of a context-aware user-adaptive system for providing useful audiological interventions.

Descriptive analysis of the collected data showed that the sound experienced when changing settings throughout the day was, on a group level, equal to that experienced when submitting self-reports (Figure 2). However, specific participants (e.g., n. 6 in Figure 2) showed a stronger deviation between the sounds experienced in the two situations. This may be either because self-reports are cognitively demanding (hence, they are completed in quiet environments only) or because reports are only submitted when problems are experienced (thus, leading to worse SNR for preference reports than for normal level changes). Nevertheless, lack of representativeness of in-situ preference and experience reports can, to some extent, be expected (Schinkel-Bielefeld et al. 2020; Ziesemer et al. 2020). Comparing the distributions of sound collected during self-report submission and everyday level changes (i.e., Figure 2) can help to validate participants' data. Consequently, more trust can be placed in the data collected from those participants who exhibit a relationship close to β = 1 between the sounds in the two situations. Moreover, we found that self-reported context both supports and differentiates the automatically logged contextual sound data (Figure 3). The self-reported listening environments were associated with different loudness of the environment (i.e., SPL). The self-reported listening intentions were associated with different quality of the environment (i.e., SNR). In particular, "focused" listening intentions (e.g., watching TV) were associated with higher SNRs, indicating that the sound signals convey clean and relevant information. Conversely, "social" listening intentions exhibited low SNR and high SPL, suggesting that such sound environments are characterized by poor signals and loud noise. This analysis documents that self-reports not only reflect subjective perceptual evaluations of the listening scene, but also convey objective information that are relevant for a hearing aid adjustment. At the same time, objectively similar acoustic environments might imply different audiological needs according to the self-reported listening environments and intentions. This means that a more fine-grained resolution of user context can be obtained by combining self-reports with objective data logging.

We also investigated how the perceived usefulness (i.e., rated “choice-usefulness”) of being offered a choice between different intervention levels of three independent audiological parameters affected the rated listening satisfaction. That is, if the intervention levels of an offered audiological parameter solve the listening needs of a user (or not), then the rated satisfaction should increase (or decrease). For the Noise Reduction parameter there was no correlation between satisfaction and choice-usefulness. This might be explained by the limited audibility of the change between the intervention levels of the parameter. Indeed, a substantial change (from 3 to 4 decibel) of acceptable noise level is required to yield a minimal clinically important and perceptual difference (Wong et al. 2018). Conversely, for the Brightness and Soft Gain parameters, significant correlations between satisfaction and choice-usefulness were observed and noted to be moderated by the quality of the sound environment (i.e., SNR). This implies that the listening experience can indeed be influenced by offering the user a choice among different intervention levels, and that the outcome depends on the parameter and on the context the user is in.

By applying statistical multi-level modeling, we examined the influence of context on the perceived usefulness of the offered intervention levels. In summary, self-reported and objective contexts have measurable and distinct impacts on the rated choice-usefulness. For instance, the Brightness parameter was significantly more useful in “Focus” listening intentions and in low intensity sound environments. This is consistent with previous studies showing that high-frequency amplification can be useful to improve speech understanding (Hornsby et al. 2011; Levy et al. 2015) and sound localization (Best et al. 2005). These findings suggest that a user-adaptive system can assign context-aware usefulness to the three audiological parameters based on previous user feedback. Contextual information would help filter the complex space of possible hearing aid settings, by providing an indication about which parameter the user should be asked to adjust. As applied in this study, the proposed mixed-effects model (Eq. 1) accounts for individual differences by including a random term for each user, which enables user-level predictions. However, the model could be expanded by simply adding random terms for relevant user features, such as hearing loss, age, measures of auditory perception, patterns of hearing aid use (Pasta et al. 2021). This would enable group-level predictions for users with similar features. In this way, users with sparse feedback (i.e., new users or users that do not supply explicit feedback) that share similar features with other users could benefit from the learned context-dependent preferences to alleviate the cold start problem (Chen et al. 2020a).

While choice-usefulness can help determine in which contexts a given audiological parameter should be adjusted, context-aware predictions of the preferred intervention level can help decide which levels of the parameter are the most important. Thus, the collected preferences for different intervention levels were modelled with contextual predictors. Across all participants, there was clear evidence that contextual data improve the prediction of the self-reported explicit level preferences. Coefficients from the statistical model (Figure 5) revealed distinct predictors for the three audiological parameters: higher levels of Noise Reduction were preferred in noisier environments, but not in social situations; higher levels of Brightness were preferred when being alone; higher levels of Soft Gain were preferred in quieter environments. These effects provide an indication on the direction that should be taken (i.e., increasing or decreasing) when adjusting the level of each parameter depending on the context. This may result in more relevant levels proposed to the user, ensuring a more effective interaction and a more engaging experience. The relevance of both logged and self-reported context suggests that it is important to model context both by directly observable features (here, SPL and SNR), as well as by hidden features reflecting the user's specific intentions (e.g., enhancing speech or ignoring voices) and situational environment. Similarly, although the differences were subtle, we found that contextual information obtained from continuous data logging (SPL and SNR) improves predictions of implicit level preferences (i.e., preferences derived from usage patterns) both on a group level (Figure 6) and user level (i.e., participant n. 4 in Figure 7). Thus, a system aimed at autonomous preference prediction will benefit from continuously logging contextual data from the hearing aid microphones and from using device interactions (e.g., level selections) to assign a preference to the offered choices. In addition, while objective data logging cannot fully capture subjective listening intentions (Figure 3) and implicit preferences are potentially less reliable (Amatriain et al. 2009), capitalizing on these data would help overcome the scarcity of user feedback and enable preference modeling of new or less engaged users.

Limitations

Real-world data has high ecological validity (Verma et al. 2017; Hicks et al. 2019) but also lacks control for when, where, and how much data are logged. In that sense, a limitation of this study is that some participants collected less data than others. However, in our statistical modeling, we specifically adjusted for effects of individual differences among participants by partial pooling (see *Methods*).

Due to specific requirements for the participants to be included in the study (being experienced hearing aid users, having a hearing loss compatible with the Oticon Opn™ S1 MiniRITE hearing aids, being iOS users), the sample size of the study is rather small. However, the aim of the study was to evaluate the impact of everyday context on hearing aid users with similar features. Thus, the repeated-measures design (i.e., continuous data logging and repeated in-situ reports) ensured that a high number of observations were acquired for statistical modeling of data representing everyday hearing aid usage. This helped compensating for the rather small sample size. A limitation remains in disentangling specific user features from the results as this would require a larger sample. Moreover, while the statistical modeling (Eq. 1) accounts for individual differences among the participants in terms of choice-usefulness and level preference random offsets, it does not account for individual effects of context on individual preferences. However, the inclusion of more participants with more repeated measures could enable an expansion of the statistical model to also account for participant-specific random slopes. That would enable modeling participant-specific sensitivity towards contextual predictors (Harrison et al. 2018; Gao et al. 2019).

The three audiological parameters were evaluated in chronological order. Thus, temporal effects (e.g., hearing aid acclimatization (Wright and Gagné 2021) or study fatigue) cannot be disentangled from the main results. Future research could assign random order of the parameters to each participant to investigate detailed differences among them without confounds from temporal effects.

Conclusions

Rethinking hearing aids as user-adaptive systems can provide a context-aware and personalized alternative to hearing aids with predefined settings. Our results show that participants' listening experience can effectively be influenced by providing a choice among different intervention levels of specific audiological parameters. Moreover, contextual data significantly improved predictions of how useful the offered choice among intervention levels were perceived to be. Additionally, contextual data significantly improved the prediction of both explicit and implicit level preferences.

We conclude that, when rethinking hearing aids as context-aware user-adaptive systems, both objective (i.e. SNR and SPL) and subjective (i.e., self-reported listening intention and environment) contextual data should be taken into consideration to optimize recommendations of the most relevant parameters and intervention levels. We propose training a proportional-odds mixed-effects model on preference and level selections data from experienced hearing aid users to provide context-aware recommendations to new users.

Acknowledgements We would like to thank Eriksholm Research Centre and Oticon A/S for providing hardware, access to test subjects, clinical approval, and clinical resources. We would like to thank the test participants for their contribution and Rikke Rossing for the clinical support.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Abrams H, Edwards B, Valentine S, Fitz K (2011) A patient-adjusted fine-tuning approach for optimizing hearing aid response. Hear Rev 18:18–27

Akehurst J, Koprinska I, Yacef K, et al (2012) Explicit and Implicit User Preferences in Online Dating. In: Cao L, Huang JZ, Bailey J, et al. (eds) New Frontiers in Applied Data Mining. PAKDD 2011 International Workshop. Springer, Berlin, Heidelberg, pp 15–27

Aldaz G, Puria S, Leifer LJ (2016) Smartphone-Based System for Learning and Inferring Hearing Aid Settings. J Am Acad Audiol 27:732–749. https://doi.org/10.3766/jaaa.15099

Amatriain X, Pujol JM, Oliver N (2009) I Like It... I Like It Not: Evaluating User Ratings Noise in Recommender Systems. In: Houben G, McCalla G, Pianesi F, Zancanaro M (eds) User Modeling, Adaptation, and Personalization. UMAP 2009. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp 247–258

Andersson KE, Andersen LS, Christensen JH, Neher T (2021) Assessing Real-Life Benefit From Hearing-Aid Noise Management: SSQ12 Questionnaire Versus Ecological Momentary Assessment With Acoustic Data-Logging. Am J Audiol 30:93–104. https://doi.org/10.1044/2020_AJA-20-00042

Arlinger S, Nordqvist P, Öberg M (2017) International Outcome Inventory for Hearing Aids: Data From a Large Swedish Quality Register Database. Am J Audiol 26:443–450. https://doi.org/10.1044/2017_AJA-16-0123

Baguley DM, Cope TE, McFerran DJ (2016) Functional auditory disorders. In: Hallett M, Stone J., Carson A (eds) Handbook of Clinical Neurology. Elsevier, pp 367–378

Bennett RJ, Laplante-Lévesque A, Eikelboom RH (2019) How Do Hearing Aid Owners Respond to Hearing Aid Problems? Ear Hear 40:77–87. https://doi.org/10.1097/AUD.0000000000000595

Best V, Carlile S, Jin C, van Schaik A (2005) The role of high frequencies in speech localization. J Acoust Soc Am 118:353–363. https://doi.org/10.1121/1.1926107

Boothroyd A, Mackersie C (2017) A "Goldilocks" Approach to Hearing-Aid Self-Fitting: User Interactions. Am J Audiol 26:430–435. https://doi.org/10.1044/2017_AJA-16-0125

Brons I, Houben R, Dreschler WA (2013) Perceptual Effects of Noise Reduction With Respect to Personal Preference, Speech Intelligibility, and Listening Effort. Ear Hear 34:29–41. https://doi.org/10.1097/AUD.0b013e31825f299f

Chalupper J, Junius D, Powers T (2009) Algorithm lets users train aid to optimize compression, frequency shape, and gain. Hear J 62:26–33.

https://doi.org/10.1097/01.HJ.0000359131.58356.8f

Chen W-H, Hsu C-C, Lai Y-A, et al (2020a) Attribute-Aware Recommender System Based on Collaborative Filtering: Survey and Classification. Front Big Data 2:1–30. https://doi.org/10.3389/fdata.2019.00049

Chen Z, Zhu S, Niu Q, Zuo T (2020b) Knowledge Discovery and Recommendation With Linear Mixed Model. IEEE Access 8:38304–38317. https://doi.org/10.1109/ACCESS.2020.2973170

Christensen JH, Pontoppidan NH, Rossing R, et al (2019) Fully Synthetic Longitudinal Real-World Data From Hearing Aid Wearers for Public Health Policy Modeling. Front Neurosci 13:1–5. https://doi.org/10.3389/fnins.2019.00850

Christensen JH, Saunders GH, Porsbo M, Pontoppidan NH (2021) The everyday acoustic environment and its association with human heart rate: evidence from real-world data logging with hearing aids and wearables. R Soc Open Sci 8:201345. https://doi.org/10.1098/rsos.201345

Clark JG (1981) Uses and abuses of hearing loss classification. ASHA 23:493–500

Condliff MK, Lewis DD, Madigan D, Posse C (1999) Bayesian Mixed-Effects Models for Recommender Systems. In: ACM SIGIR ’99 Workshop on Recommender Systems: Algorithms and Evaluation

Cunningham DR, Williams KJ, Goldsmith LJ (2001) Effects of Providing and Withholding Postfitting Fine-Tuning Adjustments on Outcome Measures in Novice Hearing Aid Users. Am J Audiol 10:13–23. https://doi.org/10.1044/1059-0889(2001/001)

Elberling C, Hansen VK (1999) Hearing instruments: Interaction with user preference. In: Rasmussen AN, Osterhammel PA, Andersen T, Poulsen T (eds) Auditory Models and Non-Linear Hearing Instruments. Proceedings of the 18th Danavox Symposium. Copenhagen, Denmark, pp 341–357

Galvez G, Turbin MB, Thielman EJ, et al (2012) Feasibility of Ecological Momentary Assessment of Hearing Difficulties Encountered by Hearing Aid Users. Ear Hear 33:497–507. https://doi.org/10.1097/AUD.0b013e3182498c41

Gao B, Zhan G, Wang H, et al (2019) Learning with Linear Mixed Model for Group Recommendation Systems. In: Proceedings of the 2019 11th International Conference on Machine Learning and Computing - ICMLC ’19. ACM Press, New York, New York, USA, pp 81–85

Gelman A, Hill J (2006) Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press, Cambridge

Harrison XA, Donaldson L, Correa-Cano ME, et al (2018) A brief introduction to mixed effects modelling and multi-model inference in ecology. PeerJ 6:e4794.

https://doi.org/10.7717/peerj.4794

Hedeker D (2008) Multilevel Models for Ordinal and Nominal Variables. In: De Leeuw J, Meijer E (eds) Handbook of Multilevel Analysis. Springer New York, New York, NY, pp 237–274

Hedeker D (2005) Generalized linear mixed models. In: Everitt BS, Howell DC (eds) Encyclopedia of Statistics in Behavioral Science. John Wiley & Sons, Ltd, Chichester, UK, pp 729–738

Hicks JL, Althoff T, Sosic R, et al (2019) Best practices for analyzing large-scale health data from wearables and smartphone apps. npj Digit Med 2:45. https://doi.org/10.1038/s41746-019-0121-1

Hockley NS, Bahlmann F, Chasin M (2010) Programming hearing instruments to make live music more enjoyable. Hear J 63:30. https://doi.org/10.1097/01.HJ.0000388538.27405.93

Hornsby BWY, Johnson EE, Picou E (2011) Effects of degree and configuration of hearing loss on the contribution of high- and low-frequency speech information to bilateral speech understanding. Ear Hear 32:543–555. https://doi.org/10.1097/AUD.0b013e31820e5028

Jawaheer G, Szomszor M, Kostkova P (2010) Comparison of implicit and explicit feedback from an online music recommendation service. In: Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems, HetRec 2010, Held at the 4th ACM Conference on Recommender Systems, RecSys 2010. pp 47–51

Jawaheer G, Weller P, Kostkova P (2014) Modeling user preferences in recommender systems: A classification framework for explicit and implicit user feedback. ACM Trans Interact Intell Syst 4:. https://doi.org/10.1145/2512208

Jensen NS, Hau O, Nielsen JBB, et al (2019) Perceptual Effects of Adjusting Hearing-Aid Gain by Means of a Machine-Learning Approach Based on Individual User Preference. Trends Hear 23:1–23. https://doi.org/10.1177/2331216519847413

Johansen B, Petersen M, Korzepa M, et al (2017) Personalizing the Fitting of Hearing Aids by Learning Contextual Preferences From Internet of Things Data. Computers 7:1. https://doi.org/10.3390/computers7010001

Keidser G, Alamudi K (2013) Real-Life efficacy and reliability of training a hearing aid. Ear Hear 34:619–629. https://doi.org/10.1097/AUD.0b013e31828d269a

Killion MC (2002) New Thinking on Hearing in Noise: A Generalized Articulation Index. Semin Hear 23:057–076. https://doi.org/10.1055/s-2002-24976

Kochkin S, Beck DL, Christensen Laurel A, et al (2010) MarkeTrak VIII: The impact of the hearing healthcare professional on hearing aid user success. Hear Rev 17:12–34

Korzepa MJ, Johansen B, Petersen MK, et al (2018) Learning preferences and soundscapes for augmented hearing. CEUR Workshop Proc 2068:

Levy SC, Freed DJ, Nilsson M, et al (2015) Extended high-frequency bandwidth improves reception of speech in s. Ear Hear 36:214–224. https://doi.org/10.1097/AUD.0000000000000161.Extended

Long M (2014a) Fundamentals of Acoustics. In: Architectural Acoustics. pp 39–79

Long M (2014b) Human Perception and Reaction to Sound. In: Architectural Acoustics. pp 81–127

Marozeau J, Florentine M (2007) Loudness growth in individual listeners with hearing losses: A review. J Acoust Soc Am 122:EL81–EL87. https://doi.org/10.1121/1.2761924

Musiek FE, Shinn J, Chermak GD, Bamiou D-E (2017) Perspectives on the Pure-Tone Audiogram. J Am Acad Audiol 28:655–671. https://doi.org/10.3766/jaaa.16061

Nelson J, Kiessling J, Dyrlund O, Groth J (2006) Integrating hearing instrument datalogging into the clinic. In: American Academy of Audiology Convention. Minneapolis

Nelson PB, Perry TT, Gregan M, VanTasell D (2018) Self-Adjusted Amplification Parameters Produce Large Between-Subject Variability and Preserve Speech Intelligibility. Trends Hear 22:1–13. https://doi.org/10.1177/2331216518798264

Ng EHN, Rudner M, Lunner T, et al (2013) Effects of noise and working memory capacity on memory processing of speech for hearing-aid users. Int J Audiol 52:433–441. https://doi.org/10.3109/14992027.2013.776181

Oetting D, Hohmann V, Appell J-E, et al (2018) Restoring Perceived Loudness for Listeners With Hearing Loss. Ear Hear 39:664–678. https://doi.org/10.1097/AUD.0000000000000521

Pasta A, Petersen MK, Jensen KJ, Larsen JE (2019) Rethinking hearing aids as recommender systems. In: The 4th Workshop on Health Recommender Systems co-located with ACM RecSys'19. Copenhagen, Denmark, pp 11–17

Pasta A, Szatmari T-I, Christensen JH, et al (2021) Clustering Users Based on Hearing Aid Use: An Exploratory Analysis of Real-World Data. Front Digit Heal 3:109. https://doi.org/10.3389/fdgth.2021.725130

RDocumentation (2019) Regression models for ordinal data via Cumulative Link (Mixed) Models: ordinal package. In: RDocumentation. https://www.rdocumentation.org/packages/ordinal/versions/2019.12-10/topics/ordinal-package. Accessed 12 Apr 2021

Schinkel-Bielefeld N, Kunz P, Zutz A, Buder B (2020) Evaluation of Hearing Aids in

Everyday Life Using Ecological Momentary Assessment: What Situations Are We Missing? Am J Audiol 29:591–609. https://doi.org/10.1044/2020_AJA-19-00075

Schum DJ (2003) Noise-reduction circuitry in hearing aids. Hear J 56:32. https://doi.org/10.1097/01.HJ.0000294049.75608.1c

Shi LF, Doherty KA, Kordas TM, Pellegrino JT (2007) Short-term and long-term hearing aid benefit and user satisfaction: A comparison between two fitting protocols. J Am Acad Audiol 18:482–495. https://doi.org/10.3766/jaaa.18.6.3

Shiffman S, Stone AA, Hufford MR (2008) Ecological Momentary Assessment. Annu Rev Clin Psychol 4:1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Verma L, Sanju HK, Scaria B, et al (2017) A Comparative Study on Hearing Aid Benefits of Digital Hearing Aid Use (BTE) from Six Months to Two Years. Int Arch Otorhinolaryngol 21:224–231. https://doi.org/10.1055/s-0036-1592117

Walker JJ, Cleveland LM, Davis JL, Seales JS (2013) Audiometry Screening and Interpretation. Am Fam Physician 87:41–47

Walravens E, Keidser G, Hickson L (2020) Consistency of Hearing Aid Setting Preference in Simulated Real-World Environments: Implications for Trainable Hearing Aids. Trends Hear 24:. https://doi.org/10.1177/2331216520933392

Wendt D, Hietkamp RK, Lunner T (2017) Impact of Noise and Noise Reduction on Processing Effort. Ear Hear 38:690–700. https://doi.org/10.1097/aud.0000000000000454

Wong LLN, Chen Y, Wang Q, Kuehnel V (2018) Efficacy of a Hearing Aid Noise Reduction Function. Trends Hear 22:1–14. https://doi.org/10.1177/2331216518782839

Wright D, Gagné J-P (2021) Acclimatization to Hearing Aids by Older Adults. Ear Hear 42:193–205. https://doi.org/10.1097/AUD.0000000000000913

Yoon SH, Nam KW, Yook S, et al (2017) A trainable hearing Aid Algorithm reflecting individual preferences for degree of noise-suppression, input sound level, and listening situation. Clin Exp Otorhinolaryngol 10:56–65. https://doi.org/10.21053/ceo.2015.01690

Ziesemer K, Maria König L, Jo Boushey C, et al (2020) Occurrence of and reasons for "missing events" in mobile dietary assessments: Results from three event-based ecological momentary assessment studies. JMIR mHealth uHealth 8:. https://doi.org/10.2196/15430

Appendix C

Designing Audiologist Bots Fusing Soundscapes and User Feedback

---

Alessandro Pasta, Michael Kai Petersen, Kasper Juul Jensen, and Jakob Eg Larsen (2020). Designing Audiologist Bots Fusing Soundscapes and User Feedback. *CHI2020, Workshop on Conversational Agents for Health and Wellbeing.*

Designing audiologist bots fusing soundscapes and user feedback

Alessandro Pasta
Technical University of Denmark
Kongens Lyngby, Denmark
alpas@dtu.dk

Michael Kai Petersen
Eriksholm Research Centre
Snekkersten, Denmark
mkpe@eriksholm.com

Kasper Juul Jensen
Oticon A/S
Smørum, Denmark
kjen@oticon.com

Jakob Eg Larsen
Technical University of Denmark
Kongens Lyngby, Denmark
jaeg@dtu.dk

CHI 2020, Workshop on Conversational Agents for Health and Wellbeing, April 26, 2020, Honolulu, HI, USA

Abstract
Despite having different audiological preferences, hearing aid users are usually provided with default settings. This lack of personalization is due to a scarcity of audiological resources and to the difficulty of optimizing hearing aid settings in the clinic. Implementing a conversational agent allows to automatically gather user feedback in real-world environments, while monitoring the soundscape, in order to recommend personalized settings. We outline a conversational agent model that interprets user utterances as audiological intents and fuses user feedback and soundscape features to predict the most likely preferred hearing aid setting. Subsequently, we propose two use cases for a conversational agent, that envisage two different interactions to address distinct user needs: troubleshooting and contextual personalization.

Author Keywords
conversational agents; recommender systems; personalization; hearing healthcare; hearing aids

CCS Concepts
•**Human-centered computing** → **Human computer interaction (HCI); Natural language interfaces;** *User centered design; Ambient intelligence;*

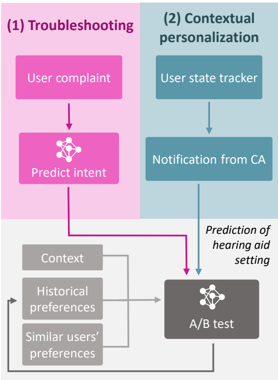


Figure 1: Overview of two different conversational agent use cases with different objectives: Troubleshooting (1) and Contextual personalization (2). If the objective is troubleshooting (1), the interaction is initiated by the user and the complaint is translated into an audiological intent. If the objective is contextual personalization (2), the CA tracks user's state and notifies the user when an alternative setting could enhance the experience. Based on the context, the historical preferences and, potentially, the audiological intent, the CA proposes an A/B setting pair. The user selects the preferred setting and the memory network is updated.

Introduction

Hearing aid users are currently prescribed amplification solely based on a hearing test, which measures hearing thresholds at different frequencies but does not capture individual differences in the ability to understand speech in noise [2] and in the loudness perception of sounds [13]. Despite having different audiological preferences [9], users are usually provided with default settings. Hearing care professionals assume that such default settings suffice [5] and subsequently, during follow-up visits, might modify the gain or noise reduction settings based on user recollections of past listening experiences [10]. The resulting lack of personalization is due to both the scarcity of audiological resources [14] and the difficulty of optimizing hearing aid settings based on user descriptions from recollection in the clinic, rather than dynamically adjusting the settings in the actual listening scenario. Furthermore, the same hearing aid user may have very contrasting preferences depending on the context [8].

Internet-connected hearing aids constitute a paradigm shift in hearing healthcare, as the devices can now potentially be complemented with a smartphone app capable of recommending truly personalized settings [11]. Previous research has shown that gathering user preferences in real-world environments makes it possible to define a device configuration which is preferred over a configuration personalized in a standard clinical workflow [15]. A conversational agent (CA) might autonomously gather user feedback in real-world environments and collect information about the soundscape, to learn the audiological preferences and eventually personalize the device settings. This requires a domain-specific mapping of user utterances to audiological intents [7], as well as the integration of contextual data into the dialogue management.

In this paper, we outline a CA model, which is trained by combining natural language understanding with sequential patterns of soundscape features, to predict the most likely preferred hearing aid setting. Subsequently, as shown in Figure 1, we propose two use cases for a CA:

1. **Troubleshooting**. By eliciting active user participation, the CA understands user complaints and fine-tunes the settings prescribed by the audiologist in the first visit.

2. **Contextual personalization**. By proactively asking the user to compare two alternative settings in specific real-world environments, the CA learns to contextually personalize the listening experience.

Model design

Recent research in CAs has progressed towards reaching human-level understanding within a general conversation framework [1]. Social CA evaluation metrics, reflecting whether responses are logically coherent and sufficiently specific, are equally relevant for the hearing healthcare domain. However, designing an audiologist CA entails domain-specific requirements for customized actions which go beyond dialogue management. Task-oriented dialogue systems often need to keep track of the context across multiple domains and store it as vectors in memory slots [12], while applying attention mechanisms to select the most relevant parameters in order to predict the next action [18].

Due to the requirement for our CA to interpret user utterances referring to audiological intents, we cannot rely on generic pretrained word embeddings. Instead, we build

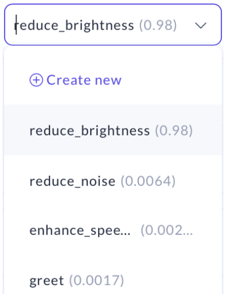


Figure 2: Interactive learning of a CA model. Prediction, based on highest probability, for mapping the word embeddings of the utterance "the sounds are too sharp" to a user intent.

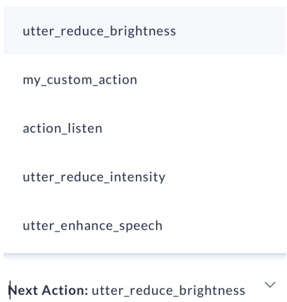


Figure 3: Interactive learning of a CA model. Prediction, based on vector similarity, for selecting the next action in response to the user utterance "the sounds are too sharp" and specific soundscape parameters.

from scratch a natural language understanding (NLU) module trained on audiological terms. In order to learn domain-specific word embeddings, we implement a neural network using frequently occurring user complaints as inputs and potential CA audiological actions as labels [6]. The model is trained by maximizing the similarity between user utterances and associated labels [20]. As shown by Figure 2, this allows to map utterances (e.g. "the sounds are too sharp") into audiological intents (e.g. "reduce high-frequency gain" or, in short, "reduce_brightness"). The most likely audiological intent is predicted based on the highest ranked probability (e.g. 0.98).

Subsequently, in order to predict the most likely next CA action, a dialogue management model is implemented in another neural network, using a transformer architecture similar to BERT [19]. Both the audiological intent embeddings learned by the NLU model, as well as feature vectors capturing the corresponding auditory environment, are forwarded to the second neural network. This dialogue management model compares the cosine similarities of the vectors to predict the most likely next CA action, based on previously learned contextual dialogue patterns. As shown in Figure 3, a user complaint (i.e. "the sounds are too sharp"), merged with the soundscape parameters stored in memory slots, results in a predicted next CA action (i.e. proposing the user an alternative setting defined by lower high-frequency amplification or, in short, "utter_reduce_brightness").

In practical terms, we implement both the NLU and the dialogue management models by training two TensorFlow Neural Networks using the Rasa open source framework [17]. As displayed in Figures 2 and 3, the interactive learning mode allows to accelerate the CA training, by providing feedback and fixing any mistakes.

Discussion

The CA model outlined above might be deployed in different ways, shaping different conversational experiences to address different user needs.

1. Troubleshooting

After the user has been prescribed a default amplification based on the hearing test, she usually tries the hearing aids for some weeks. Given that the prescriptive formula does not guarantee user satisfaction, some fine-tuning based on users' subjective reactions is usually needed [10]. A successful fine-tuning requires the user to be able to communicate her experiences and the professional to be able to interpret and translate them into an adjustment of hearing aid settings [4]. However, this is a time-consuming, yet not optimal procedure, since fine-tuning and additional hearing tests (e.g. speech-in-noise test) performed in the clinic environment do not guarantee a significant advantage over a default prescription [3, 16].

A CA might address users' issues and fine-tune newly acquired hearing aids during the trial phase. Since, in this scenario, the objective is to solve a problem experienced by the user, the interaction is initiated by the user herself. The CA maps the complaint expressed by the user into an audiological intent. In parallel, data describing the context and the current hearing aid settings are inserted into the dialogue. These contextual features, together with the user input, enable the CA to generate a setting adjustment potentially capable of solving the problem experienced by the user. Furthermore, the dialogue allows to gather immediate user feedback on the proposed setting adjustment.

As exemplified in Figure 4, a user might perceive that the sounds are too sharp and initiate a dialogue to report the complaint to the CA. The CA would translate the complaint into an audiological intent and would respond by decreas-

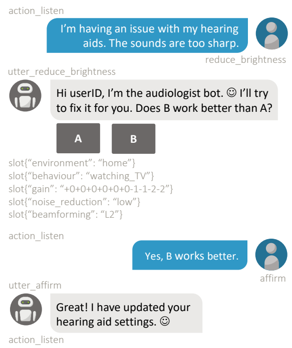


Figure 4: Troubleshooting dialogue initiated by a user complaint. The CA translates the complaint into an audiological intent, responds by decreasing high-frequency amplification and asks the user for her feedback on the new setting.

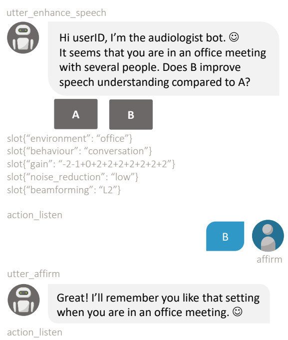


Figure 5: Contextual personalization initiated by the CA, based on tracked user state. The CA monitors the context, asks the user to compare two settings and learns from user feedback.

ing high-frequency amplification. It would ask the user for her feedback on the new setting and remember it, gradually learning the preferred fine-tuning actions in response to users' complaints. This solution enables solving audiological issues as soon as they arise, by gathering user feedback in real-world environments. It, thereby, potentially provides an effective tool for troubleshooting and reduces the clinical workload.

2. Contextual Personalization

After the hearing aid user has completed the fine-tuning process, there is still potential for improving her listening experience, as user preferences vary based on the context [8]. Although users potentially benefit from a contextually personalized hearing aid setting, collecting user preferences in real-world environments is a time-consuming process.

A CA might autonomously gather user feedback in real-world environments and learn to contextually recommend personalized hearing aid settings based on the historical choices of the user and on the preferences of other users in similar environments. In this scenario, the interaction is initiated by the CA, that tracks user's state, by monitoring the environment and the current device setting, to suggest an alternative setting. Since the user might benefit from a more effective setting, while not having an explicit complaint in mind, the interaction is simplified.

As shown in Figure 5, the CA might ask the user to compare two alternative settings: the current setting and one that, according to the algorithm, is the most likely to improve user satisfaction in that sound environment. The user would compare the two settings and choose the preferred one. This solution would allow to get an immediate user feedback on alternative settings tried in real-world environments. By iteratively gathering user feedback and learning user preferences in different situations, the CA could, eventually, autonomously personalize hearing aid settings based on the context.

In order for this solution to be successfully implemented, a challenge posed by possible user fatigue needs to be addressed. A hearing aid user might be frequently moving between different contexts and might feel overwhelmed if the CA continuously prompts her to test different settings. In order to avoid an excessive number of notifications, the CA would have to learn when hearing aid personalization is particularly relevant and only prompt the user in those situations. The relevance of personalization is potentially determined by several factors, such as the distance between the current setting and the alternative one, and the perceived usefulness of adjusting an audiological parameter in that situation.

Acknowledgements

We would like to thank Oticon A/S and Eriksholm Research Centre.

REFERENCES

[1] Daniel Adiwardana, Minh-Thang Luong, David R. So, Jamie Hall, Noah Fiedel, Romal Thoppilan, Zi Yang, Apoorv Kulshreshtha, Gaurav Nemade, Yifeng Lu, and Quoc V. Le. 2020. Towards a Human-like Open-Domain Chatbot. (2020).

[2] Mead C. Killion. 2002. New Thinking on Hearing in Noise: A Generalized Articulation Index. *Seminars in Hearing* 23 (January 2002), 057–076. DOI: `http://dx.doi.org/10.1055/s-2002-24976`

[3] David R. Cunningham, Kristie J. Williams, and L. Jane Goldsmith. 2001. Effects of Providing and Withholding Postfitting Fine-Tuning Adjustments on Outcome Measures in Novice Hearing Aid Users. *American*

Journal of Audiology 10, 1 (2001), 13–23. DOI: http://dx.doi.org/10.1044/1059-0889(2001/001)

[4] C. Elberling and K.V. Hansen. 1999. *Auditory Models and Non-Linear Hearing Instruments (Rasmussen, Arne Nørby and Osterhammel, Poul Aabo and Ture Andersen and Torben Poulsen)*. Holmens Trykkeri, Copenhagen, Denmark, Chapter Hearing Instruments: Interaction with User Preference, 341–357.

[5] David B. Hawkins and Jodi A Cook. 2003. Hearing aid software predictive gain values: How accurate are they? *The Hearing Journal* 56 (2003), 26,28,32,34.

[6] Lorienne Jenstad, Dianne Tasell, and Chiquita Ewert. 2003. Hearing aid troubleshooting based on patients' descriptions. *Journal of the American Academy of Audiology* 14 (10 2003), 347–60.

[7] Benjamin Johansen, Michael K. Petersen, Niels H. Pontoppidan, and Jakob E. Larsen. 2019. Modelling User Utterances as Intents in an Audiological Design Space. In *Workshop on Computational Modeling in Human-Computer Interaction (CHI 2019)*.

[8] Benjamin Johansen, Michael K. Petersen, Maciej J. Korzepa, Jan Larsen, Niels H. Pontoppidan, and Jakob E. Larsen. 2018. Personalizing the Fitting of Hearing Aids by Learning Contextual Preferences From Internet of Things Data. *Computers* 7, 1 (2018). DOI:http://dx.doi.org/10.3390/computers7010001

[9] Gitte Keidser and Harvey Dillon. 2006. What's New in Prescriptive Fitting Down Under?. In *Hearing Care for Adults*. 133–142.

[10] S. Kochkin, D.L. Beck, L.A. Christensen, C. Compton-Conley, P.B. Kricos, B.J. Fligor, J.B. McSpaden, H.G. Mueller, M.J. Nilsson, J.L. Northern, T.A. Powers, R.W. Sweetow, B. Taylor, and R.G. Turner. 2010. MarkeTrak VIII: The impact of the hearing healthcare professional on hearing aid user success. *Hearing Review* 17 (4 2010), 12–34.

[11] Maciej J. Korzepa, Benjamin Johansen, Michael K. Petersen, Jan Larsen, Jakob Eg Larsen, and Niels H. Pontoppidan. 2018. Modeling User Intents as Context in Smartphone-Connected Hearing Aids. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*. 151–155. DOI: http://dx.doi.org/10.1145/3213586.3226211

[12] Adarsh Kumar, Peter Ku, Anuj Goyal, Angeliki Metallinou, and Dilek Hakkani-Tur. 2019. MA-DST: Multi-Attention-Based Scalable Dialog State Tracking. In *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*.

[13] Dirk Oetting, Volker Hohmann, Jens Appell, Birger Kollmeier, and Stephan Ewert. 2017. Restoring Perceived Loudness for Listeners With Hearing Loss. *Ear and Hearing* 39 (11 2017), 1. DOI: http://dx.doi.org/10.1097/AUD.0000000000000521

[14] World Health Organization. 2013. *Multi-Country Assessment of National Capacity to Provide Hearing Care*. Geneva, Switzerland.

[15] Alessandro Pasta, Michael K. Petersen, Kasper J. Jensen, and Jakob Eg Larsen. 2019. Rethinking Hearing Aids as Recommender Systems. In *Proceedings of the 4th International Workshop on Health Recommender Systems co-located with 13th ACM Conference on Recommender Systems (HealthRecSys'19)*. 11–17. http://ceur-ws.org/Vol-2439/4-paginated.pdf

[16] Lu-Feng Shi, Karen A. Doherty, Tammy M. Kordas, and Joseph Pellegrino. 2007. Short-term and

long-term hearing aid benefit and user satisfaction: a comparison between two fitting protocols. *Journal of the American Academy of Audiology* 18 6 (2007), 482–95.

[17] Rasa Technologies. 2020. Build contextual chatbots and AI assistants with Rasa. (2020). Retrieved April 8, 2020 from `https://rasa.com/docs/rasa/`

[18] Vladimir Vlasov, Akela Drissner-Schmid, and Alan Nichol. 2018. Few-Shot Generalization Across Dialogue Tasks. *CoRR* abs/1811.11707 (2018). `http://arxiv.org/abs/1811.11707`

[19] Vladimir Vlasov, Johannes E. M. Mosig, and Alan Nichol. 2019. Dialogue Transformers. (2019).

[20] Ledell Wu, Adam Fisch, Sumit Chopra, Keith Adams, Antoine Bordes, and Jason Weston. 2017. StarSpace: Embed All The Things! *CoRR* abs/1709.03856 (2017). `http://arxiv.org/abs/1709.03856`

Appendix D

Clustering Users Based on Hearing Aid Use: An Exploratory Analysis of Real-World Data

Alessandro Pasta, Tiberiu-Ioan Szatmari, Jeppe Høy Christensen, Kasper Juul Jensen, Niels Henrik Pontoppidan, Kang Sun, and Jakob Eg Larsen (2021). Clustering Users Based on Hearing Aid Use: An Exploratory Analysis of Real-World Data. *Frontiers in digital health*, 3, 725130. https://doi.org/10.3389/fdgth.2021.725130 frontiers in Digital Health

ORIGINAL RESEARCH
published: 03 September 2021
doi: 10.3389/fdgth.2021.725130

Clustering Users Based on Hearing Aid Use: An Exploratory Analysis of Real-World Data

Alessandro Pasta[1,2]*, *Tiberiu-Ioan Szatmari*[1,2], *Jeppe Høy Christensen*[3], *Kasper Juul Jensen*[2], *Niels Henrik Pontoppidan*[3], *Kang Sun*[1,3] *and Jakob Eg Larsen*[1]

[1] Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kongens Lyngby, Denmark, [2] Demant A/S, Smørum, Denmark, [3] Eriksholm Research Centre, Oticon A/S, Snekkersten, Denmark

OPEN ACCESS

Edited by:
Jing Chen,
Peking University, China

Reviewed by:
Pasin Israsena,
National Electronics and Computer Technology Center, Thailand
Parisis Gallos,
National and Kapodistrian University of Athens, Greece

***Correspondence:**
Alessandro Pasta
alpas@dtu.dk

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

Received: *14 June 2021*
Accepted: *26 July 2021*
Published: *03 September 2021*

Citation:
Pasta A, Szatmari T-I, Christensen JH, Jensen KJ, Pontoppidan NH, Sun K and Larsen JE (2021) Clustering Users Based on Hearing Aid Use: An Exploratory Analysis of Real-World Data. Front. Digit. Health 3:725130. doi: 10.3389/fdgth.2021.725130

While the assessment of hearing aid use has traditionally relied on subjective self-reported measures, smartphone-connected hearing aids enable objective data logging from a large number of users. Objective data logging allows to overcome the inaccuracy of self-reported measures. Moreover, data logging enables assessing hearing aid use with a greater temporal resolution and longitudinally, making it possible to investigate hourly patterns of use and to account for the day-to-day variability. This study aims to explore patterns of hearing aid use throughout the day and assess whether clusters of users with similar use patterns can be identified. We did so by analyzing objective hearing aid use data logged from 15,905 real-world users over a 4-month period. Firstly, we investigated the daily amount of hearing aid use and its within-user and between-user variability. We found that users, on average, used the hearing aids for 10.01 h/day, exhibiting a substantial between-user (SD = 2.76 h) and within-user (SD = 3.88 h) variability. Secondly, we examined hearing aid use hourly patterns by clustering 453,612 logged days into typical days of hearing aid use. We identified three typical days of hearing aid use: full day (44% of days), afternoon (27%), and sporadic evening (26%) day of hearing aid use. Thirdly, we explored the usage patterns of the hearing aid users by clustering the users based on the proportion of time spent in each of the typical days of hearing aid use. We found three distinct user groups, each characterized by a predominant (i.e., experienced ∼60% of the time) typical day of hearing aid use. Notably, the largest user group (49%) of users predominantly had full days of hearing aid use. Finally, we validated the user clustering by training a supervised classification ensemble to predict the cluster to which each user belonged. The high accuracy achieved by the supervised classifier ensemble (∼86%) indicated valid user clustering and showed that such a classifier can be successfully used to group new hearing aid users in the future. This study provides a deeper insight into the adoption of hearing care treatments and paves the way for more personalized solutions.

Keywords: data logging, user clustering, ensemble classification, hearing aid use amount, hearing aid use patterns, hearing aids, personalization

INTRODUCTION

It is estimated that, globally, 430 million people have disabling hearing loss, i.e., a hearing loss greater than 35 decibels (dB) in the better hearing ear (1). By 2050 over 700 million people are expected to have disabling hearing loss (1). Untreated hearing loss has repercussions at an individual level. It is associated with poorer cognitive and psychological status, resulting in increased risk of depression, dementia, falls, and quality of life (2–4). Hearing loss negatively impacts education, employment, and household income (1, 5). Additionally, untreated hearing loss has a negative impact on society and the economy. Older adults with untreated hearing loss experience higher health care costs and utilization patterns compared with adults without hearing loss (4). The World Health Organization (1) estimates that untreated hearing loss poses an annual global cost of US$ 980 billion, including health sector costs, costs of educational support, loss of productivity, and societal costs.

The adoption of hearing aids (HAs) has been shown to have a positive impact on the quality of life of users (6, 7) and to mitigate the effect on their household income (5). The success of HA provision as a treatment for hearing loss depends on the fact that the patient is provided with a favorable change in their condition, but also on the patient compliance with the intervention program (8). Perez and Edmonds (8) conducted a systematic review to identify and evaluate how studies have measured and reported the use of HAs in older adults. A limited number of studies (5 out of 64) were found to assess HA use based on objective measures, such as data logging and battery consumption. Most of the studies assessed HA use through self-reported measures, such as standardized questionnaires, custom questionnaires, interviews, and diaries. However, self-reported measures have been shown to diverge from objective measures, leading to inaccurate and overreported HA use (9–12). In addition to avoiding such recall bias, objective data logging enables measuring HA use with a greater temporal resolution and longitudinally (13). The widespread adoption of smartphones among older adults (14) and the introduction of smartphone-connected hearing aids make it possible to objectively assess the HA use of a larger number of users than ever before (15).

When evaluating HA usage, the amount of HA use time is commonly regarded as an indicator of treatment success (16) and frequently investigated (9, 12, 17–19). Although the amount of HA use time generally correlates with HA satisfaction (20), this metric might not provide a complete picture. Indeed, frequent use does not necessarily equate with benefit (21). A previous study found that some HA users reported low HA use time and high HA satisfaction, while other users reported high HA use time and low HA satisfaction (22). Furthermore, HA use time provides information about how much the HA has been used during the day, but it is not informative of when and how the HA has been used. For instance, two users might exhibit the same amount of use time (e.g., 8 h), but use the HAs at different times of the day (e.g., from 8:00 to 16:00 and from 15:00 to 23:00) or in different ways (e.g., on-off usage or continuous usage). For these reasons, in addition to the amount of HA use time, other patterns of HA use should be analyzed (11). However, methods possessing low temporal resolution (e.g., self-reports or accumulated use time across a day or a week) do not account for the hourly and daily variability in HA use. Smartphone-connected HAs enable continuous data logging, thereby making it possible to assess the hourly HA use and more accurately identify recurrent use patterns.

Additionally, the HA industry is currently predominantly accommodating for the average user (23). However, the amount of HA use varies widely among HA users (9, 19, 24). Similarly, the pattern of HA use has been reported to vary among HA users. Laplante-Lévesque et al. (11) clustered 171 HA owners, showing that 57% had, on average, a continuous HA use during the day, while 43% had an on-off HA use. A qualitative study (16) reported that optimal HA use depends on the individual needs of the HA owners and does not necessarily correspond to wearing the HAs most of the time. Some HA users reported that they do not depend on their HAs and that they experience situations which they can successfully attend without HAs. Therefore, it is of interest to objectively measure and investigate the HA use of a large number of HA owners, in order to identify and quantify different types of users based on their HA use patterns. This potentially enables gaining deeper insight into the adoption of hearing care treatments and paves the way for more personalized solutions (25).

Finally, when comparing users based on their HA use, the average individual use is usually considered. This means that the within-user variability in HA use is often disregarded (11, 19, 24). However, HA users might exhibit different HA use patterns from one day to another and two HA users with the same average use might behave differently. For instance, two users might exhibit the same average amount of use time throughout the logged days (e.g., 8 h), but one might use the HA constantly (e.g., 8 h each day) while the other might exhibit more variation among the days (e.g., alternating days with 2 and 14 h of use). Therefore, when comparing users based on HA use, it is desirable to adopt a metric that goes beyond the average use per user and that considers the within-user variability.

In this study, we analyze the objective HA use data logged from 15,905 real-world users over a 4-month period. Firstly, we investigate the daily amount of HA use and its within-user and between-user variability. Secondly, we examine HA use hourly patterns by clustering the 453,612 logged days to identify typical days of HA use. Thirdly, we explore the usage patterns of the HA users and investigate whether we can cluster the users based on how they used the HAs during the logged days. When performing the user clustering, instead of representing each user by her average HA use pattern, we consider the proportion of time spent in each of the typical days of HA use. Finally, we validate the HA user clustering by training a supervised classifier to predict the cluster to which each user belongs.

MATERIALS AND METHODS

Participants and Apparatus

This study used data from a large-scale internal database, which logs the HA use of HA owners who have signed up for the HearingFitness™ feature (25) *via* the Oticon ON™ smartphone app. The participants were the users of Oticon Opn™ hearing aids who used the HearingFitness™ feature for at least 10 days in the period between June and September 2020.

Data and Data Pre-processing

When the HAs are connected to the smartphone, the HearingFitness™ feature logs timestamped data about the HA use every 10 min. Based on HA use time (i.e., inferred from time counters embedded in the HAs) and connection status, an estimate of hourly HA use (measured in min/h) is computed. For binaural HA users, if the HA use amount was different between the right and left ear, this study selected the larger value, as done by Laplante-Lévesque et al. (11) and Walker et al. (27). If temporary disconnections occur, replacements for the missing data are injected by analyzing the time counters embedded in the HAs. When the disconnected use is full-time use (e.g., 120 min of use during 2 h of disconnection), the HA use during disconnection is simply assigned to the hours of disconnection. When the disconnected use is on-off use (i.e., not full-time use), the minutes of use are evenly distributed among the hours of disconnection (e.g., 60 min of use during 2 h of disconnection result in 30 min/h use for 2 h). The raw data set comprised 1,160,520 days of HA use from 32,216 users. In order to preserve representative patterns of HA use throughout the day, days with on-off use during temporary disconnections longer than 2 h were removed. Additionally, 12,876 days with more than 60 min/h were removed. This is likely a consequence of use time estimation when disconnections occur. Moreover, since this study focuses on analyzing HA use, only days with at least 60 min of HA use were included. Furthermore, only data related to HA use between 6:00 and 23:59 were included. Finally, to ensure that users' behavior was inferred from a representative sample of days, only users with at least 10 days of HA use were included. The cleaned data set comprised 453,612 days of HA use from 15,905 users (28.5 days per user on average).

Data Analysis

Figure 1 provides an overview of the flow of the data analysis we performed, presenting the main steps undertaken. More details on each step are provided below.

Exploring the Amount of Hearing Aid Use

We explored the amount of HA use (measured in hours/day), by computing summary statistics of the 453,612 logged days (mean, SD) and of the amount of HA use for each user (mean, between-user SD, quartiles). Furthermore, we analyzed the within-user daily variability (SD) in HA use amount. Independent sample t-tests were performed to compare the within-user SD of medium users (i.e., users with average HA use amount between Q_1 and Q_3) with that of light and heavy users (i.e., users with average HA use amount, respectively, below Q_1 and above Q_3). Cohen's d was computed to assess the magnitude of the differences (45). A polynomial linear regression was fitted to model the relationship between average amount of HA use per user (x) and within-user SD (y).

Clustering Days of Hearing Aid Use

We examined patterns of HA use by clustering the 453,612 logged days into typical days of HA use. The input data consisted of a $453,612 \times 18$ matrix

$$\mathbf{A}_{\mathbf{r}\times\mathbf{c}} = \mathbf{A}_{453612\times 18} = \begin{bmatrix} a_{11} \cdots a_{1c} \\ \vdots \ \ \ddots \ \ \vdots \\ a_{r1} \cdots a_{rc} \end{bmatrix} = \left(a_{ij}\right) \in [0, 60] \quad i = 1, \ldots, r; j = 1, \ldots, c$$

where each row i represents a day of HA use, each column j represents an hour of the day (from 6 to 23) and a_{ij} is the amount of HA use (from 0 to 60 min) in the day i and hour j. The k-means clustering technique was applied (28), since it is suitable for large data sets. K-means aims to partition the observations in k clusters by minimizing the within-cluster variance (i.e., square Euclidean distances). The k-means++ initialization algorithm (29) was applied, which seeks to spread out the k initial clusters to avoid poor approximation. The optimal value of k was determined using the elbow method (30), which aims to select a number of clusters so that adding another cluster does not substantially increase the explained variation. The resulting clusters were evaluated by conducting a Silhouette analysis (31), which aims to evaluate the between-clusters dispersion (i.e., separation) and the within-cluster dispersion (i.e., cohesion). A Silhouette Coefficient (ranging from -1 to $+1$) was calculated for each observation and constitutes a measure of how similar an observation is to its own cluster compared to the next nearest cluster. Furthermore, principal component analysis was performed to visualize the observations in a lower dimensional space. Subsequently, for each cluster, we identified and removed observations that were abnormally distant from the other observations (i.e., below $Q_1 - 1.5 \cdot IQR$ and above $Q_3 + 1.5 \cdot IQR$). This was done in order not to include days of HA use that exhibited atypical patterns and were not well-represented by the cluster centroids. The association between the type of day of HA use and the day of the week was tested by performing a χ^2 test of independence and computing Cramer's V. The clustering and related analyses were performed in Python, using the scikit-learn library (32).

Clustering Users

We explored the behavior of HA users by clustering the 15,905 users based on the proportion of time spent in each of the typical days of HA use. The input data consisted of a $15,905 \times c$ matrix

$$\mathbf{B}_{r\times c} = \mathbf{B}_{15905\times c} = \begin{bmatrix} b_{11} \cdots b_{1c} \\ \vdots \ \ \ddots \ \ \vdots \\ b_{r1} \cdots b_{rc} \end{bmatrix} = \left(b_{ij}\right) \in [0, 1] \quad i = 1, \ldots, r; j = 1, \ldots, c$$

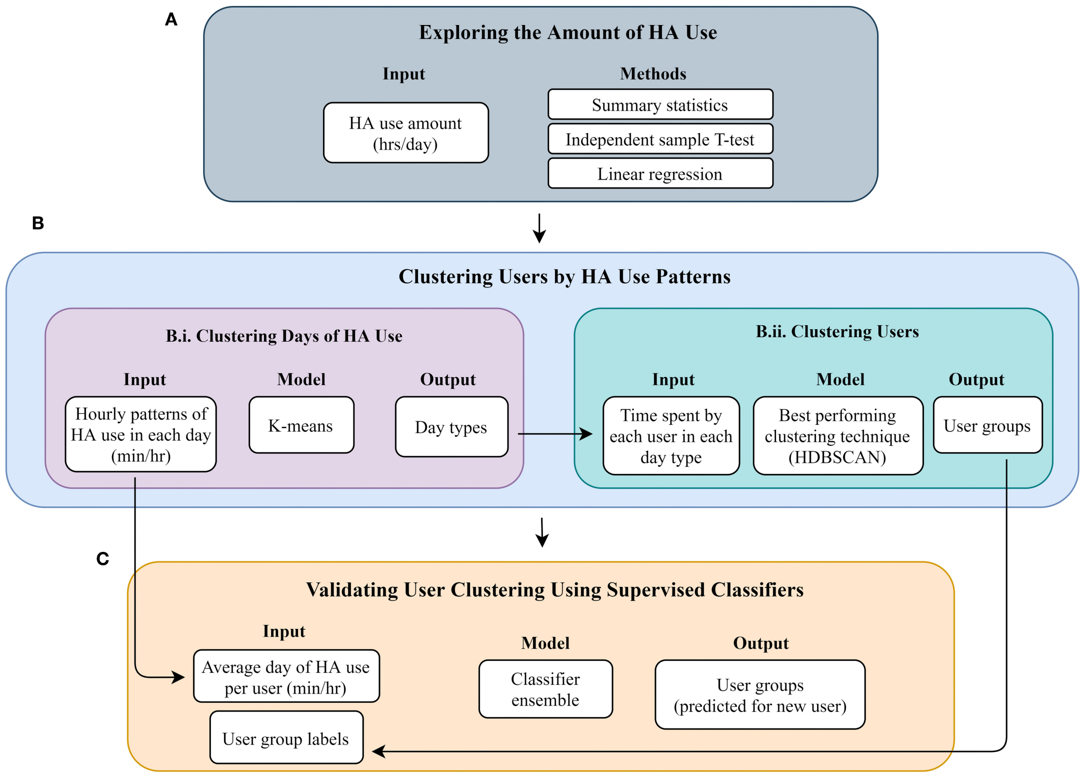


FIGURE 1 | Flow of data analysis. **(A)** Exploring the amount of HA use. **(B)** Clustering users by HA use patterns. **(C)** Validating user clustering using supervised classifiers.

where each row i represents a HA user, each column j represents one of the c typical days of HA use (referring to the clusters found *via* section Clustering Days of Hearing Aid Use) and b_{ij} is the proportion of days belonging to day type j for user i. Different clustering techniques were evaluated: k-means with k-means++ initialization algorithm, Hierarchical Agglomerative Clustering (HAC) with Ward's method, HAC with Pearson correlation and average linkage method, and Hierarchical Density-Based Spatial Clustering (HDBSCAN). HAC (33) initially treats each observation as a cluster and then builds nested clusters by successively merging pairs of the most similar clusters. HDBSCAN (34) groups observations that are in a dense region while marking the observations in sparse regions as noise. It expands on a different density-based technique, DBSCAN (35), by converting it into a hierarchical clustering technique, followed by extracting a flat clustering based on cluster stability. For k-means, the optimal value of clusters was determined using the elbow method (30). For HAC, the optimal value of clusters was determined using the dendrogram. HDBSCAN, instead, infers the optimal number of clusters based on the data. For each clustering technique, three internal validation metrics were computed: Silhouette score (31), Caliński-Harabasz score (36), and Davies-Bouldin score (37). The Caliński-Harabasz score is defined as a ratio of separation and cohesion. The Davies-Bouldin score measures the average similarity between each cluster and its most similar one, by comparing the distance between clusters with the size of the clusters themselves. Based on the three metrics, the best performing clustering technique was selected. The clustering was performed in Python, using the scikit-learn (32) and hdbscan (38) libraries.

Validating User Clustering Using Supervised Classifiers

We validated the HA user clustering by training an ensemble of supervised classifiers to predict the cluster label for individual users based on the average day of HA use for each user. The input data for classification consisted of a $15,905 \times 18$ matrix:

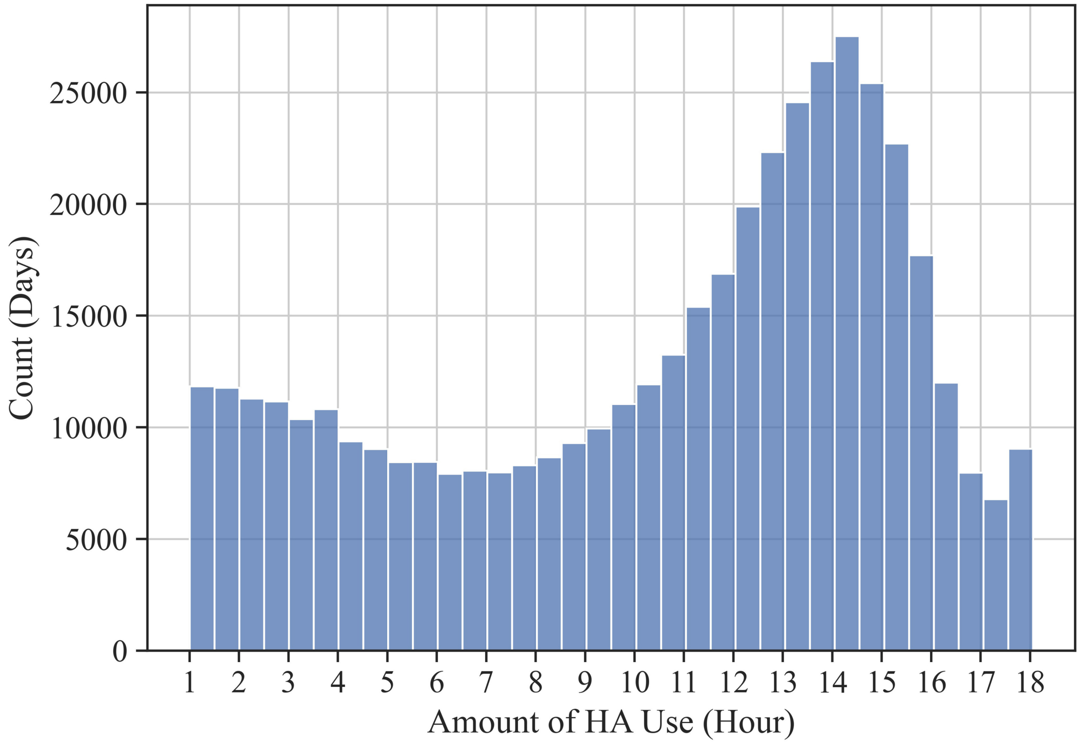


FIGURE 2 | Count of logged days by the amount of HA use time. Due to the data cleaning criteria (i.e., only days with at least 1 h of HA use were included; only data related to HA use in the 18 h between 6:00 and 23:59 were included), the amount of HA use (x-axis) ranges from 1 to 18 h.

$$\mathbf{D_{r\times c}} = \mathbf{D}_{15905\times 18} = \begin{bmatrix} d_{11} \cdots d_{1c} \\ \vdots \quad \ddots \quad \vdots \\ d_{r1} \cdots d_{rc} \end{bmatrix}$$
$$= \left(d_{ij}\right) \in [0, 60] \quad i = 1, \ldots, r; j = 1, \ldots, c$$

where each row i represents the average day of a HA user, each column j represents the hour of the day (from 6 to 23) and d_{ij} is the average amount of HA use (from 0 to 60 min) for user i in the hour j. This data was further split into separate training and testing data sets with an 80/20 split. To reduce bias (39), three classifiers were chosen from different families: multiclass logistic regression (regression), an XGBoost classifier (decision trees) (40) and a fully connected (FC) neural network classifier (41). The following individual parameters were chosen:

- Multiclass logistic regression: L2 penalty and "newton-cg" solver.
- XGBoost: estimators = 100, max depth = 5, gamma = 0, alpha = 0.1.
- FC neural network: four-layer network (128-64-32-4), ReLU activation, cross-entropy loss, Adam optimizer; trained for 25 epochs.

In order to reduce bias (39), a classification ensemble was defined, which assigns each user to a group by majority voting between the three classifiers. In cases where no majority could be defined, the group was decided by the best performing individual classifier. Two metrics were used to gauge each model's performance: accuracy, and Area Under the Receiver Operating Characteristic (ROC-AUC). Accuracy is obtained by calculating the ratio of correct test predictions to the total amount of samples in the testing set. ROC-AUC helps visualize the relationship between sensitivity (i.e., True Positive Rate) and specificity (i.e., False Positive Rate) for a binary classification problem. The ROC-AUC value ranges from 0 to 1 and represents the ability of a classifier to distinguish between classes at various thresholds. If the current classification task operates with more than two classes (i.e., multiclass classification), the individual classes are first binarized. The score of the individual classes is calculated, then a micro-average is computed by aggregating the contributions of all classes to compute the average metric. Finally, a macro-average is calculated by computing the metric independently for each class and then calculating the average. The training and evaluation of the supervised classification ensemble was performed in Python, using scikit-learn (32),

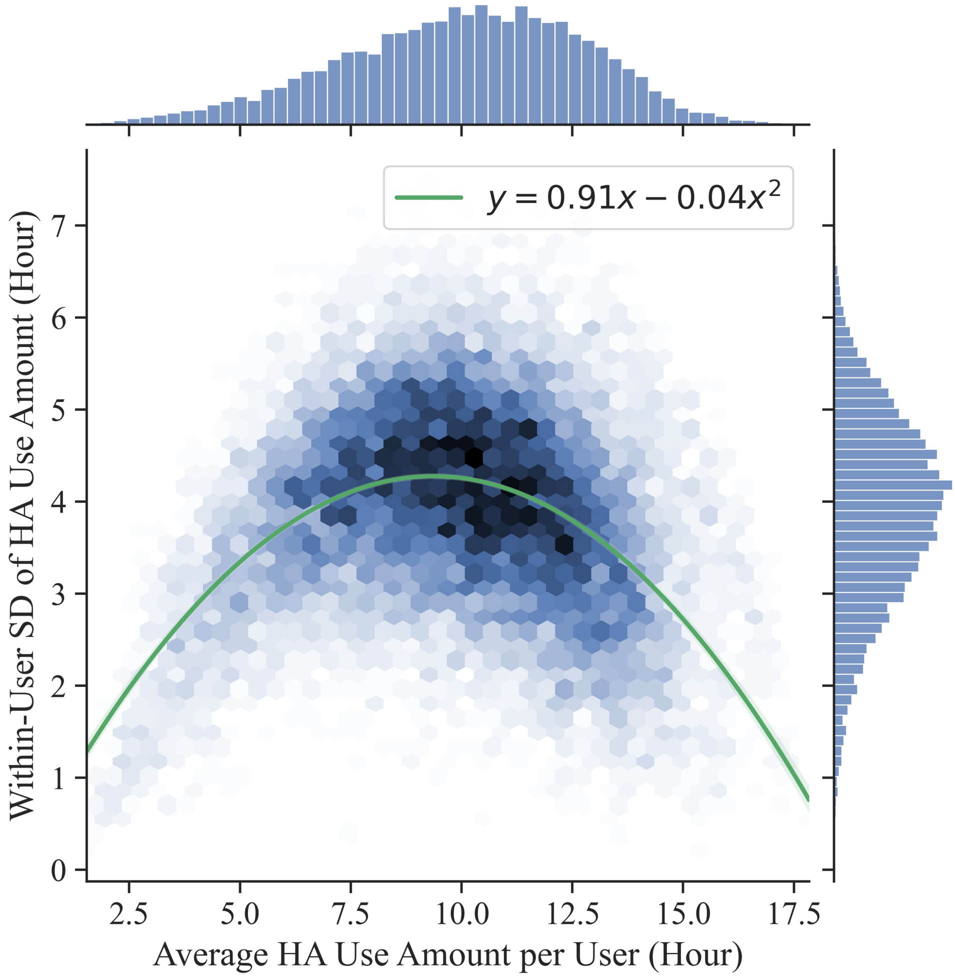


FIGURE 3 | Distribution of users by their average amount of HA use and their variability among the logged days (i.e., within-user SD). A second order linear regression model (line ± 99% confidence interval) was fitted to the data to model the relationship between average HA use (x) and within-user SD (y).

XGBoost (40), PyTorch (42), Yellowbrick (43), and scikit-plot libraries (44).

RESULTS

Exploring the Amount of Hearing Aid Use

The clean data set comprised 453,612 days of HA use from 15,905 users. The amount of HA use, defined as hours of HA use per day, was assessed to describe usage. **Figure 2** shows the frequency distribution of HA use amount during the pooled logged days. The data represents the HA use in days of connected use. On average, a day of HA use amounted to 10.55 h. However, the days were not normally distributed around the mean. The amount of HA use widely varied throughout the logged days (SD = 4.71 h), with a mode around 14 h of use and a smaller peak around 1 h of use.

On average, 28 days (SD = 18 days) were logged for each user. We investigated the extent to which users used the HAs differently among each other (i.e., between-user variability) and the extent to which the same user used the HAs consistently throughout the logged days (i.e., within-user variability). For each user, the average amount of HA use and the within-user standard deviation (SD) among the days of HA use were computed. **Figure 3** shows the distribution of the 15,905 HA users. We firstly investigated the between-user variability in the amount of HA use (*x*-axis in **Figure 3**). Users had an average amount of HA use of 10.01 h, with a SD of 2.76 h (Coefficient of variation = 0.276). The middle 50% of users (*medium users,*

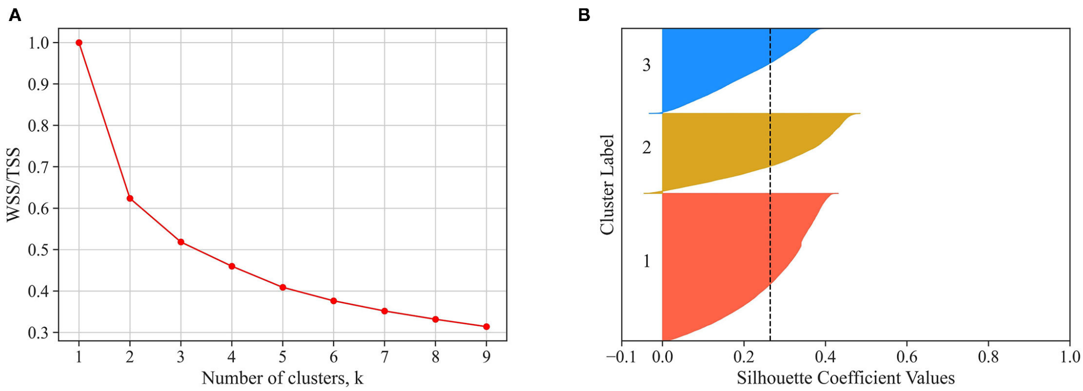


FIGURE 4 | In **(A)**, the ratio of within sum of squares (WSS) to the total sum of squares (TSS) is displayed as a function of the number of clusters. The elbow plot suggests selecting three clusters since adding an additional cluster does not substantially increase the explained variation. In **(B)**, the Silhouette coefficient value of each observation is displayed for the three-cluster solution (i.e., for each cluster, the observations are ordered by their Silhouette coefficient value and displayed in ascending order as horizontal stacked lines). The average silhouette score is reported (dashed line). The three clusters have predominantly positive scores, suggesting valid clustering.

between the first and third quartiles) ranged from 8.18 to 12.04 h (group mean = 10.16) of average HA use. The fact that the remaining 50% of the users exhibited an average HA use either below 8.18 (*light users*; group mean = 6.32) or above 12.04 (*heavy users*; group mean = 13.37) h indicates a substantial between-user variability.

Additionally, we investigated the within-user variability in the amount of HA use (y-axis in **Figure 3**). The average within-user SD was 3.88 h, indicating that the same user tended to use the hearing aids for varying durations throughout the logged days. A significantly larger within-user SD was observed for the medium users compared to both the light users (Two-sample t-test: $t = 23.06$, $p < 0.001$; Effect size: $d = 0.44$) and the heavy users (Two-sample t-test: $t = 41.85$, $p < 0.001$; Effect size: $d = 0.81$). This proves that both light users and heavy users were more consistent than medium users throughout the logged days (i.e., lower within-user SD) and constitutes an indication of users consistently displaying diverse behaviors in terms of HA use. The relationship between average HA use (x) and within-user SD (y) was modeled by fitting a second order linear regression model to the data. The line of best fit ($R^2 = 0.2$) was described by the equation $y = 0.91x - 0.04x^2$. The maximum of the curve is around 10 h of HA use, indicating that the within-user SD increases with the amount of HA use for users using the HAs up to 10 h and it decreases for users using the HAs more than 10 h.

Clustering Days of Hearing Aid Use

The substantial within-user variability in HA use suggests that a deeper analysis is warranted, which accounts for the hourly and daily variability in HA use. In addition to the amount of HA use, we also assessed patterns of HA use, defined as minutes of HA use per hour throughout the day. That was done by clustering the 453,612 logged days into typical days of HA use (see subsection Data Analysis). Based on the elbow method (**Figure 4A**), a three-cluster solution was chosen, which accounts for almost 50% of the variance among days. The Silhouette analysis (**Figure 4B**) indicated that the three clusters have predominantly positive scores, and there are no clusters with below-average silhouette scores.

Figure 5 displays the 453,612 days of HA use plotted by the two main principal components and colored by the three clusters. The eigenvectors suggest that the first principal component is negatively correlated with HA use in all hours of the day, differentiating between days of heavy use (**Figure 5A**, left) and days of light use (**Figure 5A**, right). The second principal component, instead, is positively correlated with HA use in the morning hours and negatively correlated with HA use in afternoon and evening hours, differentiating between days of morning HA use (**Figure 5A**, top) and days of HA use later in the day (**Figure 5B**, bottom). For each cluster, outliers were removed, resulting in 440,052 observations belonging to the three clusters. Looking at the hourly mean of HA use for each cluster (**Figure 5B**), it is possible to qualitatively evaluate the patterns underlying the clusters. Three distinct types of days of HA use can be identified: a full day of HA use (cluster 1, containing 204,062 days), a day of afternoon HA use (cluster 2, containing 120,810 days), and a day of sporadic evening HA use (cluster 3, containing 115,180 days).

A significant ($p < 0.001$), but negligible (Cramer's V = 0.05) association was found between the type of day of HA use and the day of the week (i.e., weekend vs. weekday). Full days of HA use occurred slightly (6%) more often during weekdays than during the weekend. Conversely, afternoon and sporadic evening days

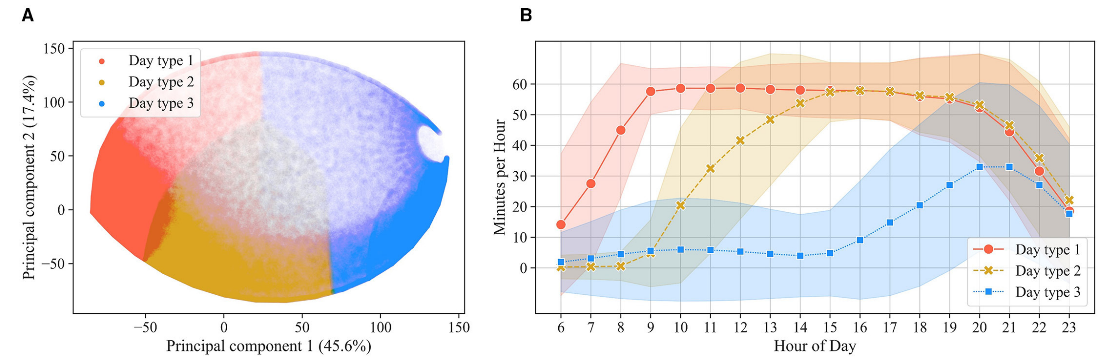


FIGURE 5 | In **(A)**, the days of HA use are displayed as scatterplot against the two main principal components and colored by the three clusters (i.e., three day types). In **(B)**, the mean (±SD) of hourly HA use for each cluster is displayed.

TABLE 1 | Comparison of four different clustering techniques [K-means, HAC (Euclidean distance), HAC (Pearson correlation), and HDBSCAN] based on three internal validation metrics (Silhouette, Davies-Bouldin, and Caliński-Harabasz).

	K-means	HAC (Euclidean distance and Ward's method)	HAC (Pearson correlation and average linkage)	HDBSCAN (Pearson correlation)
Silhouette (Higher is better)	0.4539	0.4264	0.6400	0.7604
Davies-Bouldin (Lower is better)	0.8267	0.7169	0.6001	0.4176
Caliński-Harabasz (Higher is better)	18,473	13,732	35,802	70,327

HDBSCAN is the best performing technique according to all three metrics.

occurred slightly (4 and 1%) more often during the weekend than during weekdays.

Clustering Users

Having identified three types of days of HA use enables exploring HA user behavior, thus generating personalized insights, in a way that considers the day-to-day variation of each user. We explored the behavior of HA users by clustering the 15,905 users based on how they used the HAs during the logged days (see subsection Data Analysis). Each user is represented by the proportion of time spent in each of the three types of days of HA use. Four clustering techniques were evaluated. The optimal number of clusters for k-means and both HAC techniques were determined to be three. HDBSCAN also identified three clusters, with the minimum cluster size hyperparameter set to 1,000, in addition to considering some observations as noise. Based on three internal validation metrics, HDBSCAN was chosen (**Table 1**). The Silhouette analysis (**Figure 6**) suggested that the three clusters are of different sizes and have predominantly positive and large scores.

Figure 7A displays the days of HA use experienced by the users belonging to each user group. These plots can be directly compared with **Figure 5A**, which displays all days of HA use from all users. Each user group has a distinctive distribution of days. User group A is the largest cluster (7,862 users) and exhibits a higher density in the left corner of the figure, corresponding with day type 1 (i.e., full day of HA use). User group B (2,442 users) exhibits a higher density in the lower part of the figure, corresponding with day type 2 (i.e., day of afternoon HA use). User group C (3,148 users) has a higher density in the right corner of the figure, corresponding with day type 3 (i.e., day of sporadic evening HA use). Additionally, 2,453 users exhibited atypical behavior and were labeled as noise. The distinctive behavior of the three user groups is confirmed by their average time spent in each of the typical days of HA use (**Figure 7B**). User group A is predominantly having full days of HA usage, user group B is predominantly having days of afternoon HA use, user group C is predominantly having days of sporadic evening HA use. It should be noted that the predominant day of HA use is experienced around 60% of the time by the three user groups.

Validating User Clustering Using Supervised Classifiers

We validated the HA user clustering by training an ensemble of three supervised classifiers (multiclass logistic regression, XGBoost and fully connected neural network) to predict the label of each user (user group A, B, C, or noisy user). The training input was the average day of HA use for each user, defined as minutes of HA use per hour throughout the day (from 6:00 to 23:59).

When evaluating the three individual classifiers based on accuracy and ROC-AUC score (**Table 2**), XGBoost results to

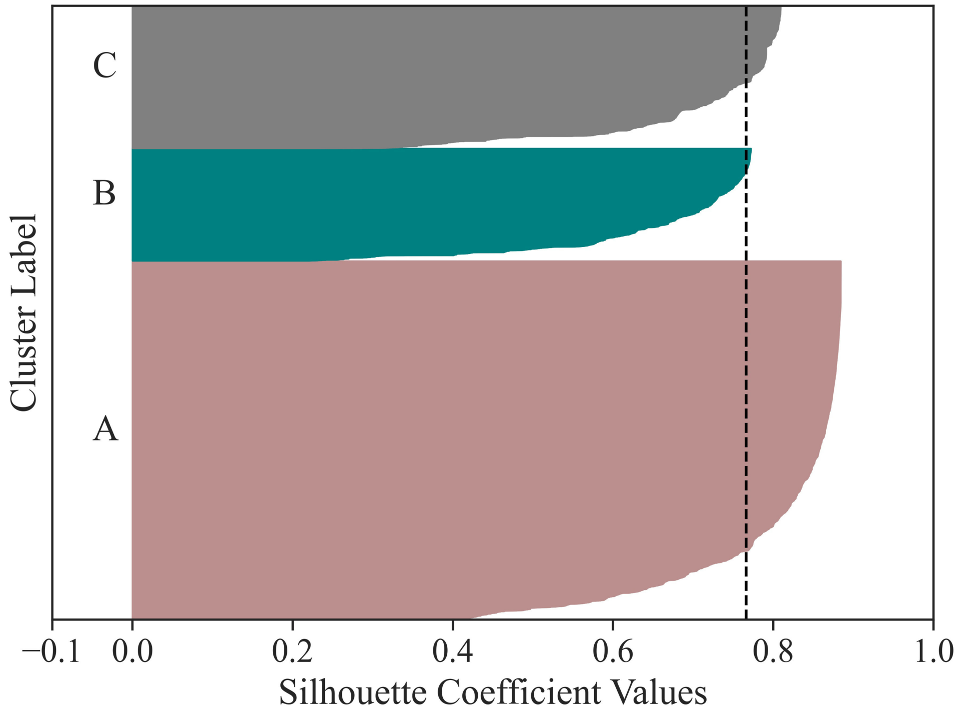


FIGURE 6 | The Silhouette coefficient value of each observation is displayed for the three user clusters (i.e., for each cluster, the observations are ordered by their Silhouette coefficient value and displayed in ascending order as horizontal stacked lines). The average silhouette score is reported (dashed line). The three clusters have predominantly positive and large scores, suggesting valid clustering.

be the best performing classifier. In order to reduce bias, an ensemble of three supervised classifiers was defined. This simulates three artificial experts coming to a decision (40). The ensemble assigns each user to a group by majority voting between the three classifiers. In case where no majority could be defined, the group was decided by the best performing individual classifier (XGBoost). The ensemble accuracy was 86.04%, while the ROC-AUC score was 0.98. While the ensemble has a slightly worse accuracy than XGBoost, relying on classifiers from different classes mitigates the effect of bias that each classifier has. The ROC curves for the ensemble of classifiers (**Figure 8**) show that noisy users exhibiting atypical behavior are the most difficult to classify (i.e., lowest AUC). Conversely, the ensemble of classifiers successfully distinguishes between the three user groups.

It is interesting to inspect the importance attributed by XGBoost to each of the 18 hours considered (**Table 3**). XGBoost values h9 and h15, indicating that these two hours are the ones that mostly differentiate the three user groups. This is consistent with the fact that each user group is characterized by a predominant day of HA use (**Figure 7**), and that h9 and h15 are the most effective hours in differentiating between the three day types (**Figure 5B**).

DISCUSSION

While HA use has been traditionally assessed through subjective self-reports, smartphone-connected HAs enable objective data logging of HA use. This study investigates the objective HA use of a large cohort of HA users. 453,612 days of HA use logged by 15,905 users were analyzed.

The amount of HA use time is informative of how long the HA has been used during a day. On average, the users used the HAs for 10.01 h/day. This value is similar (11, 17) or slightly larger (10, 19) than previous studies objectively measuring HA use. When investigating the variability between users, this study found that 25% of users used the HAs for <8.18 h. This percentage is similar to a study by Laplante-Lévesque et al. (11), but smaller than other studies (12, 18, 19) objectively or subjectively assessing the amount of HA use of several users. The inclusion criteria of this study (i.e., users of the HearingFitness™ feature *via* a smartphone app) and the data cleaning criteria (i.e., days with at least 60 min of HA use) could explain the greater average HA use and the lower percentage of light users. Moreover, a greater average HA use could be explained by the fact that, for binaural HA users this study selected the larger value between the right and left ear. While

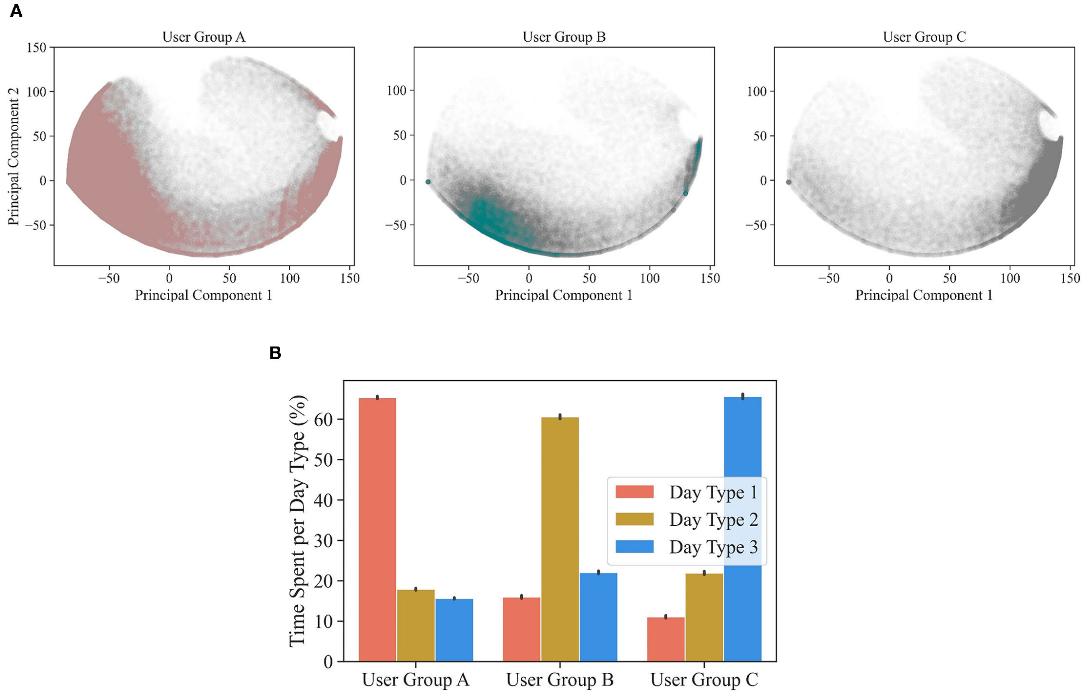


FIGURE 7 | In **(A)**, for each of the three user clusters (i.e., user group A, B, and C), the days of HA use are displayed as scatter plot against the two main principal components. The distinct densities indicate that the three user groups experienced substantially different days of HA use. In **(B)**, the average proportion of time spent (±95% confidence interval) in each day type is displayed for each user cluster.

TABLE 2 | Comparison of three individual classifiers (multiclass logistic regression, XGBoost, and fully connected neural network) and of the classification ensemble based on two performance metrics (accuracy and ROC-AUC score).

Classifier type	Accuracy (0-100 %)	ROC-AUC Score (micro-average)
Logistic regression	81.51	0.97
XGBoost	87.08	0.98
FC neural network	85.56	0.98
Ensemble	86.04	0.98

XGBoost is the best performing individual classifier according to both metrics.

HA users can either exhibit a low or high average amount of daily HA use, their day-to-day fluctuations in HA use provide a deeper understanding of HA use. The fluctuations in day-to-day HA use (i.e., within-user SD) were lower for light and heavy users compared to medium users, proving that a substantial number of users consistently displayed diverse behaviors in terms of HA use.

In addition to the amount of HA use, continuous data logging enables assessing how and when HAs were used during the day. Based on patterns of hourly use, the 453,612 days of HA use were clustered into three typical days. Forty-four percent of days were characterized by full HA use. This indicates that generally, when worn, HAs tend to be turned on in the morning (around 7), used uninterruptedly throughout the day, and turned off in the evening (around 22). Twenty-seven percent of days were characterized by afternoon use. This indicates that HAs are occasionally turned on in the late morning (around 11) and used uninterruptedly until the evening (around 22). This behavior might be due to a different individual daily rhythm or to a day encompassing different activities (e.g., weekend had a significant, but negligible effect on the day type). Twenty-six percent of days were characterized by sporadic evening HA use. This suggests that HAs are sometimes used in isolated occasions and for a limited number of hours. The remaining days (3%) were atypical days of HA use and exhibited infrequent behavior.

Based on the proportion of time spent in each of the typical days of HA use, the 15,905 users were clustered in three user groups. This method allowed to investigate users' behavior while preserving the individual day-to-day variability in HA use. Almost half of the users (group A, 49% of users) predominantly had full days of HA use. This group might include users that have an active life and engage in social interactions starting in the

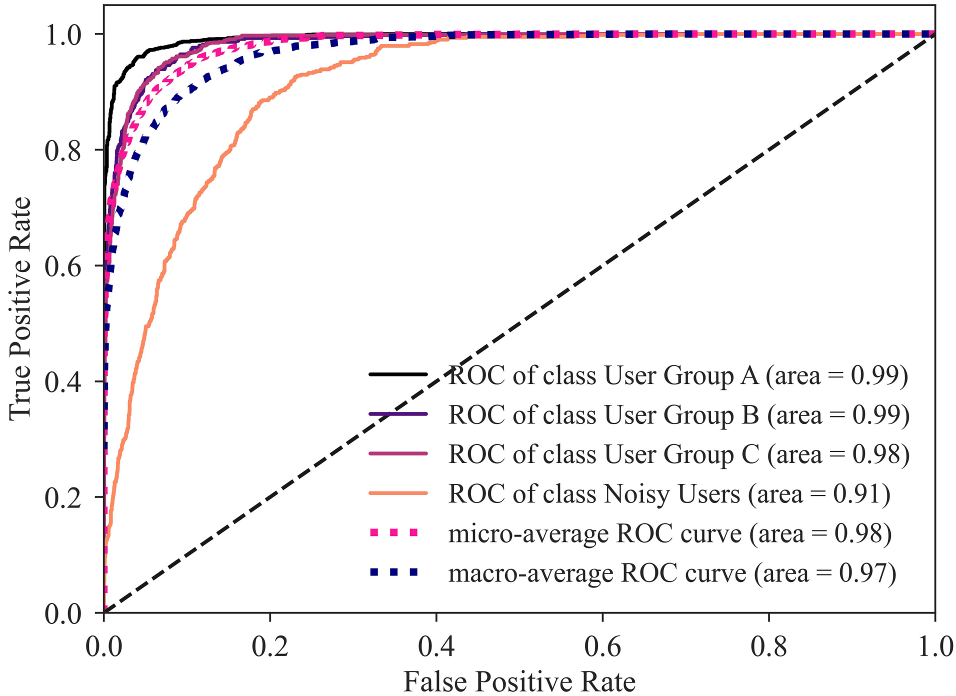


FIGURE 8 | ROC-AUC plot for the ensemble of classifiers, illustrating the tradeoff between sensitivity (True Positive Rate) and specificity (False Positive Rate). The ideal point is the top-left corner, higher AUC is better. In this multiclass scenario, the individual classes are first binarized, the individual scores are computed for each user group, then micro-averages and macro-averages are calculated for each classifier.

morning and throughout the entire day. Because of the inclusion criteria of this study (i.e., users of a smartphone app that tracks HA use), this group might be overrepresented. A smaller portion of users (group B, 15%) predominantly had days of afternoon use. This group might include users that engage in activities and social interactions later in the day. Group A and B, together, indicate that 64% of users tended to use the HAs uninterruptedly, a percentage similar to the 57% found by Laplante-Lévesque et al., (11). Twenty percent of users (Group C) predominantly had days of sporadic evening use. This group might either contain users that are not acclimatized to their HAs or users that do not depend on their HAs and only need them in specific situations (16). The remaining 15% of users were classified as noise, suggesting that some users have an uncommon behavior, more evenly alternating among the typical days of HA use. This percentage is in line with a study by Laplante-Lévesque et al., (11), according to which 23% of the subjects described their HA use to be different from day to day. Interestingly, in all three user groups, we found that the predominant day of HA use accounted for ∼60% of the time, suggesting that users exhibited a substantial within-user variability in terms of day type experienced throughout the logged days. This aspect might not emerge from self-reported assessments that suffer from recall bias, as indicated by a previous study in which most participants (77%) reported their HA use to be the same every day (11).

The user clustering was validated by training a supervised classification ensemble to predict the cluster to which each user belongs. The high accuracy achieved by the supervised classifier ensemble (∼86%) indicates valid user clustering. Indeed, this approach is based on the idea that good clustering should also support good classification, where the better the classification performance the higher the quality of the partition. As such, a high-quality partition is defined by compact clusters separated from each other to the extent that an artificial expert (i.e., a supervised classifier) can distinguish the cluster to which a new user belongs (39). This evaluation was performed to complement internal validation methods (i.e., using information of the clustering process). Internal validation methods attempt to evaluate cluster structure quality, the appropriate clustering algorithm, and the number of clusters without additional information but depend on assumptions such as the presence of underlying structure for each cluster, resulting in weaker results when they do not hold. Alternatively, cluster quality could theoretically be evaluated using external validation, which requires additional, "true" cluster labels to compare against. In real-world scenarios, finding "true" labels is often difficult as

TABLE 3 | Input feature importance returned by XGBoost.

H6	H7	H8	H9	H10	H11	H12	H13	H14	H15	H16	H17	H18	H19	H20	H21	H22	H23
0.013	0.014	0.028	**0.326**	0.017	0.019	0.021	0.039	0.064	**0.256**	0.042	0.037	0.021	0.025	0.024	0.014	0.017	0.015

*The values indicate how valuable each of the 18 features (from H6 in the morning to H23 at night) are in the construction of the boosted decision trees (internal to the model). The values greater than 0.2 are marked in bold. The model values divergence points between the three day types (H9, H15) (see **Figure 5B**). Total value is 1.*

raw data may not have reference labels, thus making external validation methods unusable.

Clustering users based on their HA use patterns provides a deeper insight into the adoption of hearing care treatments and paves the way for more personalized solutions. For instance, users that predominantly have days of sporadic evening HA use might have specific needs compared to the users that uninterruptedly use the HA for the entire day. They might only need the HAs in specific situations and thus benefit from targeted HA settings or features. Additionally, training a supervised classifier based on data labeled by a clustering technique enables future predictions for new users. Based on the average day of HA use of a new user, the classifier can predict her user group, thereby identifying users with similar behaviors and potentially leveraging on the accumulated knowledge of existing users. This can improve the clinical flow by helping audiologists make data-driven decisions.

Looking into the future, a more advanced level of personalization could improve the quality of hearing care solutions and help alleviate major challenges concerning new users, such as the cold start problem. This can be defined as the delay between starting to use the HAs and the moment when enough data was generated locally for meaningful results. Furthermore, an individual's dynamic sound environment, or soundscape, may also be an important factor for personalization. Considering the large number of soundscapes a user may be exposed to throughout the day (public transport, social events, work environments, etc.), additional features can potentially account for both the within-user and the between-user variability. An effective clustering technique for grouping similar users may serve to balance this increase in complexity, especially if advanced privacy-preserving techniques such as federated learning and differential privacy are considered. Federated learning is a machine learning framework where models are trained locally, and afterwards aggregated between participating users. This type of model development could provide access to unrivaled amounts of quality user data, as privacy concerns can only be alleviated if users never have to give away their data. Real-world implementation of such a technique could provide tangible benefits to both existing users, as well as improve the experience of new users, thus enabling next-generation privacy focused personalization systems.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is not publicly available. Requests to access the data should be directed to the corresponding author.

ETHICS STATEMENT

In the sign-up process, the participants actively gave their consent for data to be collected, stored, and used for research purposes on aggregated levels. No personal identifier was collected. No ethical approval was required for this study according to Danish National Scientific Ethical Committee (26).

AUTHOR CONTRIBUTIONS

AP and T-IS conceived and designed the study, organized the database, and performed the statistical analysis. AP wrote the manuscript. JC, KJ, JL, NP, and KS supervised the findings and revised the final manuscript. All authors contributed to the article, read, and approved the submitted version.

REFERENCES

1. World Health Organization. *Deafness and Hearing Loss.* (2021). Available online at: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss (accessed May 25, 2021).
2. Kirkwood DH. Major survey documents negative impact of untreated hearing loss on quality of life. *Hear J.* (1999) 52:32. doi: 10.1097/00025572-199907000-00004
3. Jayakody DMP, Friedland PL, Eikelboom RH, Martins RN, Sohrabi HR. A novel study on association between untreated hearing loss and cognitive functions of older adults: baseline non-verbal cognitive assessment results. *Clin Otolaryngol.* (2018) 43:182–91. doi: 10.1111/coa.12937
4. Reed NS, Altan A, Deal JA, Yeh C, Kravetz AD, Walhagen M, et al. Trends in health care costs and utilization associated with untreated hearing loss over 10 years. *JAMA Otolaryngol Head Neck Surg.* (2019) 145:27–34. doi: 10.1001/jamaoto.2018.2875
5. Kochkin S. The *Impact of Untreated Hearing Loss on Household Income. Better Hearing Institute*. (2005). Available online at: http://ihsinfo.org/IhsV2/News_Info/MarkeTrak7_ImpactUntreatedHLIncome.pdf
6. Kochkin S, Rogin CM. Quantifying the obvious: the impact of hearing instruments on quality of life. *Hear Rev.* (2000) 7:6–34. doi: 10.1097/00025572-200002000-00004
7. Vuorialho A, Karinen P, Sorri M. Effect of hearing aids on hearing disability and quality of life in the elderly. *Int J Audiol.* (2006) 45:400–5. doi: 10.1080/14992020600625007
8. Perez E, Edmonds BA. A systematic review of studies measuring and reporting hearing aid usage in older adults since 1999: a descriptive summary of measurement tools. *PLoS ONE.* (2012) 7:e31831. doi: 10.1371/journal.pone.0031831
9. Mäki-Torkko EM, Sorri MJ, Laukli E. Objective assessment of hearing aid use. *Scand Audiol.* (2001) 30:81–2. doi: 10.1080/010503901300007146
10. Gaffney P. Reported hearing aid use versus datalogging within a VA population. *Hear Rev*. (2008) 15:42–7.

11. Laplante-Lévesque A, Nielsen C, Jensen LD, Naylor G. Patterns of hearing aid usage predict hearing aid use amount (data logged and self-reported) and overreport. *J Am Acad Audiol.* (2014) 25:187–98. doi: 10.3766/jaaa.25.2.7
12. Solheim J, Hickson L. Hearing aid use in the elderly as measured by datalogging and self-report. *Int J Audiol.* (2017) 56:472–9. doi: 10.1080/14992027.2017.1303201
13. Doyle JB, Raghunathan RR, Cellum I, Li G, Golub JS. Longitudinal tracking of sound exposure and hearing aid usage through objective data logs. *Otolaryngol Head Neck Surg.* (2018) 159:110–6. doi: 10.1177/0194599818766056
14. Nelson Kakulla B. *Tech Trends of the 50+*. Washington, DC: AARP Research (2020). doi: 10.26419/res.00329.001
15. Pontoppidan NH, Li X, Bramsløw L, Johansen B, Nielsen C, Hafez A, et al. Data-driven hearing care with time-stamped data-logging. *Proc Int Sympos Audit Audiol Res*. (2018) 6:127–34. Available online at: https://proceedings.isaar.eu/index.php/isaarproc/article/view/2017-15
16. Laplante-Lévesque A, Jensen LD, Dawes P, Nielsen C. Optimal hearing aid use: focus groups with hearing aid clients and audiologists. *Ear Hear.* (2013) 34:193–202. doi: 10.1097/AUD.0b013e31826a8ecd
17. Taubman LB, Palmer CV, Durrant JD, Pratt S. Accuracy of hearing aid use time as reported by experienced hearing aid wearers. *Ear Hear.* (1999) 20:299–305. doi: 10.1097/00003446-199908000-00003
18. Bertoli S, Staehelin K, Zemp E, Schindler C, Bodmer D, Probst R. Survey on hearing aid use and satisfaction in Switzerland and their determinants. *Int J Audiol.* (2009) 48:183–95. doi: 10.1080/14992020802572627
19. Timmer BHB, Hickson L, Launer S. Hearing aid use and mild hearing impairment: learnings from big data. *J Am Acad Audiol.* (2017) 28:731–41. doi: 10.3766/jaaa.16104
20. Wong LLN, Hickson L, McPherson B. Hearing aid satisfaction: what does research from the past 20 years say? *Trends Amplif.* (2003) 7:117–61. doi: 10.1177/108471380300700402
21. Humes LE, Garner CB, Wilson DL, Barlow NN. Hearing-aid outcome measures following one month of hearing aid use by the elderly. *J Speech Lang Hear Res.* (2001) 44:469–86. doi: 10.1044/1092-4388(2001/037)
22. Laplante-Lévesque A, Knudsen LV, Preminger JE, Jones L, Nielsen C, Öberg M, et al. Hearing help-seeking and rehabilitation: perspectives of adults with hearing impairment. *Int J Audiol.* (2012) 51:93–102. doi: 10.3109/14992027.2011.606284
23. Jensen NS, Pedersen AK, Petersen EB, Theill J. When one size does not fit all: special hearing solutions for special occasions. *Hear Rev.* (2019) 26:22–6.
24. Staehelin K, Bertoli S, Probst R, Schindler C, Dratva J, Sutz EZ. Gender and hearing aids: patterns of use and determinants of nonregular use. *Ear Hear.* (2011) 32:e26–37. doi: 10.1097/AUD.0b013e3182291f94
25. Laplante-Lévesque A, Pontoppidan NH, Mazevski A, Schum D, Behrens T, Porsbo M. *Data-Driven Hearing Care With Hearing Fitness^{TM}: Oticon Shares Openly Its Innovative Vision and Roadmap*. Smørum: Oticon A/S (2017).
26. National Videnskabsetisk Komité. *What Should I Report? National Scientific Ethics Committee.* (2020). Available online at: https://www.nvk.dk/forsker/naar-du-anmelder/hvilke-projekter-skal-jeg-anmelde (accessed May 31, 2021).
27. Walker EA, Mccreery RW, Spratford M, Oleson JJ, Buren JV, Bentler R, et al. Trends and predictors of longitudinal hearing aid use for children who are hard of hearing HHS public access. *Ear Hear.* (2015) 36:38–47. doi: 10.1097/AUD.0000000000000208
28. Lloyd S. Least squares quantization in PCM. *IEEE Trans Inf Theory.* (1982) 28:129–37. doi: 10.1109/TIT.1982.1056489
29. Arthur D, Vassilvitskii S. K-means^{++}: the advantages of careful seeding. In: *18th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2007*. New Orleans, LO: Society for Industrial and Applied Mathematics (2007). p. 1027–35.
30. Thorndike RL. Who belongs in the family? *Psychometrika.* (1953) 18:267–76. doi: 10.1007/BF02289263
31. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math.* (1987) 20:53–65. doi: 10.1016/0377-0427(87)90125-7
32. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Prettenhofer P, et al. Scikit-learn: machine learning in python. *J Mach Learn Res.* (2011) 12:2825–30.
33. Rokach L, Maimon O. Clustering methods. In: Maimon O, Rokach L, editors. *Data Mining and Knowledge Discovery Handbook*. Boston, MA: Springer (2005). doi: 10.1007/0-387-25465-X_15
34. Campello RJGB, Moulavi D, Sander J. Density-based clustering based on hierarchical density estimates. In: Pei J, Tseng VS, Cao L, Motoda H, Xu G, editors. *Advances in Knowledge Discovery and Data Mining. PAKDD 2013. Lecture Notes in Computer Science, Vol. 7819*. Berlin; Heidelberg: Springer. doi: 10.1007/978-3-642-37456-2_14
35. Hahsler M, Piekenbrock M. *dbscan: Density Based Clustering of Applications with Noise (DBSCAN) and Related Algorithms*. (2021). Available online at: https://cran.r-project.org/package=dbscan
36. Caliński T, Harabasz J. A dendrite method for cluster analysis. *Commun Stat.* (1974) 3:1–27. doi: 10.1080/03610917408548446
37. Davies DL, Bouldin DW. A cluster separation measure. IEEE Trans Pattern Anal Mach Intell PAMI (1979). 1:224–7. doi: 10.1109/TPAMI.1979.4766909
38. McInnes L, Healy J, Astels S. hdbscan: Hierarchical density based clustering. *J Open Source Softw.* (2017) 2:205. doi: 10.21105/joss.00205
39. Rodríguez J, Medina-Pérez MA, Gutierrez-Rodríguez, AE, Monroy R, Terashima-Marín H. Cluster validation using an ensemble of supervised classifiers. *Knowl Based Syst.* (2018) 145:134–44. doi: 10.1016/j.knosys.2018.01.010
40. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* New York, NY: ACM (2016) p. 785–94.
41. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys.* (1943) 5:115–33. doi: 10.1007/BF02478259
42. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: an imperative style, high-performance deep learning library. In: Wallach H, Larochelle H, Beygelzimer A, et al., editors. *Advances in Neural Information Processing Systems 32.* Curran Associates, Inc., (2019). p. 8024–35.
43. Bengfort B, Danielsen N, Bilbro R, Gray L, McIntyre K, Richardson G, et al. Yellowbrick v0.6 (0.6). *Zenodo*. (2018). doi: 10.5281/zenodo.1206264
44. Nakano R. reiinakano/scikit-plot: v0.2.1 (v0.2.1). *Zenodo*. (2017). doi: 10.5281/zenodo.293191
45. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Lawrence Erlbaum Associates (1988). doi: 10.4324/9780203771587

Conflict of Interest: AP, T-IS, and KJ are employed by Demant A/S. JC, NP, and KS are employed by Oticon A/S.

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

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

Appendix E

Investigating the Provision and Context of Use of Hearing Aid Listening Programs from Real-World Data

Alessandro Pasta, Tiberiu-Ioan Szatmari, Jeppe Høy Christensen, Kasper Juul Jensen, Niels Henrik Pontoppidan, Kang Sun, and Jakob Eg Larsen. Investigating the Provision and Context of Use of Hearing Aid Listening Programs from Real-world Data.

Investigating the Provision and Context of Use of Hearing Aid Listening Programs from Real-world Data

Abstract

Background: Listening programs enable hearing aid users to change device settings for specific listening situations and thereby personalize their listening experience. However, investigations into real-world use of such listening programs to support clinical decisions and evaluate the success of hearing aid treatment are lacking.
Objective: To investigate the distribution of provided listening programs among a large group of in-market hearing aid users and the context by which the programs are typically used.
Methods: First, we analyzed how many and which programs were provided to 32,336 in-market hearing aid users. Second, we explored 396,723 program selections from 1,312 selected users to investigate the sound environments in which specific programs are used and whether such environments reflect the listening intent conveyed by the name of the used program. Our analysis is based on real-world longitudinal data logged by smartphone-connected hearing aids.
Results: 57% of hearing aid users in our sample had programs for specific listening situations, which is a higher proportion than previously reported, most likely due to the inclusion criteria. Based on association rule mining, we identified a primary additional listening program, “Speech in Noise”, usually provided when also getting other programs. We also identified two secondary additional programs (“Comfort”, “Music”), usually provided to users that also have “Speech in Noise”. Two programs (“TV”, “Remote Mic”) were instead related to the use of external accessories and not associated with other programs. On average users selected “Speech in Noise”, “Comfort” and “Music” in louder, noisier, and less modulated (all $P<.001$) environments compared with the environment in which they selected the default “General". The difference from the sound environment in which they selected “General” was significantly higher in the minutes following program selection compared to the minutes preceding it.
Conclusions: This study provides a deeper insight into the provision of listening programs on a large scale and demonstrates that additional listening programs are used as intended and according to the sound environment conveyed by the program name.

Keywords: listening programs; sound environment; personalization; data logging; multimemory hearing aids

Introduction

Background

Untreated hearing loss is a widespread condition [1] that has repercussions at an individual [2–4] and societal level [1,5,6]. Globally, 72 million people could potentially benefit from the use of a hearing device [1]. The adoption of hearing aids (HAs) has been shown to have a positive impact on the quality of life of users [7,8] and to mitigate the effect of hearing loss on household income [4]. However, one of the requisites for the widespread adoption and usage of HAs is user satisfaction [9]. HA users use the HAs and report listening difficulties in different real-life situations ranging from face-to-face conversations to coping with environmental sounds [10]. Therefore, to achieve high user satisfaction, HAs need to be able to cater for a wide range of situations. This is confirmed by previous research that found that one of the main reasons for not owning or not using HAs is that the HAs are not working well in specific situations, for instance when there is background noise [9,11,12], when listening to speech [13], or when being in a large group of people [14].

For these reasons, programmable multimemory HAs have been introduced, which enable providing the user with multiple listening programs for specific listening situations. Nowadays, 41% of HA owners have such programs [15]. Listening programs set pre-defined rules for contextually adapting different audiological parameters such as overall gain, frequency shaping of the gain, noise reduction, and directionality. Programs can be manually selected via the HA buttons, a remote control, or a smartphone app. Users are usually advised to use a program in a specific listening situation [16]. This is reflected by the name of the program, which often conveys the situation where it is meant to be used (e.g., "speech in noise", "music") [16]. Thus, programs are a way for users to contextually adapt the device settings in specific listening situations and thereby personalize their listening experience. Therefore, investigating the use of listening programs potentially enables a deeper understanding of users' behavior and needs.

Related work

Previous research has shown that HA users can benefit from certain listening programs in specific listening environments. For instance, having specific listening programs for listening in quiet and in noise has been found to improve speech recognition [17,18]. However, to benefit from programs, HA users need to be able to characterize the listening environment adequately and actively select the appropriate program [16]. Previous research conducted on eleven experienced HA users has shown that the percentage of users who selected identical programs in the same situation (repeatability) surpassed the level corresponding to pure guess under almost all listening conditions [19]. Higher repeatability has been found in demanding listening situations [19]. These results suggest that listening programs can discernibly impact the listening experience.

Although different listening programs can potentially be beneficial and discernible for HA users, little is known about their real-world use. De Graaff et al. [16] performed a scoping review on the use of multimemory devices containing several listening programs and investigated if HA users appreciate and adequately use the option to switch between programs. Remarkably few studies were found about the use of multiple programs for various listening environments. Stelmachowicz, Lewis and Carney [20] found that HA users did not tend to select different settings (in terms of frequency shaping of the gain) across simulated sound environments, although differences in the preferred overall gain were sometimes observed. Conversely, Keidser et al. [21] found that 5 out of 27 HA users preferred different frequency response characteristics in different listening conditions, mainly in noisy environments. Similarly, Banerjee [22] found that HA users preferred the default setting most often and nondefault settings mainly in difficult listening situations. Additionally, several studies found that most HA users switched between omnidirectional and directional microphone settings and that microphone preferences depend on the characteristics of the listening environment [23–26].

These studies suggest that some HA users value and use the option to switch between listening programs. However, the existing literature is sparse and dated. While listening programs investigated in older studies used to set a constant level for an audiological parameter (e.g., higher constant amount of noise reduction), nowadays listening programs set dynamic rules for contextually adapting the parameters (e.g., rule that provides earlier and higher noise reduction as the user transitions to a complex environment). Moreover, some questions remain unanswered. First, it is not clear what motivates a HA user to get a multimemory HA and manually switch between programs, and in which listening situations users particularly seek device personalization. Second, as highlighted in the aforementioned systematic review, little is known about the correct use of programs designated for a specific listening environment [16]. Indeed, establishing the need for a multimemory device does not guarantee that the user will immediately notice the benefits of multiple programs. The failure to match the multimemory HA settings to the communication and environmental needs of the individual may lead to delays in fully realizing its benefits [27]. None of the studies included in the systematic review examined whether a certain program was used in the correct listening environment (e.g., whether users selected a "Speech in noise" program in noisy environments)[16] during everyday life.

Furthermore, most of the aforementioned studies relied on self-reported measures collected over a short period of time. Indeed, they used diaries or questionnaires in which HA users reported use, preferences, and details of the listening environments. Whether the appropriate program is used in each listening environment cannot be derived from these data [16]. Moreover, most studies pay little attention to the continuation of use of listening programs after the completion of the study. On the one hand, subjects might use programs during the study period but stop using them once the study finishes. On the other hand, they might need to acclimatize to the use of programs and their preferences may only be evident after extended use [28]. In contrast to self-reported measures, data logging enables investigating the real-world behavior of a larger number of users [29]. It allows gathering objective program usage, as well as objective contextual data. Moreover, it enables assessing program use with a greater temporal resolution and longitudinally, making it possible to investigate detailed patterns of use, explore the long-term user behavior, and account for the acclimatization phase [30]. Investigating the use of listening programs by using objective data logging could unveil insights on how users select between multiple listening programs under natural conditions and thereby pave the way for more personalized hearing care solutions.

Research objective

This study aims to investigate the provision and context of use of multimemory HAs by leveraging objective data logged by smartphone-connected HAs from in-market users across several countries. First, we investigate the provision of multiple listening programs for various listening environments. Namely, we examine how many and which programs HA users have, and whether some programs are commonly provided together. Second, we explore whether HA users use specific programs in distinct listening situations and whether such situations reflect the listening intent conveyed by the name of the program. We do so by focusing on users that repeatedly use specific programs and investigating the sound environment in which such programs are selected.

Methods

Participants and Apparatus

This study used data from a large-scale internal (Oticon A/S, Smoerum) database, which stores logs of the HA use of HA owners who have signed up for the HearingFitness™ feature [31] via the Oticon ON™ smartphone app. The participants were the owners of Oticon Opn™ HAs who used the HearingFitness™ feature in the period between June and September 2020. In the sign-up process, the participants actively gave their consent for data to be collected, stored, and used for research purposes on aggregated levels. No personal identifier was collected. No ethical approval is necessary for this study according to Danish National Scientific Ethical Committee [32].

Data and Data Analysis

When the HAs are connected to the smartphone, the HearingFitness™ feature logs timestamped data about the interactions with the HAs, such as the selection of specific listening programs. To account for different phrasing or different languages adopted by hearing care professionals when naming the programs, similar program names were coded in fewer categories. Moreover, when the HAs are connected to the smartphone, timestamped data about the sound environment is collected every 10 minutes and every time a listening program is selected by the user. Namely, the sound pressure level (SPL), the noise floor (NF), and the sound modulation level (SML) in decibels are measured across a broad frequency band (0.1 – 10kHz) [33].

The SPL is the most used indicator of the sound wave strength and correlates well with human perception of loudness [34]. The NF is the level of background noise in a signal. The SML describes how much the modulated variable (e.g., speech) of the signal varies around its unmodulated level.

Provision of Listening Programs

The provision of listening programs was investigated by including users that have usage information for at least 20 hours and analyzing, for each user, the programs that have been selected at least once in the 4-month period. This resulted in a total of 32,336 users and 67,996 programs provided.

We explored the provision of listening programs by computing the number of programs available per user and the most frequently provided programs across users. Furthermore, we investigated the relationships between programs by determining association rules [35] via the Apriori algorithm [36]. Given a set of n programs $P = \{p_1, p_2, \dots, p_n\}$, and a set of users $U = \{u_1, u_2, \dots, u_m\}$, where each user is provided with a subset of the programs in P, a rule is defined as an implication of the form: $X \Rightarrow Y$, where X is the antecedent, Y is the consequent, $X, Y \subseteq P$, and $X, Y \cap \emptyset$ [36]. In determining the association rules, the default program (i.e., "General") was excluded and only users having at least one additional listening program were considered (i.e., 18,153 users). Indeed, the default program is available for nearly all users and including it in the association rules would not be of any interest. Instead, the association rules related to the five most frequent additional listening programs were inspected. The rules were evaluated based on several metrics: support, coverage, confidence, coverage, and lift [36]. The support of a rule defines how often the rule appears in the data set. The coverage refers to how often the antecedent of a rule appears in the data set and measures how often the rule can be applied [37]. The confidence of a rule is defined as $\mathrm{conf}(\mathrm{X} \Rightarrow \mathrm{Y}) = \mathrm{support}(\mathrm{X} \cup \mathrm{Y})/\mathrm{support}(\mathrm{X})$ and can be interpreted as an estimate of the probability $\mathrm{P}(\mathrm{Y}|\mathrm{X})$ [36], measuring how often a rule is correct out of the applicable cases. A potential issue with confidence is that an association rule having a very frequent consequent will always have high confidence. The lift addresses this concern by taking into account how frequent the items are in the data set. The lift of a rule is defined as $\mathrm{lift}(\mathrm{X} \Rightarrow \mathrm{Y}) = \mathrm{support}(\mathrm{X} \cup \mathrm{Y})/(\mathrm{support}(\mathrm{X})\mathrm{support}(\mathrm{Y}))$, and can be interpreted as the deviation of the support of the whole rule from the support expected if the antecedent and the consequent were independent [36]. Finally, the likelihood of a program being provided to users with respectively one, two, three or four programs was investigated. The data manipulation was performed in Python. The association rule mining was performed in R by using the arules package [38].

Use of Listening Programs vs Sound Environment

Contextual program use was evaluated by analyzing the sound environment (SPL, NF, SML) during program change selection. For each logged selection of a specific listening program, the sound environment measured in a 10-minute time-window centered on the program selection was considered (i.e., 5 minutes preceding and 5 minutes following the selection). The sound environment occurring during different selections of a specific program by the same user was averaged. For each program, only users with at least 5 selections were included. Such threshold was chosen to ensure that users' behavior was inferred from a representative sample of program selections, while, at the same time, not dropping too many users. Moreover, based on the analysis described in the previous section, only a relevant subset of listening programs was included. The resulting data set comprised 396,723 program selections from 1,312 users. The data manipulation was performed in Python.

The aforementioned computations resulted, for each program, in a distribution of users by their average sound environment when selecting the program. We visually compared the sound environment occurring when selecting a specific listening program with the sound environment occurring, for the same users, when selecting the default program (i.e., "General"). Subsequently, we tested whether the differences between the two sound environments were normally distributed. Based on the result of the normality test, we tested the significance of the differences by performing either a paired t-test or a Wilcoxon signed-rank test [39]. The latter is a non-parametric version of the paired t-test which is robust to outliers and tests whether the distribution of the differences is symmetric around zero. Additionally, the difference between the sound environment occurring when selecting "Speech in Noise" and the sound environment occurring when selecting respectively "Comfort" and "Music" was tested by performing a Mann-Whitney U test [40], a nonparametric test of the difference between the distributions underlying two samples.

Furthermore, we investigated whether the sound environment changed before or after the program selection by computing, for each program, a 5-minutes running average of the sound environment throughout the 10-minute time window (i.e., for each minute, we looked for sound environment logs occurred in the previous 5 minutes for the same user and averaged them).

Additionally, by means of a Wilcoxon signed-rank test [39], the average sound environment occurring in the 5 minutes before program selection was compared with the sound environment occurring, for the same users, in the 5 minutes after program selection.

The analysis was performed in Python using the NumPy [41], Pandas [42], Seaborn [43], Scipy [44] libraries.

Results

Provision of Listening Programs

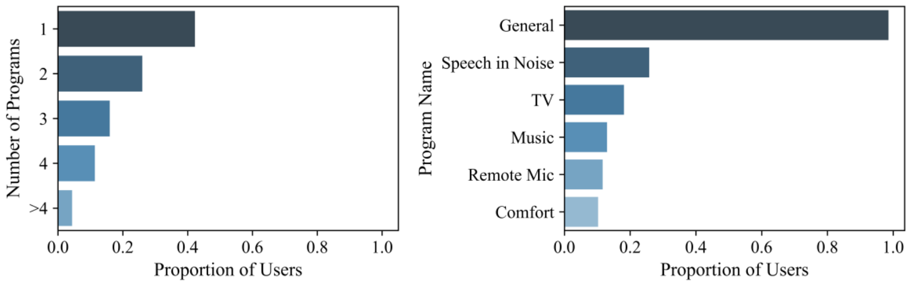


Figure 1 Number of programs available for each user and names of programs most frequently provided

Among the HA users, 57% have more than one listening program (Figure 1). Almost every user (99%) has the default program, "General" (Figure 1). This means that more than half of the users have at least one program for specific listening situations in addition to the default one. Respectively 26%, 13%, and 10% of users have a "Speech in Noise", "Music", and "Comfort" program. The names of these programs convey a specific listening intent. Compared to "General", "Speech in Noise" provides more noise reduction, less directionality and more sound details. "Music" provides an omnidirectional sound and no noise reduction. "Comfort" provides more noise reduction in complex environments, more directionality, less amplification, and less sound details.

Additionally, respectively 18% and 12% have a "TV" and "Remote Mic" program. These programs are related to the use of an accessory, such as a TV adapter and a remote microphone.

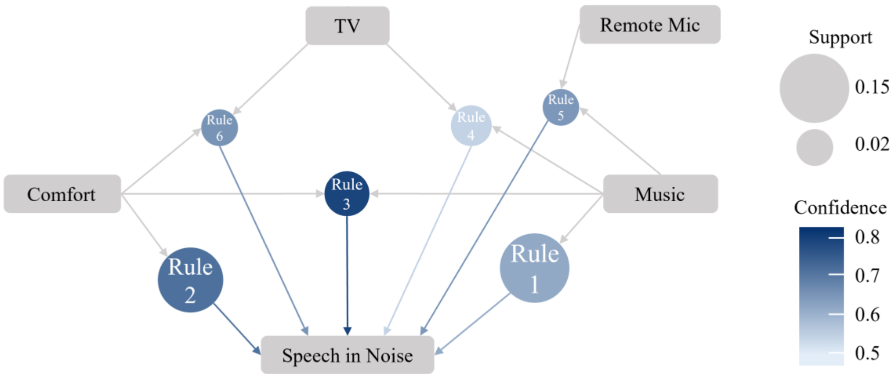


Figure 2 Association rules with support≥0.02, confidence>0.5 and lift>1. The support of each rule is indicated by the area of the circle, while the confidence is conveyed by the color intensity. "Speech in Noise" is the consequent of all rules, suggesting that it is a primary program, frequently provided when also providing secondary programs such as "Comfort" and "Music".

Table 1 Association rules with support≥0.02, confidence>0.5 and lift>1

Rule	Antecedent	Consequent	Support	Coverage	Confidence	Lift	Count
1	Music	Speech in Noise	0.14	0.23	0.62	1.35	2612
2	Comfort	Speech in Noise	0.13	0.18	0.71	1.54	2357
3	Comfort, Music	Speech in Noise	0.04	0.06	0.79	1.71	801
4	Music, TV	Speech in Noise	0.03	0.05	0.56	1.21	476
5	Music, Remote Mic	Speech in Noise	0.02	0.03	0.66	1.43	401
6	Comfort, TV	Speech in Noise	0.02	0.03	0.67	1.45	377

Investigating the association rules with support≥0.02 and confidence>0.5 (Figure 2) enables exploring the relationships between programs. In this analysis, "General" was not considered as it is uniformly provided and is not an additional listening program. The detailed metrics of the selected rules are presented in Table 1. Not only is "Speech in Noise" the most common additional listening program, but it is also a primary program that users get when also getting secondary programs. Indeed, "Speech in Noise" is the consequent of all selected rules, while either "Comfort" or "Music" are always in the antecedent set. As shown by Table 1, respectively 62% and 71% of users who have either "Music" (rule 1) or "Comfort" (rule 2) also have "Speech in Noise". Similarly, 79% of users who have both "Music" and "Comfort" (rule 3) also have "Speech in Noise". For these rules, the lift is greater than 1, indicating that users are more likely to have "Speech in Noise" when they also have "Music" and/or "Comfort". Conversely, although "TV" is a frequently provided program, users that have such program are not more likely to have other listening programs.

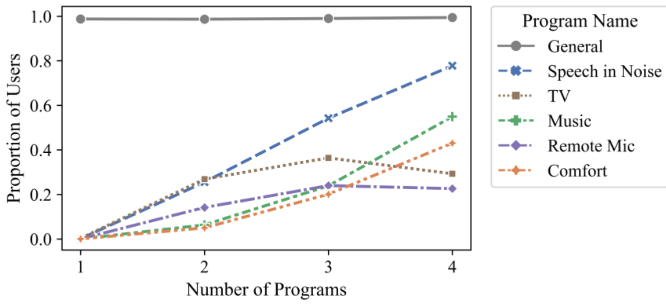


Figure 3 Likelihood of specific programs being provided to users with respectively one, two, three or four programs.

Figure 3 confirms some of the previous findings. Almost all users have the "General" program regardless of the number of additional programs. Among the users that have two programs, "Speech in Noise", "TV", and, to a lesser extent, "Remote Mic" are more likely to be available than "Music" and "Comfort". For users with three or four programs, the likelihood of having the primary program "Speech in Noise" grows linearly, while the likelihood of having "TV" or "Remote Mic" remains relatively constant and secondary programs "Music" and "Comfort" increase in likelihood.

Use of Listening Programs vs Sound Environment

Based on the findings in the previous section, we investigated the sound environment in which a relevant subset of the listening programs was used. We focused on programs that convey a specific listening intent, whether they are primary ("Speech in Noise") or secondary ("Comfort", "Music"). These three programs are meant to be used in specific listening situations and are not related to the use of an accessory.

First, we analyzed whether the primary program ("Speech in Noise") was selected in different listening situations compared to the default program ("General"). The upper graphs in Figure 4 display the distribution of users by their average sound environment respectively when selecting "Speech in Noise" and "General". Users selected "Speech in Noise" in louder (higher SPL), noisier (higher NF), and less modulated (lower SML) sound environments. Indeed, on average, users selected "Speech in Noise" at 55.1 dB SPL (SD=7.4), 46.9 dB NF (SD=7.0), 17.1 dB SML (SD=4.9), while they selected "General" at 53.0 dB SPL (SD=5.6), 44.5 dB NF (SD=5.2), 18.2 dB SML (SD=3.5). The distributions of users by their average sound environment resulted not to be normally distributed (P>.1), therefore a Wilcoxon Signed-Rank Test was performed to test their difference. Based on such test, the difference between the two distributions ("Speech in Noise" and "General") was statistically significant (P<.001) for all three parameters, as shown by Table 2. Focusing on individual users, the lower graphs in Figure 4 show that most of the users (respectively 64%, 66%, 62%) selected “Speech in Noise” in environments characterized by higher SPL, higher NF, and lower SML.

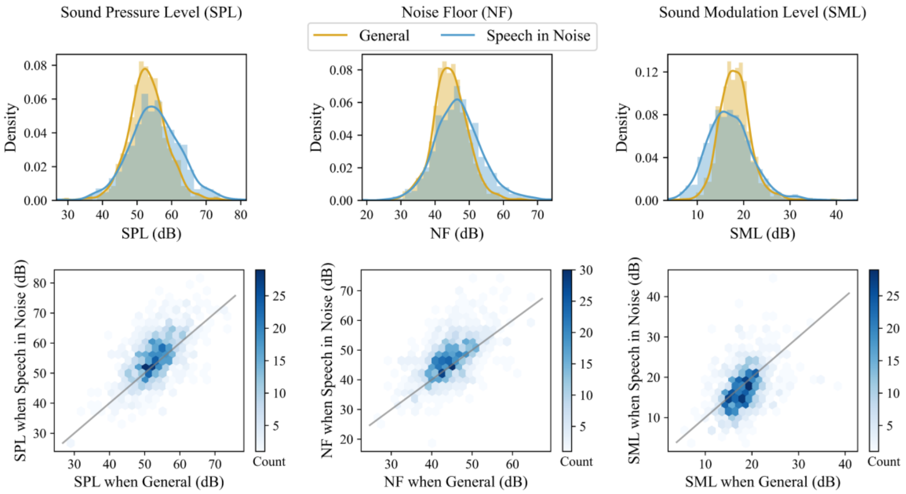


Figure 4 Analysis of the sound environment (SPL, NF, SML) in which “Speech in Noise” and “General” are selected. Compared to “General”, users select “Speech in Noise” in louder, noisier, and less modulated environments. In the upper figures, distribution of users (using histograms and kernel density estimation) by their average sound environment when selecting "General" and "Speech in Noise". In the lower figures, 2d histograms displaying, for each user, the sound environment when selecting “Speech in Noise” (y-axis) and “General” (x-axis). The color of the hexagon is determined by the number of users in the hexagon. The identity line (y=x) is drawn in grey. If a user experiences the same sound environment when selecting “Speech in Noise” and “General”, the corresponding hexagon falls exactly on the identity line.

Second, we analyzed whether the secondary programs (“Comfort” and “Music”) were selected in specific listening situations. As shown in Table 2, users selected both programs in louder, noisier, and less modulated (Wilcoxon Signed-Rank test, all P<.001) sound environments compared with the sound environment in which they selected “General”. Compared with “Speech in Noise”, “Music” was selected in less loud (Mann-Whitney U test, P=.009) and less noisy (Mann-Whitney U test, P<.001) environments. There was no difference in SML between selection of “Speech in Noise” and “Music” (P=.152). Furthermore, no difference was found between the sound environment in which “Speech in Noise” and “Comfort” were selected (SPL, P=.405; NF, P=.595; SML, P=.344).

Table 2 Average difference between the sound environment when selecting a program (i.e., "Speech in Noise", "Comfort", "Music") and the sound environment when selecting the "General" program, computed in a 10-minute interval around program selection.

Program	**N**	**Mean difference from General**		
		SPL	**NF**	**SML**
Speech in Noise	963	+2.06* (SD=6.05)	+2.35* (SD=5.83)	-1.11* (SD=4.65)
Comfort	300	+1.93* (SD=6.50)	+2.50* (SD=5.97)	-1.28* (SD=4.74)
Music	366	+1.67* (SD=6.23)	+1.59* (SD=5.77)	-1.04* (SD=4.69)
* Statistically significant at P<.001 based on the Wilcoxon Signed-Rank test.				

Finally, we investigated to what extent the sound environment changes before and after the program selection. Figure 5 displays, for a time-window near the program selection, the 5-minutes running average of the difference between the sound environment when selecting a program and when selecting "General". For all three programs ("Speech in Noise", "Comfort", "Music") and all three sound environment features (SPL, NF, SML), a sound environment difference from "General" was observed throughout the whole 10-minute time window. Moreover, as shown by Table 3, for all three programs and sound environment features, the sound environment difference from "General" was significantly higher in the 5 minutes following program selection, compared to the 5 minutes preceding program selection (Wilcoxon Signed-Rank test, all $P<.05$). By observing the magnitude of the differences (Table 3), it can be noticed that this was particularly the case for the "Speech in Noise" and "Comfort" programs and the NF and SML features.

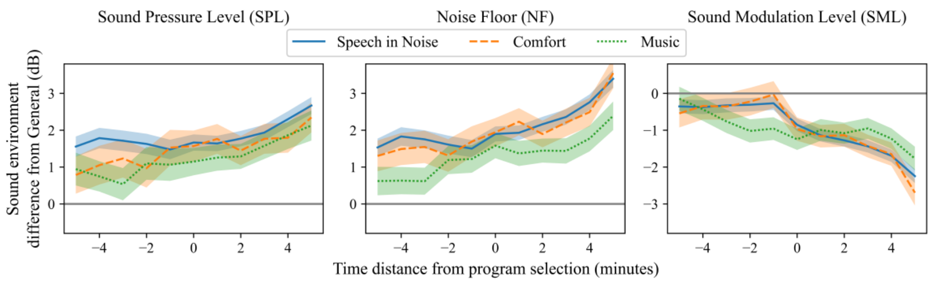


Figure 5 5-minutes running average (± SE) of the sound environment difference from "General", computed in a time-window near the program selection. The difference deviates from zero throughout the whole time-window. However, especially for NF and SML, the difference increases after program selection.

Table 3 Average difference between the sound environment when selecting a program (i.e., "Speech in Noise", "Comfort", "Music") and the sound environment when selecting the "General" program, computed for the 5 minutes before program selection ("Before") and for the 5 minutes after program selection ("After"). A Wilcoxon Signed-Rank test was conducted to test whether the sound environment differences before and after the selection are statistically different from each other.

Program	N	Mean difference from General					
		SPL		NF		SML	
		Before	After	Before	After	Before	After
Speech in Noise	583	+1.81 (SD=5.89)	+2.73 (SD=5.58)	+2.12 (SD=5.44)	+3.36 (SD=5.71)	-0.85 (SD=4.74)	-2.25 (SD=4.70)
		Diff=0.92 (P<.001)		Diff=1.24 (P<.001)		Diff=1.4 (P<.001)	
Comfort	179	+1.60 (SD=5.60)	+2.25 (SD=5.42)	+2.14 (SD=4.99)	+3.51 (SD=5.57)	-1.45 (SD=4.93)	-2.75 (SD=4.80)
		Diff=0.65 (P=.031)		Diff=1.37 (P=.002)		Diff=1.3 (P<.001)	
Music	206	+1.21 (SD=6.31)	+2.13 (SD=6.12)	+1.45 (SD=5.81)	+2.31 (SD=5.81)	-0.86 (SD=4.89)	-1.51 (SD=4.85)
		Diff=0.92 (P=.016)		Diff=0.86 (P=.001)		Diff=0.65 (P=.002)	

Discussion

This study investigates the provision and context of use of HA listening programs by leveraging on real-world data logged through smartphone-connected HAs.
The majority of HA users in our sample (57%) were found to have listening programs for specific listening situations in addition to the default program. According to a previous study analyzing self-reported data, 41% of HA owners have a program button or switch to change the HA response for different listening environments [15]. The inclusion criteria (i.e., users of the HearingFitness™ feature via a smartphone app) and the data collection method (i.e., objective data logging) of our study could explain the higher prevalence of listening programs. "Speech in Noise" was the most commonly provided program besides the "General" default program. By association rule mining, "Speech in Noise" was also found to be a primary program that users tend to get when also getting other secondary programs, such as "Comfort" and "Music". This indicates that users either request or are recommended a "Speech in Noise" program as initial step when being interested in personalizing their listening experience in specific listening situations. This finding is consistent with previous studies reporting that HA users most frequently struggle when there is background noise [9,11,12], or when being in a large group of people [14] and, consequently, they are least likely to be satisfied with their hearing when following conversations in noise and in large groups [15]. "Comfort" and "Music" resulted to be secondary programs, frequently provided to users that also have "Speech in Noise" and more likely to be provided to users having three or four programs. As with "Speech in Noise", these programs signal the interest in personalizing the listening experience in a specific listening situation, i.e., when it is

noisy but there is no need to communicate and when listening to music. Although these situations are not as prevalent as communicating in noise, users highly motivated to personalize their experience can still benefit from adopting specific listening programs for these situations [45]. Despite being common programs, "TV" and "Remote Mic" were provided differently than the other programs. They were frequently provided to users having only two programs (including "General"), but they were not frequently provided in connection with other programs and their prevalence did not increase among users having more than two programs. This might be explained by the fact that such programs are frequently provided to users that own a TV adapter (i.e., a device that enables streaming the TV sound via the HAs) or a remote microphone. Therefore, such programs show an interest in using the accessory more than in contextually adapting the HA settings through a listening program. Moreover, selecting the "TV" program might actively modify the sound environment by silencing the TV while reproducing the sound directly into the HAs. Therefore, "TV" and "Remote Mic" are not included in the following sections.

Subsequently, we analyzed the sound environment in which "Speech in Noise", "Comfort" and "Music" were selected. First, we found that, on average, users selected "Speech in Noise" in louder, noisier, and less modulated environments compared with the environment in which they selected "General". This proves that HA users select the "Speech in Noise" program in environments that possess distinct characteristics and that better resembles a conversation in noise.
Second, "Comfort" was also selected in louder, noisier, and less modulated listening environments compared with "General", suggesting that HA users select it when they want to get relief in noisy environments. Interestingly, HA users selected "Comfort" and "Speech in Noise" in the same sound environments (i.e., no significant difference was found), indicating that similar sound environments might require different HA settings because of differing listening intents, or that users fail to recognize what program is best suited for the environment. In the former case, the high-level sound environment features SPL, NF, and SML might not be sufficient to reveal different listening intents from the ambient sound. Third, "Music" was selected in louder, noisier, and less modulated listening environments compared with "General", but in less loud and less noisy environments compared with "Speech in Noise". The music playing in the environment might explain the higher loudness and noise, although not as extreme as "Speech in Noise" scenarios. Overall, considering that HA users are typically counselled to use a program in a specific listening situation [16], our findings suggest that they tend to follow such recommendations in the real-world usage of their HAs. Empowering users to personalize their listening experience by contextually adapting the HA settings can therefore result in more appropriate settings for some relevant listening situations.
Finally, we analyzed the temporal progression of the sound environment throughout a time-window close to the program selection. In the minutes following the program selection, especially for "Speech in Noise" and "Comfort", a significant increase was observed in the difference to the sound environment occurring when selecting "General". This suggests that a number of users tends to select additional listening programs in anticipation, rather than as a reaction, to a change in the

sound environment. This could indicate that users are aware of what the contextually most appropriate program is and proactively select it before entering a specific listening situation.

In terms of future work, it would be interesting to investigate to what extent the provision of listening programs depends on HA owners requesting a program or on the hearing care professional recommending it. Indeed, hearing care professionals traditionally have a great influence on the prescribed hearing solution, and data about the provision of listening programs might not only reflect the needs and preferences of HA users, but also the beliefs and knowledge of the professionals. Moreover, the role of individual predictors for the provision and use of listening programs deserves further study. Indeed, the benefit from a personalized and contextualized solution might depend on the degree of hearing loss or additional data characterizing the individuals such as age, prior experience with hearing aids, auditory cognitive capabilities, or supra-threshold hearing characteristics. Finally, the significant differences found in the sound environment occurring when using specific listening programs indicate that the analyzed sound environment features (SPL, NF, SML) are promising candidate for predicting the selection of an additional listening program over the default program. Complementing such objective sound environment features with more subjective contextual features and with an evaluation of the listening experience might also enable a deeper understanding of the provision and use of HA listening programs.

Acknowledgements

Alessandro Pasta, and Tiberiu-Ioan Szatmari conceived and designed the study, organized the database, and performed the statistical analysis. Alessandro Pasta wrote the manuscript. Jakob Eg Larsen, Jeppe Høy Christensen, Kasper Juul Jensen Niels Henrik Pontoppidan, and Kang Sun supervised the findings and revised the final manuscript. All authors contributed to the article, read, and approved the submitted version.

Conflicts of Interest

Alessandro Pasta, Tiberiu-Ioan Szatmari, Kasper Juul Jensen are employed by Demant A/S. Jeppe Høy Christensen, Niels Henrik Pontoppidan, and Kang Sun are employed by Oticon A/S. The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

HA: Hearing aid
NF: Noise floor
SML: Sound modulation level
SPL: Sound pressure level

References

1. World Health Organization. Deafness and hearing loss [Internet]. 2021 [cited 2021 May 25]. Available from: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss
2. Jayakody DMP, Friedland PL, Eikelboom RH, Martins RN, Sohrabi HR. A novel study on association between untreated hearing loss and cognitive functions of older adults: Baseline non-verbal cognitive assessment results. Clin Otolaryngol [Internet] Blackwell Publishing Ltd; 2018 Feb 1 [cited 2021 May 18];43(1):182–191. PMID:28710824
3. Kirkwood DH. Major survey documents negative impact of untreated hearing loss on quality of life. Hear J [Internet] 1999 Jul;52(7):32. [doi: 10.1097/00025572-199907000-00004]
4. Kochkin S. The Impact of Untreated Hearing Loss on Household Income [Internet]. Better Hear Inst. 2005. Available from: http://ihsinfo.org/IhsV2/News_Info/MarkeTrak7_ImpactUntreatedHLIncome.pdf
5. Reed NS, Altan A, Deal JA, Yeh C, Kravetz AD, Wallhagen M, Lin FR. Trends in Health Care Costs and Utilization Associated With Untreated Hearing Loss Over 10 Years. JAMA Otolaryngol Neck Surg [Internet] American Medical Association; 2019 Jan 1 [cited 2021 May 18];145(1):27. PMID:30419131
6. McDaid D, Park A-L, Chadha S. Estimating the global costs of hearing loss. Int J Audiol [Internet] Taylor & Francis; 2021 Mar 1 [cited 2021 Sep 20];60(3):162–170. [doi: 10.1080/14992027.2021.1883197]
7. Kochkin S, Rogin CM. Quantifying the Obvious: The Impact of Hearing Instruments on Quality of Life. Hear Rev 2000;7(1):6–34.
8. Vuorialho A, Karinen P, Sorri M. Effect of hearing aids on hearing disability and quality of life in the elderly. Int J Audiol [Internet] Taylor & Francis; 2006 Jan 7 [cited 2021 May 25];45(7):400–405. PMID:16938798
9. Bertoli S, Staehelin K, Zemp E, Schindler C, Bodmer D, Probst R. Survey on hearing aid use and satisfaction in Switzerland and their determinants. Int J Audiol [Internet] Taylor & Francis; 2009 Jan 7 [cited 2021 May 29];48(4):183–195. PMID:19363719
10. Galvez G, Turbin MB, Thielman EJ, Istvan JA, Andrews JA, Henry JA. Feasibility of Ecological Momentary Assessment of Hearing Difficulties Encountered by Hearing Aid Users. Ear Hear [Internet] 2012 Jul;33(4):497–507. [doi: 10.1097/AUD.0b013e3182498c41]
11. Hartley D, Rochtchina E, Newall P, Golding M, Mitchell P. Use of Hearing Aids and Assistive Listening Devices in an Older Australian Population. J Am Acad Audiol [Internet] 2010 Nov 6 [cited 2021 Sep 20];21(10):642–653. [doi: 10.3766/jaaa.21.10.4]
12. Aazh H, Prasher D, Nanchahal K, Moore BCJ. Hearing-aid use and its determinants in the UK National Health Service: A cross-sectional study at the Royal Surrey County Hospital. Int J Audiol [Internet] Informa Healthcare; 2015 Mar 4 [cited 2021 Sep 20];54(3):152–161. [doi: 10.3109/14992027.2014.967367]

13. Takahashi G, Martinez CD, Beamer S, Bridges J, Noffsinger D, Sugiura K, Bratt GW, Williams DW. Subjective Measures of Hearing Aid Benefit and Satisfaction in the NIDCD/VA Follow-Up Study. J Am Acad Audiol [Internet] American Academy of Audiology; 2007 Apr 7 [cited 2021 Sep 20];18(04):323–349. [doi: 10.3766/jaaa.18.4.6]
14. Gallagher NE, Woodside J V. Factors Affecting Hearing Aid Adoption and Use: A Qualitative Study. J Am Acad Audiol [Internet] 2018 Apr 29 [cited 2021 Sep 16];29(04):300–312. [doi: 10.3766/jaaa.16148]
15. Picou EM. MarkeTrak 10 (MT10) Survey Results Demonstrate High Satisfaction with and Benefits from Hearing Aids. Semin Hear [Internet] Thieme Medical Publishers, Inc.; 2020 Feb 10 [cited 2021 Oct 21];41(01):021–036. PMID:32047346
16. de Graaff F, Huysmans E, Ket JCF, Merkus P, Goverts ST, Leemans CR, Smits C. Is there evidence for the added value and correct use of manual and automatically switching multimemory hearing devices? A scoping review. Int J Audiol [Internet] Informa UK Limited, trading as Taylor & Francis Group; 2018;57(3):176–183. PMID:29017358
17. van den Heuvel J, Goverts ST, Kapteyn TS. Evaluation of Fitting Rules with a Programmable Hearing Aid. Int J Audiol [Internet] Audiology; 1997 Jan [cited 2021 Dec 12];36(5):261–278. PMID:9305523
18. Ricketts TA, Bentler RA. Comparison of two digitally programmable hearing aids. J Am Acad Audiol [Internet] 1992 Mar 1 [cited 2021 Dec 12];3(2):101–12. PMID:1600212
19. Berninger E, Nordström I. Multiple Memory Hearing Aid. Consistency of Program-Usage in Real-World Listening Situations. Scand Audiol [Internet] Taylor & Francis; 1997 Jan 12 [cited 2021 Sep 14];26(4):252–256. [doi: 10.3109/01050399709048016]
20. Stelmachowicz PG, Lewis DE, Carney E. Preferred Hearing-Aid Frequency Responses in Simulated Listening Environments. J Speech, Lang Hear Res [Internet] American Speech-Language-Hearing Association; 1994 Jun [cited 2021 Sep 23];37(3):712–718. [doi: 10.1044/jshr.3703.712]
21. Keidser G, Pellegrino A, Delifotis A, Ridgway J, Clarke M. The use of different frequency response characteristics in everyday environments. Aust J Audiol 1997;19(1):9–21.
22. Banerjee S. Hearing Aids in the Real World: Use of Multimemory and Volume Controls. J Am Acad Audiol [Internet] 2011 Jun 6;22(06):359–374. [doi: 10.3766/jaaa.22.6.5]
23. Cord MT, Surr RK, Walden BE, Olson L. Performance of Directional Microphone Hearing Aids in Everyday Life. J Am Acad Audiol [Internet] 2002 Jun 7 [cited 2021 Sep 23];13(06):295–307. [doi: 10.1055/s-0040-1715973]
24. Walden BE, Surr RK, Cord MT, Dyrlund O. Predicting Hearing Aid Microphone Preference in Everyday Listening. J Am Acad Audiol [Internet] 2004 May 7 [cited 2021 Sep 27];15(05):365–396. [doi: 10.3766/jaaa.15.5.4]
25. Palmer C, Bentler R, Mueller GH. Evaluation of a Second-Order Directional Microphone Hearing Aid: II. Self-Report Outcomes. J Am Acad Audiol [Internet] 2006 Mar 7 [cited 2021 Sep 27];17(03):190–201. [doi:

10.3766/jaaa.17.3.5]
26. Ricketts T, Henry P, Gnewikow D. Full Time Directional versus User Selectable Microphone Modes in Hearing Aids. Ear Hear [Internet] Ear Hear; 2003 Oct [cited 2021 Sep 27];24(5):424–439. PMID:14534412
27. Kuk FK. Clinical Considerations in Fitting a Multimemory Hearing Aid. Am J Audiol 1993;2(3):23–27. [doi: 10.1044/1059-0889.0203.23]
28. Kuk FK. Evaluation of the Efficacy of a Multimemory Hearing Aid. J Am Acad Audiol. 1992.
29. Pasta A, Szatmari T-I, Christensen JH, Jensen KJ, Pontoppidan NH, Sun K, Larsen JE. Clustering Users Based on Hearing Aid Use: An Exploratory Analysis of Real-World Data. Front Digit Heal [Internet] Frontiers; 2021 Sep 3 [cited 2021 Sep 16];3:109. [doi: 10.3389/fdgth.2021.725130]
30. Wright D, Gagné J-P. Acclimatization to Hearing Aids by Older Adults. Ear Hear [Internet] NLM (Medline); 2021 Jan 31 [cited 2021 Apr 28];42(1):193–205. PMID:32769437
31. Laplante-Lévesque A, Pontoppidan NH, Mazevski A, Schum D, Behrens T, Porsbo M. Data-driven hearing care with HearingFitness™: Oticon shares openly its innovative vision and roadmap [Internet]. Oticon, Inc; 2017. Available from: https://wdh01.azureedge.net/-/media/oticon/main/pdf/master/whitepaper/27896uk_datadriven_hearing_care_with-hearingfitness_kitenet.pdf?la=en&rev=FF6A&hash=01557C708E53021DFE4E192A272AFA78
32. National Videnskabsetisk Komité. What should I report? | National Scientific Ethics Committee [Internet]. 2020 [cited 2021 May 31]. Available from: https://www.nvk.dk/forsker/naar-du-anmelder/hvilke-projekter-skal-jeg-anmelde
33. Christensen JH, Pontoppidan NH, Rossing R, Anisetti M, Bamiou D-E, Spanoudakis G, Murdin L, Bibas T, Kikidiks D, Dimakopoulos N, Giotis G, Ecomomou A. Fully Synthetic Longitudinal Real-World Data From Hearing Aid Wearers for Public Health Policy Modeling. Front Neurosci 2019;13:1–5. [doi: 10.3389/fnins.2019.00850]
34. Long M. Fundamentals of Acoustics. Archit Acoust 2014. p. 39–79. [doi: 10.1016/b978-0-12-398258-2.00002-7]
35. Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. ACM SIGMOD Rec [Internet] ACM PUB27 New York, NY, USA; 1993 Jun 1 [cited 2021 Oct 22];22(2):207–216. [doi: 10.1145/170036.170072]
36. Hahsler M, Grün B, Hornik K. arules - A Computational Environment for Mining Association Rules and Frequent Item Sets. J Stat Softw [Internet] American Statistical Association; 2005 Sep 29 [cited 2021 Oct 22];14(15):1–25. [doi: 10.18637/jss.v014.i15]
37. coverage: Calculate coverage for rules in arules: Mining Association Rules and Frequent Itemsets [Internet]. [cited 2022 Jan 8]. Available from: https://rdrr.io/cran/arules/man/coverage.html
38. Hahsler M, Buchta C, Gruen B, Hornik K. arules: Mining Association Rules and

Frequent Itemsets [Internet]. 2021. Available from: https://cran.r-project.org/package=arules
39. Wilcoxon F. Individual Comparisons by Ranking Methods. Biometrics Bull [Internet] JSTOR; 1945 Dec [cited 2021 Oct 22];1(6):80. [doi: 10.2307/3001968]
40. Mann HB, Whitney DR. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. Ann Math Stat [Internet] Institute of Mathematical Statistics; 1947 Mar 1 [cited 2021 Oct 26];18(1):50–60. [doi: 10.1214/aoms/1177730491]
41. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, del Río JF, Wiebe M, Peterson P, Gérard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C, Oliphant TE. Array programming with NumPy. Nat 2020 5857825 [Internet] Nature Publishing Group; 2020 Sep 16 [cited 2021 Dec 12];585(7825):357–362. PMID:32939066
42. The pandas development team. pandas-dev/pandas: Pandas. Zenodo; 2020.
43. Waskom ML. seaborn: statistical data visualization. J Open Source Softw [Internet] The Open Journal; 2021;6(60):3021. [doi: 10.21105/joss.03021]
44. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat \.Ilhan, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P, SciPy 1.0 Contributors. {SciPy} 1.0: Fundamental Algorithms for Scientific Computing in Python. Nat Methods 2020;17:261–272. [doi: 10.1038/s41592-019-0686-2]
45. Hockley NS, Bahlmann F, Chasin M. Programming hearing instruments to make live music more enjoyable. Hear J [Internet] 2010 Sep;63(9):30. [doi: 10.1097/01.HJ.0000388538.27405.93]

Bibliography

[1] Hashir Aazh et al. “Hearing-aid use and its determinants in the UK National Health Service: A cross-sectional study at the Royal Surrey County Hospital”. In: *International Journal of Audiology* 54.3 (Mar. 2015), pp. 152–161. ISSN: 1499-2027. DOI: 10.3109/14992027.2014.967367.

[2] Harvey Abrams et al. “A patient-adjusted fine-tuning approach for optimizing hearing aid response”. In: *Hearing Review* 18.3 (2011), pp. 18–27. URL: https://www.hearingreview.com/practice-building/practice-management/a-patient-adjusted-fine-tuning-approach-for-optimizing-the-hearing-aid-response.

[3] Jayne B. Ahlstrom, Amy R. Horwitz, and Judy R. Dubno. “Spatial separation benefit for unaided and aided listening”. In: *Ear and hearing* 35.1 (Jan. 2014), pp. 72–85. ISSN: 1538-4667. DOI: 10.1097/AUD.0B013E3182A02274.

[4] Gabriel Aldaz, Sunil Puria, and Larry J. Leifer. “Smartphone-Based System for Learning and Inferring Hearing Aid Settings”. In: *Journal of the American Academy of Audiology* 27.09 (Oct. 2016), pp. 732–749. ISSN: 1050-0545. DOI: 10.3766/jaaa.15099.

[5] American Speech-Language-Hearing Association. *Conductive Hearing Loss*. URL: https://www.asha.org/public/hearing/conductive-hearing-loss/.

[6] American Speech-Language-Hearing Association. *Guidelines for Manual Pure-Tone Threshold Audiometry*. Tech. rep. 2005. DOI: 10.1044/policy.GL2005-00014.

[7] American Speech-Language-Hearing Association. *Mixed Hearing Loss*. URL: https://www.asha.org/public/hearing/mixed-hearing-loss/.

[8] American Speech-Language-Hearing Association. *Sensorineural Hearing Loss*. URL: https://www.asha.org/public/hearing/sensorineural-hearing-loss/.

[9] American Speech-Language-Hearing Association. *Types of Hearing Loss*. URL: https://www.asha.org/public/hearing/types-of-hearing-loss/.

[10] Hélène Amieva et al. "Death, Depression, Disability, and Dementia Associated With Self-reported Hearing Problems: A 25-Year Study". In: *The Journals of Gerontology: Series A* 73.10 (Sept. 2018), pp. 1383–1389. ISSN: 1079-5006. DOI: 10.1093/gerona/glx250.

[11] Stig Arlinger, Peter Nordqvist, and Marie Öberg. "International Outcome Inventory for Hearing Aids: Data From a Large Swedish Quality Register Database". In: *American Journal of Audiology* 26.3S (Oct. 2017), pp. 443–450. ISSN: 1059-0889. DOI: 10.1044/2017{_}AJA-16-0123.

[12] Nicole M. Armstrong et al. "Associations of Hearing Loss and Depressive Symptoms With Incident Disability in Older Adults: Health, Aging, and Body Composition Study". In: *The Journals of Gerontology: Series A* 75.3 (Feb. 2020). Ed. by Jay Magaziner, pp. 531–536. ISSN: 1079-5006. DOI: 10.1093/gerona/gly251.

[13] Nidhi Arora et al. *The value of getting personalization right - or wrong - is multiplying*. URL: https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/the-value-of-getting-personalization-right-or-wrong-is-multiplying.

[14] D.M. Baguley, T.E. Cope, and D.J. McFerran. "Functional auditory disorders". In: *Handbook of Clinical Neurology*. Ed. by M Hallett, Stone J., and A. Carson. Vol. 139. 94. Elsevier, 2016. Chap. 32, pp. 367–378. ISBN: 9780128017722. DOI: 10.1016/B978-0-12-801772-2.00032-1.

[15] Robert W. Baloh, Sarah H. Ying, and Kathleen M. Jacobson. "A Longitudinal Study of Gait and Balance Dysfunction in Normal Older People". In: *Archives of Neurology* 60.6 (June 2003), p. 835. ISSN: 0003-9942. DOI: 10.1001/archneur.60.6.835.

[16] Shilpi Banerjee. "Hearing Aids in the Real World: Use of Multimemory and Volume Controls". In: *Journal of the American Academy of Audiology* 22.06 (June 2011), pp. 359–374. ISSN: 1050-0545. DOI: 10.3766/jaaa.22.6.5.

[17] Rebecca J. Bennett, Ariane Laplante-Lévesque, and Robert H. Eikelboom. "How Do Hearing Aid Owners Respond to Hearing Aid Problems?" In: *Ear & Hearing* 40.1 (Jan. 2019), pp. 77–87. ISSN: 0196-0202. DOI: 10.1097/AUD.0000000000000595.

[18] Ruth Bentler et al. "Digital noise reduction: Outcomes from laboratory and field studies". In: *International Journal of Audiology* 47.8 (Aug. 2008), pp. 447–460. ISSN: 14992027. DOI: 10.1080/14992020802033091.

[19] Erik Berninger and Inger Nordström. "Multiple Memory Hearing Aid. Consistency of Program-Usage in Real-World Listening Situations". In: *Scandinavian Audiology* 26.4 (Jan. 1997), pp. 252–256. ISSN: 0105-0397. DOI: 10.3109/01050399709048016.

[20] Sibylle Bertoli et al. "Survey on hearing aid use and satisfaction in Switzerland and their determinants". In: *International Journal of Audiology* 48.4 (Jan. 2009), pp. 183–195. ISSN: 1499-2027. DOI: 10.1080/14992020802572627.

[21] Virginia Best et al. "The role of high frequencies in speech localization". In: *The Journal of the Acoustical Society of America* 118.1 (July 2005), pp. 353–363. ISSN: 0001-4966. DOI: 10.1121/1.1926107.

[22] Arthur Boothroyd and Carol Mackersie. "A "Goldilocks" Approach to Hearing-Aid Self-Fitting: User Interactions". In: *American Journal of Audiology* 26.3S (Oct. 2017), pp. 430–435. ISSN: 1059-0889. DOI: 10.1044/2017{_}AJA-16-0125.

[23] The Editors of Encyclopaedia Britannica. *deafness*. 2018. URL: https://www.britannica.com/science/deafness.

[24] Inge Brons, Rolph Houben, and Wouter A. Dreschler. "Acoustical and Perceptual Comparison of Noise Reduction and Compression in Hearing Aids". In: *Journal of Speech, Language, and Hearing Research* 58.4 (Aug. 2015), pp. 1363–1376. ISSN: 1092-4388. DOI: 10.1044/2015{_}JSLHR-H-14-0347.

[25] Simon Carlile and Daviid Schonstein. "Frequency Bandwidth and Multi-talker Environments". In: *120th Convention of the Audio Engineering Society Convention*. Audio Engineering Society, May 2006.

[26] Wade Chien and Frank R Lin. "Prevalence of hearing aid use among older adults in the United States". In: *Archives of internal medicine* 172.3 (Feb. 2012), pp. 292–3. ISSN: 1538-3679. DOI: 10.1001/archinternmed.2011.1408.

[27] John Greer Clark. "Uses and abuses of hearing loss classification". In: *ASHA* 23.7 (July 1981), pp. 493–500. ISSN: 0001-2475. URL: http://www.ncbi.nlm.nih.gov/pubmed/7052898.

[28] Mary T. Cord et al. "Performance of Directional Microphone Hearing Aids in Everyday Life". In: *Journal of the American Academy of Audiology* 13.06 (June 2002), pp. 295–307. ISSN: 1050-0545. DOI: 10.1055/s-0040-1715973.

[29] David R. Cunningham, Kristie J. Williams, and L. Jane Goldsmith. "Effects of Providing and Withholding Postfitting Fine-Tuning Adjustments on Outcome Measures in Novice Hearing Aid Users". In: *American Journal of Audiology* 10.1 (June 2001), pp. 13–23. ISSN: 1059-0889. DOI: 10.1044/1059-0889(2001/001).

[30] Dayna S Dalton et al. "The Impact of Hearing Loss on Quality of Life in Older Adults". In: *The Gerontologist* 43.5 (Oct. 2003), pp. 661–668. ISSN: 1758-5341. DOI: 10.1093/geront/43.5.661.

[31] Dana David and Perla Werner. "Stigma regarding hearing loss and hearing aids: A scoping review". In: *Stigma and Health* 1.2 (May 2016), pp. 59–71. ISSN: 2376-6972. DOI: 10.1037/SAH0000022.

[32] Piers Dawes et al. "Hearing Loss and Cognition: The Role of Hearing Aids, Social Isolation and Depression". In: *PLOS ONE* 10.3 (Mar. 2015). Ed. by Blake Johnson, e0119616. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0119616.

[33] Jamie L. Desjardins and Karen A. Doherty. "The effect of hearing aid noise reduction on listening effort in hearing-impaired adults". In: *Ear and hearing* 35.6 (2014), pp. 600–610. ISSN: 1538-4667. DOI: 10.1097/AUD.0000000000000028.

[34] John B. Doyle et al. "Longitudinal Tracking of Sound Exposure and Hearing Aid Usage through Objective Data Logs". In: *Otolaryngology - Head and Neck Surgery (United States)* 159.1 (July 2018), pp. 110–116. ISSN: 10976817. DOI: 10.1177/0194599818766056.

[35] C Elberling and Vejby K Hansen. "Hearing instruments: Interaction with user preference". In: *Auditory Models and Non-Linear Hearing Instruments. Proceedings of the 18th Danavox Symposium.* Ed. by Arne Nørby Rasmussen et al. Copenhagen, Denmark, 1999, pp. 341–357.

[36] Claus Elberling and Kirsten Worsoe. *Fading sounds*. Herlev: Bording A/S, 2006. ISBN: 8799130106.

[37] Melanie A. Ferguson et al. "Hearing aids for mild to moderate hearing loss in adults". In: *Cochrane Database of Systematic Reviews* 2017.9 (Sept. 2017). ISSN: 14651858. DOI: 10.1002/14651858.CD012023.pub2.

[38] Danielle M. Foley, Kevin D. Frick, and Frank R. Lin. "Association Between Hearing Loss and Healthcare Expenditures in Older Adults". In: *Journal of the American Geriatrics Society* 62.6 (June 2014), pp. 1188–1189. ISSN: 00028614. DOI: 10.1111/jgs.12864.

[39] Food and Drug Administration. *Types of Hearing Aids*. URL: https://www.fda.gov/medical-devices/hearing-aids/types-hearing-aids.

[40] Andrew Freeland, Jane Jones, and Naufal Kassim Mohammed. "Sensorineural deafness in Tanzanian children—Is ototoxicity a significant cause? A pilot study". In: *International Journal of Pediatric Otorhinolaryngology* 74.5 (May 2010), pp. 516–519. ISSN: 01655876. DOI: 10.1016/j.ijporl.2010.02.010.

[41] Patricia Gaffney. "Reported Hearing Aid Use Versus Datalogging Within a VA Population". In: *Hearing Review* 15.6 (Jan. 2008), pp. 42–47. URL: https://nsuworks.nova.edu/hpd_aud_facarticles/37%20https://www.hearingreview.com/practice-building/practice-management/reported-hearing-aid-use-versus-datalogging-in-a-va-population.

[42] Nicola E Gallagher and Jayne V Woodside. "Factors Affecting Hearing Aid Adoption and Use: A Qualitative Study". In: *Journal of the American Academy of Audiology* 29.04 (Apr. 2018), pp. 300–312. ISSN: 1050-0545. DOI: 10.3766/jaaa.16148.

[43] Shaoqing Ge et al. "Longitudinal Association Between Hearing Loss, Vision Loss, Dual Sensory Loss, and Cognitive Decline". In: *Journal of the American Geriatrics Society* 69.3 (Mar. 2021), pp. 644–650. ISSN: 15325415. DOI: 10.1111/JGS.16933.

[44] Dane J. Genther et al. "Association Between Hearing Impairment and Risk of Hospitalization in Older Adults". In: *Journal of the American Geriatrics Society* 63.6 (June 2015), pp. 1146–1152. ISSN: 00028614. DOI: 10.1111/jgs.13456.

[45] Feike de Graaff et al. "Is there evidence for the added value and correct use of manual and automatically switching multimemory hearing devices? A scoping review". In: *International Journal of Audiology* 57.3 (2018), pp. 176–183. ISSN: 17088186. DOI: 10.1080/14992027.2017.1385864.

[46] Michael Hahsler, Bettina Grün, and Kurt Hornik. "arules - A Computational Environment for Mining Association Rules and Frequent Item Sets". In: *Journal of Statistical Software* 14.15 (Sept. 2005), pp. 1–25. ISSN: 1548-7660. DOI: 10.18637/jss.v014.i15.

[47] Lydia M. Haile et al. "Hearing loss prevalence and years lived with disability, 1990–2019: findings from the Global Burden of Disease Study 2019". In: *The Lancet* 397.10278 (Mar. 2021), pp. 996–1009. ISSN: 01406736. DOI: 10.1016/S0140-6736(21)00516-X.

[48] David Hartley et al. "Use of Hearing Aids and Assistive Listening Devices in an Older Australian Population". In: *Journal of the American Academy of Audiology* 21.10 (Nov. 2010), pp. 642–653. ISSN: 1050-0545. DOI: 10.3766/jaaa.21.10.4.

[49] J. van den Heuvel, S. Th Goverts, and Th S. Kapteyn. "Evaluation of Fitting Rules with a Programmable Hearing Aid". In: *International Journal of Audiology* 36.5 (Jan. 1997), pp. 261–278. ISSN: 1499-2027. DOI: 10.3109/00206099709071979.

[50] Benjamin W.Y. Hornsby, Earl E. Johnson, and Erin Picou. "Effects of degree and configuration of hearing loss on the contribution of high- and low-frequency speech information to bilateral speech understanding". In: *Ear and hearing* 32.5 (Sept. 2011), pp. 543–555. ISSN: 1538-4667. DOI: 10.1097/AUD.0B013E31820E5028.

[51] Søren Hougaard and Stefan Ruf. "EuroTrak I: A Consumer Survey About Hearing Aids in Germany, France, and the UK". In: *Hearing Review* 18.2 (2011), pp. 12–28. URL: https://hearingreview.com/hearing-products/eurotrak-i-a-consumer-survey-about-hearing-aids-in-germany-france-and-the-uk.

[52] Larry E. Humes et al. "Hearing-Aid Outcome Measures Following One Month of Hearing Aid Use by the Elderly". In: *Journal of Speech, Language, and Hearing Research* 44.3 (June 2001), pp. 469–486. ISSN: 1092-4388. DOI: 10.1044/1092-4388(2001/037).

[53] Institute for Health Metrics and Evaluation. *GBD Compare*. Tech. rep. Seattle, WA: University of Washington. URL: https://vizhub.healthdata.org/gbd-compare/.

[54] L. Jenstad and J. Moon. "Systematic Review of Barriers and Facilitators to Hearing Aid Uptake in Older Adults". In: *Audiology Research* 1.1 (Mar. 2011), pp. 91–96. ISSN: 2039-4349. DOI: 10.4081/audiores.2011.e25.

[55] Benjamin Johansen et al. "Personalizing the Fitting of Hearing Aids by Learning Contextual Preferences From Internet of Things Data". In: *Computers* 7.1 (2017), p. 1. DOI: 10.3390/computers7010001.

[56] Amisha Kanji, Katijah Khoza-Shangase, and Nomfundo Moroe. "Newborn hearing screening protocols and their outcomes: A systematic review". In: *International Journal of Pediatric Otorhinolaryngology* 115 (Dec. 2018), pp. 104–109. ISSN: 01655876. DOI: 10.1016/j.ijporl.2018.09.026.

[57] G Keidser et al. "The use of different frequency response characteristics in everyday environments". In: *Australian Journal of Audiology* 19.1 (1997), pp. 9–21. ISSN: 01571532.

[58] G. Keidser et al. "The NAL-NL2 Prescription Procedure". In: *Audiology Research* 1.1 (Mar. 2011), pp. 88–90. ISSN: 2039-4349. DOI: 10.4081/audiores.2011.e24.

[59] Gitte Keidser and Karima Alamudi. "Real-Life efficacy and reliability of training a hearing aid". In: *Ear and Hearing* 34.5 (2013), pp. 619–629. ISSN: 01960202. DOI: 10.1097/AUD.0b013e31828d269a.

[60] Mead C Killion. "The SIN report: Circuits haven't solved the hearing-in-noise problem". In: *The Hearing Journal* 50.10 (1997), pp. 28–30.

[61] Mead C. Killion. "New Thinking on Hearing in Noise: A Generalized Articulation Index". In: *Seminars in Hearing* 23.1 (2002), pp. 057–076. ISSN: 07340451. DOI: 10.1055/s-2002-24976.

[62] S Kochkin. *The Impact of Untreated Hearing Loss on Household Income.* Tech. rep. August. 2005. URL: http://ihsinfo.org/IhsV2/News_Info/MarkeTrak7_ImpactUntreatedHLIncome.pdf.

[63] Sergei Kochkin et al. "MarkeTrak VIII: The impact of the hearing healthcare professional on hearing aid user success". In: *Hearing Review* 17.4 (2010), pp. 12–34. ISSN: 1074-5734.

[64] Francis K Kuk. *Evaluation of the Efficacy of a Multimemory Hearing Aid.* Tech. rep. 1992, pp. 338–348.

[65] Ariane Laplante-Lévesque et al. "Hearing help-seeking and rehabilitation: Perspectives of adults with hearing impairment". In: *International Journal of Audiology* 51.2 (Feb. 2012), pp. 93–102. ISSN: 1499-2027. DOI: 10.3109/14992027.2011.606284.

[66] Ariane Laplante-Lévesque et al. "Optimal hearing aid use: Focus groups with hearing aid clients and audiologists". In: *Ear and Hearing* 34.2 (2013), pp. 193–202. ISSN: 01960202. DOI: 10.1097/AUD.0b013e31826a8ecd.

[67] Ariane Laplante-Lévesque et al. "Patterns of hearing aid usage predict hearing aid use amount (data logged and self-reported) and overreport". In: *Journal of the American Academy of Audiology* 25.2 (2014), pp. 187–198. ISSN: 10500545. DOI: 10.3766/jaaa.25.2.7.

[68] Suzanne Carr Levy et al. "Extended High-Frequency Bandwidth Improves Speech Reception in the Presence of Spatially Separated Masking Speech". In: *Ear & Hearing* 36.5 (Sept. 2015), e214–e224. ISSN: 0196-0202. DOI: 10.1097/AUD.0000000000000161.

[69] Frank R. Lin and Luigi Ferrucci. "Hearing Loss and Falls Among Older Adults in the United States". In: *Archives of Internal Medicine* 172.4 (Feb. 2012), pp. 369–271. ISSN: 0003-9926. DOI: 10.1001/archinternmed.2011.728.

[70] Frank R. Lin et al. "Hearing Loss and Cognitive Decline in Older Adults". In: *JAMA Internal Medicine* 173.4 (Feb. 2013), p. 293. ISSN: 2168-6106. DOI: 10.1001/jamainternmed.2013.1868.

[71] Frank R. Lin et al. "Hearing Loss and Incident Dementia". In: *Archives of Neurology* 68.2 (Feb. 2011), pp. 214–220. ISSN: 0003-9942. DOI: 10.1001/archneurol.2010.362.

[72] Xuezhong Liu et al. "Prevalence and aetiology of profound deafness in the general population of Sichuan, China". In: *The Journal of Laryngology & Otology* 107.11 (Nov. 1993), pp. 990–993. ISSN: 0022-2151. DOI: 10.1017/S0022215100125071.

[73] Gill Livingston et al. "Dementia prevention, intervention, and care: 2020 report of the Lancet Commission". In: *The Lancet* 396.10248 (Aug. 2020), pp. 413–446. ISSN: 01406736. DOI: 10.1016/S0140-6736(20)30367-6.

[74] Marshall Long. "Fundamentals of Acoustics". In: *Architectural Acoustics*. 2014, pp. 39–79. DOI: 10.1016/b978-0-12-398258-2.00002-7.

[75] Marshall Long. "Human Perception and Reaction to Sound". In: *Architectural Acoustics*. 2014, pp. 81–127. DOI: 10.1016/b978-0-12-398258-2.00003-9.

[76] David G. Loughrey et al. "Association of Age-Related Hearing Loss With Cognitive Function, Cognitive Impairment, and Dementia: A Systematic Review and Meta-analysis". In: *JAMA otolaryngology– head & neck surgery* 144.2 (Feb. 2018), pp. 115–126. ISSN: 2168-619X. DOI: 10.1001/jamaoto.2017.2513.

[77] Kristy Jones Lowery and Patrick N. Plyler. "The effects of noise reduction technologies on the acceptance of background noise". In: *Journal of the American Academy of Audiology* 24.8 (Sept. 2013), pp. 649–659. ISSN: 1050-0545. DOI: 10.3766/JAAA.24.8.2.

[78] Asri Maharani et al. "Longitudinal Relationship Between Hearing Aid Use and Cognitive Function in Older Americans". In: *Journal of the American Geriatrics Society* 66.6 (July 2018), pp. 1130–1136. ISSN: 00028614. DOI: 10.1111/jgs.15363.

[79] Elina M. Mäki-Torkko, Martti J. Sorri, and Einar Laukli. "Objective assessment of hearing aid use". In: *Scandinavian Audiology* 30.1 (Jan. 2001), pp. 81–82. ISSN: 0105-0397. DOI: 10.1080/010503901300007146.

[80] Jeremy Marozeau and Mary Florentine. "Loudness growth in individual listeners with hearing losses: A review". In: *The Journal of the Acoustical Society of America* 122.3 (Sept. 2007), EL81–EL87. ISSN: 0001-4966. DOI: 10.1121/1.2761924.

[81] Abby McCormack and Heather Fortnum. "Why do people fitted with hearing aids not wear them?" In: *International Journal of Audiology* 52 (2013), pp. 360–368. ISSN: 14992027. DOI: 10.3109/14992027.2013.769066.

[82] David McShefferty, William M. Whitmer, and Michael A. Akeroyd. "The Just-Meaningful Difference in Speech-to-Noise Ratio". In: *Trends in Hearing* 20 (Jan. 2016). ISSN: 23312165. DOI: 10.1177/2331216515626570.

[83] David McShefferty, William M. Whitmer, and Michael A. Akeroyd. "The just-noticeable difference in speech-to-noise ratio". In: *Trends in Hearing* 19 (Jan. 2015). ISSN: 23312165. DOI: 10.1177/2331216515572316.

[84] Brian C. J. Moore, Christian Füllgrabe, and Michael A. Stone. "Effect of spatial separation, extended bandwidth, and compression speed on intelligibility in a competing-speech task". In: *The Journal of the Acoustical Society of America* 128.1 (July 2010), p. 360. ISSN: 0001-4966. DOI: 10.1121/1.3436533.

[85] Brian C. J. Moore and Chin-Tuan Tan. "Perceived naturalness of spectrally distorted speech and music". In: *The Journal of the Acoustical Society of America* 114.1 (July 2003), pp. 408–419. ISSN: 0001-4966. DOI: 10.1121/1.1577552.

[86] H. Gustav Mueller, Jennifer Weber, and Benjamin W.Y. Hornsby. "The Effects of Digital Noise Reduction on the Acceptance of Background Noise". In: *Trends in Amplification* 10.2 (2006), pp. 83–93. ISSN: 19405588. DOI: 10.1177/1084713806289553.

[87] Cynthia D. Mulrow. "Quality-of-Life Changes and Hearing Impairment". In: *Annals of Internal Medicine* 113.3 (Aug. 1990), p. 188. ISSN: 0003-4819. DOI: 10.7326/0003-4819-113-3-188.

[88] Frank E. Musiek et al. "Perspectives on the Pure-Tone Audiogram". In: *Journal of the American Academy of Audiology* 28.07 (July 2017), pp. 655–671. ISSN: 1050-0545. DOI: 10.3766/jaaa.16061.

[89] National Institute on Deafness and Other Communication Disorders. *Cochlear Implants*. 2021. URL: https://www.nidcd.nih.gov/health/cochlear-implants.

[90] National Institute on Deafness and Other Communication Disorders. *Hearing Aids*. 2017. URL: https://www.nidcd.nih.gov/health/hearing-aids.

[91] National Institute on Deafness and Other Communication Disorders. *Quick Statistics About Hearing*. 2021. URL: https://www.nidcd.nih.gov/health/statistics/quick-statistics-hearing.

[92] National Institutes of Health (US). *Information about Hearing, Communication, and Understanding*. 2007. URL: https://www.ncbi.nlm.nih.gov/books/NBK20366/.

[93] Peggy B. Nelson et al. "Self-Adjusted Amplification Parameters Produce Large Between-Subject Variability and Preserve Speech Intelligibility". In: *Trends in Hearing* 22 (Jan. 2018), pp. 1–13. ISSN: 2331-2165. DOI: 10.1177/2331216518798264.

[94] Brittne Nelson Kakulla. *2020 Tech Trends of the 50+*. Tech. rep. Washington, DC: AARP, Jan. 2020. DOI: 10.26419/res.00329.001.

[95] Elaine Hoi Ning Ng et al. "Effects of noise and working memory capacity on memory processing of speech for hearing-aid users". In: *International Journal of Audiology* 52.7 (2013), pp. 433–441. ISSN: 14992027. DOI: 10.3109/14992027.2013.776181.

[96] Dirk Oetting et al. "Restoring Perceived Loudness for Listeners With Hearing Loss". In: *Ear & Hearing* 39.4 (July 2018), pp. 664–678. ISSN: 0196-0202. DOI: 10.1097/AUD.0000000000000521.

[97] Bolajoko O. Olusanya, Adrian C. Davis, and Howard J. Hoffman. "Hearing loss grades and the International classification of functioning, disability and health". In: *Bulletin of the World Health Organization* 97.10 (Oct. 2019), pp. 725–728. ISSN: 0042-9686. DOI: 10.2471/BLT.19.230367.

[98] Oticon Inc. *Where do our prescribed values come from?* Aug. 2017. URL: https://www.oticon.com/professionals/audiology-students/education-and-information/fitting-tips/tips/2017/august-2017---fitting-tip.

[99] Catherine Palmer, Ruth Bentler, and Gustav H. Mueller. "Evaluation of a Second-Order Directional Microphone Hearing Aid: II. Self-Report Outcomes". In: *Journal of the American Academy of Audiology* 17.03 (Mar. 2006), pp. 190–201. ISSN: 1050-0545. DOI: 10.3766/jaaa.17.3.5.

[100] Alessandro Pasta et al. "Clustering Users Based on Hearing Aid Use: An Exploratory Analysis of Real-World Data". In: *Frontiers in Digital Health* 3 (Sept. 2021), p. 109. ISSN: 2673-253X. DOI: 10.3389/fdgth.2021.725130.

[101] Alessandro Pasta et al. "Designing audiologist bots fusing soundscapes and user feedback". In: *Workshop on Conversational Agents for Health and Wellbeing*. Honolulu, 2020.

[102] Alessandro Pasta et al. *Investigating the Provision and Context of Use of Hearing Aid Listening Programs from Real-World Data*. 2021.

[103] Alessandro Pasta et al. *Measuring and Modeling Context-Dependent Preferences for Hearing Aid Settings*. 2021.

[104] Alessandro Pasta et al. "Rethinking hearing aids as recommender systems". In: *The 4th Workshop on Health Recommender Systems co-located with ACM RecSys'19*. Vol. 2439. Copenhagen, Denmark, Sept. 2019, pp. 11–17. URL: http://ceur-ws.org/Vol-2439/4-paginated.pdf.

[105] Jonathan E. Peelle and Arthur Wingfield. "The neural consequences of age-related hearing loss". In: 39.7 (2016), pp. 486/497. DOI: 10.1016/j.tins.2016.05.001.

[106] Elvira Perez and Barrie A. Edmonds. *A systematic review of studies measuring and reporting hearing aid usage in older adults since 1999: A descriptive summary of measurement tools*. Mar. 2012. DOI: 10.1371/journal.pone.0031831.

[107] Erin M. Picou. "MarkeTrak 10 (MT10) Survey Results Demonstrate High Satisfaction with and Benefits from Hearing Aids". In: *Seminars in Hearing* 41.01 (Feb. 2020), pp. 021–036. ISSN: 0734-0451. DOI: 10.1055/s-0040-1701243.

[108] Jaydip Ray, Gurleen Popli, and Greg Fell. "Association of Cognition and Age-Related Hearing Impairment in the English Longitudinal Study of Ageing". In: *JAMA Otolaryngology–Head & Neck Surgery* 144.10 (Oct. 2018), p. 876. ISSN: 2168-6181. DOI: 10.1001/jamaoto.2018.1656.

[109] T. A. Ricketts and R. A. Bentler. "Comparison of two digitally programmable hearing aids". In: *Journal of the American Academy of Audiology* 3.2 (Mar. 1992), pp. 101–12. ISSN: 1050-0545. URL: https://europepmc.org/article/med/1600212%20http://www.ncbi.nlm.nih.gov/pubmed/1600212.

[110] Todd Ricketts, Paula Henry, and David Gnewikow. "Full Time Directional versus User Selectable Microphone Modes in Hearing Aids". In: *Ear and Hearing* 24.5 (Oct. 2003), pp. 424–439. ISSN: 0196-0202. DOI: 10.1097/01.AUD.0000094555.89110.0A.

[111] Jorge Rodríguez et al. "Cluster validation using an ensemble of supervised classifiers". In: *Knowledge-Based Systems* 145 (Apr. 2018), pp. 134–144. ISSN: 09507051. DOI: 10.1016/j.knosys.2018.01.010.

[112] Thomas Rossing. *Springer Handbook of Acoustics*. Ed. by Springer. 2007, p. 747. ISBN: 9780387304465.

[113] Johanna Elisabeth Ruusuvuori et al. "Studies on stigma regarding hearing impairment and hearing aid use among adults of working age: a scoping review". In: *Disability and Rehabilitation* 43.3 (Jan. 2021), pp. 436–446. ISSN: 0963-8288. DOI: 10.1080/09638288.2019.1622798.

[114] Donald J Schum. *Fitting Rationale*. Mar. 2017. URL: https://www.oticon.com/professionals/audiology-students/education-and-information/fitting-tips/tips/2017/032017--fitting-rationale.

[115] Lu Feng Shi et al. "Short-term and long-term hearing aid benefit and user satisfaction: A comparison between two fitting protocols". In: *Journal of the American Academy of Audiology* 18.6 (2007), pp. 482–495. ISSN: 10500545. DOI: 10.3766/jaaa.18.6.3.

[116] Niels Søgaard Jensen et al. "Perceptual Effects of Adjusting Hearing-Aid Gain by Means of a Machine-Learning Approach Based on Individual User Preference". In: *Trends in hearing* 23 (Jan. 2019), pp. 1–23. ISSN: 2331-2165. DOI: 10.1177/2331216519847413.

[117] Jorunn Solheim and Louise Hickson. "Hearing aid use in the elderly as measured by datalogging and self-report". In: *International Journal of Audiology* 56.7 (July 2017), pp. 472–479. ISSN: 1499-2027. DOI: 10.1080/14992027.2017.1303201.

[118] Katharina Staehelin et al. "Gender and Hearing Aids: Patterns of Use and Determinants of Nonregular Use". In: *Ear & Hearing* 32.6 (Nov. 2011), e26–e37. ISSN: 0196-0202. DOI: 10.1097/AUD.0b013e3182291f94.

[119] Patricia G. Stelmachowicz, Dawna E. Lewis, and Edward Carney. "Preferred Hearing-Aid Frequency Responses in Simulated Listening Environments". In: *Journal of Speech, Language, and Hearing Research* 37.3 (June 1994), pp. 712–718. ISSN: 1092-4388. DOI: 10.1044/jshr.3703.712.

[120] Gretchen Stevens et al. "Global and regional hearing impairment prevalence: an analysis of 42 studies in 29 countries". In: *European Journal of Public Health* 23.1 (Feb. 2013), pp. 146–152. ISSN: 1101-1262. URL: https://academic.oup.com/eurpub/article/23/1/146/460112.

[121] Gail Takahashi et al. "Subjective Measures of Hearing Aid Benefit and Satisfaction in the NIDCD/VA Follow-Up Study". In: *Journal of the American Academy of Audiology* 18.04 (Apr. 2007), pp. 323–349. ISSN: 1050-0545. DOI: 10.3766/jaaa.18.4.6.

[122] Lisa B. Taubman et al. "Accuracy of Hearing Aid Use Time as Reported by Experienced Hearing Aid Wearers". In: *Ear and Hearing* 20.4 (Aug. 1999), pp. 299–305. ISSN: 0196-0202. DOI: 10.1097/00003446-199908000-00003.

[123] Barbra H B Timmer, Louise Hickson, and Stefan Launer. "Hearing Aid Use and Mild Hearing Impairment: Learnings from Big Data". In: *Journal of the American Academy of Audiology* 28.08 (Sept. 2017), pp. 731–741. ISSN: 1050-0545. DOI: 10.3766/jaaa.16104.

[124] Dana P. Turner. "Experimental Study Designs". In: *Headache: The Journal of Head and Face Pain* 60.8 (Sept. 2020), pp. 1501–1507. ISSN: 0017-8748. DOI: 10.1111/head.13928.

[125] Dana P. Turner and Timothy T. Houle. "Observational Study Designs". In: *Headache: The Journal of Head and Face Pain* 59.7 (July 2019), pp. 981–987. ISSN: 0017-8748. DOI: 10.1111/head.13572.

[126] Anne Viljanen et al. "Hearing as a Predictor of Falls and Postural Balance in Older Female Twins". In: *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 64A.2 (Feb. 2009), pp. 312–317. ISSN: 1079-5006. DOI: 10.1093/gerona/gln015.

[127] Brian E. Walden et al. "Predicting Hearing Aid Microphone Preference in Everyday Listening". In: *Journal of the American Academy of Audiology* 15.05 (May 2004), pp. 365–396. ISSN: 1050-0545. DOI: 10.3766/jaaa.15.5.4.

[128] Jennifer Junnila Walker et al. "Audiometry Screening and Interpretation". In: *American Family Physician* 87.1 (2013), pp. 41–47. ISSN: 15320650.

[129] Margaret I Wallhagen et al. "Comparative impact of hearing and vision impairment on subsequent functioning." In: *Journal of the American Geriatrics Society* 49.8 (Aug. 2001), pp. 1086–92. ISSN: 0002-8614. DOI: 10.1046/j.1532-5415.2001.49213.x.

[130] Els Walravens, Gitte Keidser, and Louise Hickson. "Consistency of Hearing Aid Setting Preference in Simulated Real-World Environments: Implications for Trainable Hearing Aids". In: *Trends in Hearing* 24 (2020), pp. 1–19. ISSN: 23312165. DOI: 10.1177/2331216520933392.

[131] Dorothea Wendt, Renskje K. Hietkamp, and Thomas Lunner. "Impact of Noise and Noise Reduction on Processing Effort". In: *Ear and Hearing* 38.6 (2017), pp. 690–700. ISSN: 0196-0202. DOI: 10.1097/aud.0000000000000454.

[132] Lena L. N. Wong, Louise Hickson, and Bradley McPherson. "Hearing Aid Satisfaction: What Does Research from the Past 20 Years Say?" In: *Trends in Amplification* 7.4 (Jan. 2003), pp. 117–161. ISSN: 1084-7138. DOI: 10.1177/108471380300700402.

[133] World Health Organization. *Deafness and hearing loss*. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss.

[134] World Health Organization. "Hearing aids and services for developing countries". In: *Pan American Journal of Public Health (PAHO)* 9.3 (Mar. 2001), pp. 139–42. URL: http://www.paho.org/English/DBI/es/v10n2-TEMA-Hearing.pdf.

[135] World Health Organization. *World report on hearing*. Tech. rep. Geneva, 2021. URL: https://www.who.int/publications/i/item/world-report-on-hearing.

[136] Humberto Yévenes-Briones et al. "Association Between Hearing Loss and Impaired Physical Function, Frailty, and Disability in Older Adults". In: *JAMA Otolaryngology–Head & Neck Surgery* 147.11 (Nov. 2021), p. 951. ISSN: 2168-6181. DOI: 10.1001/jamaoto.2021.2399.

[137] Sung Hoon Yoon et al. "A trainable hearing Aid Algorithm reflecting individual preferences for degree of noise-suppression, input sound level, and listening situation". In: *Clinical and Experimental Otorhinolaryngology* 10.1 (2017), pp. 56–65. ISSN: 20050720. DOI: 10.21053/ceo.2015.01690.