Clinical auditory profiling and profile-based hearing-aid fitting

Sanchez Lopez, Raul

Publication date:
2020

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

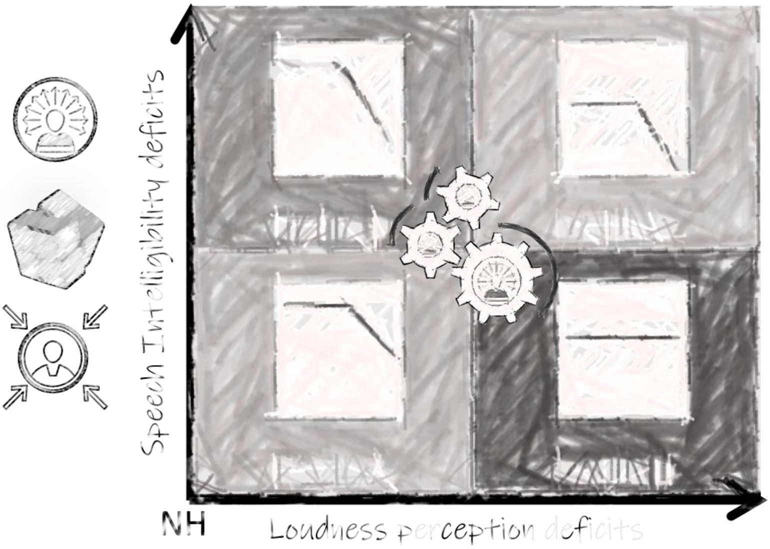
Citation (APA):
Sanchez Lopez, R. (2020). *Clinical auditory profiling and profile-based hearing-aid fitting*. DTU Health Technology. Contributions to Hearing Research Vol. 43

CONTRIBUTIONS TO
HEARING RESEARCH

Volume 43

Raul Sanchez-Lopez

Clinical auditory profiling and profile-based hearing-aid fitting



DTU Health Tech
Department of Health Technology

Clinical auditory profiling and profile-based hearing-aid fitting

PhD thesis by
Raul Sanchez-Lopez

Submitted version: May 31, 2020

HEARING SYSTEMS

Technical University of Denmark

2020

This PhD dissertation is the result of a research project carried out at the Hearing Systems Section, Department of Health Technology, Technical University of Denmark. Part of the project was carried out at the Institute of Clinical Research at the University of Southern Denmark (Odense, Denmark), Odense University Hospital (Odense, Denmark), Department of Otorhinolaryngology, Head and Neck Surgery & Audiology at Rigshospitalet (Copenhagen, Denmark) and the Hearing Sciences Scotish Section at the University of Nottingham (Glasgow, Scotland, UK).

The project was partly financed by Innovation Fund Denmark Grand Solutions 5164-00011B (”Better Hearing Rehabilitation” project) (61%) the industrial partners of the BEAR project: GN Hearing A/S, Oticon A/S and WSAudiology A/S (32%) and the Technical University of Denmark (7%).

Supervisors

Main supervisor

Prof. Torsten Dau
Hearing Systems Section
Department of Health Technology
Technical University of Denmark
Kgs. Lyngby, Denmark

Co-supervisors

Assoc. Prof. Sébastien Santurette
Dr. Michal Fereczkowski
Dr. Federica Bianchi
Hearing Systems Section
Department of Health Technology
Technical University of Denmark
Kgs. Lyngby, Denmark

External Advisor

Assoc. Prof. Tobias Neher
Institute of Clinical Research
Faculty of Health Sciences
University of Southern Denmark
Odense, Denmark

... to the intelligent ear

Abstract

In this thesis, the basis for "precision audiology" was explored. The prerequisites for implementing precision treatments are 1) that the diseases must be heterogenous, 2) that there exist multiple options for treatment and 3) that there are "markers" that associate certain characteristics of the patient to specific treatments. The sources and consequences of a sensorineural hearing loss (SNHL) are diverse and the hearing devices, especially hearing aids, have multiple configurations that can be adjusted for specific needs. The present thesis focused on the investigation of auditory biomarkers that allow the link between auditory perceptual deficits and hearing-aid settings. Data-driven auditory profiling has the potential to identify hidden patterns in the data and subpopulations of hearing-impaired people with distinct differences in terms of their perceptual hearing deficits.

The first study of this thesis showed that a method for auditory profiling can provide meaningful results. The results supported the hypothesis of having four profiles along two independent dimensions of perceptual deficits resulting from different auditory distortions. However, the main limitation for drawing strong conclusions was the selection of the analyzed datasets. In the second study, a new test battery for characterizing auditory deficits was proposed and tested on 75 listeners with various degrees of hearing loss and speech discrimination. The test battery assessed the auditory processing abilities of the listeners covering aspects such as audibility, loudness perception, binaural processing abilities, speech perception, spectro-temporal modulation sensitivity and spectro-temporal resolution. The dataset was then analyzed with an iterative data-driven profiling method based on the aforementioned profiling method. This robust auditory profiling yielded four clinically relevant subgroups of listeners. Importantly, the results were consistent with previous approaches of hearing loss characterization leading to the following conclusions: 1) The listeners' hearing deficits were characterized by two independent auditory distortions, a “speech intelligibility related distortion”, that affected listeners with audiometric thresholds above 50 dB hearing level (HL) at high-frequencies, and a “loudness perception related distortion”, exhibited by listeners with audiometric thresholds above 30 dB HL at low frequencies; and 2)

The four profiles (A-B-C-D), showed similarities to the audiometric phenotypes provided by Dubno et al. (2013) suggesting that Profile B might be considered a sensory loss and Profile D a metabolic loss.

Finally, a proof-of-concept study was performed, where listeners evaluated different compensation strategies using a realistic hearing-aid simulator. The results suggested that listeners belonging to different profiles might prefer different compensation strategies. Listeners with high degree of loudness-related deficits might benefit from different forms of gain prescription, whereas listeners with speech intelligibility-related deficits might benefit from signal-to-noise ratio improvement. Different approaches for precision audiology may be implemented in the near future, which should have implications for hearing-aid development, hearing loss characterization and the quality of service in the hearing-care clinic towards an evidence-based audiological practice.

Resumé

I denne afhandling blev grundlaget for "præcisions audiologi"undersøgt. Forudsætningerne for implementering af præcisions-behandling er 1), at sygdommene er heterogene, 2) at der findes flere behandlingsmuligheder, og 3) at der er "markører", der forbinder visse egenskaber hos patienten til specifikke behandlinger. Årsagerne til og konsekvenserne af et sensorineural høretab (SNHT) er forskellige, og hørehjælpemidler, især høreapparater, har flere konfigurationer, der kan tilpasses specifikke behov. Denne afhandling fokuserede på undersøgelse af auditive biomarkører, der tillader sammenhængen mellem perceptuelle mangler og høreapparatindstillinger. Datadrevet auditiv profilering har potentialet til at identificere skjulte mønstre i data samt underpopulationer af hørehæmmede med tydelige forskelle med hensyn til deres perceptuelle høretab.

Den første undersøgelse i denne afhandling viste, at en metode til auditiv profilering kan give meningsfulde resultater. Resultaterne bekræfter hypotesen om at have fire profiler langs to uafhængige dimensioner af perceptuelle høretab, der skyldes forskellige auditive forvrængninger. Ikke desto mindre, var en hovedbegrænsning i at drage stærke konklusioner valget af det analyserede datasæt. I den anden undersøgelse blev et nyt testbatteri, der karakteriserer høretab udarbejdet og testet på 75 lyttere med forskellige grader af høretab og talediskrimineringsevne. Testbatteriet testede lytternes auditive bearbejdningsevner, der dækker aspekter som hørbarhed, lydstyrkeopfattelse, binaural behandlingsevne, taleforståelse, spektro-temporal moduleringssensitivitet og spektro-tidsmæssig opløsning. Datasættet blev derefter analyseret med en iterativ datadrevet profileringsmetode baseret på den førnævnte profileringsmetode. Denne robuste auditive profilering gav fire klinisk relevante undergrupper af lyttere. Resultaterne var i overensstemmelse med tidligere studier og førte til følgende konklusioner: 1) Lytterenes høretab var karakteriseret ved to uafhængige auditive forvrængninger, en "taleforståelhedsrelateret forvrængning", der påvirkede lyttere med audiometriske tærskler over 50 dB høreniveau (HL) ved høje frekvenser og en "lydstyrke-opfattelsesrelateret forvrængning"forekommende hos lyttere med audiometriske tærskler over 30 dB HL ved lave frekvenser; og 2) De fire profiler (A-B-C-D) viste ligheder med de audiometriske fænotyper foreslået af Dubno et al. (2013), hvilket antydede, at Profil B kan betragtes som et sensorisk tab

og Profil D som et metabolisk tab.
Endelig blev der udført en proof-of-concept-undersøgelse, hvor lyttere evaluerede forskellige kompensationsstrategier ved hjælp af en realistisk høreapparatssimulator. Resultaterne antydede, at lyttere, der hører til forskellige profiler, muligvis foretrækker forskellige kompensationsstrategier. Lyttere med stor grad af lydstyrkerelaterede høreproblemer kan drage fordel af forskellige former for forstærkning, mens lyttere med taleforståelsesrelaterede høreproblemer muligvis kan drage fordel af forbedring af signal-til-støjforholdet. Forskellige tilgange til præcisions audiologi kan implementeres i den nærmeste fremtid, hvilket bør have konsekvenser for udviklingen af høreapparater, karakterisering af høretab og kvaliteten af servicen i høreapparatklinikken imod en evidensbaseret audiologisk praksis.

Acknowledgments

A PhD thesis is an important achievement for any researcher. The thesis is not only "work", but it contains a lot of ourselves. Writing these *Acknowledgments*, I have revisited many important moments and people who have shaped this work and the way I am as a researcher and as a "technical audiologist". In 2015, I became a Hearing Care Professional (HCP). From that time, a found my passion in the understanding of the hearing loss its remediation. Some people such as J.L. Blanco, E. Avila, S. Alvarez and R. Borrego, and the patients that I saw during my time as a HCP contributed to that in those first steps.

At the "House of Acoustics", I have shared "space" and "time" with interesting people, nice colleagues, visitors, and collected a ton of memories. I would like to thank: M.L. Jepsen for the opportunity to do my MSc Thesis with him at Widex. It was an important step in my career where I learned a lot in an intense semester. The Hearing Systems (Section), with all its current and former members. I want to give special thanks to Caroline for her great support. J. Zaar and G. Encina for countless spontaneous interesting discussions. My "fellow-friends" (Axel, Helia, Antoine, Andreu, Borys, Wiebke,...) , Charlotte and Andrew and my office-mates. Also, the colleagues from Acoustic Technology, and the Interacoustics Research Unit.

During the PhD I had the chance to attend many events and conferences. I want to thank the Hearing Science Scottish Section for your inspiration, especially B Whitmer, G Naylor and L Hadley. I really enjoyed my short time there and I appreciate your warm hospitality. I also thank the DTAS and AEDA communities and the people that run these associations. Besides, I have met some researchers who inspired me, not necessarily in aspects related to this work, but about "research" in general, for example, L Carney, M. Schädler, C. Hauth, W. Nogueira, I. Holube, P. von Gablenz, T. Jürgens, P. Johannesen, etc.

The BEAR is a big animal and I have had the chance to develop, not only my research skills but also other valuable skills. Besides, I have interacted with many interesting people and in close collaboration with colleagues from academia, industry, and clinics. I appreciate the work of the initiators of the "Better hEAring Rehabilitation" project, D. Hammershøi, N. Bisgaard and all the people that are involved in different activities. I can't write all the names here. Some parts of this work could not be possible without Silje G Nielsen and Mouhamad El-Haj-Ali, you did an excellent work. Also, the participants and volunteers in these studies. I also thank O. Cañete, M. Wu and J.B. Nielsen.

I deeply acknowledge the "plot twists" of my PhD (i.e., the moments when I changed my perception of the preceding ideas). These twists made me reconsider some of my conclusions and improved substantially my research. These happened after giving a presentation, a discussion in front of a poster, in a meeting, or even a short chat before a meeting. Examples of that, happened with Bernd T. Meyer, J. Harte, G. Naylor, and O. Strelcyk among others. There were also "plot twists" on the other end, which made substantial changes in my plans. I must say that all these moments have contributed to this thesis and I appraise any single one of them.

With special affection, I want to thank my supervisors . F. Bianchi for being always caring, constructive, and consistent. You were crucial in my first year of the PhD. M. Fereczkowski for countless interesting discussing sometimes beyond the scope of the PhD, I have enjoyed working with you. S. Santurette, for your trust and advice, I appreciate your supervision style. T. Neher for being an unofficial supervisor, for your involvement, perseverance, and important critical view. This thesis would be different without you. And of course, T. Dau, for your trust, respect, and enthusiasm. I could always be honest and critical in different aspects and you have always responded. I appreciate the last intense months.

I would like to acknowledge with gratitude the help of Marta ("naming" things and giving to this thesis the importance it deserves). Thanks to my friends here and there and my family including the ones we miss. And last but not least, I want to thank *Ariadni* because you have followed all the process of my PhD. For your support, help, patience and for this nice time being office-mates, while the pandemic invites us to reconsider many things. Gracias!

Raúl H Sánchez López, May 31, 2020.

Related publications

Journal papers

- Sanchez Lopez, R., Bianchi, F., Fereczkowski, M., Santurette, S., and Dau, T. (**2018**). "Data-Driven Approach for Auditory Profiling and Characterization of Individual Hearing Loss." Trends in Hearing, **22**, 12. [2331216518807400]. `https://doi.org/10.1177/2331216518807400`
- Sanchez-Lopez, R., Nielsen, S. G., El-Haj-Ali, M., Bianchi, F., Fereckzowski, M., Cañete, O., Wu, M., Neher, T., Dau, T. and Santurette, S. (**2020**). "Auditory tests for characterizing hearing deficits: The BEAR test battery". Submitted to International Journal of Audiology.
  Preprint available in medRxiv from: `https://doi.org/10.1101/2020.02.17.20021949`
- Sanchez-Lopez, R., Fereczkowski, M., Neher, T., Santurette, S., and Dau, T. (**2020**). "Robust data-driven auditory profiling for precision audiology." Submitted to Trends in Hearing .
  Preprint available in medRxiv from: `https://doi.org/10.1101/2020.04.09.20036442`
- Sanchez-Lopez, R., Fereczkowski, M., Santurette, S., Dau, T., and Neher, T. (**2020**). "Auditory Profile-based Hearing-aid Fitting: A Proof-of-concept study." Submitted to Ear and Hearing.
  Preprint available in medRxiv from: `https://doi.org/10.1101/2020.04.14.20036459`
- Sanchez-Lopez, R., Dau, T., and Whitmer, W. M. (**2020**). "Audiometric profiles and patterns of benefit. A data-driven analysis of subjective hearing difficulties and handicaps." Submitted to International Journal of Audiology.

Preprint available in medRxiv from: https://doi.org/10.1101/2020.04.20.20045690

Conference papers

- Sanchez Lopez, R., Bianchi, F., Fereczkowski, M., Santurette, S., and Dau, T. (**2017**). Data-driven approach for auditory profiling." In S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen (Eds.), Proceedings of the International Symposium on Auditory and Audiological Research: Vol. 6: Adaptive Processes in Hearing The Danavox Jubilee Foundation.
- Sanchez Lopez, R., Fereczkowski, M., Bianchi, F., Piechowiak, T., Hau, O., Pedersen, M. S., Behrens, T., Neher, T., Dau, T., Santurette, S. (**2018**). Technical evaluation of hearing-aid fitting parameters for different auditory profiles." In Euronoise 2018 - Conference Proceedings (pp. 381-388). European Acoustics Association.
- Sanchez Lopez, R., Nielsen, S. G., Cañete, O., Fereczkowski, M., Wu, M., Neher, T., Dau, T., and Santurette, S. (**2019**). "A clinical test battery for Better hEAring Rehabilitation (BEAR): Towards the prediction of individual auditory deficits and hearing-aid benefit," In Proceedings of the 23rd International Congress on Acoustics (pp. 3841-3848). Deutsche Gesellschaft für Akustik e.V. https://doi.org/10.18154/RWTH-CONV-239177
- Wu, M., Sanchez Lopez, R., El-Haj-Ali, M., Nielsen, S. G., Fereczkowski, M., Dau, T., Santurette, S., and Neher, T. (**2019**). Assessing the interaction between different auditory profiles and benefit from six hearing aid processing strategies: Insights from the BEAR project." In Proceedings of the 23rd International Congress on Acoustics (pp. 3849-3856). Deutsche Gesellschaft für Akustik e.V. https://doi.org/10.18154/RWTH-CONV-239059
- Wu, M., Sanchez-Lopez, R., El-Haj-Ali, M., Nielsen, S.G., Fereczkowski, M., Dau, T., Santurette, S., and Neher, T. (**2019**) Perceptual evaluation of six hearing-aid processing strategies from the perspective of auditory profiling: Insights from the BEAR project " In S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, & T. Poulsen (Eds.), Proceedings of the International Symposium on Auditory and Audiological Research: Vol. 7: Auditory Learning in Biological and Artificial Systems. The Danavox Jubilee Foundation.

Datasets and repositories

- Sanchez-Lopez, R., Nielsen, S. G., El-Haj-Ali, M., Bianchi, F., Fereckzowski, M., Cañete, O., … Santurette, S. (**2019**). Data from "Auditory tests for characterizing hearing deficits: The BEAR test battery." [Zenodo]. `https://doi.org/10.5281/zenodo.3459579`
- Sanchez-Lopez, R., Fereczkowski, M., Santurette, S., Dau, T., Neher, T. (**2020**). Data and materials from: "Auditory Profile-based Hearing-aid Fitting: A Proof-of-concept study." [Zenodo]. `https://doi.org/10.5281/zenodo.3732320`
- Sanchez-Lopez, R., Nielsen, S. G. (**2019**). BEAR Test Battery. [Bitbucket repository]. `https://bitbucket.org/hea-dtu/bear-test-battery/src/master/bear-clinical-test-battery/`

Contents

1

General introduction

> *"For millions of years, mankind lived just like the animals. Then something happened which unleashed the power of our imagination. We learned to talk and we learned to listen. Speech has allowed the communication of ideas, enabling human beings to work together to build the impossible. Mankind's greatest achievements have come about by talking, and its greatest failures by not talking. It doesn't have to be like this. Our greatest hopes could become reality in the future. With the technology at our disposal, the possibilities are unbounded. All we need to do is make sure we keep talking."*
>
> *Stephen Hawking (1942-2018)*

Communication is an important ability that shapes our social life. Communication allows us to express complex thoughts, ideas and feelings, and defines our capacity to develop relationships. Oral communication is a co-operative process that involves speaking and listening. Speech is a complex signal that varies in frequency and time producing different sounds. The listener's ears decode these particular sounds into vowels and consonants and interpret them as meaningful words and sentences. A simple conversation constitutes a combination of different complex processes. For example, "listening" consists of not only audition (i.e. the transformation of the acoustic stimuli into informative sensations) but also cognition (i.e. the interpretation of the sensation as significant information) and can be affected by different factors, such as the presence of noise, room reflections or other distortions in the transmission line (Plomp, 2001). An important factor that affects the successful oral communication is hearing loss. If the verbal information is not properly perceived, the interpretation can be flawed leading to an unsatisfactory and probably tiring and uncomfortable conversation. Besides, a hearing loss entails significant activity limitations and participation restrictions (Simeonsson, 2000; Wilson et al., 2017).

The World Health Organization estimated that 466 million people live nowadays with a disabling hearing loss and one-third of older adults (>65 years) are affected by this problem (WHO, 2018). Hearing loss is currently among the top ten of the global burdens of diseases[a] (Graydon et al., 2019). The associated costs of the untreated hearing loss can entail an important impact on society in the following years. This is, in part, because of the aging population and the impact on the quality of life of individuals with hearing loss in general. Furthermore, recent research associates the undertreated hearing loss with a higher risk for developing dementia and cognitive decline (Dawes et al., 2019; Lin et al., 2013). Therefore, it is of great interest to provide high quality hearing rehabilitation in the public and private health care clinics.

1.1 Current hearing health practice

Hearing rehabilitation focuses on the remediation of sensory impairments to reduce hearing difficulties and participation restrictions due to hearing loss. Sensory management implies the use of adequate treatment to optimize auditory function. The intervention often involves instructions for the use of hearing devices, perceptual training, and counseling (Boothroyd, 2007).

The typical clinical flow involves the correct diagnosis and remediation of the hearing deficits (Goldstein and Stephens, 1981). However, the first step occurs when a person with hearing difficulties sets an appointment with a hearing care professional (HCP) and actively seeks for help. This is the starting point of the "patient's journey"(Manchaiah et al., 2011).

---

[a] Burden of disease is concept to describe death and loss of health due to diseases, injuries and risk factors for all regions of the world. [https://tinyurl.com/y77k9cw9]

In the hearing-care clinic, the HCP will carefully examine the person's hearing, but before this, qualitative information of the hearing difficulties is gathered in an informal short interview. The hearing test battery performed in the hearing-care clinic usually starts with a simple inspection of the outer ear with the help of an otoscope. This device allows seeing whether any obstacle is located in the ear canal. Then, a measurement of the middle ear function, a tympanometry, is performed (Jerger et al., 1974). This test evaluates the impedance of the middle ear, which transmits the vibrations of the sound into the cochlea, the "sensory analyzer". Finally, to assess the degree of hearing loss, the pure-tone audiometry is performed. The test consists of the detection of pure tones. The patient is asked to press a button or raise the hand every time the examiner presents a sound. This is done at different levels and for different frequencies until the hearing threshold, the minimum intensity perceived by the patient, is estimated. The pure-tone audiometry is usually performed via headphones (air-conduction) and by using a vibrator placed on the bone behind the pinna (bone-conduction).

The hearing thresholds are annotated on a graph, the so-called "audiogram" (Figure 1.1). The frequency is presented on the horizontal axis. The low frequencies are located on the left of the graph and the high frequencies on the right. In terms of every-day sounds, low-frequency sounds would correspond, for example, to a car's engine and high-frequency sounds would correspond to a birdsong. In terms of speech sounds, vowels (e.g. /a/, /e/, etc.) contain energy mainly at low frequencies, whereas consonants (e.g. /s/, /f/, etc.) contain energy mainly at high frequencies. The level is represented in the vertical axis, in dB hearing level (HL). This unit takes as a reference the average hearing threshold of normal-hearing listeners, depicted in the audiogram as the horizontal line at 0 dB HL. Soft sounds, such as a whisper, are located on the top and loud sounds, like the engine of a plane, are on the bottom. The deviation between the 0 dB HL line and the individual hearing thresholds represents the hearing loss. The air-conduction thresholds (red circles for the right ear) represent the sensitivity of the entire auditory system (i.e. outer, middle and inner ear) to pure tones at different frequencies. The bone conduction thresholds (>) correspond to the sensitivity of

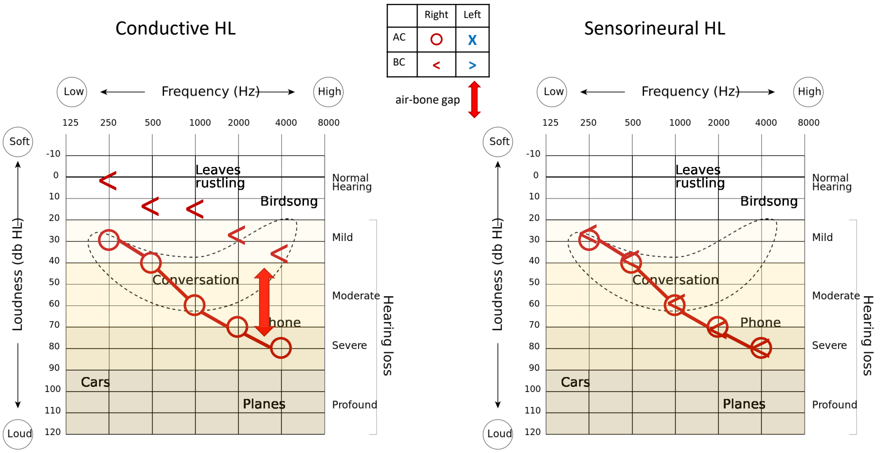


Figure 1.1: Audiograms presenting the resuts of pure-tone audiometry. The graph is annotated with examples daily sounds placed in the graph according to their frequency-level characteristics. The so-called "speech banana" represents the typical frequency-level characteristics of the speech signal. The qualitative description of the degrees of hearing loss (from mild to profound) is shown in the right Y axis. Left panel: conductive hearing loss in the right ear, where the bone-conduction and air-conduction thresholds are not coincing. Right panel: Sensorineural hearing loss with similar air-conduction thresholds as in left panel but with no air-bone gap.

the inner ear, since the outer and middle ear are bypassed by applying the stimuli directly to the cochlea by the mechanical vibrations transmitted through the bone. Figure 1.1 shows two audiograms. The left panel corresponds to a conductive hearing loss in the right ear. The conductive (or transmission) hearing loss is characterized by the difference between the air conduction and bone conduction thresholds (the air-bone gap). The right panel corresponds to a sensorineural hearing loss. This type of hearing loss is attributed to cochlear dysfunction or impairments in the auditory nerve. Therefore, the curves corresponding to the air- and bone-conduction thresholds will lay on similar values since the middle-ear function is normal.

When the audiometric test battery is concluded, the patient receives a diagnosis and professional advice about the possible remediation of the hearing problem. Usually, the type (conductive vs sensorineural) and the degree of the hearing loss are estimated by the audiogram and explained to the patient. To diagnose the hearing loss and investigate its etiology, it is important to consider the patient's medical history which might require further testing for differential diagnosis. This is particularly important when other pathologies are present, the hearing loss has appeared suddenly or together with other symptoms (e.g. vertigo). However, it is often the case that an older adult shows an audiogram that resembles a typical age-related hearing loss or a noise-induced hearing loss (Kujawa and Liberman, 2019). These are sensorineural hearing losses that are irreversible and have an increased prevalence from the age of 65 years. Age-related hearing losses are a major concern for health-related disability (Graydon et al., 2019). In those cases, hearing aids are the most common solution to remediate the hearing loss. These devices capture and process the incoming sound by applying amplification and advanced signal processing such as noise suppression. The selection of the hearing aid is a task that the HCP and the patient do together, considering different advantages and disadvantages of the distinct styles of hearing aids (behind-the-ear, in-the-ear) and the available technology.

The patient's journey continues when the hearing-aid and the custom ear-mold (an earpiece fabricated based on the ear impression of the patient) are ready to be fitted to the patient. The hearing-aid fitting is usually guided by a fitting software that integrates the audiometry of the patient and the characteristics of the hearing aid to adjust the hearing-aid parameters to a target gain prescription (e.g. NAL; Keidser et al., 2011). This target prescription must be verified by electroacoustic measurements usually done in the real ear. Once the initial fit is verified, then the patient receives instructions about their new hearing aids and some professional advice.

One would think that the patient journey has concluded here, but this is not necessarily the case. The patient is sent into the real world with a “new way of hearing” that requires an “acclimatization” process (Gatehouse, 1993) that is affected by the auditory ecology of the patient (i.e. the patient's frequent sound environments; Gatehouse et al., 1999). Ideally, this period is overpassed with no major challenges. However, it is often the case that the patient needs additional follow-up visits where the hearing aid needs to be readjusted in a trial-and-error process. Besides, other aspects of hearing rehabilitation beyond the sensory management (e.g. auditory training or counseling) might be considered as part of the intervention. Although the HCP strives to provide the best hearing rehabilitation spending several follow-up sessions, sometimes the hearing intervention remains suboptimal, leading to unsatisfied patients who do not find an adequate solution to their hearing problems. This unfortunate situation might happen because sensorineural hearing loss is complex and similar audiograms can be caused by different impaired mechanisms (Wong and Ryan, 2015). Therefore, the reason for a suboptimal rehabilitation might be that the sensory management is exclusively based on the audibility loss (e.g. the audiogram) and not on the perceptual deficits at supra-threshold sound levels.

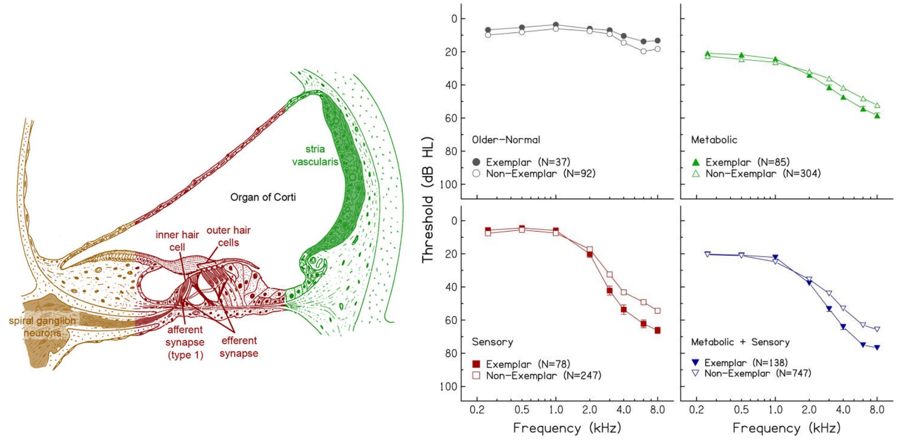
The present thesis focuses on exploring a novel approach to characterize the perceptual auditory deficits of the sensorineural hearing loss. In an ideal scenario, a complementary test battery, able to examine the auditory function at supra-threshold levels, can be implemented in the clinical hearing-care practice and used to enhance the efficacy of the hearing-aid fitting process.

1.2 The complexity of sensorineural hearing loss

Sensorineural hearing loss is caused by cochlear dysfunction or by the loss of neural fibers in the auditory nerve (Wong and Ryan, 2015). The auditory receptor is in a structure called ‘organ of Corti’ placed on the basilar membrane along the cochlea (Figure 1.2a). The Organ of Corti consists of two types of hair cells: the outer hair cells, which are activated by the mechanical vibrations of the basilar membrane; and the inner hair cells that are depolarized, in part, by the action of the outer hair cells and produce the neural activation of the auditory nerve. When an acoustic signal is delivered into the ear, it is transmitted through the outer and middle ear into the inner ear, where the basilar membrane vibrates in response to the incoming sound. The basilar membrane has tonotopic properties, which means that motion is spatially distributed along the membrane depending on the frequency content of the incoming signal, operating as a frequency analyzer. In a healthy organ of Corti, the outer hair cells act as a "cochlear amplifier" that emphasizes the basilar membrane motion at low input levels, whereas the inner hair cells act as a transducer that transforms the mechanical vibrations into receptor potentials. A cochlear hearing loss occurs when there is a loss of hair cells (sensory loss) or when there is a cochlear dysfunction often associated with metabolic processes that affect the cochlear amplifier (Mills et al., 2006). A retrocochlear hearing loss is caused either by a loss of neuro-fibers, a dysfunctional auditory nerve or central auditory lesions (Lidén and Korsan-Bengtsen, 1973). This can produce a loss of audibility, but it is typically associated with impairments that are not reflected in the detection but in the discrimination of the incoming sounds (Pauler et al., 1986; Wu et al., 2019). In summary, a sensorineural hearing loss can be caused by loss of outer or inner hair cells, by metabolic processes affecting the cochlear amplifier, and/or by neurodegeneration.

Schuknecht (1964) conducted studies with human temporal bones and identified connections between audiometric thresholds and different types of age-related sensorineural hearing loss. Later, Dubno et al. (2013) proposed an approach to classify clinical audiograms into four audiometric phenotypes: older normal, sensory, metabolic and sensory+metabolic (Figure 1.2 b). However, a classification only based on the audiogram may not reflect other important aspects of auditory processing that might better characterize the hearing deficits. Besides, the quantification of the degree of inner hair cell and outer hair cell losses to the audiometric thresholds is not considered in such approach and would require additional measurements of the auditory function (Lopez-Poveda and Johannesen, 2012).

The differential diagnostic of cochlear and retrocochlear hearing losses is based on supra-threshold and physiological tests, some of which can roughly



a) Organ of Corti

b) Audiometric phenotypes

Figure 1.2: a) Mechanisms of sensorineural hearing loss in the cochlea. The spiral ganglion that contains the neuro-fibers (yellow) can be affected by neurodegeneration. The loss of outer and inner hair cells (red) can affect the cochlear amplifier and the cochlear transduction. The dysfunctional stria vascularis (green), usually produced by metabolic processes, leads to an endocochlear potential loss. This provokes a cochlear dysfunction that affects the cochlear amplifier along the entire basilar membrane. Figure modified with permission from Wong and Ryan (2015). b) Audiometric phenotypes form the classification of clinical audiograms in connection to animal studies where either metabolic or sensory hearing losses are induced. The figure shows the phenotypical exemplars selected by clinical experts and the average audiometric thresholds of the audiograms classified using machine learning techniques. Figure taken from Dubno et al. (2013) with permission.

identify the contributions of different types of impairments. Audiological and psychoacoustical tests showed the potential to disentangle the effect of cochlear and retrocochlear hearing loss already in the 70s (Jerger and Jerger, 1967, 1974). More recently, non-invasive physiological tests, such as otoacoustic emissions, have been also shown to be useful for identifying the origin of sensorineural hearing loss (Patuzzi, 1993). The most recent contribution to the clinical assessment of cochlear hearing loss has been the incorporation of tone-in-noise detection tests that can estimate the presence of dead cochlear regions (i.e. with a substantial inner-hair cell loss; Moore et al., 2000). Some of the aforementioned tests are currently available in commercial audiometric devices. However, they are only used in special cases, and no systematic evaluation of the sensorineural hearing loss beyond the audiogram has been considered in current clinical practice.

In contrast, in the last three decades, several studies have focused on the perceptual consequences of hearing loss (e.g. Houtgast and Festen, 2008; Lopez-Poveda et al., 2017; Moore et al., 1999; Strelcyk and Dau, 2009) rather than on its clinical diagnosis. Sensorineural hearing loss has been found to not only affect the hearing thresholds but also loudness perception, spectral and temporal resolution, pitch perception, intensity discrimination, spatial hearing and speech intelligibility (Moore, 2007). Although the experiments used in hearing research are time-consuming and require systematic training of the participants, there is potential in some of the tasks to be adapted to the clinical practice to better characterize the hearing deficits associated with the sensorineural hearing loss.

The present thesis focuses on the identification of clinical subpopulations of individuals with sensorineural hearing loss based on their perceptual deficits. The use of new diagnostic measures to pinpoint specific deficits that can allow such classification will be explored throughout the thesis.

1.3 Data-driven profiling and precision medicine

Precision medicine provides personalized treatments for a specific disease that are targeted to the needs of the patient. The targeted treatments are based on biomarkers, genotypic, phenotypic, or psychosocial characteristics of the individual (Jameson and Longo, 2015) that are associated with an optimal response for the treatment. The aim is to distinguish a given patient from other patients with similar clinical presentations and to improve clinical outcomes. One of the main advantages of precision medicine is that it can predict the treatment that provides the best response for a subgroup of patients and minimizes unnecessary side effects for those who do not show a response. Dividing the patients into subpopulation is known as stratification (see Figure 1.3). Thus, the terms “stratified” and “precision” or “personalized” medicine are often used interchangeably. Importantly, the stratification of patients in different clinically relevant subpopulations (phenotypes) can reduce the complexity of characterizing heterogenous diseases and has a reasonable cost-effectiveness compared to the personalization of the treatment for a single individual (Trusheim et al., 2007).

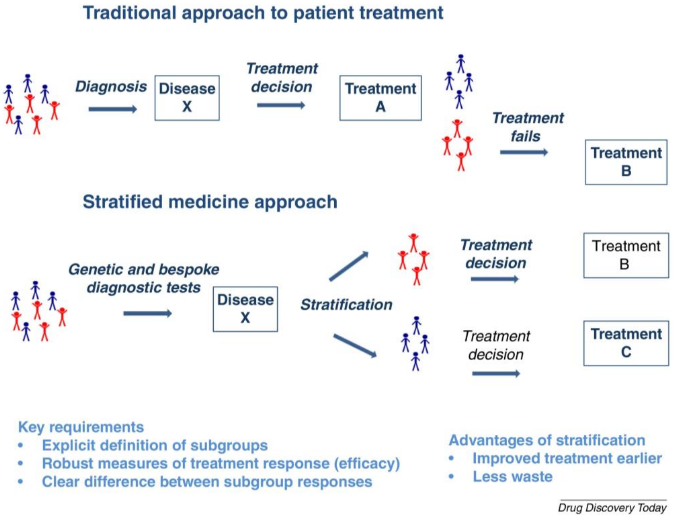


Figure 1.3: Sketch of the differences between traditional medicine and stratified medicine. Top panel shows different treatments are applied to the entire population with the disease in a trial and error approach. Bottom panel shows the stratified medicine approach where the patients are stratified intro sub groups that are more likely to respond to specific treatments. Figure taken from Lonergan et al. (2017) with permission.

The first criterion for implementing stratified medicine to treat a medical condition is that the identification of the patient subpopulations must be technically feasible. In the clinical practice, the patient’s phenotype is usually obtained by clinical biomarkers, which are measurable characteristics that associate the optimal treatment to a patient subpopulation. However, before the clinical biomarkers can be established, patient subpopulations with a likely different response to different treatments must be identified (Lonergan et al., 2017). In genetics, this is often done by using computational data-analysis or machine learning for gene expression profiling which is determined by patterns among the genes (Gligorijević et al., 2016). The diversity in these patterns can lead to subtypes of a disease that are susceptible to receive different treatments.

In this thesis, principles of stratification and patient classification will be explored for the hearing impaired population using data-driven approaches.

1.4 Overview of the thesis

The present thesis is inspired by the idea of implementing precision medicine in the field of rehabilitative audiology.

> First, the complexity of the hearing deficits needs to be reduced to some essential dimensions where extreme patterns might be identified. In *Chapter 2*, behavioral data obtained in psychoacoustic tasks are used for evaluating a data-driven method aiming to identify different auditory profiles. The method of "auditory profiling" is tailored to the hypothesis that the hearing deficits can be described as a combination of independent perceptual distortions.
>
> Second, potential auditory tasks that can be used in the hearing-care clinic as auditory "markers" are explored. In *Chapter 3*, a test battery is implemented and tested in a clinical population of listeners with different hearing abilities. The tests with the potential for their implementation in the clinical practice were prioritized towards a future adoption of a clinical test battery in the hearing-care clinics.
>
> Third, clinical subpopulations characterized by different hearing deficits are identified. In *Chapter 4*, the data set generated in the study presented in *Chapter 3* are analyzed with a refined version of the method evaluated in *Chapter 2*. The results are discussed with previous approaches of hearing loss characterization to define and characterize the auditory profiles.

Fourth, compensation strategies, tailored to the different hearing deficits observed in the auditory profiles are tested. *Chapter 5*, explores whether listeners belonging to different auditory profiles show substantial differences in terms of preferred hearing-aid settings This chapter lays on the foundations of a possible “precision audiology” in a proof-of-concept study.

Furthermore, the thesis explores the differences in terms of the benefit that hearing-aid users experience in their daily life. This might help set priorities in hearing rehabilitation in patients belonging to different audiometric groups. In *Chapter 6*, a data set containing subjective data of disabilities and handicaps is analyzed in a data-driven approach. The goal of the study is to identify “patterns of benefit” in the subjective responses from questionnaires in connection to audiometric groups that approximate the auditory profiles.

Finally, the main findings of each chapter are summarized in *Chapter 7*. The implications in terms of hearing loss characterization, hearing aid-fitting and auditory modelling are discussed. Moreover, the perspectives of a future implementation of “precision audiology” and the possibilities for the hearing-aid industry and the hearing-care centers to adopt such an approach are discussed.

2

A data-driven approach for auditory profiling and characterization of individual hearing loss [a]

Abstract

Pure-tone audiometry still represents the main measure to characterize individual hearing loss and the basis for hearing-aid fitting. However, the perceptual consequences of hearing loss are typically not only associated with a loss of sensitivity, but also with a loss of clarity that is not captured by the audiogram. A detailed characterization of a hearing loss may be complex and needs to be simplified to efficiently explore the specific compensation needs of the individual listener. Here, it is hypothesized that any listener's hearing profile can be characterized along two dimensions of distortion: type I and type II. While type I can be linked to factors affecting audibility, type II reflects non-audibility-related distortions. To test this hypothesis, the individual performance data from two previous studies were re-analyzed using an unsupervised-learning technique to identify extreme patterns in the data, thus forming the basis for different auditory profiles. Next, a decision tree was determined to classify the listeners into one of

---

[a] This chapter is based on Sanchez-Lopez, Bianchi, Fereczkowski, Santurette, and Dau (2018a) *"Data-Driven Approach for Auditory Profiling and Characterization of Individual Hearing Loss"*. Trends in Hearing, and Sanchez, Bianchi, Fereczkowski, Santurette, and Dau (2017) *"Data-driven approach for auditory profiling"*, ISAAR2017

the profiles. The analysis provides evidence for the existence of four profiles in the data. The most significant predictors for profile identification were related to binaural processing, auditory non-linearity, and speech-in-noise perception. This approach could be valuable for analyzing other data sets to select the most relevant tests for auditory profiling and propose more efficient hearing-deficit compensation strategies.

2.1 Introduction

Currently, the pure-tone audiogram is the main tool used to characterize the degree of hearing loss and for hearing-aid fitting. However, the perceptual consequences of hearing loss are typically associated not only with a loss of sensitivity, as reflected by the audiogram, but also with a loss of clarity that is not captured by the audiogram (e.g. Killion and Niquette, 2000). This loss of clarity may be associated with distortions in the auditory processing of supra-threshold sounds. While amplification can effectively compensate for loss of sensitivity, supra-threshold distortions may require more advanced signal processing to overcome the loss of clarity and improve speech intelligibility, particularly in complex acoustic conditions (e.g. Kollmeier and Kiessling, 2018; Plomp, 1978). Plomp (1978) suggested that a hearing loss can be divided into two components: an attenuation component and a distortion component. When a pure attenuation loss, also referred to as audibility loss or sensitivity loss, is compensated for by amplification, the speech reception threshold in stationary noise ($\mathrm{SRT_N}$) is similar to that of a normal-hearing (NH) listener. In contrast, when a distortion component is present, $\mathrm{SRT_N}$ remains elevated despite amplification.

Several studies have attempted to shed light on the potential mechanisms underlying supra-threshold distortions (e.g Glasberg and Moore, 1989; Houtgast and Festen, 2008; Johannesen et al., 2016; Strelcyk and Dau, 2009; Summers et al., 2013). In these studies, different psychoacoustic tests were considered in listeners with various degrees of hearing loss in an attempt to explain the variance observed in the listeners' speech-in-noise intelligibility performance. It was suggested that, beyond pure-tone audiometry, an elevated SRT_N could be associated with outcome measures related to spectral and/or temporal processing deficits. The supra-threshold distortions relevant for speech intelligibility in noise may thus reflect inaccuracies in the coding and representation of spectral and/or temporal stimulus features in the auditory system. To achieve the optimal compensation strategy for the individual hearing-impaired listener, a characterization of the listener's hearing deficits in terms of audibility loss, as well as clarity loss, thus seems essential.

Large-scale studies have attempted to establish a new hearing profile based on test batteries involving supra-threshold outcome measures in addition to pure-tone audiometry. As a part of the European project HEARCOM (Esch et al., 2013; Van Esch and Dreschler, 2015; Vlaming et al., 2011) new screening tests were proposed, as well as a test battery designed for assessing the specific hearing deficits of the patients. The factor analysis performed in a study with 72 hearing-impaired subjects revealed that the test outcomes can be grouped in four dimensions: audibility, high-frequency processing, low-frequency processing, and recruitment (Vlaming et al., 2011). Rönnberg et al. (2016) also used factor analysis to explore the relations between hearing, cognition, and speech-in-noise intelligibility in a large-scale study with 200 listeners. Even though the sensitivity loss was still a dominant factor, the new test battery provided information about supra-threshold processing using new outcome measures. However, these additional outcome measures were highly cross-correlated, which complicated the factor analysis. Although the above studies explored the relative importance of diverse factors in the individual subject, the interpretation of an individual hearing profile became more complex, particularly because the clinical tests were highly inter-related.

Other studies have suggested strategies for classifying hearing-impaired (HI) listeners based on a characterization of their hearing deficits. In one approach, (Lopez-Poveda, 2014) reviewed the mechanisms associated with hearing loss and their perceptual consequences for speech. In a two-dimensional space, the hearing loss was considered to represent the sum of an outer-hair-cell (OHC) loss and an inner-hair-cell (IHC) loss (Lopez-Poveda and Johannesen, 2012). The importance of this distinction is related to the way these mechanisms affect speech. While the OHC loss has been associated with audibility loss and reduced frequency selectivity, the IHC loss may yield a loss of clarity and temporal processing deficits (Killion and Niquette, 2000). However, since OHC and IHC loss can only be estimated by indirect outcome measures (Jürgens et al., 2011; Lopez-Poveda and Johannesen, 2012), and since pure-tone audiometry only reflects the mixed effects of OHC and IHC loss (Moore et al., 1999) , this approach seems limited in terms of an individual hearing-loss characterization in a clinical setting. Another approach was presented by Dubno et al. (2013), who suggested four audiometric phenotypes to account for age-related hearing loss. The phenotypes were proposed based on animal models with either a metabolic or a sensory impairment. Using a large database of audiograms from older humans, the corresponding human exemplars of the four audiometric phenotypes were identified by an expert researcher. Finally, a classifier trained on these exemplars was used to classify the remaining audiograms into the audiometric phenotypes. Although the audiometric phenotypes can be linked to the underlying mechanism of the hearing loss, a limitation of this approach is that it is fully based on the information provided by the audiogram. Hence, supra-threshold distortions are not, or only partly, reflected in this classification.

Inspired by the studies of Lopez-Poveda (2014) and Dubno et al. (2013), a two-dimensional approach was also considered in the present study. However, in contrast to these previous approaches, the classification of the listeners was mainly based on perceptual outcome measures, rather than physiological indicators of hearing loss. While the physiological indicators, such as OHC and IHC loss, cannot be assessed directly in humans, their corresponding perceptual distortions can likely be quantified using psychoacoustic tests. The aims of the present study were 1) to achieve a new hearing-loss characterization strategy

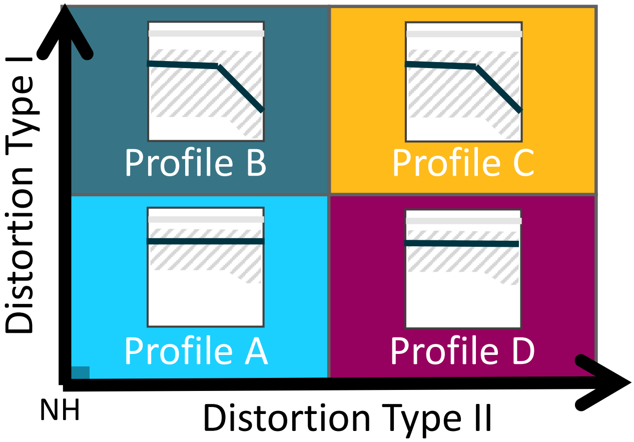


Figure 2.1: Sketch of the hypothesis. Hearing deficits arise from two independent types of distortions. Distortion type I: distortions that accompany loss of sensitivity. Distortion type II: distortions that do not covary with sensitivity loss. Profile A: Low distortion for both types. Profile B: High type I distortion and low type II distortion. Profile C: High distortion for both types. Profile D: Low type I distortion and high type II distortion.

that takes supra-threshold hearing performance into account and is based on functional tests reflecting auditory perception, and 2) to propose a new statistical analysis protocol that can be used to re-analyze existing data sets to improve and optimize such a characterization.

It was hypothesized here that any listener's hearing can be characterized along two independent dimensions: distortion type I and distortion type II, as indicated in Figure 2.1. Distortion type I was hypothesized to reflect deficits that have been found to co-vary with a loss of audibility, such as a loss of frequency selectivity and of cochlear compression (Moore et al., 1999). Distortion type II was hypothesized to reflect deficits that typically do not co-vary with audibility loss and may be related to inaccuracies in terms of temporal coding according to the conclusions from other studies (Johannesen et al., 2016; Strelcyk and Dau, 2009; Summers et al., 2013). The two dimensions can be roughly defined as audibility related and non-audibility related distortions. In this two-dimensional space, NH listeners are placed in the bottom-left corner and defined as not exhibiting any type of distortion. Then, four profiles may thus be identified, depending on the extent to which each type of distortion is present in the individual listener (Figure 2.1).

To test this hypothesis, a new data-driven statistical method is proposed here and used to re-analyze two existing data-sets and exploit the individual differences of HI listeners in terms of perceptual outcome measures. In line with the hypothesis, the method divides different perceptual measures into two independent dimensions. Next, the method identifies patterns in the data, hence the analysis is considered data-driven. The approach is similar to the one used to identify the four audiometric phenotypes in Dubno et al. (2013) but considers additional outcome measures beyond audiometry for the classification of the listeners into the four auditory profiles. The proposed statistical analysis is based on an archetypal analysis (Cutler and Breiman, 1994), an unsupervised learning method that is particularly useful for identifying patterns in data, and has been suggested for prototyping and benchmarking purposes (Ragozini et al., 2017). The main advantage of using unsupervised learning in terms of auditory profiling is that the analysis involves the performance of the listener in different tests, in contrast to correlations between single tests or regression analyses (e.g. Glasberg and Moore, 1989; Houtgast and Festen, 2008; Summers et al., 2013), which explore relations between various hearing disabilities in a HI population, rather than in an individual listener. The novel method was evaluated by re-analyzing the data from two previous studies presented in Thorup et al. (2016) and Johannesen et al. (2016). In both studies, an extensive auditory test battery was proposed and tested in HI listeners, to better characterize hearing deficits. While the analysis of those studies focused on finding correlates of speech intelligibility in noise and hearing-aid benefit, the goal here was to further define the two hypothesized distortion types and identify which outcome measures are most relevant to classify listeners into the four suggested auditory profiles.

2.2 Methods

The data-driven approach was conducted in two stages (Figure 2.2). First, unsupervised learning was used to identify the trends in the data that could be used to estimate the amount of each distortion type in individual listeners and thus categorize the listeners into different profiles. The second stage consisted of supervised learning, i.e., once the subjects were assigned to a profile, the data were analyzed again to find the best classification structure that could predict the identified profile.

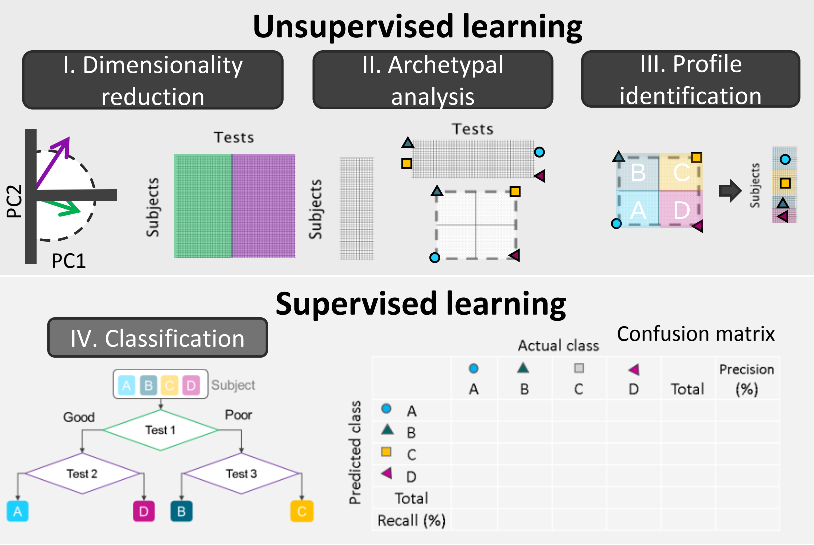


Figure 2.2: Sketch of the method considered in the present study. The upper panel shows the unsupervised learning techniques applied to the whole dataset. The bottom panel shows the supervised learning method, which uses the original data as the input and the identified profiles from the archetypal analysis as the output.

Unsupervised learning

Unsupervised learning aims to identify patterns occurring in the data, where the output is unknown and the statistical properties of the whole dataset are explored (Cutler and Breiman, 1994). In contrast to regression analysis, unsupervised learning does not aim to predict a specific output, for example, speech intelligibility. In the present approach, the identified auditory profiles were eventually inferred using various unsupervised learning techniques. First, a list of outcome measures obtained from different tests in the re-analyzed study of interest was selected as the input to the unsupervised learning stage.

Since two types of distortions to characterize the individual hearing loss were assumed, a principal component analysis (PCA) was run as the first step of the data analysis to reduce the dimensionality of the data. The reduction was done by keeping the variables that were strongly correlated to the first principal component (PC1) or the second principal component (PC2). These variables were used in further analysis rather than using directly the PCs (Figure 2.2-I). Therefore, a dimensionality reduction algorithm was implemented as follows: the optimal subset of variables that suggest a strong relationship with each of the two principal components was chosen using a leave-one-out cross-validation in an iterative PCA. In each iteration, a single variable was left out of the subset and the variance explained by the two principal components was re-calculated for the remaining set of variables. If the variance increased, the outcome measure that was left-out in this iteration was discarded. This process was repeated until either the variance explained was higher than 90% or the number of variables was lower than eight (four in each dimension). This reduction in the number of variables ensures that the variables are balanced with regard to both distortion types and the chosen variables are connected with the hypothesis in an unsupervised process.

An archetypal analysis (Cutler and Breiman, 1994) was then performed on the output of the dimensionality reduction stage (Figure 2.2-II). This technique combines the characteristics of matrix factorization and cluster analysis. In the present study, an algorithm similar to the one described in Mørup and Hansen (2012) was used. This analysis aimed to identify extreme patterns in the data (archetypes). As a result, the listeners were no longer defined by the performance in each of the tests, but by their similarity to the extreme exemplars contained in the data, i.e., the archetypes.

Based on the archetypal analysis, the subjects were placed in a simplex plot (square visualization) to perform profile identification (Figure 2.2-III). In such a plot, the archetypes are located at each corner and the listeners are placed in the two-dimensional space according to the distance to each archetype. In the present analysis, it was assumed that the subjects placed close to an archetype would belong to the same cluster. Consequently, each subject was labeled according to the nearest archetype.

Supervised learning

Once the profiles were identified, supervised learning could be performed. The purpose of this stage was to explore the accuracy of a classification scheme that makes use of only a few variables (here, outcome measures from different tests of auditory function). The joint probability density of the dataset and the output (i.e., the identified profiles) could then be used to select the most relevant tests for the classification of the subjects into the four auditory profiles.

Decision trees were used to classify each individual observation (Figure 2.2 IV). Here, each relevant outcome measure was used in the nodes forming a logical expression and dividing the observations accordingly. Since a decision tree needs to be trained with a subset of the data and a known output, the identified auditory profiles (Figure 2.2 III) were used as the response variable and a five-fold cross-validation was used to train the classifier. In the cross-validation procedure, the data were randomly divided into five segments. Four segments were used to train the classifier and the remaining one was used for testing. This was done iteratively ten times. The decision tree which provided the minimum test-error was used as the “optimal classifier”. Additionally, the decision tree was pruned to only have three nodes. This ensured that an efficient classification of the listeners based on only three tests could be considered in future clinical protocols.

Description of the data sets

In the present study, the data from Thorup et al. (2016) (study 1) and Johannesen et al. (2016) (study 2) were reanalyzed with the unsupervised and supervised learning techniques described above.

The dataset from study 1 contained 59 listeners, among which 26 listeners had normal hearing thresholds (NHx), 29 listeners were hearing impaired (HIx) and 4 had been previously identified as suffering from obscure dysfunction (ODx), i.e., with normal hearing thresholds but self-reports of hearing difficulties. The total number of variables (outcome measures from the different tests) considered in the analysis was 27. The variables used in the analysis were as follows (see Thorup et al., 2016, for details):

- Audiometric thresholds at low (HL_{LF}) and high frequencies (HL_{HF}).
- Spectral (MR_{spec}) and temporal (MR_{temp}) resolution at low and high frequencies.
- Binaural temporal fine structure (TFS) processing measured by interaural phase difference (IPD) frequency thresholds.
- Speech recognition thresholds in stationary (SRT_N) and fluctuating (SRT_{ISTS}) noise.
- Reading-span test of working memory (RS).

The results of additional tests, not reported in Thorup et al. (2016) but collected in the same listeners, were also included in the present analysis:

- Binaural pitch test, using a procedure adapted from Santurette and Dau (2012) measuring the detection of pitch contours presented either diotically or dichotically. The variables used here were $\mathrm{Bp}_{\mathrm{dicho}}$ (percent correct for dichotic presentations only) and $\mathrm{Bp}_{\mathrm{total}}$ (percent correct for the total number of presentations, i.e., diotic and dichotic).
- Speech reception threshold in quiet ($\mathrm{SRT}_{\mathrm{Q}}$) and discrimination scores (DS) using the Dantale I (Elberling et al., 1989) speech material.
- Adaptive categorical loudness scaling (ACALOS; Brand and Hohmann, 2002), with the most comfortable level (MCL) and the lower slope of the growth of loudness ($\mathrm{ACALOS}_{\mathrm{Slope}}$) at low and high frequencies used here as variables.

The dataset from Johannesen et al. (2016) (study 2) contained 67 HI listeners (HIx). The total number of variables considered in the analysis was 11:

- Audiometric thresholds at low ($\mathrm{HL}_{\mathrm{LF}}$) and high frequencies ($\mathrm{HL}_{\mathrm{HF}}$).
- Aided speech recognition thresholds in stationary noise ($\mathrm{HINT}_{\mathrm{SSN}}$) and reversed two-talker masker ($\mathrm{HINT}_{\mathrm{R2TM}}$).
- Frequency modulation detection threshold (FMDT).
- Basilar membrane compression (BM comp) and OHC and IHC loss estimated from the results of a temporal masking curve experiment (Nelson et al., 2001). These three variables were each divided into a high-frequency and a low-frequency estimate.

Pre-processing of the data sets

For both data sets, the performance in each outcome measure was normalized such that the 25th percentile equaled −0.5 and the 75th percentile equaled +0.5. To more easily compare the tests, a good performance thus always corresponded to a positive number and a poor performance corresponded to a negative number.

The tests that corresponded to measures taken at different frequencies, e.g., pure-tone audiometry, were reduced to the mean at low frequencies ($\leq$ 1 kHz) and at high frequencies ($>$ 1 kHz)[b]. Additionally, when the tests were performed in more than one ear, the average between the two ears was used as the outcome measure[c]. Listeners that did not complete more than three of the considered tests were excluded from the analysis. Furthermore, an artificial observation with an optimal performance (+1) in all tests was created, which served as an ideal NH reference that did not exhibit any type of distortion. This observation was always the archetype A, located in the origin of coordinates of the hypothesis stated in Figure 2.1. The pre-processing was performed identically for both data sets.

2.3 Results

The two studies were analyzed using an identical method. For convenience, results corresponding to the re-analysis of the data from Thorup et al. (2016) are referred to using the sub-index 1, e.g., profile A_1. The results from the re-analysis of the Johannesen et al. (2016) data are referred to using sub-index 2, e.g., profile A_2. For general mentions of an auditory profile, no sub-index is added. The whole dataset was reduced to the variables that were strongly correlated to Dimension I (PC1) or Dimension II (PC2), as summarized in Table 2.1.

For study 1, the dimensionality reduction revealed that the performance in binaural tests was largely independent of hearing thresholds, suggesting that Dimension II may be related to binaural processing abilities and Dimension I to audibility at low and high frequencies. The PCA could explain 80.3% of the variance in the performance for different hearing tests with only two components, with 63.01% explained by PC1 and 17.3% by PC2.
For study 2, Dimension II was more dominated by low-frequency processing

---

[b] Since the data from study 1 were already in that form, the data from study 2 were processed accordingly.

[c] Since the data from study 2 were collected only in the better ear, the data from study 1 were processed accordingly for a better comparison.

Table 2.1: Results from the dimensionality reduction of the two datasets. The table includes variables strongly correlated to PC1 (distortion type I, top four rows) and PC2 (distortion type II, bottom four rows) and their correlation coefficient obtained from the loadings of the PCA

Study I: Thorup et al. (2016)				**Study II: Johannesen et al. (2016)**			
Variable	**Test**	**PC1**	**PC2**	**Variable**	**Test**	**PC1**	**PC2**
HL_{LF}	Hearing loss at low frequencies	0.45	-0.03	HL_{HF}	Hearing loss at high frequencies	0.52	0.08
HL_{HF}	Hearing loss at high frequencies	0.41	-0.22	OHC $loss_{HF}$	Outer hair cell loss estimated at high frequencies	0.55	0.05
SRT_{Q}	Speech reception threshold (SRT) in quiet	0.46	-0.01	IHC $loss_{HF}$	Inner hair cell loss estimated at high frequencies	0.37	-0.03
SRT_{ISTS}	SRT in noise using international speech test signal	0.47	-0.17	BM $comp_{HF}$	Basilar membrane compression at high frequencies	0.51	-0.01
DS	Word discrimination score	0.33	-0.24	HL_{LF}	Hearing loss at low frequencies	-0.00	0.62
MCL_{LF}	Most comfortable level at low frequencies	0.14	0.46	FMDT	Frequency modulation discrimination threshold	-0.02	0.42
Bp_{dicho}	BP dichotic condition	0.2	0.53	OHC $loss_{LF}$	Outer hair cell loss estimated at low frequencies	0.03	0.45
Bp_{tot}	BP diotic + dichotic	0.16	0.61	IHC $loss_{LF}$	Inner hair cell loss estimated at low frequencies	-0.11	0.45

abilities and Dimension I by high-frequency processing abilities. The PCA could explain 67.8% of the variance in the performance for the behavioral tasks with the chosen variables, with 37.2% explained by PC1 and 30.6% by PC2.

The archetypal analysis was used to identify four archetypes using the variables from Table 2.1. As shown in Figure 2.3, in both studies, Profile A (archetype A) exhibited the best performance in both dimensions and Profile C the worst. Profile

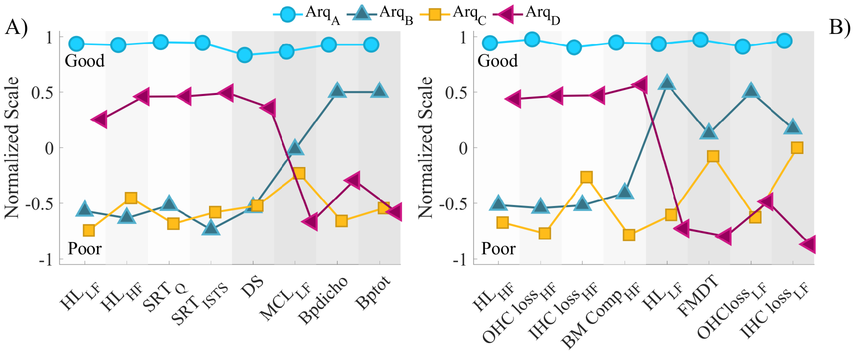


Figure 2.3: Archetypes (Artyp): Extreme exemplars of the different patterns found in the data. A) Normalized performance of each of the four archetypes from study 1. B) The same for study 2. The variables are divided according to Table 1. The first four variables correspond to distortion type I and the remaining four to distortion type II.

B showed poor performance only in Dimension I, while Profile D showed poor performance only in Dimension II.

Figure 2.3 A) illustrates the four archetypes from study 1. The performance in the tests related to distortion type I was good for archetypes A_1 and D_1 and poor for B_1 and C_1, in line with the hypothesis of the present study. However, the performance in the tests corresponding to distortion type II was less consistent. Archetypes A_1 and B_1, with an expected low degree of distortion type II, exhibited good performance in the binaural tests. Archetypes C_1 and D_1 showed poor performance in $\mathrm{Bp}_{\mathrm{dicho}}$, $\mathrm{Bp}_{\mathrm{tot}}$, and $\mathrm{MCL}_{\mathrm{LF}}$, but not in DS. This is because DS was correlated to both principal components. As described in the Method section, the number of variables per dimension was set to four. Hence, DS should not be considered as a representative variable of distortion type II. Panel B of Figure 2.3 depicts the four archetypes from study 2. The performance in the tests related to distortion type I was good for archetypes A_2 and D_2 and poor for B_2 and C_2, in line with the hypothesis of the present study. Besides, the performance in the tests related to distortion type II was better for archetypes A_2 and B_2 than for D_2 and C_2, also in line with the hypothesis of the existence of four auditory profiles along with two independent types of distortion.

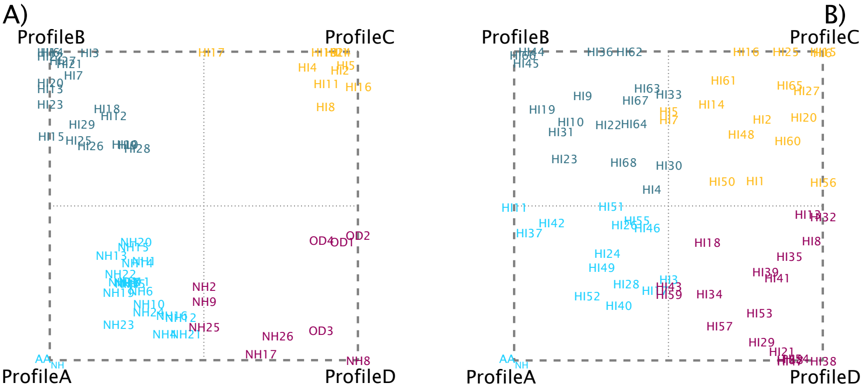
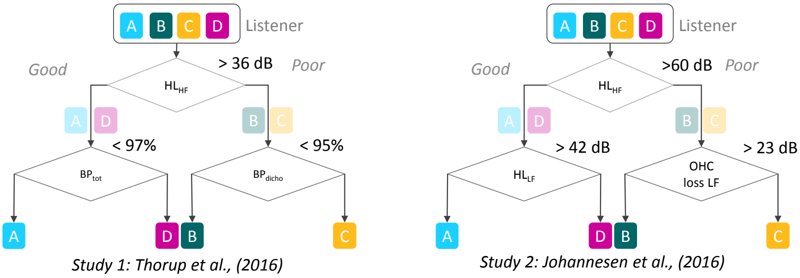


Figure 2.4: Simplex plots for (A) study 1 and (B) study 2. Representation of the listeners in a two-dimensional space. The four archetypes are located at the corners and the remaining observations are placed in the simplex plot depending on their similarity with the archetypes.

Based on the archetypes presented in Figure 2.3, each listener was assigned to the auditory profile defined by the nearest archetype. Results from study 1 are depicted in Figure 2.4 A). The simplex representation shows how the listeners could be divided into clear clusters in the two-dimensional space. In the case of study 2 (Figure 2.4 B), the listeners were more spread out across the two-dimensional space and no clear groups could be identified. It should be noted that, in this case, archetype Q_2, labeled as AA_{NH}, corresponded to the artificial observation with a good performance in all tests. It is located in the bottom-left corner in the simplex plot and is far from the rest of the observations because, in contrast to study 1, the data set did not contain any data from NH listeners.

Figure 2.5 depicts the results of the supervised learning analysis. Decision trees were obtained by using the raw data as an input and the identified auditory profiles as the output. In study 1, the classification tree based on HL_{HF} and binaural pitch showed a very high sensitivity (95% true positives). In study 2, the classification was based not only on the audibility loss at high and low frequencies but also



Actual class (Predicted class in rows)

	● A	▲ B	■ C	◀ D	Total	Precision (%)
● A	20	0	0	3	23	87
▲ B	0	19	0	0	19	100
■ C	0	0	10	0	10	100
◀ D	0	0	0	7	7	100
Total	20	19	10	10	59	
Recall (%)	100	100	100	70		

Actual class (Predicted class in rows)

	● A	▲ B	■ C	◀ D	Total	Precision (%)
● A	17	1	0	0	18	94
▲ B	1	15	1	0	17	88
■ C	0	0	17	2	19	89
◀ D	0	1	0	12	13	92
Total	18	17	18	14	67	
Recall (%)	94	88	94	85		

Figure 2.5: Decision trees and confusion matrices of the classifiers from the analysis of both data sets. For each study, the resulting classifier has three nodes. The right branch corresponds to a poor performance and the left branch to a good performance. The accuracy of the classifier is shown in the form of a confusion matrix where the correspondence of the actual classes and predicted classes are evaluated.

the estimate of OHC loss at low frequencies. The sensitivity of this classifier was slightly lower (91%). Although $\mathrm{HL_{HF}}$ is in the first node of both classifiers, the amount of hearing loss at high frequencies required to divide the listeners along the distortion type I dimension (i.e., subgroups A-D and B-C) was lower for study 1 than for study 2. This is mainly due to the differences in the distribution of the hearing thresholds among the participants of each study. The differences observed between the two classification trees are further discussed in the following section.

2.4 Discussion

Two types of distortion to characterize individual hearing loss

The present study proposed a new data-driven statistical analysis protocol, which was applied to two existing data sets. The goal was to determine the nature of the two main independent dimensions for individual hearing loss characterization. Based on existing literature findings, it was hypothesized that one dimension (distortion type I) would reflect audibility-related distortions, while the other dimension would reflect non-audibility-related distortions (distortion type II). The analysis performed on the two data sets provided different results, which need to be interpreted taking the differences between the two studies into account.

The analysis of the data set in study 1, with a population of both near-NH and HI listeners, revealed that binaural processing tests were highly sensitive for the classification of the listeners and a main contributor to the distortion type II. In that study, the listeners presented a mild low-frequency hearing-loss (24 dB HL ± 6 dB) and a higher degree of high-frequency hearing loss (55 dB HL ± 6 dB). As shown in Figure 2.5, the HI listeners were classified into Profile B or C according to their high-frequency hearing loss and were divided along the distortion type II dimension according to their binaural processing abilities. The analysis of the data set in study 2, with only HI listeners, suggested that distortion type I was also related to high-frequency processing, while distortion type II was related to low-frequency processing. In study 2, the listeners presented a higher degree of low-frequency hearing loss (37 dB HL ±12 dB) compared to study 1 and a similar degree of high-frequency hearing loss (58 dB HL ± 12 dB) but with a larger variance. Although the listeners of study 2 were distributed across the four profiles, they were not clearly divided into clusters as in study 1 (Figure 2.4). This suggests that, although the hearing loss at low vs. high frequencies may, in this case, be considered as a good indicator of distortion type I vs. type II, the corresponding auditory profiles were less clearly separated than in study 1. This is probably due to the lack of NH or near-NH listeners in study 2.

Study 1 and 2 also differed in terms of test batteries. Study 2 did not consider any test of binaural temporal fine structure processing, which may partly account for the difference in the reduced variance explained in the analysis of study 2 compared to study 1.

Although studies 1 and 2 differed both in terms of listeners and test batteries, the analysis performed here revealed that in both cases distortion type I was dominated by high-frequency hearing loss. This was observed also in previous studies in which the sensitivity loss, particularly at high frequencies, was the main predictor of the differences among listeners (e.g Vlaming et al., 2011). The loss of sensitivity at high frequencies can be ascribed to a loss of sensory cells, specifically OHC loss, which yields loss of cochlear compression and a reduced frequency selectivity (Moore et al., 1999). Other relevant dimensions suggested in previous studies were related to temporal fine structure (TFS) processing (Rönnberg et al., 2016) and low-frequency processing (Vlaming et al., 2011). In agreement with this, the analysis of study 1 pointed towards measures of binaural TFS processing abilities (IPD detection frequency limit and binaural pitch test) for the second dimension, measures that may be correlated to FMDTs (Strelcyk and Dau, 2009), a measure assumed to involve monaural TFS processing abilities. In contrast, study 2 contained tests that estimated OHC and IHC loss as well as BM compression. As shown in Table 1, HL_{HF} and $BMComp_{HF}$ were strongly correlated to PC1, and HL_{LF} was strongly correlated to PC2 together with FMDTs. This suggests that, while HL_{HF} can be ascribed to a compression loss, HL_{LF} is most likely related to temporal coding deficits, as reflected by FMDTs. Despite the different outcome measures used in the two studies, the analysis of both studies is consistent with distortion type II being related to temporal fine structure processing. In summary, the outcomes of this study support the hypothesis that distortion type I may be more related to functional measures of spectral auditory processing deficits and distortion type II may be more related to functional measures of temporal auditory processing deficits.

Auditory profiling and the audibility-distortion model

In the present study, it was assumed that there are two independent types of distortion that affect the overall listening experience and functional performance of the listener. Although it was hypothesized, based on earlier literature findings, that distortion type I involved deficits that covaried with the loss of sensitivity, audibility itself was not a priori considered as a fully separate dimension as in previous approaches, according to the proposal of the present study of two types of distortions that are, ideally, fully independent. In Plomp's model, besides the attenuation component, a distortion component related to the supra-threshold deficits was proposed to account for the elevated SRTs in speech-in-noise intelligibility tests (Plomp, 1994). However, Humes (1994) argued that the distortion component can, in fact, appear as a consequence of a non-optimal compensation of the spectral configuration of the audibility loss and not because of additional and independent supra-threshold deficits. They also stated that the effective compensation of the attenuation component should be performed prior to further investigation of the origin of the supra-threshold distortions. Humes (2007) reviewed previous studies of aided speech and concluded that the main factors that explained the individual differences in speech intelligibility in older adults were audibility and cognitive factors. In the analysis presented here, both re-analyzed studies included hearing threshold and speech intelligibility outcomes and study 1 included a cognitive test of working memory. As audibility and cognitive factors are known to indirectly influence the performance in some of the other functional tests used in the analysis (e.g. Humes, 2007), it was decided to not treat them as independent dimensions, as this would have biased the analysis and the aim was to take advantage of a data-driven statistical method to neutrally define the assumed two independent dimensions. While audibility was reflected as a contributor to the two distortion types in the present analysis, cognition did not emerge as a key variable. This is consistent with Lopez-Poveda et al. (2017), where it was found that working memory was only weakly related to outcome measures of hearing-aid benefit. However, as only one cognitive test was included in the present analysis, the findings do not allow for a clear conclusion about the role of cognition. Applying the present statistical method to test batteries that include more extensive cognitive measures might help clarify this aspect.

In contrast to the current study, Kollmeier and Kiessling (2018) explained the factors contributing to hearing loss by three components: an attenuation component that produces a loss of sensitivity due to OHC and IHC loss, a distortion component associated with a reduced frequency selectivity, and a neural component related to degradations presented in the neural representation of the stimulus and associated to binaural processing deficits. The three components were not assumed to be independent such that the loss of sensitivity (the attenuation component) could covary with reduced frequency selectivity (the distortion component) and with IHC loss (the neural component). Despite the important difference of the assumption of an independent attenuation component between the two approaches, the present findings do reconcile rather well with Kollmeier and Kiessling (2018) approach. While distortion type I was found to be related to compression loss and elevated speech-in-noise recognition thresholds, distortion type II was associated with temporal and binaural processing deficits. Distortion type I in the present study can thus be compared to the distortion component from Kollmeier and Kiessling (2018) and distortion type II to their neural component. The two approaches thus share some similarities, except for the assumption of independence of the two distortion components in the current approach vs. the assumption of an additional attenuation component in Kollmeier and Kiessling (2018).

Sensitivity loss as a consequence of hair cell loss

Pure-tone audiometric thresholds are used to quantify the hearing loss but they can, in fact, be the consequence of different factors. As shown in Figure 2.3 B), dimension I does not only contain the high-frequency hearing loss but also estimated cochlear compression. Dimension II contains the low-frequency hearing loss and the outcome of the frequency modulation detection task which has been suggested to reflect temporal processing abilities. Therefore, it is important to bear in mind that there are interactions between the audibility and the two types of distortions proposed here. One approach to disentangle this interaction may be made based on the effects related to the OHC vs IHC processing. If a substantial population of IHC or neural fibers is affected, the thresholds can be elevated (Lobarinas et al., 2013), leading to temporal distortions as well as degraded binaural processing (Profiles D and C). However, the temporal acuity can also be compromised while audiometric thresholds are normal or close-to-normal (Zeng et al., 1999)(Profile D). OHC loss is typically associated with basilar membrane (BM) compression loss (reduced frequency selectivity) as well as elevated audiometric thresholds (Ahroon et al., 1993). Although reduced compression leads to a threshold elevation (Profile B), listeners with elevated thresholds can still have a nearly-normal BM compression (Profile A).

Evaluation of the data-driven method

The method used in the present study was designed based on the hypothesis that the listeners could be divided into four auditory profiles according to the results from their perceptual outcome measures. First, two independent types of distortions were assumed to characterize the individual hearing deficits of the listeners. Second, the extreme exemplars, i.e., the archetypes, contained in the data were identified and the listeners were defined according to their similarity to the nearest exemplar. Third, the outcome measures that were the most relevant for the classification of the listeners were identified. Other methods, such as linear regression, make use of the outcome measures to predict the performance in specific tests. This is typically done to explore the effects of different outcome measures on speech intelligibility. The novelty of this method lies in the fact that the characterization of the hearing deficits was carried out by analyzing the whole data set with the goal of achieving an individual hearing loss characterization. Since this is a data-driven method, the results are highly influenced by the data included in the analysis. Therefore, one should be cautious when interpreting the results.

The method considered only two principal dimensions for explaining the data.

The number of variables was reduced to have only four tests in each dimension. This decision makes the archetypes strongly connected to the hypothesis and keeps the number of variables per dimension balanced. However, if fewer than four variables are representative of one of the dimensions, the current algorithm also considers variables that can be correlated to both principal components. In the present study, this was the case for only one variable of study 1 (DS), which did not yield significant changes in the analysis. This limitation can be solved by imposing the assumption of orthogonality in the selection of the variables instead of using cross-validation. In this case, all the variables that are considered to belong to both dimensions are initially discarded. However, the explained variance might be lower and the number of the representative variables might change when using that method instead of iterative cross-validation.

The archetypes, representing extreme exemplars, were used here as prototypes of the auditory profiles. Therefore, the rest of the listeners were assumed to belong to the same category as the nearest archetype. This has two main disadvantages. First, if outliers are present in the data, these will be most likely used as archetypes. Second, subgroups of listeners that are not well represented by any of the four auditory profiles are not considered here. In contrast, in Dubno et al. (2013), the identification of the exemplars corresponding to each audiometric phenotype was done by an experienced researcher. That method is, however, not feasible for large data sets and may also be prone to judgement bias from the researchers. The use of unsupervised learning provides a solution to this potential problem. To better define the auditory profiles, alternative clustering, as well as other advanced pattern-recognition techniques, may also be explored instead of an archetypal analysis for profile identification and benchmarking (Ragozini et al., 2017).

The proposed method showed a potential for re-analyzing other existing data sets. The new exploratory approach can help test specific hypotheses by dividing the listeners into meaningful groups before analyzing the data. However, some requirements about the data are needed to reach consistent conclusions about a general characterization of hearing deficits. The data set should contain a representative sample of different degrees of hearing loss and a normal hearing reference, as well as a substantial variability in performance in other tests, such as speech-in-noise intelligibility, which should be performed unaided. In this way, both audibility and non-audibility related factors would influence the performance of the listeners. Since the method is sensitive to the input variables, a representative number of supra-threshold outcome measures should also be considered, including measures of loudness perception, binaural processing abilities, as well as outcome measures of spectral and temporal resolution, as it has been suggested in this and previous studies. Besides, cognition, as well as physiological indicators of hearing loss, such as auditory brainstem responses or middle-ear response, may be included to further characterize a listener's auditory profile. If these requirements are not fulfilled, the method would still categorize listeners into four subgroups, but the results may be misleading and difficult to interpret.

Overall, the present method provided results in line with the initial hypotheses. The two types of distortions were found to be related to spectral and temporal auditory processing deficits, which supports the idea of considering two independent dimensions instead of previous models based on audibility and additional factors. The analysis of further and more extensive existing data sets with the data-driven method proposed here, provided that they contain a representative population of listeners and outcome measures, may help refine the definition of the two distortion types and improve future characterization of individual hearing loss. Such a characterization may be useful in future clinical practice towards a better classification of patients in terms of hearing-aid rehabilitation.

2.5 Conclusion

The data-driven statistical analysis provided consistent evidence of the existence of two independent sources of distortion in hearing loss and, consequently, different "auditory profiles" in the data. While distortion type I was more related to audibility loss at high frequencies, the origin of distortion type II was connected to reduced binaural and temporal fine-structure processing abilities. The most informative predictors for profile identification beyond the audiogram were related to temporal processing, binaural processing, compressive peripheral nonlinearity, and speech-in-noise perception. The current approach can be used to analyze other existing data sets and may help define an optimal test battery to achieve efficient auditory profiling towards more effective hearing-loss compensation.

Acknowledgments

We thank R. Ordoñez, T. Behrens, M. L. Jepsen, M. Gaihede, and the rest of BEAR partners for their input during the realization of this study. Additionally, we want to thank E. López-Poveda and P. Johannesen for proving access to their data (study 2), as well as G. Lőcsei, O. Strelcyk, A. Papakonstantinou, and T. Neher, who provided data for preliminary analyses not presented here. We would also like to thank the two anonymous reviewers who suggested some improvements in an earlier version of the manuscript. This work was supported by Innovation Fund Denmark Grand Solutions 5164-00011B (BEAR project), Oticon, GN Resound, Widex and other partners. The funding and collaboration of all partners is sincerely acknowledged.

3

Auditory tests for characterizing hearing deficits: The BEAR test battery [a]

Abstract

The Better hEAring Rehabilitation (BEAR) project aims to provide a new clinical profiling tool – a test battery – for hearing loss characterization. Whereas the loss of sensitivity can be efficiently measured using pure-tone audiometry, the assessment of supra-threshold hearing deficits remains a challenge. In contrast to the classical ‘attenuation-distortion’ model, the proposed BEAR approach is based on the hypothesis that the hearing abilities of a given listener can be characterized along two dimensions reflecting independent types of perceptual deficits (distortions). A data-driven approach provided evidence for the existence of different auditory profiles with different degrees of distortions. Ten tests were included in a test battery, based on their clinical feasibility, time efficiency and related evidence from the literature. The tests were divided into six categories: audibility, speech perception, binaural processing abilities, loudness perception, spectro-temporal modulation sensitivity and spectro-temporal resolution. Seventy-five listeners with symmetric, mild-to-severe sensorineural hearing loss were se-

---

[a] This chapter is based on:
Sanchez-Lopez, Nielsen, El-Haj-Ali, Bianchi, Fereckzowski, Cañete, Wu, Neher, Dau, and Santurette (2020d) “Auditory tests for characterizing hearing deficits: The BEAR test battery” Submitted to the International Journal of Audiology. Preprint at medRxiv: 021949.

lected from a clinical population. The analysis of the results showed interrelations among outcomes related to high-frequency processing and outcome measures related to low-frequency processing abilities. The results showed the ability of the tests to reveal differences among individuals and their potential use in clinical settings.

3.1 Introduction

In current clinical practice, hearing loss is diagnosed mainly on the basis of pure-tone audiometry (ISO 8253-1, 2010). The audiogram helps differentiate between conductive and sensorineural hearing losses and can characterize the severity of the hearing loss from mild to profound. However, the pure-tone audiogram only assesses the sensitivity to simple sounds, which is not necessarily related to listening abilities at supra-threshold sound pressure levels (e.g. a person's ability to discriminate speech in noise)

Pure-tone audiometry is often complemented by speech audiometry (ISO 8253-3, 2012), which is a test typically performed in the form of word recognition performance in quiet (Anderson et al., 2018). Although this test can provide information about supra-threshold deficits (Gelfand, 2009), measurements of speech understanding in noise have been found more informative (Killion et al., 2004; Nilsson et al., 1994). Since improving speech intelligibility is usually the main goal of successful hearing rehabilitation, several auditory factors affecting speech intelligibility in noise have been investigated (e.g. Glasberg and Moore, 1989; Houtgast and Festen, 2008; Strelcyk and Dau, 2009). Audibility (in conditions with fluctuating maskers), frequency selectivity (in conditions with stationary noise), and temporal processing acuity (in conditions with speech interferers), have been identified as important factors affecting speech reception thresholds in noise when using meaningful sentences as speech material (e.g. Desloge et al., 2017; Johannesen et al., 2016; Oxenham and Simonson, 2009; Rhebergen et al., 2006)[b] Thus, a hearing evaluation that goes beyond pure-tone sensitivity and speech intelligibility in quiet would be expected to provide a more accurate characterization of a listener's hearing deficits.

In Denmark, the Better hEAring Rehabilitation (BEAR) project was initiated with the aim of developing new diagnostic tests and hearing-aid compensation strategies for audiological practice. Although the assessment of individual hearing deficits can be complex, new evidence suggests that the perceptual consequences of a hearing loss can be characterized effectively by two types of hearing deficits, defined as "auditory distortions" (Sanchez-Lopez et al., 2018a). By analysing the outcomes of two previous studies (Johannesen et al., 2016; Thorup et al., 2016) with a data-driven approach, Sanchez-Lopez et al. (2018a) identified high-frequency hearing loss as the main predictor of one of the distortions, whereas the definition of the second type of distortion was inconclusive. The inconclusiveness in the prediction of the second distortion was most likely due to differences between the two studies in terms of hearing loss profiles and outcome measures. Here, a new dataset was therefore collected based on a heterogeneous group of listeners with audiometric hearing losses ranging from very mild to severe and with a large range of audiometric profiles. To that end, the most informative tests resulting from the analysis of Sanchez-Lopez et al. (2018a) were included, together with additional auditory tests that had shown potential for hearing profiling in other previous studies. The tests included in the current study are referred to as the BEAR test battery.

The characterization of hearing deficits beyond the audiogram was considered in several earlier studies (e.g., Brungart et al., 2014; Lecluyse et al., 2013; Rönnberg et al., 2016; Santurette and Dau, 2012; Saunders et al., 1992; Vlaming et al., 2011).

---

[b] The factors identified correspond to the authors' conclusions based on cited references. For example, Johannesen et al. (2016) identified the basilar membrane compession as a predictor of speech intelligibility in stationary noise and temporal processing as a predictor of speech-in-speech intelligibility. Desloge et al. (2017), Oxenham and Simonson (2009), and Rhebergen et al. (2006) identified the audibility of the soft speech sounds in the presence of fluctuating maskers as a crucial factor for speech intelligibility.

Among them, the HEARCOM project (Vlaming et al., 2011) proposed an extended hearing profile formed by the results of several behavioural tests. These tests targeted various auditory domains, such as audibility, loudness perception, speech perception, binaural processing, and spectro-temporal resolution, as well as a test of cognitive abilities. Importantly, while the auditory domains considered in the BEAR test battery are similar to the ones considered in the HEARCOM project, the BEAR project aims to additionally classify the patients in subcategories and to create a link between hearing capacities and hearing-aid parameter settings.
The tests included in the BEAR test battery were chosen based on the following criteria: 1) There is evidence from the hearing research literature that the considered test is informative (i.e., it provides information about the individual hearing deficits) and reliable (i.e., the result of the test does not vary over time); 2) The outcomes of the test may be linked to a hearing-aid fitting strategy; 3) The outcome measures are easy to interpret and to explain to the patient; 4) The task is reasonably time-efficient or can be suitably modified to meet this requirement (e.g., by changing the test paradigm or developing an out-of-clinic solution); 5) The test implementation can be done with equipment available in clinics; 6) The tasks are not too demanding for patients and clinicians; 7) Tests with several outcome measures are prioritized, and 8) The tests are language independent are also prioritized.

The selected test battery included measures of audibility, loudness perception, speech perception, binaural processing abilities, spectro-temporal modulation (STM) sensitivity and spectro-temporal resolution. It was implemented and tested in older listeners with different hearing abilities (from near-normal to severe hearing losses). The goals of the study were: 1) To collect reference data from a representative sample of HI listeners for each of the selected tests, 2) to analyse the test-retest reliability of these tests, 3) to analyse the relationships between the different outcome measures, and 4) to propose a version of the test battery that can be implemented in hearing clinics.

3.2 General methods

Participants and general setup

Seventy-five listeners (38 of them females) participated in the study, who were aged between 59 and 82 years (median: 71 years). Five participants were considered NH with thresholds below 25 dB Hearing Level (HL) in the frequency range between 0.25 and 4 kHz in both ears (PTA ≤ 22 dB HL). PTA was defined as the pure-tone average between 0.5, 1 and 2 kHz. Two of these participants were not usual hearing-aid users. The HI group consisted of 70 participants with symmetric sensorineural hearing losses. Symmetric sensorineural hearing loss was defined as an interaural difference (ID) ≤ 15 dB HL at frequencies below 8 kHz and ID ≤ 25 dB HL at 8 kHz and air-bone gap < 10 dB HL. The pure-tone audiograms of the participants are shown in 3.1.

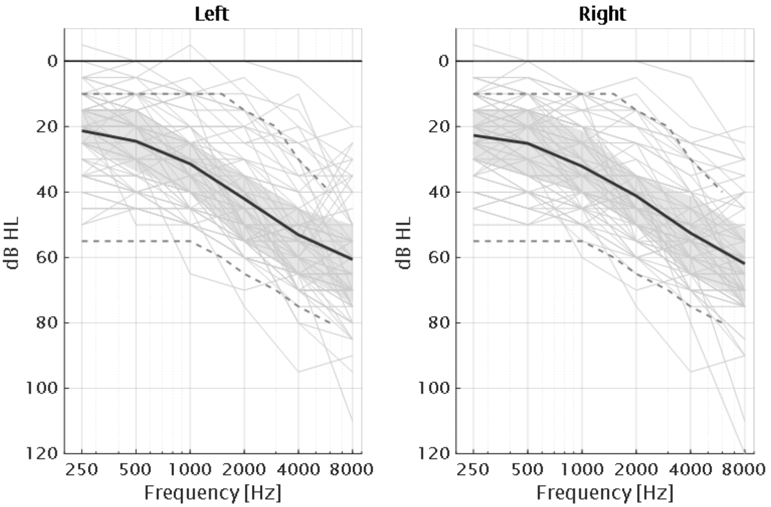


Figure 3.1: Audiograms of the 75 participants of the study together with the average for each ear (dark solid lines) and interquartile ranges (grey areas). The grey dashed lines correspond to the standard audiograms N1 and N4 from Bisgaard et al. (2010).

---

[c] Despite other listeners presented PTA ≤ 22 dB, the individual thresholds did not fulfil this criteria.

The participants were recruited from the BEAR database (Wolff et al., 2020) at Odense University Hospital (OUH), from the patient database at Bispebjerg Hospital (BBH), and from the database at the Hearing Systems Section at the Technical University of Denmark (DTU). The basic audiological assessment consisted of pure-tone audiometry, wideband tympanometry (Rosowski et al., 2013) and middle ear muscle reflex, and was conducted in the facilities of OUH, BBH and DTU. The rest of the tests were performed via PC in a double-walled sound-insulated booth (BBH and DTU) or in a small anechoic chamber (OUH). The tests were implemented in Matlab with a graphical user interface (GUI) that the examiner could operate without programming experience. Most of the tests were implemented using a modular framework for psychoacoustic experiments (AFC; Ewert, 2013), except for HINT, provided by Jens Bo Nielsen and Binaural Pitch test which was a reimplementation of the Binaural Pitch Test v1.0, Bispebjerg hospital, 2008. The participants were seated in the room and the stimuli were presented through headphones (Sennheiser HDA200) connected to a headphone-amplifier (SPL phonic) and an audio interface (RME Surface 24-bit). The equipment was calibrated using an artificial ear according to IEC 60318-1:2009. The tests consisting of threshold estimation using the AFC framework were repeated at least two times and the mean of the two measurements was considered as the final value. Tu ensure the quality of the data collected, a repetition was considered as an outlier if it was greater than three scaled median absolute deviations and additional repetitions were suggested by the framework until certain standard deviation across measures was achieved. The study was approved by the Science-Ethics Committee for the Capital Region of Denmark H-16036391. All participants gave written informed consent and received financial compensation for their participation.

3.2.1 Analysis of test reliability

The test-retest reliability of the test battery was assessed using intraclass correlation coefficients (ICC; Koo and Li, 2016) and the standard error of measurement (SEM; Stratford and Goldsmith, 1997) . It was of special interest to test the reliability in older listeners with different hearing abilities. Therefore, test-retest measurements were performed with a subgroup consisting of 11 participants for all tests of the test battery. The seven listeners had bilateral hearing loss with a mean PTA of 31 dB HL. The participants were aged between 59 and 82 years (median 69 years). The retest session was conducted within four months after the first visit.

3.3 Overview of the test battery

Table 3.1: List of the tests included in the BEAR test battery and their corresponding auditory domains.

Test Name	Category	Variables
Pure-tone audiometry	Audibility	AUD_x
Fixed level frequency threshold (eAUD-HF)		FLFT
Word recognition scores (WRS-4UFC)	Speech Perception	SRT_Q, maxDS
Hearing in noise test		SRT_N, $SScore^{4dB}$
Adaptive categorical loudness scaling	Loudness perception	MCL_x $Slope_x$ $DynR_x$
Spectro-temporal modulation test	Spectro-temporal processing	$sSTM_8$, $fSTM_8$
Extended audiometry in noise (eAUD-N, eAUD-S, eAUD-T)		TiN_x SMR_x, TMR_{xF}.
Maximum frequency for IPD detection	Binaural processing abilities	IPD_{fmax}
Binaural pitch test		BP20
Extended binaural audiometry in noise		BMR

AUDx: Pure-tone average at low (x=LF; f ≤ 1kHz) or high (x=HF; f > 1kHz) frequencies. // ACALOS outcome variabless are averaged for low (x=LF; f ≤ 1kHz) and high (x=HF; f > 1kHz) frequencies. // Extended audiometry outcome measures were measured at 0.5 kHz (x=LF) and at 2 kHz (x=HF).

The proposed tests are divided into six categories. Table 3.1 shows the tests and the corresponding auditory domains. The following sections present all tests individually and the experimental method. The summary statistics of the outcome measures presented in Table 3.2. The dataset is publicly available in a Zenodo repository (Sanchez-Lopez et al., 2019). More details about the method can be found in the supplementary material in the data repository.

Table 3.2 – Summary statistics of the outcome measures of the BEAR test battery for the NH and HI group. The results are presented in terms of mean, standard deviation (SD) and the 1st (Q1) and 3rd quantiles (Q3) for the right ear (RE), left ear (LE) or both ears (Bin). In the case of frequency-specific examination, the frequency range is either low (LF) or high (HF).

			NH			***HI***		
Outcome measure	***Freq. Range***	***Ear***	***Mean (SD)***	***Q1***	***Q3***	***Mean (SD)***	***Q1***	***Q3***
SRTQ (dB)		*LE*	*19.9 (7.1)*	*16.5*	*19.2*	*41.5 (13.5)*	*31.8*	*50.6*
		RE	*23.3 (8.9)*	*17.2*	*29.0*	*42.7 (12.6)*	*33.9*	*51.1*
Max DS (%)		*LE*	*99.2 (1.6)*	*100*	*100*	*97.2 (4.1)*	*95.3*	*100*
		RE	*97.2 (1.8)*	*95.5*	*97.6*	*93.9 (6.4)*	*92.1*	*98.4*
SRT_N *(dB)*		*LE*	*1.0 (0.7)*	*0.4*	*1.5*	*4.1 (3.4)*	*1.4*	*6.7*
		RE	*-0.5 (1.1)*	*-1.0*	*0.0*	*2.6 (3.8)*	*0.0*	*4.2*
$SScore^{+4dB}$ *(%)*		*LE*	*85.0 (11.7)*	*85*	*90*	*60.0 (26.6)*	*40*	*85*
		RE	*91.0 (9.6)*	*90*	*95*	*62.3 (24.0)*	*48.7*	*80*
MCL (dB HL)	*LF*	*LE*	*81.5 (14.8)*	*73.3*	*84.1*	*80.6 (8.4)*	*76.4*	*85.8*
		RE	*76.5 (13.2)*	*70*	*80*	*79.1 (7.9)*	*74.7*	*84.1*
	HF	*LE*	*79.0 (17.6)*	*66.6*	*90.8*	*82.7 (12.3)*	*75.8*	*90*
		RE	*73.8 (17.2)*	*65*	*80*	*80.3 (9.9)*	*74.7*	*87.5*
Slope (CU/dB)	*LF*	*LE*	*0.35 (0.1)*	*0.3*	*0.4*	*0.45 (0.1)*	*0.3*	*0.5*
		RE	*0.36 (0.1)*	*0.3*	*0.4*	*0.48 (0.2)*	*0.3*	*0.5*
	HF	*LE*	*0.45 (0.1)*	*0.3*	*0.4*	*0.84 (0.5)*	*0.5*	*0.9*
		RE	*0.41 (0.1)*	*0.3*	*0.4*	*0.81 (0.4)*	*0.5*	*0.9*
DynR (dB HL)	*LF*	*LE*	*91.5 (16.8)*	*78.3*	*97.5*	*76.7 (15.8)*	*64.5*	*88.3*
		RE	*91.1 (18.8)*	*79.1*	*100*	*73.9 (16.0)*	*61.6*	*86.8*
	HF	*LE*	*77.6 (18.2)*	*72.5*	*85.8*	*50.8 (15.1)*	*40.6*	*60.2*
		RE	*78.6 (17.9)*	*67.5*	*90.8*	*50.7 (15.5)*	*38.9*	*60.4*
sSTM -3dB (d')	*LF*	*Bin*	*2.6 (0.6)*	*2.4*	*3*	*1.7 (1.3)*	*0.4*	*3*
	HF		*1.6 (0.8)*	*1.1*	*2.4*	*0.6 (1.1)*	*-0.3*	*1.4*
fSTM (dB)	*LF*	*LE*	*-7.7 (1.8)*	*-9*	*-7.6*	*-2.8 (2.1)*	*-3.5*	*-0.8*
		RE	*-5.1 (3.1)*	*-7.2*	*-1.6*	*-1.6 (1.3)*	*-2*	*-0.6*
	HF	*LE*	*-8.0 (2.0)*	*-8.6*	*-6.2*	*-2.6 (2.4)*	*-3.8*	*-0.6*

Continued on next page...

			NH			*HI*		
Outcome measure	***Freq. Range***	***Ear***	***Mean (SD)***	***Q1***	***Q3***	***Mean (SD)***	***Q1***	***Q3***
		RE	*-5.6 (3.6)*	*-8.6*	*-2.1*	*-1.9 (1.5)*	*-2*	*-1*
eAUD-HF FLFT (kHz)		*LE*	*10.9 (1.2)*	*10.2*	*11.9*	*7.57 (2.7)*	*5.3*	*10*
		RE	*11.7 (1.1)*	*10.9*	*12.5*	*8.12 (2.3)*	*6.7*	*10.2*
eAUD-N (dB HL)	*LF*	*LE*	*70.4 (4.5)*	*68*	*71.5*	*71.8 (2.6)*	*70.2*	*73.2*
		RE	*69.2 (4.6)*	*65.2*	*72.5*	*72.0 (2.8)*	*69.6*	*74.3*
	HF	*LE*	*71.1 (2.5)*	*69.7*	*72.7*	*74.7 (3.4)*	*72.5*	*76.1*
		RE	*70.8 (3.6)*	*70.5*	*71.7*	*74.2 (3.1)*	*72*	*76.2*
TMR (dB) eAUD (N -T)	*LF*	*LE*	*7.5 (3.4)*	*6*	*7.5*	*7.7 (4.0)*	*6.1*	*10.1*
		RE	*5.2 (3.3)*	*4*	*7.6*	*8.3 (2.7)*	*6.5*	*10.3*
	HF	*LE*	*13.0 (0.6)*	*12.7*	*13.2*	*7.9 (5.0)*	*5*	*11.6*
		RE	*10.7 (3.1)*	*9.1*	*10.2*	*8.1 (5.2)*	*5.1*	*10.7*
SMR (dB) eAUD (N -S)	*LF*	*LE*	*19.3 (3.6)*	*16.5*	*21.7*	*19.6 (17.7)*	*17.7*	*23.2*
		RE	*18.8 (4.6)*	*17*	*21.2*	*20.0 (5.2)*	*16.5*	*23.8*
	HF	*LE*	*26.8 (4.5)*	*27.5*	*29*	*19.3 (9.5)*	*12.1*	*26.3*
		RE	*27.2 (3.7)*	*26.2*	*29.5*	*19.5 (9.9)*	*12*	*26.8*
IPD fmax (kHz)		*Bin*	*0.76 (0.26)*	*0.59*	*0.98*	*0.69 (0.27)*	*0.52*	*0.88*
Bin Pitch 20 (%)		*Bin*	*87.5 (25.0)*	*87.5*	*100*	*80.7 (30.9)*	*70*	*100*
BMR (dB) ($S_0N_0 - S_\pi N_0$)		*Bin*	*16.5 (4.7)*	*13.5*	*17.5*	*14.7 (4.6)*	*12.2*	*17.5*

SRTQ: Speech reception threshold in quiet / Max DS: Maximum speech discrimination score. // SRTN: Speech reception threshold in noise / Score +4: Sentence recognition score at +4 dB SNR // MCL: Most comfortable level / Slope: Slope of the loudness function / DynR: Dynamic range // sSTM: Sensitivity for detecting a spectro-temporally modulated noise at 20log(m) = -3 dB, where m is the modulation depth / fSTM: Fast version of the STM test (Bernstein et al., 2016) // eAUD-HF: Fixed-level frequency threshold (FLFT) at 80 dB SPL // eAUD-N: Tone detection in TEN noise // TMR: Temporal masking release // SMR: Spectral masking release // IPD fmax: Frequency threshold for detecting an interaural phase difference of 180°. // Bin pitch: Binaural pitch detection scores for 20 presentations // BMR: Binaural masking release.

Time efficiency of the test battery

The examiners kept track of the time used by each of the participants in completing the test battery. In the case of unexpected events (e.g., unexpected or incongruent results), these events were cautiously annotated for later investigation. Regarding the test procedure, additional repetitions of the threshold estimations were need if: 1) a repetition was considered as an outlier if a given threshold was greater than three scaled median absolute deviations of the two repetitions; or 2) the responses of the listeners during the tracking procedure were inconsistent or reached the maximum or minimum possible values. In that case, the measurement was considered an invalid or "missing" data point.
The timing annotated of the individual tests are shown in Figure 3.2. Besides, the probability of needing an additional measurement and mean number of extra repetitions per listener are shown in Table 3.3. The repetitions were only suggested when the test was done using the AFC framework, i.e. the IPD test, the STM test and the eAUD test in all the conditions.

The total testing time was approximately 2'5 hours excluding the initial interview, information about the study and preparations.

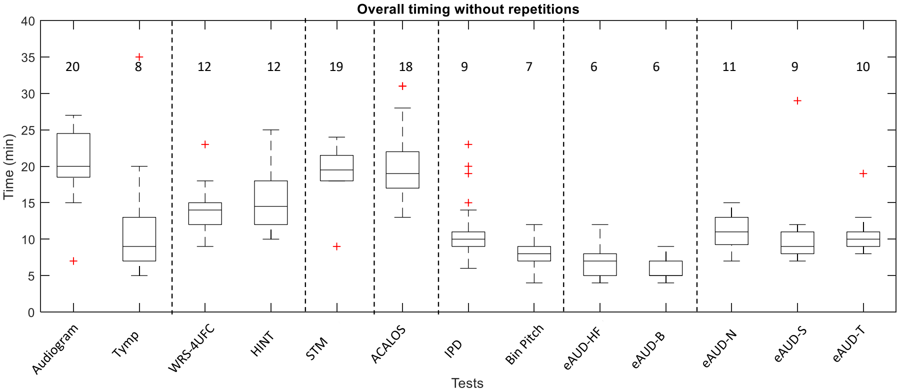


Figure 3.2: The overall time of the different tests in the test battery including the instructions. The data corresponds to the annotations of the examiners. The basic examination with the audiometry and the tympanometry (Tymp) are included. The numbers represent the rounded median in minutes.

Table 3.3: Table with the probability of needing repetitions (PR), and the probability of having missing values (PM). The total probability of repetitions (PT). The mean number of extra repetitions (E. Rep).

Test	PR (%)	PM (%)	PT (%)	E.Rep.
STM	42.86	90.79	88.16	4.32
IPD	10.77	10.97	20.55	1.87
eAUD-HF	5.63	4.05	9.46	1.85
eAUD-N	66.67	46.58	82.19	3.00
eAUD-S	48.57	52.70	75.68	3.07
eAUD-T	53.85	46.58	75.34	3.27
S_0N_0	42.59	27.03	58.11	2.00
$S_\pi N_0$	20.59	9.11	27.03	1.85

3.4 Speech perception in quiet

Method

The word recognition score (WRS-4UFC) test was proposed as a systematic and self-administered procedure that allows the estimation of supra-threshold deficits in speech perception in quiet. The speech material was the same as the one used for standard speech audiometry (Dantale I; Elberling et al., 1989) in Danish. The self-administered procedure consisted of a 4-interval-unforced-choice paradigm (4UFC). After the acoustical presentation of each word, the target written word was placed randomly in one of four intervals. The other three written words were also taken from the Dantale-I corpus. They were chosen based on the lowest Levenshtein phonetic distance (Sanders and Chin, 2009) from the target. Four lists of 25 words were presented at 40, 30, 20 and 10 dB above the individual PTA, in this order. A logistic function was fitted to the results from each individual ear and the speech reception threshold (SRTQ) and maximum speech discrimination score (Max DS) were estimated using psignifit 4 software (Schütt et al., 2016).

Results and discussion

The HI listeners' SRT_Q were, on average, 20 dB higher than the ones of the NH group. The interquartile range for the HI group was about 19 dB whereas for the NH group it was 3 dB for the left ear (LE) and 11.8 dB for the right ear (RE). The Max DS for both groups was close to 100%. However, the HI listeners showed larger variability, especially in the right ear (SD= 6.42%). In the analysis of the test-retest variability, the WRS-4UFC test showed poor to moderate reliability especially at low levels (PTA + 10 dB; ICC = 0.25). However, at the higher presentation levels (i.e. individual PTA + 40 dB) the standard error of the measurement was only 4% (1 word). Regarding clinical applicability, the WRS-4UFC needs to be compared to traditional speech audiometry to explore the influence of using closed- vs. open-set and forced- vs. unforced-choice test procedures on the results.

3.5 Speech perception in noise

The Hearing in Noise Test (HINT; Nilsson et al., 1994) is an adaptive sentence recognition test carried out with speech-shaped noise. The following assumptions are considered in HINT (based on Plomp, 1978): 1) Speech materials made of meaningful sentences yield a steep psychometric function; 2) Stationary noise with the same spectral shape as the average spectrum of the speech material makes the speech reception threshold in noise ($\mathrm{SRT_N}$) less dependent of the spectral characteristics of the speaker's voice. Furthermore, the signal-to-noise ratio (SNR) between the target and masker is better defined across the frequency range; 3) The ($\mathrm{SRT_N}$) is independent of the absolute noise level as long as the noise level is above the "internal noise" level. Therefore, it is recommended to present the noise at least 30 dB above the "internal noise". The internal noise is defined as the sum of the SRT in quiet of the tested listener and the SRT in noise for NH listeners, for a given speech material.

Methods

The Danish HINT was used as in Nielsen and Dau (2011) to obtain the $\mathrm{SRT_N}$. Additionally, a 20-sentence list was presented at a fixed signal-to-noise ratio of +4 dB and scored to obtain a sentence recognition score ($\mathrm{SScore^{+4dB}}$). The presentation level of the noise was set between 65 and 85 dB SPL to ensure that the noise was always presented 30 dB above the individual PTA. Each ear was tested individually. All participants were tested using the same list with the same ear. Since small differences across lists were found in Nielsen and Dau (2011), this was done to ensure that all the listeners were tested with an equally difficult list. However, for the test-retest reliability study, the list and ear presented were randomized, only using lists 6-10. The listeners did not report previous experience with the test.

Results and discussion

The SRT_N for NH listeners were, on average, 2 dB higher than the ones reported (Nielsen and Dau, 2011). However, this might be explained by the fact that they used diotic presentation which can lead to a 1.5 dB improvement (Plomp and Mimpen, 1979). The results also showed a lower SRT_N (1.5 dB) and higher $\text{SScore}^{+4\text{dB}}$ (4%) for the right ear in both groups of listeners. According to (Nielsen and Dau, 2011), there was a significant main effect of test list. Such differences are seen mainly for lists 1-4, which were the lists used here. Therefore, the observed interaural difference can be ascribed to a list effect, however, it might be ascribed to other factors. The ICC values (SRT_N: ICC = 0.61; $\text{SScore}^{+4\text{dB}}$: ICC = 0.57) indicated only moderate reliability of the HINT. The SRT_N showed an SEM = 1.02 dB, which is below the step size of the test (2 dB). The $\text{SScore}^{+4\text{dB}}$ showed an SEM value of 7.94%, which corresponds to an error in one of the sentences.

3.6 Loudness perception

Loudness perception can substantially differ between NH and HI listeners and has been connected to the peripheral non-linearity (e.g. Jürgens et al., 2011). While the growth of loudness shows a non-linear behaviour in a healthy ear, the results from HI listeners suggest that loudness perception becomes linear when outer-hair cell (OHC) function is affected (e.g Moore, 2007). Besides, the possibilities of characterizing hearing deficits, loudness function can be used for fitting hearing aids (e.g. Oetting et al., 2018). Adaptive categorical loudness scaling (ACALOS; Brand and Hohmann, 2002) is the reference method for the current standard (ISO 16832, 2006) for loudness measurements.

Methods

According to the ACALOS method, 1/3-octave band noise were presented sequentially, and the participant had to judge the perceived loudness using a 11-category scale ranging from “not heard” to “extremely loud”. The presentation level of the next stimulus is calculated based on the previous trials. The raw results, which correspond to categorical units (CU) spanned between 0 and 50, were fitted to a model of loudness as described in (Oetting et al., 2014). The outcome measures of the ACALOS presented here are the most comfortable level (MCL), the slope of the loudness function (Slope), and the dynamic range (DynR) defined as the difference between uncomfortable level (50 CU) and the hearing threshold (0.5 CU).

Low-frequency (LF) average corresponds to frequencies below 1.5 kHz, high-frequency (HF) average correspond to frequencies above 1.5 kHz

Results and discussion

The average MCL estimate ranged between 73 and 82 dB HL in both groups and for both frequency ranges. The average slope of the loudness growth was slightly steeper for the HI listeners in the low-frequency range (0.45 CU/dB for HI vs. 0.35 CU/dB for NH) and substantially steeper in the high-frequency range (0.8 CU/dB for HI vs 0.45 CU/dB for NH). The average dynamic range was between 80 and 90 dB HL for the NH listeners, and smaller for the HI listeners, especially at high frequencies (50.8 dB). Regarding the test-retest reliability, ACALOS showed an excellent reliability for estimating the hearing thresholds (ICC = 0.94; SEM = 4.5 dB), good reliability for estimating the MCL (ICC = 0.68, SEM = 6.5 dB) and very good reliability for estimating the slope (ICC = 0.82; SE M = 0.07 CU/dB). Overall, these results supported the inclusion of ACALOS in a clinical test battery, as it provides several outcomes (hearing thresholds, growth of loudness, MCL and dynamic range). ACALOS also showed a high time efficiency (around 10 min. per ear).

3.7 Spectro-temporal modulation sensitivity

A speech signal can be decomposed into spectral and temporal modulations. While speech-in-noise perception assessment leads to some confounds due to the variety of speech corpora, noise maskers, and test procedures that can all affect the results, the assessment of the sensitivity of simpler sounds might be of interest for characterizing a listener's spectro-temporal processing abilities. Bernstein et al. (2013) showed significant differences between NH and HI listeners for detecting STM in random noise. These differences corresponded to specific conditions that were also useful for the prediction of speech-in-noise performance in the same listeners. Lately, the assessment of STM sensitivity in these specific conditions gained an increasing interest due to its potential for predicting speech intelligibility (Bernstein et al., 2016; Gallun et al., 2018; Zaar et al., 2019) and for assessing cochlear-implant candidacy (Choi et al., 2016). Here, STM sensitivity was assessed using a new test paradigm that may be more suitable for a clinical implementation. The test was performed in two conditions: a low-frequency condition (similar to the one previously used in Bernstein et al., 2016) and a high-frequency condition (Mehraei et al., 2014).

Methods

The stimuli were similar to those of Bernstein et al. (2016) and Mehraei et al. (2014), but a different presentation paradigm was employed. A sequence of four noises was presented in each trial. The first and third stimulus always contained unmodulated noise, whereas the second and fourth stimuli could be either modulated or unmodulated. The stimuli were presented at 75 dB sound pressure level (SPL). After the sequence was presented, the listener had to respond whether the four sounds were different ('yes') or the same ('no'). Two procedures involving catch trials were evaluated. The first test (sSTM -3 dB) was a screening test consisting of 10 stimuli modulated at 20log(m) = -3 dB level, where m is the modulation depth, and five unmodulated ones presented in random order. The outcome measure was the listener's sensitivity (d')[c] in the task. The second test (fSTM) tracked the 80% threshold using the single-interval adjusted matrix (SIAM; Kaernbach, 1990) paradigm.

Results and discussion

The screening STM test shows the sensitivity in terms of d', where the maximum value is d' = 3, i.e. 10 modulated and 5 unmodulated stimuli correctly detected. In the hypothetical case when all the catch trials are detected, the lowest d' value can be -0.3. The NH listeners showed a high sensitivity in the low-frequency condition (d' = 2.6) and a somewhat lower sensitivity in the high-frequency condition (d' = 1.63) corresponding to 65% correct responses. The HI listeners showed a higher variability and a lower sensitivity in the low-frequency condition (≈ 70% correct) and substantially lower sensitivity in the high-frequency condition (0-50% correct responses). The threshold tracking procedure (fSTM) showed results between -9 and -6 dB in the NH group, whereas the HI listeners showed thresholds between -3.50 and -0.5 dB. Although the results of the fSTM low-frequency condition were consistent with Bernstein et al. (2016), the results in the high-frequency condition showed higher thresholds than the ones in Mehraei et al. (2014). This can be ascribed to the higher presentation level used in Mehraei et al. (2014) than in the current test procedure. The fSTM showed an excellent reliability (ICC = 0.91; SEM = 0.93 dB) in the LF condition. However, several HI listeners were not able to complete the procedure for the HF condition. Overall, the use of the SIAM tracking procedure allowed us to obtain accurate thresholds, although additional repetitions were required, especially in the HF condition. This might be because the psychometric

---

[c] d' was defined as $Z(\frac{N_H+0.5}{H+1}) - Z(\frac{N_{FA}+0.5}{FA+1})$ where Z refers to the z-score transformation, H is the total number of target presentations and FA the total number of catch trials.

function for detecting the stimulus can be shallower in this condition or because the 100% detection could not be reached even in the fully-modulated trials. Therefore, a Bayesian procedure being able to estimate the threshold and slope of the psychometric function, such as the Bayes Fisher information gain (FIG: Remus and Collins, 2008), might be more suitable for this type of test.

3.8 Extended audiometry in noise (eAUD)

The extended audiometry in noise (eAUD) is a tone detection test intended to assess different aspects of auditory processing by means of a task similar to pure-tone audiometry. The tone is presented either in noise or in quiet and the listener has to indicate whether the tone was perceived or not. The aspects of auditory processing assessed here are 1) high-frequency audibility, 2) spectral and temporal resolution.

High-frequency audibility

Recently, elevated thresholds at high frequencies (> 8 kHz) have been linked to the concept of "hidden hearing loss" and synaptopathy (Liberman et al., 2016). However, the measurement of audiometric thresholds above 8 kHz is not part of the current clinical practice. The fixed-level frequency threshold (FLFT) has been proposed as a quick and efficient alternative to high-frequency audiometry (Rieke et al., 2017). The test is based on the detection of a tone presented at a fixed level. The frequency of the tone is varied towards high frequencies and the maximum audible frequency at the given level is estimated in an adaptive procedure. Here, a modified version of FLFT, using warble tones presented at 80 dB SPL, was used as the extended audiometry at high frequencies (eAUD-HF).

Spectro-temporal resolution

Frequency and temporal resolution are aspects of hearing that are fundamental for the analysis of perceived sounds. While NH listeners exhibit a frequency selectivity on the order of one third of an octave when using isoinput levels (from Eustaquio-Martín and Lopez-Poveda, 2011; Glasberg and Moore, 1990), HI listeners have typically broader auditory filters leading to impaired frequency selectivity (Moore, 2007). Temporal resolution can be characterized by the ability to "listen in the dips" when the background noise is fluctuating based on the so-called masking release (Festen and Plomp, 1990). Schorn and Zwicker (1990) proposed an elaborated technique for assessing both spectral and temporal resolution using two tests: 1) Psychoacoustical tuning curves and 2) temporal resolution curves. In both cases, the task consists of detecting a pure tone that is masked by noise or another tone while the spectral or temporal characteristics of the masker are varied. Later, Larsby and Arlinger (1998) proposed a similar paradigm, the F-T test, which was successfully tested in HI listeners (Esch and Dreschler, 2011). Here, the spectro-temporal resolution was assessed using a new test. This test is a tone-in-noise detection task consisting of three conditions as sketched in 3.3.

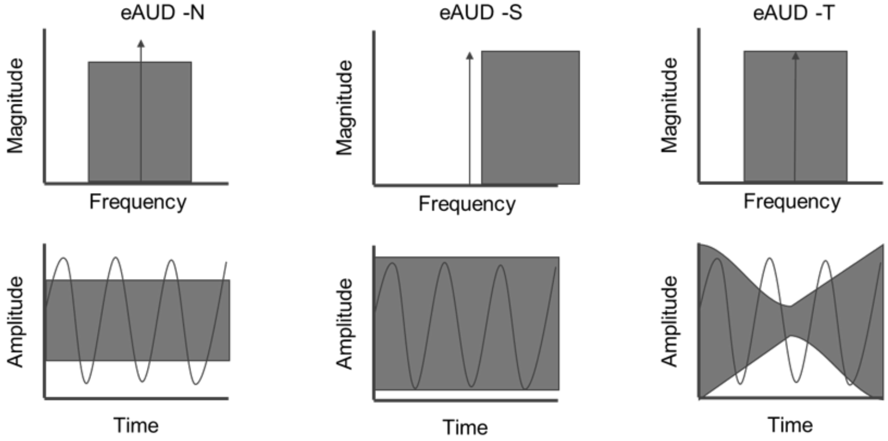


Figure 3.3: Sketch of the conditions of the spectro-temporal resolution measures of the extended audiometry in noise (eAUD). The top panel shows the spectrum of the noise and target pure-tone (delta), the bottom panel shows both signals in the time domain. Left panel: Tone in noise condition (eAUD-N). Middle panel: Spectral condition (eAUD-S). Right panel: Temporal condition (eAUD-T).

1. eAUD-N: The tone is embedded in a 1-octave-wide threshold equalizing noise (TEN; Moore, 2001). Because of the properties of the TEN, the tone detection threshold is comparable to the level of the noise in dB HL.
2. eAUD-S: The tone is embedded in a TEN that has been shifted up in frequency. In the spectral domain, this yields spectral unmasking of the tone, so the detection threshold is lower than in eAUD-N.
3. eAUD-T: The tone is embedded in a temporally-modulated noise with the same spectral properties as the one in eAUD-N. In the temporal domain, the modulations

of the noise yield temporal unmasking, so the tone can be detected in the dips.

The outcome measures were focused on the temporal and spectral benefits expected in the eAUD-S and eAUD-T conditions compared to the eAUD-N condition. While in the noise condition (eAUD-N) the threshold is expected to be approximately at the level of the noise, in the temporal and spectral conditions the thresholds should be lower showing temporal masking release (TMR) and spectral masking release (SMR).

Methods

The procedure used here was a yes/no task using a SIAM procedure (Kaernbach, 1990). As in traditional up-down procedures, the target can be presented in a given trial or not. If the target was detected, the target-presentation level is decreased according to a given step size; if it was not detected, the level is increased. If the stimulus was not presented (catch trial) but the listener provided a positive response, the level is decreased compared to the previous trial. The target stimulus for all the conditions tested here was a warble tone. For each run, the first two reversals were discarded, and the threshold of each trial was calculated as the average of the four subsequent reversals. The noise was presented at 70 dB HL. The low-frequency condition (LF) corresponds to the detection of a 0.5-kHz warble tone, whereas the high-frequency (HF) condition corresponded to a 2-kHz warble tone. The final threshold was calculated as the mean threshold of two repetitions. In the eAUD-S condition the center frequency of the noise was $f_{c,\text{noise}} = 1.1 f_{\text{tone}}$. In the eAUD-T condition the modulation frequency of the noise was set to, $f_m = 4$ Hz. The outcome measures of the eAUD are 1) the high-frequency threshold (eAUD-HF), 2) the tone-in-noise threshold (eAUD-N), 3) the SMR, 4) the TMR.

Results and discussion

The maximum frequency threshold for a tone presented at 80 dB SPL (eAUD-HF) was 11 kHz for the NH listeners and 8 kHz for the HI listeners. The HI group showed larger variability compared to the NH group (interquartile range: 6 kHz vs. 10 kHz). In contrast, the eAUD-N condition showed a larger variance for the NH group (SD = 4.5 dB HL) at low frequencies. The detection thresholds were in line with previous work with thresholds close to the noise presentation level (70 dB HL) (Vinay et al., 2017). The TMR shown by the NH group was larger at high frequencies (10 dB) than at low frequencies (7 dB). The HI group showed, on average, similar TMR only at low frequencies. The SMR shown by the NH listeners was

19 dB for low frequencies and 26 dB for high frequencies. In contrast, for the HI listeners, the SMR was 7 dB lower only in the high-frequency condition. The reliability of the eAUD was moderate for most of the conditions (ICC < 0.75). The eAUD-HF test showed very good reliability (ICC = 0.89; SEM = 495 Hz), and the eAUD-S at low frequencies showed good reliability (ICC = 0.85; SEM = 1.78 dB). The masking release estimates showed good reliability only for the high-frequency condition. The reason for this might be that masking release is a differential measure, and the cumulative error is, therefore, higher than that of each individual measure. The reduced reliability can be explained to some extent by the method used. To have a similar procedure as in pure-tone audiometry, the parameters of the SIAM tracking procedure were set accordingly. However, this made the test challenging and the listeners consistently missed several catch trials. Thus, extra trials were required to improve measurement accuracy. However, the standard error of the measurement was in most cases larger than the final step size (2 dB). As in the case of the fSTM, a different procedure, such as Bayesian adaptive methods, might increase measurement reliability.

3.9 Binaural processing abilities

Binaural hearing is useful for sound localization and the segregation of complex sounds (Darwin, 1997). Interaural differences in level or timing are processed for spatial hearing purposes in the auditory system. In the case of hearing loss, the neural signal at the output of the cochlea can be degraded which may lead to reduced binaural abilities typically connected to temporal fine structure (TFS) processing. Based on a method estimating the upper-frequency limit for detecting an interaural phase difference (IPD) of 180° (IPD_{fmax} Neher et al., 2011; Ross et al., 2007; Santurette and Dau, 2012), Füllgrabe and Moore (2017) recently proposed a refined test as a feasible way to evaluate TFS sensitivity. This paradigm was used in recent research that suggested that IPD_{fmax} might be related to non-auditory factors (Strelcyk et al., 2019) and affected by factors beyond hearing loss, such as musical training (Bianchi et al., 2019). Therefore, the IPD_{fmax} might be a task that requires auditory and non-auditory processing abilities beyond TFS sensitivity. In contrast, binaural pitch detection assesses binaural processing abilities in a different manner. This test requires the detection of pitch contours embedded in noise, which are diotically or dichotically evoked. While the diotic condition can be resolved monoaurally, the dichotic condition requires the binaural processing abilities to be sufficiently intact to detect the contour. Previous studies showed that some listeners were unable to detect binaural pitch, regardless of the audiometric configuration (Sanchez-Lopez et al., 2018a; Santurette and Dau, 2012). Therefore, it was of interest to compare the results of these two binaural processing tests.

Besides the binaural tests presented previously, another approach for evaluating the binaural processing abilities is assessing binaural masking release (Durlach, 1963), which has been used in several studies (Neher, 2017; Strelcyk and Dau, 2009) and implemented in some commercial audiometers (Brown and Musiek, 2013). In this paradigm, a tone-in-noise stimulus is presented in two conditions: (1) a diotic condition where the tone is in phase in the two ears, and (2) a dichotic condition where the tone is in antiphase in the two ears. The difference between the two yields the benefit for tone detection due to binaural processing, the so-called binaural masking release (BMR).

Methods

The maximum frequency for detecting an IPD of 180° with pure-tones was obtained using a 2-AFC tracking procedure similar to the one used in Füllgrabe and Moore (2017). The frequency threshold (IPD_{fmax}) was obtained from the average of two runs. Binaural pitch detection scores were obtained using a clinical implementation of the test proposed by Santurette and Dau (2012). A 3-minute sequence of noise was presented bilaterally. Ten diotic and ten dichotic pitch contours, embedded in the noise, had to be detected by the listener. The tones forming the pitch contours were generated by adding frequency-specific IPDs to the presented noise (Cramer and Huggins, 1958). The outcome measure of the binaural pitch test was the percentage score averaged across two repetitions (BP20). The BMR was assesed using the same method as the extended audiometry. Two measurements were required: 1) tone-in-noise detection presented diotically (S_0N_0) and tone-in-noise detection presented dichotically, i.e., with the tone in anti-phase across the two ears ($S_\pi N_0$).

Results and discussion

The listeners in the NH and HI groups showed IPD_{fmax} thresholds around 700 Hz with a standard deviation (≈ 270 Hz) and interquartile range (≈ 370 Hz) similarly in both groups. These results are in line with the ones reported in Füllgrabe and Moore (2017). The IPD_{fmax} test showed excellent reliability (ICC = 0.95; SEM = 65.4 Hz), and the median time needed for two repetitions was 10 minutes. This suggests that IPD_{fmax} is a reliable measure of binaural processing abilities that can reveal substantial variability among both NH and HI listeners, which is valuable for highlighting individual differences among patients. The overall results from the binaural pitch test for the NH listeners showed > 87.5% correct detection, whereas the HI listeners' results showed a higher variability with an interquartile range from 70-100%. The test showed excellent reliability (ICC = 0.98; SEM = 4%). Listeners reported a positive experience due to the test being short and easy to understand. The BMR shown by both groups was around 15 dB, as expected from previous studies (Durlach, 1963).

3.10 Exploratory analysis

The collection of tests included in the test battery was intended to explore different and potentially independent aspects of hearing to obtain an auditory profile with controlled interrelations among the tests. A factor analysis performed in the HEARCOM study (Vlaming et al., 2011) based on data from 72 HI subjects revealed auditory dimensions: 1) high-frequency processing, 2) audibility, 3) low-frequency processing and 4) recruitment. In the current study, the results of the behavioural tests were analysed further in order to explore possible interrelations between the various outcome measures.

Methods

First, the data were pre-processed as in Sanchez-Lopez et al. (2018a) to reduce the number of variables. The outcome variables of the frequency-specific tests were divided into LF (≤1 kHz) and HF (>1 kHz) variables. This decision was supported by a correlation analysis performed on the complete set of outcome variables, where the outcomes corresponding to 2, 4 and 6 kHz as well as the ones corresponding to 0.25, 0.5 and 1 kHz were highly intercorrelated. For the tests performed monaurally, the mean of the two ears was taken as the resulting outcome variable. The resulting dataset (BEAR3 dataset[a]) contained 26 variables, divided into six groups corresponding to the six aspects of auditory processing considered here. The exploratory analysis consisted of a correlation analysis using Spearman correlations and factor analysis. The factor analysis was performed using an orthogonal rotation ("varimax") and the method of maximum likelihood. The number of components was chosen using parallel analysis, the resulting number of components was four.

---

[a] The BEAR3 available at Zenodo contains an observation labelled '0', which corresponds to the results of one of the examiners and it is not used in the present analysis

Results

Figure 3.4 shows the results from the correlation analysis performed on the BEAR3 dataset. For convenience, the absolute value of the correlation was used when visualizing the data

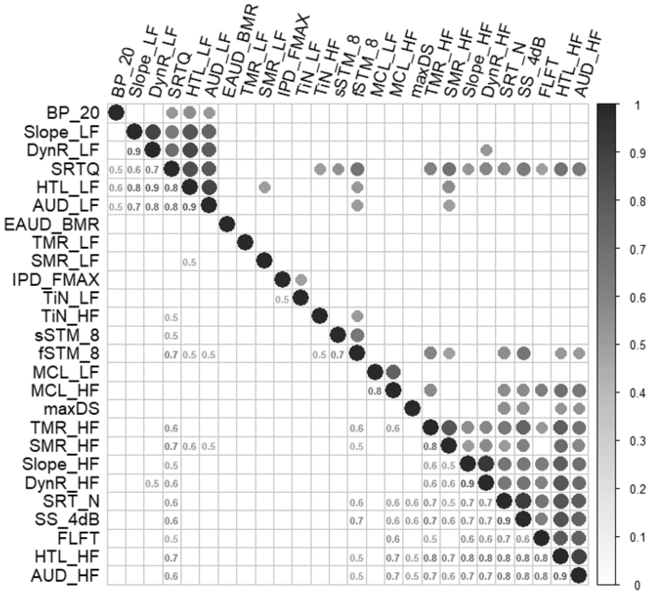


Figure 3.4: Correlation plot of the data set BEAR3. The upper part shows the significantly correlated variables as coloured circles. The lower panel shows the numeric correlation value.

to show the strength of the correlation. The circles on the left-hand side of the figure depict significant correlations (p < 0.00001), and the correlation values are presented on the left-hand side of the figure. Two groups of correlated variables can be observed. The upper-left corner shows variables related to LF processing (dynamic range, the slope of the loudness function, and hearing thresholds) and speech intelligibility in quiet. The bottom-right corner shows a larger group of correlated variables including HF processing, speech intelligibility in noise, and spectro-temporal resolution at high frequencies. The variables that are not significantly interrelated are shown in the middle part of Figure 3, including the three variables related to binaural processing abilities (IPD$_{\text{fmax}}$, BP20 and BMR) which were not significantly correlated to each other. The speech reception threshold in quiet (SRT$_{\text{Q}}$) and the STM detection were correlated to various variables such as tone-in-noise detection, HF spectro-temporal resolution, LF hearing thresholds and speech-in-noise perception.

The four factors resulting from the factor analysis showed 63% of explained cumulative variance. The variables with higher loadings (> 0.65) for each of the factors are shown in Table 3.4. The first factor, in terms of the amount of variance explained (19%), was associated with LF loudness perception and speech intelligibility in quiet, whereas the second factor (18% of variance explained) was associated with HF loudness perception. Despite loudness perception being associated with the first and second factor, the MCL was associated, both at high and low frequencies, with the third factor, while the fourth factor was associated with speech intelligibility in noise.

Table 3.4: Variables correlated to the four latent orthogonal factors resulting from the factor analysis with the method of maximum likelihood (ML). Columns are sorted in terms of the variance explained by each factor.

	ML2(19%)	***ML1(18%)***	***ML3(14%)***	***ML4(12%)***
HTL_LF	*0.93*			
DynR_LF	*-0.9*			
AUD_LF	*0.82*			
Slope_LF	*0.81*			
SRTQ	*0.67*			
DynR_HF		*-0.93*		
Slope_HF		*0.82*		
HTL_HF		*0.79*		
AUD_HF		*0.73*		
MCL_HF			*0.92*	
MCL_LF			*0.85*	
SRT_N				*0.77*
SScore_4dB				*-0.78*

3.11 General discussion

The first goal of the present study was to collect data of a heterogeneous population of HI listeners, reflecting their hearing abilities in different aspects of auditory processing. The current study was motivated by the need for a new dataset to refine the data-driven approach for auditory profiling. The dataset should contain a representative population of listeners and outcome measures (Sanchez-Lopez et al., 2018a) to allow a refined definition of the two types of auditory distortions and to identify subgroups of listeners with clinical relevance. To refine the data-driven auditory profiling, the BEAR3 dataset fulfils all the requirements discussed in Sanchez-Lopez et al. (2018a). Other datasets containing a large number of listeners (e.g., Gieseler et al., 2017; Rönnberg et al., 2016) or physiological measures (e.g., Kamerer et al., 2019) could also be interesting for complementing the auditory profiling beyond auditory perceptual measures.

Relationships across different aspects of auditory processing

The proposed test battery considers outcomes divided into six dimensions of auditory processing. One of the objectives of the study was to investigate the interrelations of different dimensions and measures. The present analysis showed two interesting findings. First, the correlation analysis shows two clusters of variables related to either low- or high-frequency audiometric thresholds. Speech-in-noise perception was associated with high-frequency sensitivity loss, temporal, and spectral masking release whereas speech-in-quiet was correlated with both low- and high-frequency hearing loss. Several outcomes were not interrelated, especially the outcomes associated with binaural processing abilities. Second, factor analysis yielded latent factors related to low- and high-frequency processing, most comfortable level and speech in noise. Vlaming et al. (2011) showed four dimensions in the factor analysis of the HEARCOM project data corresponding to high-and low-frequency spectro-temporal processing, MCL and recruitment. In contrast, the current study showed that the slopes of the loudness growth, both at low and high frequencies, were not interrelated and contributed to the first and second latent factors. Additionally, the speech-in-noise test performed in HEARCOM was associated with the low-frequency processing, whereas, in the present study, speech-in-noise dominates the fourth factor and is significantly correlated with high frequencies. The reason for this discrepancy might be the use of different types of noise and test procedures in the two studies.

Overall, the data of the present study seem to be dominated by the audiometric profiles, with low- and high-frequency processing reflecting the main sources of variability in the data. However, binaural processing abilities, loudness perception and speech-in-noise outcomes showed a greater contribution to the variability of the supra-threshold measures than spectro-temporal processing outcomes.

Towards clinical feasibility of the tests

The test-retest reliability of the test battery was investigated based on the results of a subset of listeners who participated 2-5 months after the first visit. The analysis was based on the ICC and the SEM. Some of the tests, such as IPD_{fmax}, binaural pitch and eAUD-HF (FLFT) showed good to excellent test-retest reliability with all ICC values above 0.9, while other tests, such as the extended audiometry in noise and speech intelligibility in quiet, showed poor reliability. The selected tests were conducted in two sessions and the total time was, on average, three hours including the instuctions and interview. In realistic clinical setups, a subset of tests with high reliability and a reasonably low difficulty would need to be prioritized. For a clinical version of the test battery, other tracking procedures such as Bayesian Functional information (Remus and Collins, 2008) might be adopted to improve the reliability and time-efficiency in some tasks such as STM and tone detection in noise. Moreover, if time-efficiency is crucial, testing some aspects of auditory processing out of the clinic, as other proposed test batteries for auditory research (Gallun et al., 2018), might be a solution for completing the patient's hearing profile. The use of speech-in-noise tests can be a useful tool for the characterization of the listener's hearing deficits that can be performed under different conditions, including monaural, binaural, unaided and aided stimuli presentations. While here the tests were performed monaurally and unaided, a binaural condition as well as at least one aided measure (i.e., with hearing aids), could also be included in clinical practice. A clinical test battery with the subset of tests that showed a good or excellent test-retest reliability should be evaluated in a large scale study. This should include several aspects of auditory processing and provide detailed information on the supra-threshold deficits of the patient. The tests that showed potential for the clinical implementation were ACALOS, HINT, fSTM (LF condition), Bin. Pitch and IPD_{fmax}. Such a test battery could serve to identify clinically relevant subset of patients (auditory profiles) that may benefit from specific types of hearing rehabilitation towards a “stratified approach” (Lonergan et al., 2017) for audiology practice.

3.12 Conclusion

The analysis of the data showed that a reduced BEAR test battery has the potential for clinical implementation, providing relevant and reliable information reflecting several auditory domains. The proposed test battery showed good reliability, was reasonably time-efficient and easy to perform. The implementation of a clinical version of the test battery is publicly available and can be evaluated in future research, e.g. in a larger field study to further refine the auditory profiling approach. Moreover, the current data will be re-analysed in a continuation study to better define the auditory profiles proposed in the data-driven approach and the two types of auditory distortions.

Data availability and suplemental material

The data that support the findings of this study are openly available in Zenodo at `http://doi.org/10.5281/zenodo.3459579`. A clinical implementation of the test battery is publicly available at `https://bitbucket.org/hea-dtu/bear-test-battery/src/master/`. Supplemental material can be found at `http://doi.org/10.1101/2020.02.17.20021949` (Sanchez-Lopez et al., 2020d).

Acknowledgements

We thank the staff from OUH, BBH and HEA, especially JH Schmidt, SS Houmøller, E Kjærbøl, RS Sørensen, and the student helpers from the MSc of Audiology at SDU. We also want to show our gratitude to all the participants in the study. We want to thank I. Holube and two annonymous reviewers for their comments in an earlier version of this chapter, as well as SD Ewert for the AFC framework. This work was supported by Innovation Fund Denmark Grand Solutions 5164-00011B (Better hEAring Rehabilitation project), Oticon, GN Resound, Widex and other partners (Aalborg University, University of Southern Denmark, the Technical University of Denmark, Force, Aalborg, Odense and Copenhagen University Hospitals).The funding and collaboration of all BEAR partners are sincerely acknowledged. No potential conflict of interest was reported by the authors.

4

Robust data-driven auditory profiling for precision audiology [a]

Abstract

Data-driven profiling allows uncovering complex hidden structures in a dataset and has been used as a diagnostic tool in various fields of work. In audiology, the clinical characterization of hearing deficits for hearing-aid fitting is typically based on the pure-tone audiogram only. Implicitly, this relies on the assumption that the audiogram can predict a listener's supra-threshold hearing abilities. Sanchez-Lopez et al. [Trends in hearing vol. 22 (2018), *Chapter 2*] hypothesized that the hearing deficits of a given listener, both at hearing threshold and at supra-threshold sound levels, result from two independent types of "auditory distortions". The authors performed a data-driven analysis of two large datasets with results from numerous tests, which led to the identification of four distinct "auditory profiles". However, the definition of the two types of distortion was challenged by differences between the two datasets in terms of the selected tests and type of listeners included in the datasets. Here, a new dataset was generated with the aim of overcoming those limitations. A heterogeneous group of listeners (N = 75) was tested using measures of speech intelligibility, loudness perception, binaural processing abilities and spectro-

---

[a] This chapter is based on: Sanchez-Lopez, Fereczkowski, Neher, Santurette, and Dau (2020b) *"Robust data-driven auditory profiling for precision audiology"*. Submitted to Trends in Hearing. Preprint at medRxiv: 036442

temporal resolution [a]. The subsequent data analysis allowed refining the auditory profiles proposed by Sanchez-Lopez et al. (2018a). Besides, a robust iterative data-driven method is proposed here to reduce the influence of the individual data in the definition of the auditory profiles. The updated auditory profiles may provide a useful basis for improved hearing rehabilitation, e.g., through profile-based hearing-aid fitting.

4.1 Introduction

Currently, "profiling" has gained broad attention as a tool for typifying groups of observations (e.g. users, recordings or patients) that follow similar patterns. Data-driven profiling can uncover complex structures that are "hidden" in the data. It has been used as a diagnostic tool in various fields (Shah et al., 2019) such as functional imaging (Krohne et al., 2019), genetics (Li et al., 2004), psychology (Gerlach et al., 2018) or logopedics (Sharma et al., 2019). The idea of using computational data analysis that applies principles of the knowledge discovery from databases (KDD; Frawley et al., 1992) has recently gained attention in the field of audiology in connection with hearing-aid features (Lansbergen and Dreschler, 2020; Mellor et al., 2018). As in stratified medicine (Trusheim et al., 2007), which pursues the identification of subgroups of patients (phenotypes) for the purpose of implementing more targeted treatments, it is of interest to identify subgroups of hearing-impaired (HI) listeners who might benefit from targeted hearing-aid fittings. As such, data-driven auditory profiling could help identify groups of listeners that are characterized by specific hearing disabilities and support precision audiology.

Hearing devices are the usual treatment for a hearing loss (Cunningham and Tucci, 2017). Hearing-aid fitting mainly consists of the adjustment of amplification parameters to compensate for audibility loss and impaired loudness perception. Advanced hearing-aid signal processing features such as beamforming and

---

[a] Sanchez-Lopez et al. (2020d) , see *Chapter 3*

noise reduction are typically not individually adjusted in this process, even though they could, in principle, be targeted towards the compensation of supra-threshold hearing deficits (Kiessling, 2001; Neher and Wagener, 2016). However, the characterization of individual supra-threshold hearing deficits can be complex and requires more testing than standard audiometry. The definition of supra-threshold auditory deficits is commonly based on Plomp's (Plomp, 1978) model, where hearing deficits affecting speech intelligibility are comprised of an "attenuation" and a "distortion" component. Whereas the attenuation component is assumed to affect speech intelligibility only in quiet, the distortion component is assumed to do so also in noise, yielding elevated speech reception thresholds in both cases. Kollmeier and Kiessling (2018) extended Plomp's approach and suggested a model that includes an attenuation component (affecting pure-tone sensitivity), a distortion component (affecting speech intelligibility in noise), and a neural component (affecting binaural processing abilities). Their model assumes that a sensorineural hearing loss is characterized by several factors: an "audibility loss", a "compression loss", a "central loss" and a "binaural loss". In general, these modelling approaches (Kollmeier and Kiessling, 2018; Plomp, 1978) are rather conceptual and do not pinpoint specific underlying impairment factors nor specific measures to quantify these types of losses.

There have been some attempts to stratify HI listeners based on the shapes of their audiograms. Several classification schemes have been proposed in earlier studies, some of which were based on data-driven approaches (Bisgaard et al., 2010; Chang et al., 2019; Parthasarathy et al., 2020). Based on results from human temporal bone studies, Schuknecht and Gacek (1993) proposed four different types of age-related hearing loss: sensory presbycusis, neural presbycusis, metabolic presbycusis and mechanical presbycusis. Sensory presbycusis was related to alterations in the Organ of Corti and typically associated with basilar membrane compression loss, reduced frequency selectivity and elevated audiometric thresholds. This type of age-related hearing loss was considered to reflect the loss of outer hair cells (OHC; Ahroon et al., 1993) and/or inner hair cells (IHC; Lobarinas et al., 2013) and was characterized by sloping audiograms.

Neural presbycusis was related to a substantial loss of nerve fibers in the spiral ganglion. This type of presbyacusis was characterized by a progressive loss of speech discrimination performance, even though the audiometric thresholds remained unchanged over the same time period. Metabolic presbycusis was related to the atrophy of the stria vascularis that affects the transduction of the sensory cells because of a decreased endo-cochlear potential (EP). This type of impairment was associated with flat audiograms and did not affect speech discrimination (Pauler et al., 1986). Finally, conductive presbycusis corresponded to a gently sloping hearing loss at high frequencies, not reflecting morphological alterations in the sensory cells or stria vascularis but yielding elevated thresholds. This type of presbyacusis might reflect an atypical organization in the organ of Corti that affects its mechanical properties (Motallebzadeh et al., 2018; Raufer et al., 2019). However, recent results obtained with new techniques developed for histopathological analysis suggested that OHC dysfunction might have been underestimated in the case of conductive presbycusis and for some of the other types of age-related hearing loss (Wu et al., 2020).

Animal studies, where selective damage to the sensory cells or a change of the EP was induced, have allowed a consistent definition of the metabolic and sensory types of impairments in terms of audibility loss (Ahroon et al., 1993; Lobarinas et al., 2013; Mills et al., 2006). Dubno et al. (2013) proposed a classification into sensory and metabolic audiometric phenotypes based on an approach that combined findings from animal models, expert medical advice and data-driven techniques. The main goal of their study was to analyze a large database of audiograms of HI individuals, and to identify connections between the findings from the animal studies with induced hearing losses and those based on human data. Whereas Schuknecht and Gacek (1993) characterized the metabolic and sensory types of presbyacusis in terms of physiological impairments observed in humans, Dubno et al. (2013) proposed a phenotypical classification of the audiograms of HI listeners. Dubno *et al.*'s classification was thus solely based on the shape of the pure-tone audiogram. While this may help predict the possible origin of a listener's audibility loss, supra-threshold auditory processing deficits cannot be inferred from their phenotypes. The perceptual consequences of sensory or metabolic presbyacusis beyond audibility loss have not yet been studied.

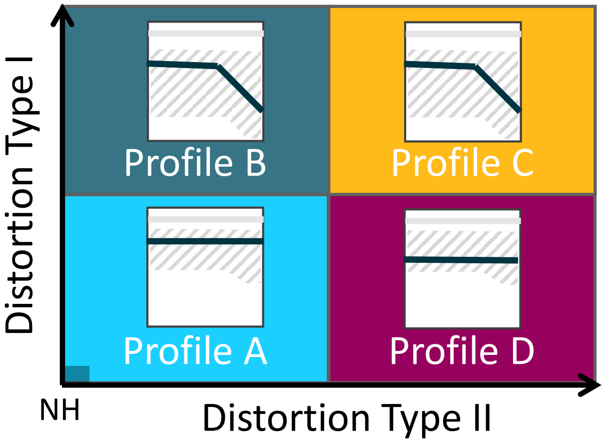


Figure 4.1: Sketch of the hypothesis. The hearing deficits of a given listener can be described as a combination of two independent perceptual distortions. In a two-dimensional space, there would be four subgroups of listeners (Profiles A-D), which exhibit different degrees of the two distortion types.

We hypothesize that a listener's hearing deficit can be characterized by two independent types of "auditory distortions", type-I and type-II, as illustrated in Figure 4.1. In this two-dimensional space, a normal-hearing (NH) listener would be placed at the origin whereas other listeners, with hearing losses that differ in the degree of the two types of distortions would be placed at different positions along the two dimensions. Each type of distortion would then be defined by specific deficits observed in behavioral tasks that covary together and define a given auditory profile. While Profile C represents a high degree of both types of distortion, profiles B and D reflect hearing deficits dominated by one of the two distortions. Profile A, the group with a low degree of distortions, represents near-normal hearing and thus only mild hearing deficits.

Recently, Sanchez-Lopez et al. (2018a) proposed a data-driven method for auditory profiling that was tested and verified by analyzing two datasets from previous experimental studies (Johannesen et al., 2016; Thorup et al., 2016). The method was tailored to the hypothesis of the four auditory profiles. In their study, it was hypothesized that distortion type-I covaries with a loss of audibility, whereas distortion type-II was assumed to be unrelated to audibility. However, the results of the analysis of two different datasets did not support this hypothesis. In fact, the analysis of the two datasets showed that distortion type-I was connected to high-frequency hearing loss and reduced speech intelligibility. Regarding distortion type-II, the analysis of one of the datasets (Thorup et al., 2016) provided a link to reduced binaural processing abilities, whereas the analysis of the other dataset (Johannesen et al., 2016) was linked to low-frequency hearing loss. These mixed results were attributed to differences between the two datasets in terms of the selection of the listeners and chosen behavioral tests. The authors concluded that a new dataset that included a larger variability of impairment factors across listeners was needed to better characterize the listeners' auditory distortions and, thus, the auditory profiles. Furthermore, they suggested that the chosen tests should investigate several aspects of auditory processing while at the same time be clinically feasible.

In this study, a new dataset was therefore generated with the aim of overcoming these limitations. Seventy-five listeners were tested in a clinical environment. The behavioural tasks included measures of audibility, loudness perception, binaural processing abilities, speech perception, spectro-temporal modulation sensitivity and spectro-temporal resolution (Sanchez-Lopez et al., 2020d). These outcomes include several measures that can be connected to previous approaches such as the attenuation-distortion model (regarding speech perception measures) and the neural component (regarding binaural processing abilities). Therefore, it was of interest to further investigate the connections between outcome measures and the two distortion types in a data-driven approach. The analysis of the new dataset was performed with a refined version of the data-driven method provided in Sanchez-Lopez et al. (Sanchez-Lopez et al., 2018a). Importantly, the current study did not aim to disentangle the effects of audibility and supra-threshold deficits but to identify four robust listener subpopulations based on the data-driven analysis of the new dataset. The outcomes of the analysis were discussed in relation to previous classification approaches as well as in terms of implications towards profile-based rehabilitation strategies. Moreover, a decision tree consisting of the auditory measures that best classified the listeners into the four profiles was generated.

4.2 Method

The development of the data-driven method for auditory profiling was based on two premises: 1) the identifaction of relevant outcome measures that tap into two independent sources of variation, and 2) the identification of extreme exemplars that can serve as “prototypes” of different subgroups of listeners.

Description of the dataset

Seventy-five listeners participated in the study. Seventy of the listeners presented various degrees and shapes of symmetrical, sensorineural hearing losses, while five showed near-normal audiometric thresholds. Additionally, one young normal-hearing listener (participant 0) with experience with the tests was included for the analysis as suggested in Sanchez-Lopez et al. (2018a). The listeners were recruited from the clinical databases at Odense University Hospital (OUH), Odense, Denmark and Bispebjerg Hospital (BBH), Copenhagen, Denmark and the Hearing System Section of the Technical University of Denmark (DTU), Kgs Lyngby, Denmark. All listeners completed the “BEAR test battery” (Sanchez-Lopez et al., 2020d). This test battery consists of a total of 10 psychoacoustic tests. The tests are divided into six aspects of auditory processing: audibility, speech perception, loudness perception, binaural processing abilities, spectro-temporal modulation sensitivity and spectro-temporal resolution.

The tests were carried out in a double-walled booth (at BBH and DTU) or in a small anechoic chamber (at OUH). The stimuli were presented via headphones (Sennheiser HDA200). For each of the tests, the outcome measures of interest were extracted from the raw results. For example, the speech reception threshold (SRT) in quiet was estimated from the word discrimination scores obtained at different speech levels. When the tests contained frequency-specific measures, the results were grouped into low-frequency ($\leq$ 1 kHz) and high-frequency ($>$ 1 kHz) averages. This decision was motivated by previous studies (Bernstein et al., 2016; Sanchez-Lopez et al., 2018a). In the case of monaural measures, the mean values across ears were used. The data were cleaned following the principles of KDD, to remove outliers or unreliable data before the analysis. For example, some of the listeners performed the speech-in-noise test at lower levels than the level recommended for the hearing-in-noise measurements (Nielsen and Dau, 2011). Since speech-in-noise perception is of great interest in the present analysis, unreliable measurements of speech reception thresholds in noise ($\mathrm{SRT_N}$) and sentence recognition scores ($\mathrm{SS^{4dB}}$) were considered as missing data. In the next step, the data were normalized between the 25th and 75th percentiles, such that the 25th percentile corresponded to a value of -0.5 and the 75th to a value of 0.5. In total, 26 variables were selected from the outcome measures, as shown in Table 4.1. The resulting dataset ('BEAR3') is publicly available (Zenodo `doi:10.5281/zenodo.3459579;` Sanchez-Lopez et al., 2019).

Stages of the data-driven method

As in Sanchez-Lopez et al. (2018a), the data-driven analysis used here was based on unsupervised learning and was divided into three main steps illustrated in the top panel of Figure 4.2:

I. Dimensionality reduction: Based on principal component analysis (PCA), a subset of variables that were highly correlated with the first two principal components, PC1 and PC2, was kept for the following steps (II and III). The subset could consist of 3, 4 or 5 variables per PC. Hence, up to 10 variables could be kept for the next step. The to-be-kept variables were chosen in an iterative process using a leave-one-out cross-validation. In each iteration, one variable was removed according to the variance explained by the remaining variables, i.e., the subset of variables that explained the

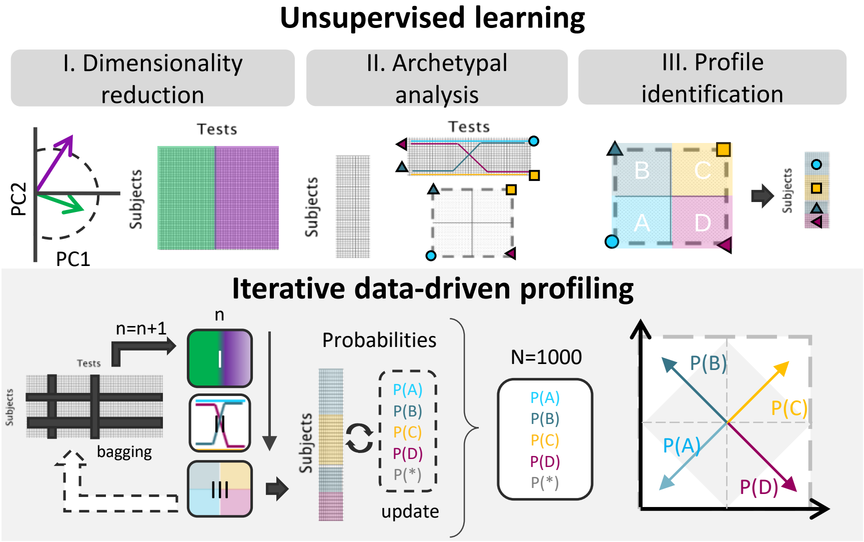


Figure 4.2: Sketch of the refined data-driven method for auditory profiling. Top panel: The unsupervised learning stages of Sanchez-Lopez et al., 2018a: I) dimensionality reduction; II) archetypal analysis; III) profile identification. Bottom panel: In each iteration, a subset of the dataset was processed using dimensionality reduction, archetypal analysis and profile identification. The profile identification stage was two-fold: 1) In each of the iterations, the profiles were identified based on the archetypal analysis. 2) After 1000 iterations, the probability was calculated based on the prevalence of each observation and the number of identifications as each of the profiles. Listeners with higher probabilities of belonging to an auditory profile were placed close to the corners in the square representations and the ones with lower probability ($P < 0.5$) were located inside the grey square.

largest amount of variance was kept and the left-out variable was discarded. Additionally, since the use of several intercorrelated variables in PCA can bias the results, highly correlated variables were removed. If two variables resulting from step I were highly correlated (Pearson's correlation coefficient, $r > 0.85$), one of them was dropped and this step was repeated.

II. Archetypal analysis: The data were decomposed into two matrices – the 'test matrix', which contained the extreme patterns of the data (archetypes) and the 'subject matrix', which contained the weights for each archetype. A given subject was then represented as a convex combination of the archetypes (Cutler and Breiman, 1994). The specific method used here was similar to the one proposed in Mørup and Hansen (2012). The analysis was limited to four archetypes to improve the interpretability of the results on the scope of the hypothesis.

III. Profile identification: The subject matrix was used to estimate the distance between observations and the four archetypes. Each listener (subject) was then assigned to an auditory profile group based on their weights in the subject matrix. The sum of weights for each listener was always 1. Listeners with a weight above 0.51 for one of the four archetypes were identified as belonging to that auditory profile (Ragozini et al., 2017). Otherwise, they were left "unidentified" ('U').

Iterative data-driven profiling

The robust data-driven auditory profiling method aimed to improve the previous method proposed in Sanchez-Lopez et al. (2018a) by reducing the influence of the data on the definition of the auditory profiles. In any data-driven analysis, and especially in unsupervised learning, individual data points can highly influence the results and lead to misinterpretations. Resampling techniques such as bagging are commonly used for supervised learning. Moreover, bagging can improve cluster analysis, making the results less sensitive to the type and number of variables (Dudoit and Fridlyand, 2003). The three unsupervised learning steps were repeated 1000 times, as illustrated in the bottom panel of Figure 2. Before each repetition, the full dataset was decimated randomly in terms of subjects and tests in each iteration. The analysis was performed with only 83% of the data (69 out of 75 listeners and 24 out of 26 variables) in each repetition. In the case of missing data, an algorithm based on spring metaphor was used to predict those data points. Furthermore, in step I (dimensionality reduction), the number of selected variables (6, 8 or 10) was also randomly selected in each iteration to further randomize the procedure. Steps II (archetypal analysis) and III (profile identification) yielded a pre-classification of the subjects contained in the subset of the data corresponding to each iteration. The probability of each listener of being identified as a given auditory profile depended on the number of times a given listener was “out-of-bag” in individual repetitions and the profile identification result from step III. In each iteration, the profile probabilities [P(A), P(B), P(C) or P(D)] and the probability of being unidentified [P(U)] were updated.

After 1000 repetitions, the listeners were divided into four subgroups based on the computed probabilities. If a given listener showed a probability above 0.5 of belonging to any of the auditory profiles, the listener was assigned to that profile. However, if the highest probability was below 0.5, the criterion for being included in one of the four clusters was that the difference between the two highest probabilities had to be above 0.1 to be considered significant. The projection of the probabilities on a two-dimensional space was done by considering four vectors, one for each profile probability, pointing towards each of the corners in a squared representation, as depicted in the right-bottom panel of Figure 4.2. Graphically, the listeners belonging to an auditory profile were then placed close to the corners.

Distortion estimation from the square representation

The final output of the refined data-driven method was the probability, P, of being identified as belonging to an auditory profile (A-D). Regarding the square representation or convex hull, which resembled the hypothesis shown in Figure 4.1, the probabilities of belonging to an auditory profile were depicted as vectors with the origin at the center of the square and oriented towards each of the four corners (Figure 4.2). Assuming that P(B) and P(C) are proportional to auditory distortion type-I (AD_I) and that assuming that P(C) and P(D) are proportional to auditory distortion type-II (AD_{II}), this yields:

$$AD_I \approx \frac{1}{2}\big(1 + P(B \cup C) \; - \; P(A \cup D)\big), \tag{4.1}$$

$$AD_{II} \approx \frac{1}{2}\big(1 + P(B \cup C) \; - \; P(A \cup D)\big). \tag{4.2}$$

Each listener was placed in the two-dimensional space along with the two estimated distortions. In addition, prototypes reflecting extreme exemplars, equivalent to the archetypes yielded by the archetypal analysis, were estimated by averaging the results of the five listeners with the highest probabilities of belonging to a given auditory profile (A-D). The relations between the $\mathrm{AD_I}$ and $\mathrm{AD_{II}}$ with the variables considered in the study were investigated using stepwise linear regression models. The variables included in the model fitting were the outcome variables resulting from the supra-threshold tests, except for $\mathrm{AUD_{LF}}$ and $\mathrm{AUD_{HF}}$, and listeners with a high probability of not being identified as any of the four profiles ($P(U) > 0.5$) were discarded. The criterion for adding a variable as a predictor of one of the distortions was an improvement of the adjusted R^2 by more than 0.01.

Decision trees

A decision tree was fitted to the entire dataset following the splitting criterion of weighted impurity (Breiman et al., 2017). Since it was of interest to obtain a decision tree with outcome measures beyond audiometry, the variables from the pure-tone audiometry were excluded from this analysis. The resulting decision tree was pruned to only have three levels and a maximum of seven binary splits. Because of the missing data, the decision tree was surrogated, i.e., it ignored the missing data to facilitate its interpretability.

4.3 Results

Overview of the dataset

The dataset (Sanchez-Lopez et al., 2019) consisted of 26 outcome variables corresponding to 75 listeners with different hearing abilities. Table 4.1 summarizes the outcome variables used in the analysis.

There were six variables related to audibility (AUD) and loudness perception (LOUD): 1) pure-tone average at low frequencies ($\mathrm{AUD_{LF}}$; $f \leq 1\mathrm{kHz}$) and at high frequencies ($\mathrm{AUD_{HF}}$; $f > 1\mathrm{kHz}$); 2) fixed-level frequency threshold (FLFT) measured at 80 dB sound pressure level (SPL); 3) hearing threshold levels (HTL) estimated from

Table 4.1: Description of the tests, dimensions and outcome measures contained in the BEAR3 dataset (Sanchez-Lopez et al., 2019). For each test, a reference is included. The tests are divided by categories, and the outcome variables are presented in the right column.

Test Name	Category	Variables
Pure-tone audiometry[1]	Audibility	$\mathrm{AUD_x}$
Fixed level frequency threshold[2]		FLFT
Adaptive categorical loudness scaling[3]		$\mathrm{HTL_x}$
	Loudness perception	$\mathrm{MCL_x}$
		$\mathrm{DynR_x}$, $\mathrm{Slope_x}$
Word recognition scores[4]	Speech Perception	$\mathrm{SRT_Q}$, maxDS
Hearing in noise test[5]		$\mathrm{SRT_N}$, $\mathrm{SS^{4dB}}$
Maximum frequency for IPD detection[6]	Binaural processing abilities	$\mathrm{IPD_{fmax}}$
Binaural pitch[7]		BP20
Extended binaural audiometry in noise[8]		BMR
Spectro-temporal modulation test[9]	Spectro-temporal processing	$\mathrm{sSTM_8}$, $\mathrm{fSTM_8}$
Extended audiometry in noise [10,11,12]		$\mathrm{TiN_x}$
		$\mathrm{SMR_x}$, $\mathrm{TMR_{xF}}$.

AUDx: Pure-tone average at low (x=LF; f ≤ 1kHz) or high (x=HF; f > 1kHz) frequencies. // ACALOS outcome variabless are averaged for low (x=LF; f ≤ 1kHz) and high (x=HF; f > 1kHz) frequencies. // Extended audiometry outcome measures were measured at 0.5 kHz (x=LF) and at 2 kHz (x=HF).

[1]ISO 8253-1 (2010); [2] Rieke et al. (2017); [3]Brand and Hohmann (2002), [4]ISO 8253-3 (2012) ; [5]Nielsen and Dau (2011); [6]Füllgrabe et al. (2017); [7]Santurette and Dau (2012); [8]Durlach (1963); [9]Bernstein et al. (2016); [10]Moore et al. (2000) ; [11]Schorn and Zwicker (1990) ; [12]Esch and Dreschler (2011)

the loudness function, averaged for low ($\mathrm{HTL_{LF}}$) and high ($\mathrm{HTL_{HF}}$) frequencies; 4) most comfortable level (MCL) estimated from the loudness function, averaged for low ($\mathrm{MCL_{LF}}$) and high ($\mathrm{MCL_{HF}}$) frequencies; 5) dynamic range (DynR) estimated as the difference between the uncomfortable level and HTL, estimated from the loudness function for low ($\mathrm{DynR_{LF}}$) and high ($\mathrm{DynR_{HF}}$) frequencies; and 6) slope of the loudness function at low ($\mathrm{Slope_{LF}}$) and high ($\mathrm{Slope_{HF}}$) frequencies. For the outcome measures estimated from the loudness function, the low-frequency average corresponded to the center frequencies 0.25, 0.5 and 1 kHz and the high-frequency average corresponded to the center frequencies 2, 4 and 6 kHz. There were four variables related to speech perception. Two of them related to speech-in-quiet (SiQ): 1) speech reception threshold in quiet ($\mathrm{SRT_Q}$); 2) maximum word recognition score (Max DS); and two of them related to speech-in-noise (SiN): 3) speech reception threshold in noise ($\mathrm{SRT_N}$); and 4) sentence recognition score at +4 dB SNR ($\mathrm{SS^{4dB}}$). There were three variables related to binaural processing abilities (BIN): maximum frequency for detecting an interaural phase difference of 180° ($\mathrm{IPD_{fmax}}$); binaural pitch detection performance, estimated as the percent correct of the dichotic presentations ($\mathrm{BP_{20}}$); and binaural masking release (BMR). BMR was estimated as the difference between the threshold in the diotic tone-in-noise detection condition (N_0S_0) and the threshold in the dichotic tone-in-noise detection condition where the tone was out of phase between the ears (N_0S_π). The frequency of the tone presented in the two conditions was 0.5 kHz. The spectro-temporal modulation (STM) and processing (STP) variables included: 1) spectro-temporal modulation sensitivity at +3 dB modulation depth ($\mathrm{sSTM_8}$) and 2) the "fast" spectro-temporal modulation detection threshold ($\mathrm{fSTM_8}$); 3) the tone-in-noise detection threshold at 500 Hz ($\mathrm{TiN_{LF}}$) and at 2 kHz ($\mathrm{TiN_{HF}}$); 4) the spectral masking release (SMR) estimated as the difference between the tone-in-noise detection threshold (TiN) and the corresponding threshold with the noise shifted towards high frequencies (center frequency of the noise, $f_{c,\mathrm{noise}} = 1.1 f_{\mathrm{tone}}$); and 5) the temporal masking release (TMR) estimated as the difference between the TiN masked threshold and the corresponding threshold with the tone presented in temporally-modulated noise (modulation frequency, $f_m = 4$ Hz).

Data-driven auditory profiling

The BEAR3 dataset was analyzed with an iterative data-driven auditory profiling method. The main results can be summarized by the probabilities of the listeners of belonging to a given auditory profile (A-D) and the expected performance of the listeners identified in each of the four groups (prototypes).

Figure 4.3 shows the results of the analysis where each listener is located in the two-dimensional space according to their degree of type-I and type-II distortion. The degree of distortion was calculated based on the probability of belonging to any of the four auditory profiles. Listeners located close to a corner exhibited a

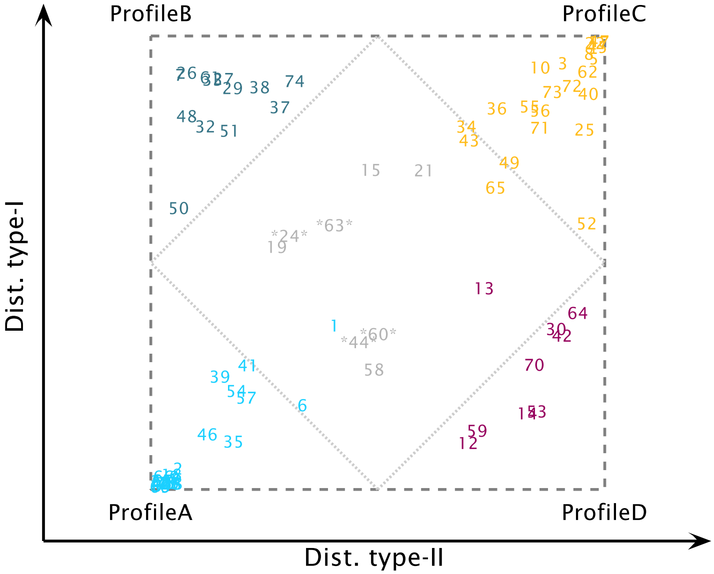


Figure 4.3: Square representation of the auditory profiles. The listeners are placed in the square representation based on their probability of belonging to one of the subgroups. The inner rhombus delimits the area where the listeners showed P < 0.50 of belonging to any subgroup. The listeners labeled as ** showed P(U) > 0.5.

high probability of belonging to a corresponding profile. Uncategorizable listeners are placed in-between the four quadrants and are marked with grey. Profile A (n = 24) and Profile C (n = 22) represented the most populated groups. The five normal-hearing listeners were placed at the bottom-left corner in Profile A. Profiles B (n = 13) and Profile D (n = 9) represented smaller subgroups. The probabilities indicated in the Profile B listeners (0.58 < P(B) < 0.89) and in the Profile D listeners (0.36 < P(D) < 0.73) were, on average, lower than for the Profile A listeners (0.30 < P(A) < 0.97) and the Profile C listeners (0.43 < P(C) < 0.98). Four listeners showed a high probability of being uncategorizable (0.33 < P(U) < 0.77), and four other listeners were “inconclusive” as reflected in similar probabilites of belonging to two profiles. The five listeners showing the largest probabilities of belonging to one of the auditory profiles (excluding the normal-hearing listeners) were considered to represent the prototypes shown in Figure 4.4.

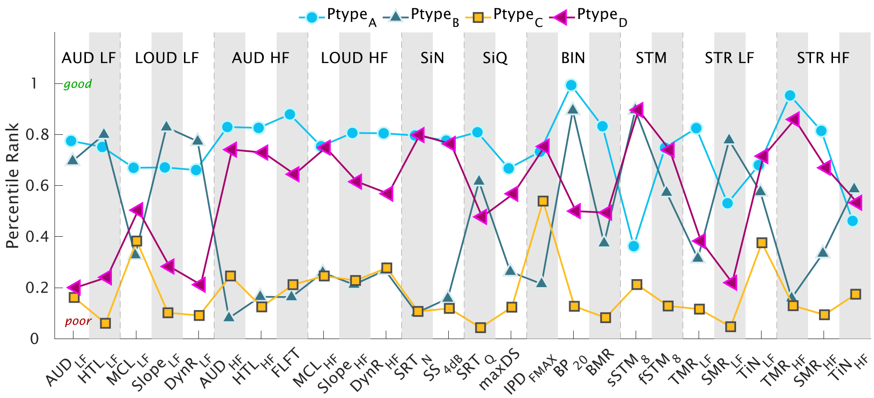


Figure 4.4: Prototypes (Ptype): Percentile rank across variables corresponding to the extreme exemplars of the different patterns found in the data. The ranks are shown for the 26 outcomes corresponding to the different aspects of auditory processing. AUD: Audibility, LOUD: Loudness, SiN: Speech-in-noise perception, SiQ: Speech-in-quiet, BIN: Binaural processing abilities. STM: Spectro-temporal modulation sensitivity, STP: Spectro-temporal processing abilities, divided into temporal and spectral masking release as well as tone in noise detection. Subgroups of measures with frequency-specific outcomes were divided into low (LF) and high (HF) frequencies.

The prototypes show archetypal patterns in the data associated with the performance obtained by the four different groups. A higher percentile rank corresponds to a higher percentile of the overall data distribution and thus to a “good” performance. Each point in Figure 4 corresponds to the mean of the listeners forming the corresponding prototype. Likewise, a low percentile rank corresponds to a “poorer” performance. Prototype A (blue circles in Figure 4.4) showed a good performance in most of the outcome measures. However, the outcomes of the tests related to sSTM_8 and $\mathrm{TiN}_{\mathrm{LF}}$ were below the 50th percentile. Prototype C (yellow squares) showed the poorest performance for most outcome measures, with only $\mathrm{MCL}_{\mathrm{LF}}$ and $\mathrm{IPD}_{\mathrm{fmax}}$ above the 30th percentile. Prototype B (dark-green upwards-pointing triangles), with a high degree of distortion type-I and a low degree of distortion type-II, showed a good performance for the outcome measures obtained at low frequencies and for BP_{20}, whereas performance was poor for the outcomes obtained at high frequencies, IPD_{fmax} and for the speech-in-noise perception tests. In contrast, prototype D (magenta left-pointing triangles), with a high-degree of distortion type-II and a low degree of distortion type-I, showed a good performance in terms of SiN and IPD_{fmax}, and a relatively good performance (above the 60th percentile) for most outcomes measures obtained at high frequencies, whereas the performance was poor for outcome measures obtained at low frequencies, especially in terms of loudness, TMR_{LF} and SMR_{LF}. The prototypes showed opposite results for the profiles located in opposite corners of Figure 4.3 (A vs. C and B vs. D).

Relations between auditory distortion types and outcome measures

Table 4.2: Stepwise regression analysis of auditory distortion (AD) type-I and type-II. The priority was established based on the accumulated adjusted $R^2 > 0.01$. Columns show the predictor name, the estimate, standard deviation (SE), t-value and probability of a significant contribution (p).

Priority	*Predictor*	*Estimate*	*SE*	*t*	*p*	*Adj R^2*
Model	AD type-I					
n/a	(Intercept)	250.0	109.0	2.3	<0.05	-
1	HTL_{HF}	-9.7	2.5	-3.9	<0.0001	0.79
2	TMR_{HF}	-1.6	0.6	-2.6	<0.05	0.82
3	TiN_{LF}	-3.5	1.5	-2.3	<0.05	0.83
4	HTL_{HF}:TiN_{LF}	0.2	0.1	4.6	<0.0001	0.87
5	TMR_{LF}	-1.8	0.6	-2.8	<0.01	0.88
Model	AD type-II					
n/a	(Intercept)	-18.7	3.9	-4.7	<0.0001	-
1	HTL_{LF}	2.4	0.14	17.4	<0.0001	0.84

The relations between the two types of distortions and outcome measures were studied using stepwise regression analysis (Table 4.2). Distortion type-I was found to be associated with elevated hearing thresholds at high frequencies, a reduced temporal masking release and increased tone-in-noise detection thresholds at low frequencies. Furthermore, distortion type-I was significantly correlated with SRT_N ($r = 0.76; p < 0.0001$), even when the effects of audibility were partialled out ($r = 0.33; p < 0.01$). In contrast, the correlations found between distortion type-I and speech recognition in quiet ($r = 0.71; p < 0.0001$) were not significant when partialling out audibility ($r = 0.15; p > 0.1$). Distortion type-II was only associated with hearing thresholds at low frequencies. The restrictive criterion (increase of $R^2 > 0.01$) did not include other variables in the model. However, distortion type-II was significantly correlated with the slope of the loudness function ($r = 0.72; p < 0.0001$) and with the amount of spectral masking release at low frequencies ($r = 0.61; p > 0.0001$). In addition, distortion type-II was correlated with $\mathrm{SRT_Q}$ ($r = 0.83; p < 0.0001$) but not with $\mathrm{SRT_N}$ ($r = 0.21; p > 0.05$). However, the correlation between $\mathrm{SRT_Q}$ and distortion type-II was weaker when controlling for the effects of audibility ($r = 0.30; p < 0.05$). Moreover, the majority of the auditory outcomes were not significantly correlated with distortion type-II when hearing thresholds were partialled out, except for $\mathrm{TMR_{HF}}$($r = 0.35; p < 0.01$).

The outcome measures related to binaural processing abilities (Figure 4.4) gave unexpected results. Indeed, the prototypes showed opposite trends for $\mathrm{IPD_{fmax}}$ and $\mathrm{BP_{20}}$, which could indicate that they reflect different auditory distortions. Distortion type-I was significantly correlated with both $\mathrm{IPD_{fmax}}$ and $\mathrm{BP_{20}}$, but only $\mathrm{BP_{20}}$ remained significant after controlling for audibility ($r = -0.36; p < 0.01$). In contrast, distortion type-II was only correlated with $\mathrm{BP_{20}}$ ($r = -0.58; p < 0.0001$) before partialling out the effects of audibility ($r = -0.1; p = 0.5$). Besides, $\mathrm{IPD_{fmax}}$ was neither correlated with any of the two distortion types when controlling for audibility nor with any of the other binaural processing abilities outcome measures ($r << 0.1; p > 0.15$). Instead, $\mathrm{IPD_{fmax}}$ was highly correlated with the tone-in-noise detection threshold at low frequencies ($r = -0.53; p < 0.0001$) — one of the main predictors of distortion type-I — even when audibility was partialled out ($r = -0.56; p < 0.0001$).

Decision tree for the identified auditory profiles

Figure 4.5 shows the decision tree fitted to the BEAR3 dataset using the identified auditory profiles as well as the uncategorizable listeners. The decision tree has three levels. The first level corresponds to high-frequency hearing loss as estimated using

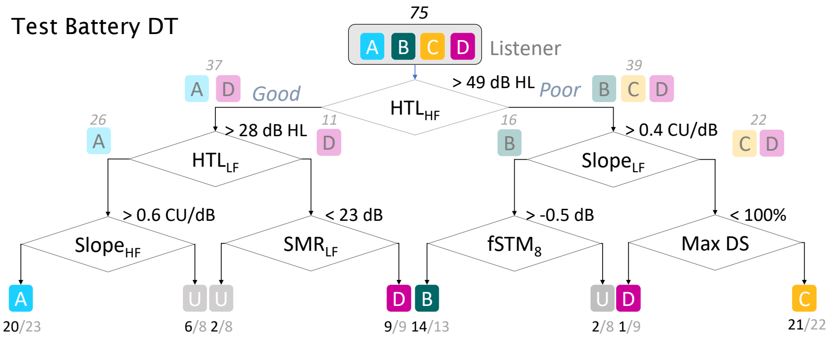


Figure 4.5: Decision tree fitted to the dataset using the auditory profiles as the output. For each binary split, the right branch corresponds to a “poor” result and the left branch to a “good” result. In each binary split, the number of listeners assigned to each branch are shown together with the most likely outputs. The classes (A-D) are together with the number of listeners belonging to that class and the number of identified listeners for a given profile.

ACALOS, which splits the listeners into two branches: Profiles A and D ($\mathrm{HTL}_{\mathrm{HF}} <$ 49 dB HL) are separated from Profiles B and C ($\mathrm{HTL}_{\mathrm{HF}} >$ 49 dB HL), together with one listener from Profile D. Thus, this first level makes a classification based on the degree of distortion type-I. The second level corresponds to outcomes measured at low frequencies and estimated using the loudness functions, which divide the listeners according to their degree of distortion type-II. Profile D ($\mathrm{HTL}_{\mathrm{LF}} >$ 28 dB HL) and Profile C ($\mathrm{Slope}_{\mathrm{LF}} >$ 0.4 CU/dB and maxDS < 100%). The third level makes use of outcomes related to loudness, spectro-temporal modulation and spectral masking release for classifing the uncategorizable listeners.

4.4 Discussion

The data-driven method for auditory profiling presented here provides new knowledge about hearing loss characterization. Regarding previous data-driven auditory profiling (Sanchez-Lopez et al., 2018a), the present results are in good agreement with the analysis performed on the data of Johannesen et al. (2016) data set. This suggests that the use of data from a representative sample of different degrees of hearing loss (e.g. in Johannesen et al., 2016) and a normal-hearing reference (e.g. in

Thorup et al., 2016) is crucial for robust profile-based hearing-loss characterization.

Two types of distortion to characterize individual hearing loss

The term "distortion" in hearing science has typically been associated with elevated $\mathrm{SRT_N}$, as reflected in Plomp's SRT model (Plomp, 1978). Here, we introduced the term "auditory distortions" to describe the perceptual consequences of sensory hearing impairment, including (but not limited to) loss of sensitivity. The two types of perceptual distortions considered here should thus be considered as consequences, and not sources of, sensory impairments. An interesting aspect of our data-driven profiling method is that the auditory distortions reflect two fairly independent dimensions of perceptual deficits associated with sensorineural hearing impairments. To reiterate, distortion type-II was associated with low-frequency hearing loss and steep loudness functions. However, listeners with a high degree of distortion type-II and a low degree of distortion type-I (Profile D) did not exhibit exclusive audibility loss, as they also exhibited an abnormal loudness growth and a reduced spectral masking release. Distortion type-I was associated with elevated hearing thresholds at high frequencies and was significantly correlated with elevated $\mathrm{SRT_N}$. Furthermore, for this distortion type, $\mathrm{TMR_{HF}}$ and $\mathrm{TiN_{LF}}$ were poorer even when the effect of the audiometric thresholds was controlled for.

Although Plomp's attenuation and distortion components are often assumed to be independent, some impairment mechanisms may, in fact, affect both speech-in-noise perception and audiometric thresholds, especially at high frequencies (Moore, 2016), which is consistent with distortion type I. Schädler et al. (2020) attempted to model supra-threshold auditory deficits that are independent of audibility loss. Their results suggested that reduced speech intelligibility represents an auditory perceptual deficit that may be associated with reduced tone-in-noise detection which is in agreement with the results from the current study. However, as demonstrated here, speech-in-noise perception can also be affected by deficits that covary with audiometric thresholds (distortion type-I), which should not be underestimated, especially when the high-frequency hearing loss exceeds 50 dB HL (Profiles B and C), as depicted in Figure 4.6.

Regarding the 'neural component' associated with reduced binaural processing abilities (Kollmeier and Kiessling, 2018), the BIN measures considered in the present study provided contradictiory results in connection to the proposed auditory profiles. Even though $\mathrm{IPD}_{\mathrm{fmax}}$ represents a test that has been proposed to reveal binaural disabilities related to the disruption of temporal fine structure (TFS) coding (Füllgrabe and Moore, 2017), a recent study linked the detection of interaural phase differences to outcomes from cognitive tests (Strelcyk et al., 2019). This suggests that $\mathrm{IPD}_{\mathrm{fmax}}$ might not reflect a purely auditory process but might also depend on top-down processes such as processing speed. Since $\mathrm{IPD}_{\mathrm{fmax}}$ and $\mathrm{TiN}_{\mathrm{LF}}$ were strongly correlated, the two tasks might be affected by either cognitive or auditory processes, which should be investigated further.

The two types of auditory distortions shown here were consistent with Plomp's approach (Plomp, 1978). However, the two auditory distortion types presented here are, in fact, the result of a data-driven analysis of a large multi-dimensional dataset rather than the conceptual interpretation of speech intelligibility deficits. Distortion type-I may then be considered as a "speech intelligibility related distortion" and distortion type-II as a "loudness perception related distortion". Nevertheless, the listeners with higher degrees of the two types of distortions showed perceptual deficits with respect to spectro-temporal processing and binaural processing abilities thus reflecting deficits that are beyond a simple combination of loudness and speech-intelligibility deficits.

Auditory profiles and hearing-loss phenotypes

Figure 4.6 shows the audiometric thresholds corresponding to the four robust auditory profiles. Profile A corresponds to a mild, gently sloping high-frequency hearing loss; Profile B corresponds to a steeply sloping high-frequency hearing loss; Profile C corresponds to a low-frequency hearing loss between 30 and 50

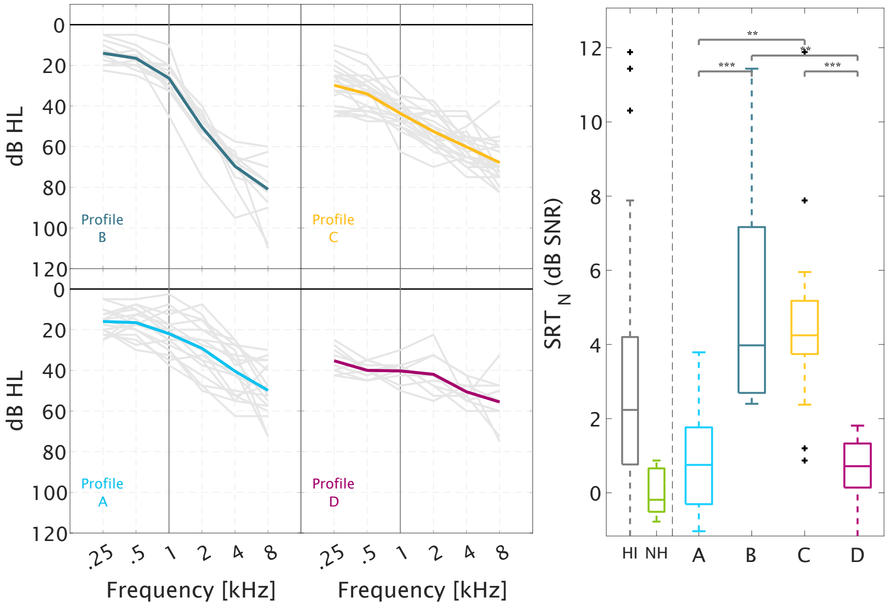


Figure 4.6: Audiometric thresholds of the four auditory profiles and speech intelligibility in noise. Left panel: The averaged audiometric thresholds of each profile are shown together with the individual audiograms. Right panel: Speech reception thresholds in noise (SRT_N), with boxplots of the HI and NH data (left) and the four auditory profiles (right). The multicomparison analysis revealed significant differences between the groups (*** $p < 0.0001$, ** $p < 0.001$).

dB HL and above 50 dB HL at high frequencies; and Profile D corresponds to a fairly flat hearing loss with audiometric thresholds between 30 and 50 dB HL. Interestingly, these four audiometric configurations look similar to the audiometric phenotypes (Dubno et al., 2013), which are based on Schuknecht's metabolic and sensory types of presbyacusis (Schuknecht and Gacek, 1993). The main difference between the two approaches is that the audiometric thresholds shown here correspond to four subgroups of HI listeners, which are the result of a data-driven analysis involving various auditory measures and not only the audiometric thresholds.

In previous studies, metabolic hearing loss (MHL) yielded flat elevated audiometric thresholds, but did not affect speech intelligibility in noise (e.g., Pauler et al., 1986), which is consistent with the results of the present study for Profile D listeners. In MHL, the atrophy of the stria vascularis produces a reduction of the EP in the scala media (Schmiedt et al., 2002). The EP loss mainly affects the electromotility properties of the OHC (i.e., the cochlear amplifier). Therefore, metabolic hearing loss can be considered as a cochlear gain loss that impairs OHC function across the entire cochlea. This, in turn, affects the hearing thresholds and is associated with a reduced frequency selectivity (Henry et al., 2019). In the present study, Profile D was characterized by an abnormal loudness function, particularly at low frequencies, and a significantly reduced spectral masking release, although speech-in-noise intelligibility and binaural TFS sensitivity were near-normal. However, one needs to bear in mind that the results observed for the listeners in Profile D might also be compatible with other types of impairments. Sensory hearing loss (SHL) is typically associated with OHC dysfunction, which yields elevated thresholds at more specific frequency regions, a loss of cochlear compression and reduced frequency selectivity (Ahroon et al., 1993). However, audiometric thresholds above about 50 dB HL at high frequencies cannot be attributed only to OHC due to the limited amount of gain induced by the OHC motion, which implies additional IHC loss or a loss of nerve fibers (Hamernik et al., 1989; Stebbins et al., 1979; Wolak et al., 2019). Therefore, listeners classified as Profile B or Profile C (i.e., with a higher degree of distortion type-I and a high-frequency hearing loss) may exhibit a certain amount of IHC dysfunction that might produce substantial supra-threshold deficits. Animal studies have shown that audiometric thresholds seem to be insensitive to IHC losses of up to about 80% (Lobarinas et al., 2013). This suggests that hearing thresholds > 50 dB HL might indicate the presence of hearing deficits that may distort the internal representation, not only in terms of frequency tuning but also in terms of a disruption of temporal coding due to the lack of sensory cells (Moore, 2001; Stebbins et al., 1979).

Profile B's audiometric thresholds are characterized by a sloping hearing loss with normal values below 1 kHz. However, Profile B exhibited the poorest performance in the IPD_{fmax} test, which cannot be explained by an audibility loss. Neural presbyacusis is characterized by a loss of nerve fibers in the spiral ganglion that is not reflected in the audiogram. Furthermore, primary neural neurodegeneration, recently termed cochlear synaptopathy (Kujawa and Liberman, 2009; Wu et al., 2019) or deafferentation (Lopez-Poveda, 2014), might be reflected in the results of some of the supra-threshold auditory tasks used here. However, the perceptual consequences of primary neural degeneration are still unclear due to the difficulty of assessing auditory nerve fibers loss in living humans (Bramhall et al., 2019). This makes it difficult to link the effects of deafferentation to the reduced binaural processing abilities observed in listeners in Profile B and Profile C.

As suggested in Dubno et al. (2013), the audiometric phenotype characterized by a severe hearing loss (similar to the one corresponding to Profile C) might be ascribed to a combination of MHL and SHL. In the present study, Profile C listeners performed similarly to Profile B listeners in supra-thresholds tasks related to distortion type-I (e.g. SRT_{N} and TMR_{HF}) and also similarly to Profile D listeners in tasks related to distortion type-II (e.g., loudness perception). In contrast, Profile C listeners also showed poorer performance in tests such as binaural pitch detection, tone-in-noise detection and spectro-temporal modulation sensitivity, which is not consistent with the idea of a simple superposition of the other profiles. As mentioned above, these deficits observed in Profile C listeners might be a consequence of auditory impairments that are unrelated to the loss of sensitivity, such as deafferentation, which can be aggravated by the presence of MHL and SHL. However, it has been found that spectro-temporal modulation sensitivity could be a good predictor of aided speech perception only in the cases of a moderate high-frequency hearing loss (Bernstein et al., 2016). They suggested that cognitive factors might be involved in the decreased speech intelligibility performance when the high-frequency hearing loss is >50 dB HL. Therefore, Profile C listeners might be affected by both auditory and non-auditory factors that worsen their performance in some demanding tasks.

Stratification in hearing research and hearing rehabilitation

In the present study, the two principal components of the dataset seemed to be dominated by the listeners' low- and high-frequency hearing thresholds. Therefore, another supra-threshold hearing deficit might be hidden in the four auditory profiles that could explain the individual differences across listeners belonging to the same profile. To explore these "additional deficits" not covered by the present approach, a stratification of the listeners might be necessary. Lőcsei et al. (2016) investigated the influence of TFS on speech perception for different interferers. In their study, the HI listeners were divided into groups based on the degree of hearing loss at high frequencies. Consequently, stratification of the listeners into two subgroups helped reduce the potential effect of audibility on speech intelligibility. In another study, Papakonstantinou et al. (2011) studied the correlation of different perceptual and physiological measures with speech intelligibility in stationary noise. In their study, all the listeners had a steeply sloping high-frequency hearing loss consistent with Profile B. Both studies (Lőcsei et al., 2016; Papakonstantinou et al., 2011) included measures of frequency discrimination thresholds and speech-in-noise perception. However, Papakonstantinou et al. (2011) tested a larger group of HI listeners with fairly similar audiograms in only one speech condition that led to a highly significant correlation between frequency discrimination and speech intelligibility in stationary noise. This suggests that the stratification of the listeners and the investigation of certain phenomena in separated auditory profiles might reveal new knowledge about hearing impairments that are not generalized to the entire population of HI listeners.

Other approaches have attempted to identify why listeners with similar audiograms present substantial differences in suprathreshold performance. Recently, Souza et al. (2020) showed how older HI listeners vary in terms of their ability to use specific cues (either spectral or temporal cues) for speech identification. Their results showed a so-called "profile cue" that characterizes the listener's abilities in terms of spectro-temporal processing. Some listeners utilized temporal envelope cues and showed good temporal discrimination abilities, whereas other listeners relied on spectral cues and were able to discriminate spectral modifications in a speech signal. The “profile cue” was associated with the spectral discrimination task; however, this test seemed to be influenced by the audiometric thresholds. Since their participants presented audiograms similar to the ones observed in Profiles B, C and D (Figure 4.6), it is possible that the categorization of the listeners based on auditory profiling could help identify “profile cues” in connection to the supra-threshold auditory deficits observed in each of the four auditory profiles reported here.

Sensory rehabilitation of the hearing deficits involves the use of hearing devices. The criteria for candidacy of implantable technology (Irving et al., 2014; Kirkby-Strachan and Que-Hee, 2016), which are based on the benefit observed by using acoustical devices, are sometimes insufficient. New findings with electro-acoustic stimulation (Imsiecke et al., 2019) suggest that this technology might benefit patients with a high degree of speech-intelligibility related deficits, i.e. with $\mathrm{HL}_{\mathrm{HF}} > 50$ dB HL as shown here. Therefore, the present characterization into auditory profiles might support a revision of the candidacy of implantable devices that might include less severe hearing losses at high frequencies. Furthermore, auditory profiling showed potential for hearing diagnostic that can help disentangle the effects of different types of impairments. This might be particularly useful for the development of therapeutics for hearing loss (Kujawa and Liberman, 2019) supporting precision medicine.

Overall, the listeners contained in the data analyzed here would be candidates for hearing aids. Hearing-aid users often show a large variability in terms of benefit and preference to specific forms of hearing-aid processing (Neher and Wagener, 2016; Picou et al., 2015; Souza et al., 2019). In some studies, the HI listeners were stratified based on their audiograms (Gatehouse and Akeroyd, 2006; Keidser et al., 1995; Keidser and Grant, 2001; Larson et al., 2002). However, the existing hearing-aid fitting rules do not make use of supra-threshold auditory measures that might help tune the large parameter space of modern hearing technology. In fact, the HA parameters are still adjusted based on the audiogram and empirical findings that provide some fine-tuning according to the HA user experience or the gender of the patient (Keidser et al., 2012). The four auditory profiles presented here showed significant differences in supra-threshold measures related to two independent dimensions, a “speech intelligibility related distortion” and a “loudness perception related distortion”. Therefore, auditory profiling allows stratification of the listeners beyond what can be done with an audiogram, which may help optimize hearing-aid parameters for a given patient using existing HA technology. Recently, it has been suggested that different advanced signal processing strategies should be used to compensate for different cochlear pathologies (Henry et al., 2019). Since the four auditory profiles showed interesting similarities to the sensory and metabolic phenotypes (Dubno et al., 2013), new forms of signal processing, dedicated to overcoming the hearing deficits in the two identified dimensions, might be developed and evaluated towards a profile-based compensation strategy. The current approach may inspire different forms of model-based hearing loss compensation (Bondy et al., 2004) to restore auditory function based on biologically inspired technology. This can lay the foundations of precision medicine (Jameson and Longo, 2015) applied to the perceptual rehabilitation of the hearing deficits.

Limitations of the data-driven auditory profiling approach

The definition of the auditory profiles reflected the main sources of hearing deficits in a relatively large and heterogeneous population of HI listeners. However, this group only contained older adults (>60 years) with symmetric sensorineural hearing losses. An extension of the auditory profiling method proposed here might contain an even more heterogeneous group, which might require different data-analysis techniques for proper analysis and interpretation (Hinrich et al., 2016). The insights from the current method could then be applied mainly to a population of mild-to-severe age-related hearing losses and to some extent to other types of non-syndromic hearing losses, e.g. noise-induced hearing loss, but cannot be generalized to the whole variability of existing auditory pathologies.

4.5 Conclusion

Using a data-driven approach, four auditory profiles (A-B-C-D) were identified that showed distinct differences in terms of supra-threshold auditory processing capabilities.. The listeners' hearing deficits could be characterized by two independent types of auditory distortion, a “speech intelligibility-related distortion” affecting listeners with audiometric thresholds >50 dB HL at high frequencies, and a “loudness perception-related distortion” affecting listeners with audiometric thresholds >30 dB HL at low frequencies. The four profiles showed similarities to the audiometric phenotypes proposed by Dubno et al. (2013), suggesting that Profile B may be resulting from a sensory loss and Profile D may be resulting from a metabolic loss. Profile C may reflect a combination of a sensory and metabolic loss, or a different type of hearing loss that results in substantially poorer supra-threshold auditory processing performance. The success of this approach provides new methods to identify homogeneous sub-populations to better investigate the perceptual consequences of different etiologies. The current results enable "precision audiology" and provide new avenues for developing auditory-profile based compensation strategies for hearing rehabilitation.

4.6 Acknowledgements

We thank F Bianchi, SG Nielsen, M El-haj-Ali, OM Cañete, and M Wu for their contribution to the design of the BEAR test battery, and work in the data collection. Also, we appreciate the valuable input from N Bisgaard, JB Nielsen, J Harte, W. Whitmer, G Naylor and O. Strelcyk among others. We would also like to thank G Encina-Llamas, AA Kressner and EN MacDonald who suggested some improvements in an earlier version of this manuscript. This work was supported by Innovation Foundation Denmark Grand Solutions 5164-00011B (Better hEAring Rehabilitation project) Oticon, GN Hearing, WSAudiology and other partners (Aalborg University, University of Southern Denmark, the Technical University of Denmark, Force, Aalborg, Odense and Copenhagen University Hospitals). The funding and collaboration of all partners are sincerely acknowledged.

5

Auditory Profile-based Hearing-aid Fitting: A Proof-Of-Concept Study [a]

Abstract

Currently, the clinical characterization of hearing deficits for hearing-aid fitting is based on the pure-tone audiogram only. Recently, Sanchez-Lopez et al. [“Robust data-driven auditory profiling for precision audiology”. *Chapter 4*] proposed that the sensory impairments of a given listener result from two independent types of auditory deficits: speech intelligibility- and loudness perception-related deficits. This proposal was based on a large dataset collected with a heterogeneous group of listeners who were tested using measures of speech perception, loudness perception, binaural processing abilities, and spectro-temporal resolution. A data-driven analysis of the collected data yielded four clinically relevant patient subpopulations or “auditory profiles”. In the same way that stratified medicine applies targeted therapies to specific patient populations, a profile-based hearing-aid fitting strategy is proposed here. Using a hearing-aid simulator, four candidate settings were evaluated by a subset of the participants tested previously. Listeners belonging different auditory profiles differed in terms of preference and favored the targeted hearing-aid setting. The results from this proof-of-concept study support further investigations with clinically fitted

---

[a] This chapter is based on Sanchez-Lopez, R., Fereczkowski, M., Santurette, S., Dau, T. and Neher, T. (2020). *"Auditory Profile-based Hearing-aid Fitting: A Proof-of-concept study"*. Submitted to Ear and Hearing. Preprint at medRxiv:036459.

hearing aids that enable the clinical implementation of a stratified, profile-based approach to hearing-aid fitting.

5.1 Introduction

Hearing loss (HL) is typically treated with hearing aids (HA). The primary purpose of HAs is to provide gain to the input signal to compensate for reduced audibility. In addition, modern HAs incorporate advanced signal processing algorithms for noise suppression (Chung, 2004). As a consequence, numerous parameters need to be adjusted as part of the hearing-aid fitting process.

In current clinical practice, the assessment of the hearing deficits of a patient relies mainly on pure-tone audiometry. Based on a fitting rule that typically uses the audiogram of the patient as the only information, the HA amplification is then adjusted. For example, the “National Acoustic Laboratories – Nonlinear 2” fitting rule (NAL-NL2; Keidser et al., 2011) is commonly used. This rule relies on a combination of empirical knowledge and modelling aimed at maximizing the effective audibility of the speech signal. While NAL-NL2 can be expected to provide a reasonable overall solution, there are also patients whose hearing difficulties are not captured by the audiogram and who may therefore benefit from other fitting strategies (Henry et al., 2019; Keidser and Grant, 2001; Oetting et al., 2018). Such fitting strategies could include the adjustment of advanced HA features, which are not yet incorporated into existing fitting rules. For example, noise reduction and directional processing are currently activated based on “life-style” considerations rather than audiological factors. Although advanced HA features can improve the signal-to-noise ratio (SNR), the individual preference for these settings diverges substantially across listeners, possibly because of unwanted speech distortions that are typically also introduced by these algorithms (Neher and Wagener, 2016). Therefore, it is possible that the individualized adjustment of noise suppression algorithms could improve the outcome, for example for patients with poor speech intelligibility in challenging environments.

In a recent study, we identified four clinically relevant subgroups of hearing-impaired (HI) listeners using a data-driven approach (Sanchez-Lopez et al., 2020b). The listeners were characterized by their degree of perceptual deficits or "distortions", which were estimated using a battery of auditory measures tapping into loudness and speech perception, binaural processing abilities and spectro-temporal resolution (Sanchez-Lopez et al., 2020d). Four archetypal patterns of perceptual deficits – referred to as "auditory profiles" – were uncovered. These profiles varied along two primary dimensions, or types, of deficits: speech intelligibility (SI) and loudness perception (LP) related deficits.

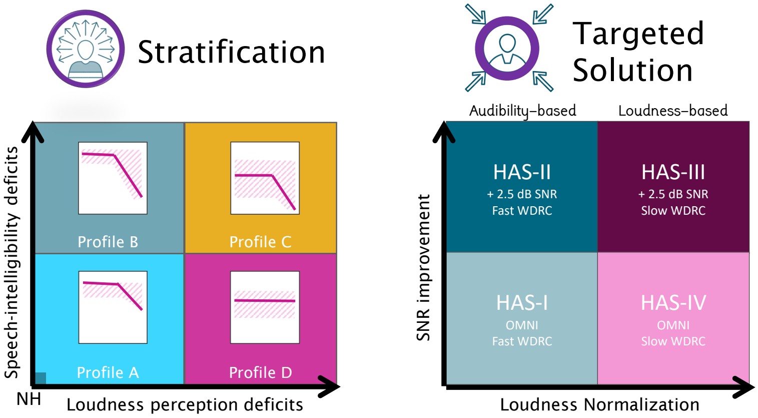


Figure 5.1: Illustration of the profile-based hearing-aid fitting strategy. Left: Data-driven auditory profiling (Sanchez-Lopez et al., 2020b). In a two-dimensional space with speech intelligibility-related (SI) deficits on one axis and loudness perception-related (LP) deficits on the other axis, listeners differing in the degree of the two types of perceptual deficits would be placed at different positions along the two dimensions. While Profile C represents a high degree of both types of deficits, profiles B and D reflect hearing deficits dominated by one deficit type. Profile A has a low degree of deficits and thus near-normal hearing abilities. Each type of deficits would then covary with specific deficits observed in a number of behavioral tasks that define a given auditory profile. Right: Proposed candidate hearing-aid settings (HAS) for the different profiles, which are intended to compensate for the specific auditory deficits. Signal-to-noise ratio (SNR) improvement as a hearing solution for SI deficits and loudness normalization as a solution for LP deficits.

In the medical field, personalized treatments aims at providing tailored solutions to clinically relevant subgroups of patients (Trusheim et al., 2007). Here, a profile-based fitting strategy including a number of candidate hearing-aid settings (HAS) was evaluated. Listeners with a high degree of LP-related deficits (Profiles C and D) were expected to prefer a gain prescription aimed at loudness normalization (Oetting et al., 2018), whereas listeners with a high degree of SI-related deficits (Profiles B and C) were expected to prefer HAS with advanced signal processing (Figure 5.1). As such, the present study examined the validity of auditory profile-based HA fitting in terms of subjective preference. A multi-comparison evaluation was performed with a group of participants who had previously been classified into the four auditory profiles. This made it possible to explore whether listeners belonging to different auditory profiles would exhibit different patterns of HA outcome.

5.2 Methods

Seven listeners participated in the current study. All of them had previously completed a comprehensive auditory test battery (Sanchez-Lopez et al., 2020d), based on which they had been classified as belonging to one of the four auditory profiles (Sanchez-Lopez et al., 2020b). For the experiment, a hearing-aid simulator (HASIM), which consisted of three stages: A beamforming stage, a noise reduction stage and an amplitude compression stage (see Table 5.1 for details). The beamformer and noise reduction settings were selected based on the achievable SNR improvement (Sanchez-Lopez et al., 2018b).

Nine sound scenarios were tested. In each scenario, a fragment of a realistic conversation taken from a publicly available database (Sørensen et al., 2018) was used for engaging the listener in the sound scene. The participant was instructed to listen actively to the conversation. The tested sound scenarios differed in terms of the background noise. Three noise conditions were included: 1) cafeteria noise (input level 65 dB SPL), 2) traffic noise (input level 75 dB SPL), and 3) quiet.

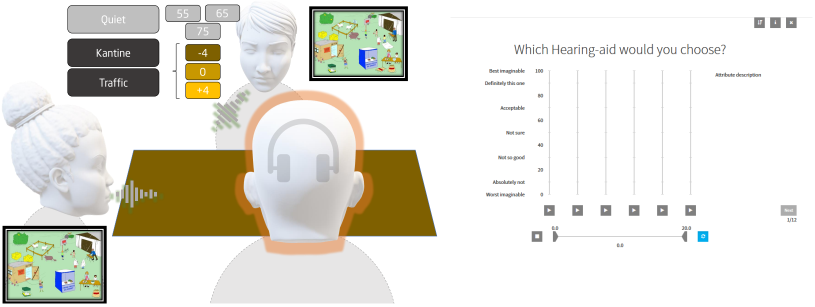


Figure 5.2: Illustration of the multi-comparison assessment. Left: Sketch of the used sound scenarios. The listeners were engaged in a simulated scene consisting of a dialog between two talkers in a quiet or noisy environment. The conversation revolved around finding differences between two Diapix figures (Baker and Hazan, 2011). Three sound scenarios and three speech levels or SNRs were used. Right: Graphical user interface of the SenseLabOnline software. The four candidate settings were tested in a multi-comparison paradigm that included an anchor and a reference. The sound corresponding to a given setting was played back in a loop when the corresponding play button was pressed. The preference judgements were provided using the sliders. The listeners were instructed to 1) identify the anchor, 2) provide a first set of coarse preference ratings between “Acceptable” and “Not good”, 3) reorganize their ratings using the ranking functionality built into the SenseLabOnline software, 4) listen to all stimuli again and refine their ratings before storing the final judgements. For each sound scenario, three repetitions were made.

Furthermore, there were three SNR or level conditions. That is, in the case of the cafeteria and traffic scenarios the target was scaled in level to achieve SNRs of -4, 0 or +4 dB. In quiet, the target input level was either 55, 65 or 75 dB SPL. The multi-comparison of the HAS was realized using the SenselabOnline software (SenseLab, 2017). On a given trial, six stimuli were presented to the listener: An anchor resembling a ‘broken’ hearing aid, a ‘commercial’ HAS, and the four candidate HAS (I, II, III and IV). Details are provided in Table 5.1 The multi-comparisons were performed sequentially across several trials. In each case, a 20-sec audio file corresponding to a given sound scenario that had been processed using the HASIM, was played back (Figure 5.2). The participant then used a slider ranging from 0 to 100 to rate the sound of each HAS. The question posed to the listeners was “Which hearing aid would you choose?”. When giving their ratings, they were instructed to focus on their overall preference rather than on specific attributes such as noise annoyance or speech clarity.

Table 5.1: Hearing-aid settings (HAS) evaluated in the multi-comparison assessment. The directionality (DIR) setting could be either omnidirectional (omni) or a fixed forward-facing cardioid setting. The noise reduction (NR) could provide an attenuation of 5, 9 or 15 dB following the estimation of the speech signal. For the anchor stimulus, errors were introduced into the speech signal estimation. The attack and release times of the amplitude compressor were similar to those used in previous studies. The HAS were characterized in terms of SNR improvement and spectral and temporal signal distortion in a complex noisy environment (Sanchez-Lopez et al., 2018).

HAS	Anchor	HAS-O	HAS-I	HAS-II	HAS-III	HAS-IV
DIR setting	Omni	Cardioid	Omni	Cardioid	Cardioid	Omni
NR (dB)	15*	5	Off	9	9	Off
Attack time (ms)	5	250	5	5	5	5
Release time (ms)	10	1250	40	40	1250	1250
SNR improvement (dB)	0	2	0	2.5	2.5	0

*(errors artificially introduced)

5.3 Results and discussion

Four candidate HAS (HAS-I, HAS-II, HAS-III and HAS-IV) were evaluated together with a standard clinical HA fitting (HAS-O). In HAS-I and HAS-II, fast-acting compression was applied to provide non-linear gain according to an audibility-based prescription formula. In HAS-III and HAS-IV, slow-acting compression was applied based on the principle of loudness normalization. Furthermore, in HAS-II and HAS-III advanced HA features were activated to provide about 2.5 dB of SNR improvement under noisy conditions (see Method, Table 5.1, Table 5.2 for more details).

Figure 5.3 shows the mean preference ratings for profiles A, B, C and D under quiet conditions. Profile-A listeners preferred HAS-O over the two HAS with fast-acting compression across a range of presentation levels (55, 65 and 75 dB sound pressure level, SPL). Profile-B listeners preferred HAS-I over HAS-III at 65 and 75 dB SPL; at the low input level (55 dB SPL) they provided the highest rating to HAS-IV. Profile-C and -D listeners showed a preference for HAS-IV and consistently disliked HAS-I.

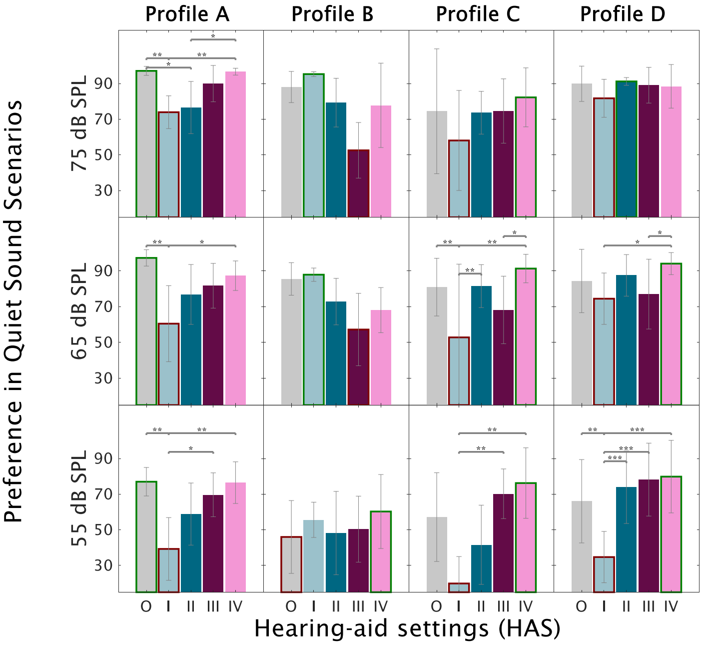


Figure 5.3: Mean preference ratings for the evaluated HAS (O-IV) in the "Quiet" sound environment across three level conditions: 55 dB SPL (bottom panels), 65 dB SPL (middle panels) and 75 dB SPL (top panels). The highest (best) ratings are highlighted in green and the lowest in red. The columns represent the results of the listeners belonging to profile A (left), B (mid-left), C (mid-right) and D (right). Error bars show ±1 standard deviation. Signficant differences according to a two-way ANOVA with repetition, and participant as factors followed by Tukey's honest significant differences tests are marked by asterisks. (***) $p < 0.001$, (**) $p < 0.01$, (*) $p < 0.05$.

Figure 5.4 shows the mean preference ratings under noisy conditions. Profile-A listeners preferred HAS-III and HAS-O over HAS-I. Profile-B listeners consistently disliked HAS-III and showed a preference for HAS-O, HAS-I and HAS-II. Profile-C listeners preferred HAS-III over the other HAS at higher SNRs (0 and +4 dB). However, HAS-O was also preferred at lower SNRs. Profile-D listeners only showed significant differences at +4 dB SNR, with HAS-IV receiving the highest ratings and HAS-I the lowest ratings.

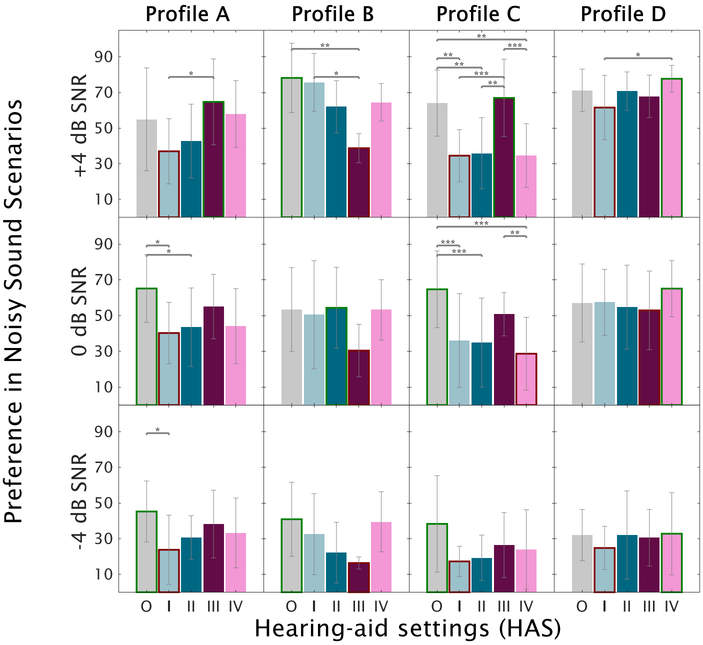


Figure 5.4: Mean preference ratings for the evaluated HAS (O-IV) under noisy conditions across the three SNR conditions: -4 dB SNR (bottom panels), 0 dB SNR (middle panels) and +4 dB SNR (top panels). The highest (best) ratings are highlighted in green and the lowest in red. Each column represents the results of the listeners belonging to profile A (left), B (mid-left), C (mid-right) and D (right). Error bars show ±1 standard deviation. Signficant differences according to a three-way ANOVA with repetition, noise type and participant as factors followed by Tukey's honest significant differences tests are marked by asterisks. (***) $p < 0.001$, (**) $p < 0.01$, (*) $p < 0.05$

The current study aimed to identify patterns of HAS preference in listeners belonging to four distinct auditory profiles. The results suggest that Profile-A and Profile-C listeners based their judgements on similar criteria, especially under noisy conditions. In contrast, Profile-B and Profile-D listeners showed significantly different patterns. While Profile-B listeners disliked the HAS with loudness-based gain prescription and SNR improvement, Profile-D listeners favored loudness-based gain prescription and showed no preference for SNR improvement. The results obtained for the quiet condition support the use of loudness-based gain prescriptions for profiles with a high degree of LP-related deficits. In contrast, SNR improvement was only preferred by one of the two profiles with a high degree of SI-related deficits when tested at positive SNRs (Profile-C). Importantly, Profile-B listeners showed a preference for fast-acting compression, consistent with previous research (Gatehouse et al., 2006a,b).

In summary, Profile-A and Profile-B listeners therefore preferred audibility-based gain prescriptions, whereas Profiles-C and Profile-D preferred loudness-based gain prescriptions. Besides, SNR improvement might improve the outcome of listeners with a high degree of SI-related deficits. Overall, these initial findings provide a useful basis for further investigations into profile-based HA fitting strategies that will include field studies with wearable devices and objective evaluations such speech intelligibility tests.

Acknowledgements

We thank SenseLab (Force technology) the personnel of Bispebjerg Hospital, and S.G. Nielsen for their input and support and the listeners for their participation. This work was supported by Innovation Fund Denmark Grand Solutions 5164-00011B (Better hEAring Rehabilitation project) and the rest of BEAR partners. The funding and collaboration of all partners is sincerely acknowledged.

Appendix: Gain prescription

Table 5.2: Gain prescription for the four candidate hearing-aid settings (HAS). Non-linear gain was calculated for inputs of 50, 65 and 80 dB SPL. The gain was calculated based on the hearing thesholds (HL) and the HAS tested. For each HAS, a correction factor β was applied that reflected different fitting rules based on audibility maximization (Ching et al., 2001) and loudness normalization (Oetting et al., 2018).

Insertion Gain = $0.31HL(f) + \beta(HAS, f)$						
β(HAS-I)	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	>6 kHz
Target 50	-	-	3	7	7	5
Target 65	-	-	-2	0	0	0
Target 80	-	-	-5	-5	-5	-5
β(HAS-II)	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	>6 kHz
Target 50	-3	-3	3	7	7	10
Target 65	-3	-3	-2	0	0	0
Target 80	-6	-6	-9	-9	-9	-9
β(HAS-III)	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	>6 kHz
Target 50	2	3	4	6	10	10
Target 65	-10	-10	-5	0	0	0
Target 80	-14	-14	-14	-14	-14	-14
β(HAS-IV)	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	>6 kHz
Target 50	2	3	4	6	5	5
Target 65	-6	-6	-6	-3	-3	-3
Target 80	-10	-10	-10	-10	-14	-14

6

Audiometric profiles and patterns of benefit. Data-driven analysis of subjective hearing difficulties and handicaps. [a]

Abstract

Hearing rehabilitation attempts to compensate for auditory dysfunction, reduce hearing difficulties and minimize participation restrictions that can lead to social isolation. However, there is no systematic approach to assess the quality of the intervention at an individual level that might help to evaluate the need of further hearing rehabilitation in the hearing care clinic. Here, a large-scale data-driven analysis on subjective behavioral data reflecting hearing disabilities and handicap was chosen to explore normative “patterns of benefit” as a result of rehabilitation in different audiometric listener groups. The method was based on five steps: 1) Dimensionality reduction; 2) Stratification in four audiometric groups; 3) Archetypal analysis to identify archetypal benefit patterns; 4) Clustering in benefit profiles; and 5) Item importance estimation. 572 hearing-aid users were interviewed and completed a questionnaire of hearing difficulties (speech, spatial and

---

[a] This chapter is based on Sanchez-Lopez, R., Dau, T., Whitmer, W.M. (2020). *"Audiometric profiles and patterns of benefit. A data-driven analysis of subjective hearing difficulties and handicaps"*. Submitted to the International Journal of Audiology. Preprint at medRxiv:20045690.

qualities hearing scale; SSQ) and hearing handicap (HHQ). The data-driven approach revealed four benefit profiles that were different for each audiometric group. The patterns of benefit and the stratification approach might guide the selection of the clinical intervention strategy and improve the efficacy and quality of service in the hearing care clinic.

6.1 Introduction

The consequences of hearing loss entail activity limitations and participation restrictions (Simeonsson, 2000). The hearing rehabilitation process aims to minimize both aspects and involves two main steps: diagnosis and remediation (Boothroyd, 2007; Goldstein and Stephens, 1981). After the diagnosis of a hearing loss, an intervention strategy that involves a hearing solution, such as a hearing aid (HA), is commonly proposed and selected by the clinician. The compensation strategy chosen in the HA fitting process is, to a large degree, based on the shape of the pure-tone audiogram. The intervention is verified and validated to ensure the quality of the device and the service (Jorgensen, 2016). However, the hearing rehabilitation often requires further interaction, such as follow-up visits to improve the HA adjustments based on patient complaints or personal preferences, as well as counseling, focused on communication programs and professional advice (Laplante-Lévesque et al., 2010). Overall, each of the steps of the intervention (diagnosis, adjustment and verification) is influenced by technical, personal and social factors (Vestergaard Knudsen et al., 2010).

The evaluation of the efficacy of the hearing rehabilitation process is typically assessed by questionnaires as outcome measures. The questionnaires can be designed to evaluate the individual benefit, the clinical practice or the inclusion of a new device or strategy (Cox, 2003; Cox et al., 2000). Some outcome measures include specific items related to benefit or satisfaction (SADL: Cox and Alexander, 1999; APHAP: Cox and Alexander, 1995; IOI-HA:

Cox, 2003; GHABP: Gatehouse, 1999), whereas others are focused on hearing disabilities and handicaps (Ronde-Brons et al., 2019; Gatehouse and Noble, 2004; Newman et al., 1990; Hallberg, 1998). These questionnaires aim to capture the overall experience or some specific aspects of the hearing rehabilitation, such as the speech, spatial and qualities hearing scale (SSQ Gatehouse and Noble, 2004). The SSQ reflects HA listeners' current difficulties with respect to speech perception, spatial sound perception and qualities of hearing, e.g. the ability to follow a conversation, to localize a sound source or to identify a sound. Although the assessment of hearing disabilities is crucial for a successful hearing rehabilitation, the overall benefit does not only depend on auditory disabilities but also handicaps experienced by the listener (Whitmer et al., 2016), such as the effects on social participation derived from the hearing loss.

One of the primary aims of outcome measures is to quantify the efficiency of hearing rehabilitation. However, no systematic method to evaluate the "quality" of the intervention exists nor a successful outcome measure at the individual listener's level. Usually, the hearing care professional (HCP) addresses the individual complaints and gathers information about the patient's experiences during the follow-up visits (Tecca, 2018). The definition of an optimal intervention is then evaluated subjectively by the two parties whereby no clear guidelines have yet been broadly established. The common goal of an "optimal" intervention would be to minimize the activity limitations and participation restrictions by applying the most suitable hearing technology and professional advice. Thus, the ability of the HCP to understand the patient's needs and to implement a suitable intervention is crucial for successful rehabilitation (Boothroyd, 2007).

The characterization of the hearing deficits of a person in terms of his/her audibility loss does typically not capture the person's performance in real-life situations. Therefore, information about supra-threshold auditory deficits, such as speech intelligibility in noise, might help to better understand the scale and scope of an individual's sensory impairments. Recently, Sanchez-Lopez et al. (2020b) proposed a stratification of hearing-impaired individuals into four clinically relevant subgroups, referred to as “auditory profiles.” The auditory profiles were the result of a data-driven analysis of a relatively large and heterogeneous group of individuals of varying hearing ability who performed several supra-threshold auditory tasks. The approach allowed the identification of four archetypal patterns of perceptual deficits along two largely independent dimensions. The two dimensions were related to speech intelligibility deficits and loudness perception deficits, respectively. Listeners presenting similar deficits were classified as belonging to the same subgroup. Since each auditory profile showed different degrees of deficits, listeners associated with a given profile are likely to experience similar distinct hearing disabilities. Furthermore, the audiometric thresholds associated with the different auditory profiles showed significant differences. Therefore, an audiometry-based stratification of a given listener into an “audiometric” profile might provide an initial classification of the perceptual deficits of the listener. This approach can be useful when applying the stratification to studies without supra-threshold measures retrospectively. However, this pre-classification in audiometric groups does not guarantee that the listener is correctly classified and supra-threshold tests should confirm the listener’s auditory profile (Sanchez-Lopez et al., 2020b).

In the present study, subjective data from a questionnaire of hearing difficulties (SSQ) and a questionnaire of hearing handicap (HHQ) were analyzed using a data-driven approach and following the principles of the “knowledge discovery from databases” (KDD; Frawley et al., 1992; Mellor et al., 2018). Here, data mining methods were applied to identify patterns in subjective responses to provide new insights about the disabilities and handicaps associated with different audiometric profiles. A data set of a clinical population of hearing-aid users was analyzed in an attempt to uncover archetypal “benefit profiles” reflected in the subjective data. The participants were divided into four audiometric groups based on the average audiometric thresholds of the four auditory profiles according to Sanchez-Lopez et al. (2020b). The goal of the study was to identify the priorities of hearing rehabilitation in terms of particular hearing difficulties and handicaps that need to be improved. These patterns of benefit could be valuable for implementing a personalized clinical rehabilitation strategy and to minimize the activity limitations and participation restrictions of patients with hearing loss.

6.2 Method

The data analysis consisted of five steps, as shown in Figure 6.1. First, the data (described in the next section) were transformed using factor analysis. Second, the participants were stratified into four groups based on their degree of low- (HL_{LF}) and high-frequency (HL_{HF}) hearing loss as an approximation of the audiometric profiles of Sanchez-Lopez et al. (2020b). Third, to identify extreme "benefit profiles," the overall data, as well as the data belonging to each of the four subgroups, were processed using an archetypal analysis. Fourth, the participants were identified as belonging to a cluster of participants showing a similar "benefit profile", based on their similarity to the archetypal benefit patterns. Finally, the identified benefit profiles were predicted using supervised learning and the importance of individual questionnaire items was analyzed.

Description of the dataset

The data-driven analysis presented here is a retrospective study performed on the dataset of Hearing Science Scottish Section (HSSS; formerly Institute of Hearing Research) of the University of Nottingham. The data were collected between the years 2002 and 2011 and most of the patients were referred from the NHS Audiology of the Glasgow Royal Infirmary. The total dataset consisted of 1220 participants. The HSSS dataset had previously been analyzed by Akeroyd et al. (2014) and Whitmer et al. (2014) where a thorough description of the normative data was provided. The variables of interest for the present study were the audiometric thresholds, the raw scores of the Speech, Spatial and Qualities hearing scale (SSQ) and the Hearing Handicap (HHQ) questionnaires (Gatehouse and Noble, 2004). Only hearing-aid users (unilateral and bilateral) were selected for the present analysis.

The subset of the HSSS dataset used here consisted of 880 observations (participants), and 62 variables. The speech-related items of SSQ (14 items), the spatial-related items (17 items), the qualities-related items (19 items) and the hearing handicap-related items (12 items). The SSQ questionnaire is scored on a 0-10 scale (in steps of 0.5 in this particular dataset), whereby a low score corresponds to high difficulty and a high score corresponds to low difficulty. If the item corresponds to a situation that the listener has not experienced, the response "not applicable" can be chosen. The HHQ was scored on a discrete 1-5 scale, with 5 representing the largest handicap. The items related to the specific restrictions on participation were based on the Hearing, Disabilities and Handicaps Scale (Hétu et al., 1994).

Data cleaning was performed by removing participants with more than 36 missing responses. The data were standardized prior to analysis. The HHQ data were multiplied by -1, such that a higher value corresponded to a lower handicap, consistent with the scale considered in the SSQ data. Since the data-analysis involved stratification of the participants in audiometry-based auditory profiles, the audiometric thresholds were also retrieved from the dataset but not used in the analysis. The audiometric thresholds were grouped into low-frequency ($\leq$ 1 kHz) and high-frequency (> 1 kHz) averages and only the better ear was used for further analysis. Since Sanchez-Lopez et al. (2020b) did not include participants with average low-frequency hearing thresholds above 65 dB hearing level (HL) in their data-driven profiling approach, here, the participants with a severe-to-profound low-frequency hearing loss (HL_{LF}>65dB) were excluded. The final number of observations considered for the analysis was 572 participants.

Data-drive pattern identification

I **Dimensionality reduction:** Based on factor analysis (Cattell, 1988), the multi-dimensional dataset was reduced to four latent factors. The number of factors was selected by parallel analysis (Horn, 1965) with

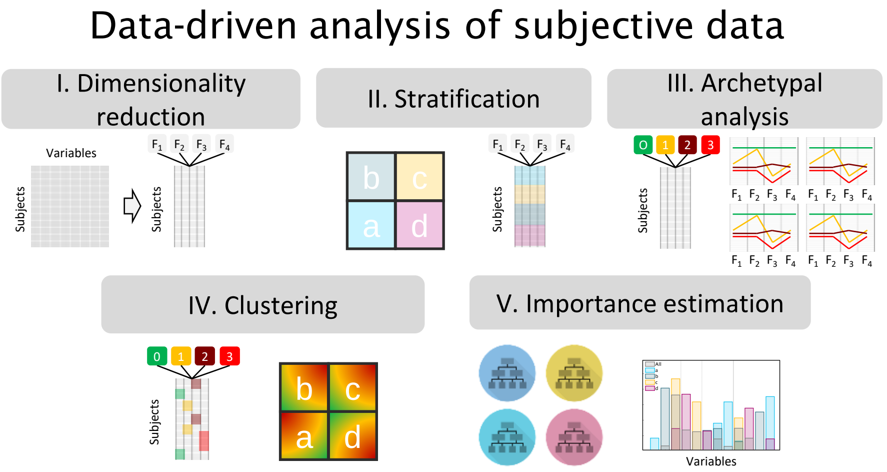


Figure 6.1: Sketch of the data-driven method for the analysis of the subjective data. Top panel: The unsupervised learning exploratory stages included: I) Dimensionality reduction into four factors (F1, F2, F3, F4); II) Stratification, where the subjects were divided into audiometric groups (a, b, c, d); III) Archetypal analysis where the data were decomposed into a matrix with the “benefit patterns” [ABP0, ABP1, ABP2 and ABP3] and the weights of each pattern resembling each subject’s observation. Bottom panel: IV) Clustering, where the participants were clustered based on the similarity of their scores with the benefit profiles derived from stage II; and V) Importance estimation, where a random forest was trained with to classify the participants into the “benefit profiles” and the importance of the predictors were estimated.

subsequent iterative resampling ($K = 200$). The factors were obtained using oblique Procrustes rotation as in (Akeroyd et al., 2014). The factor scores corresponding to each of the subjects, and obtained by Bartlett’s method, were used for further analysis.

II **Stratification:** The listeners were divided into four groups corresponding to the audiometry-based auditory profiles a, b, c and d. The binary rules used here are

a. Audiometric group-a: $HL_{HF} < 50$ dB HL, and $HL_{LF} < 30$ dBHL.

b. Audiometric group-b: $HL_{HF} > 50$ dB HL, and $HL_{LF} < 30$ dBHL.

c. Audiometric group-c: $HL_{HF} > 50$ dB HL, and $HL_{LF} > 30$ dBHL.

d. Audiometric group-d: $HL_{HF} < 50$ dB HL, and $HL_{LF} > 30$ dBHL.

III **Archetypal analysis:** Matrix factorization was applied to the results of the dimensionality reduction step. A given observation was represented as a convex combination of the archetypal patterns (Cutler Breiman, 1994). The analysis retrieves two matrices – the 'pattern matrix', which contained archetypal patterns represented in the data and the 'subject matrix', consisted of the weights corresponding to each pattern that resemble each of the observations. The specific implementation of the method used here was similar to Mørup and Hansen (2012). The identified patterns were ranked based on the degree of disability and handicap in each archetypal benefit pattern and labeled based on rehabilitation needs as in clinical triage as ABP0, ABP1, ABP2, and ABP3, being ABP0 the optimal benefit profile and ABP3 the suboptimal.

IV **Clustering:** The distance between observations and the four archetypal patterns was estimated using the weights contained in the subject matrix. The criteria used here was the nearest archetype (Ragozini et al., 2017). Each observation (subject) was then assigned to a cluster based on their weights. The "benefit profiles" were labeled as the archetypal benefit patterns (BP0, BP1, BP2, and BP3)

V **Importance estimation:** Once the subjective data corresponding to each of the subjects were analyzed with unsupervised learning techniques, supervised learning was used for estimating the importance of the specific items of the dataset. A decision tree ensemble was trained with a subset of the data corresponding to the items of the SSQ12 and the HHQ and the identified clusters as the output. The ensemble was trained with 200 surrogated trees using curvilinear prediction. The importance was obtained by the permutation of out-of-bag features, which provides the minimum square error averaged for each tree over the standard deviation across the trees.

6.3 Results

Factor analysis

The results of the factor analysis are summarized in Table 6.1. The parallel analysis revealed four factors as the optimal number of factors. The factor analysis was repeated in each of the stratified groups to check their similarity before further analysis. The loadings in each group were similar to the ones from the analysis of the entire dataset. Overall, the four factors corresponded to the four subdomains: speech understanding (SU), spatial perception (SP), qualities of hearing (QH) and hearing handicap (HH). These factors, taken together, explained a total of 50% of the variance.

Table 6.1: Variance explained by the rotated factors. The factors are labeled as the subdomain that reflects the highest loadings similar to (Akeroyd et al., 2014). The order of the factors in the table has been modified manually to match the labels instead of being sorted by the amount of variance explained.

Group	F_{SU} (%)	F_{SP} (%)	F_{QH} (%)	F_{HH} (%)
All dataset	14.4	15.0	12.4	13.4
Group-a	17.4	14.8	11.8	11.3
Group-b	17.6	15.2	10.4	14.4
Group-c	15.2	14.3	12.2	13.7
Group-d	14.1	15.8	12.2	13.2

F_{SU}: Speech understanding factor / F_{SP}: Spatial perception factor / F_{QH}: height Qualities of hearing factor / F_{HH}: Hearing Handicap factor.

Data-driven analysis Figure 6.2 shows the results of the data-driven analysis. The left panel corresponds to the patterns resulting from the archetypal analysis of the latent factors for each of the audiometric groups (a-d; lowercase to distinguish from the Sanchez-Lopez et al. (2020b) auditory profiles). The analysis of the overall data is indicated by the dotted lines. The left panel reflects the “archetypal patterns” (ABP) with respect to the latent factors. The middle panel represents the estimated importance of the individual items of the questionnaire. In the figure, only the three most important predictors in a given subdomain are shown for simplification. The right panel of Figure 6.2 combines the above findings and shows the resulting “patterns of benefit”. Each row shows the median scores and interquartile distances of the participants belonging to a given benefit profile (BP0-BP3) derived from the ABP (Figure 6.2 left panel). Each row represents a different audiometric group. The scores are shown for the most important items. These were derived from the unsupervised learning stage and are shown in the middle panel of Figure 6.2.

Archetypal benefit patterns

Figure 6.2 (left panel) shows the result of the archetypal analysis. The optimal profile (ABP0) showed a high score for all factors (green), which was similar for all of the audiometric groups, as well as for the entire data set (dotted lines). The near-optimal pattern (ABP1, yellow) was different for the different audiometric group. For group-a and group-d, the pattern showed high scores for the difficulties-related factors (F_{SU}, F_{SP}, F_{QH}) but a lower score for the handicap-related scores (F_{HH}). In contrast, for group-b and group-c, ABP1 showed lower scores reflecting the difficulties in speech understanding (F_{SU}). The four groups differed substantially in terms of the near-suboptimal pattern (ABP2, deep-red). For group-a, the lowest score corresponded to quality-related difficulties (F_{QH}), whereas for group-b, the lowest score reflected an increased handicap. For group-c and group-d, the lowest score corresponded to difficulties with spatial hearing, while the handicap-related scores were higher than in ABP1. The suboptimal pattern (ABP3, red) showed the lowest scores for all factors in group-c and group-d. However, for group-a the scores reflecting speech understanding and spatial hearing factors were lower, while for group-b the scores reflecting qualities were lower but not for handicap-related scores. The archetypal benefit patterns corresponding to the analysis of the overall data (dotted lines) resembled, to a large extent, the patterns observed in group-d.

Item importance estimation

The importance of the subdimensions in terms of difficulties and restrictions were estimated by the predictor importance of the individual items. The importance was considered here as indicative of priorities for hearing rehabilitation. Figure 6.2 (middle panel) shows the predictor importance for a subset of items corresponding to the three questions with the highest importance in each of the four domains

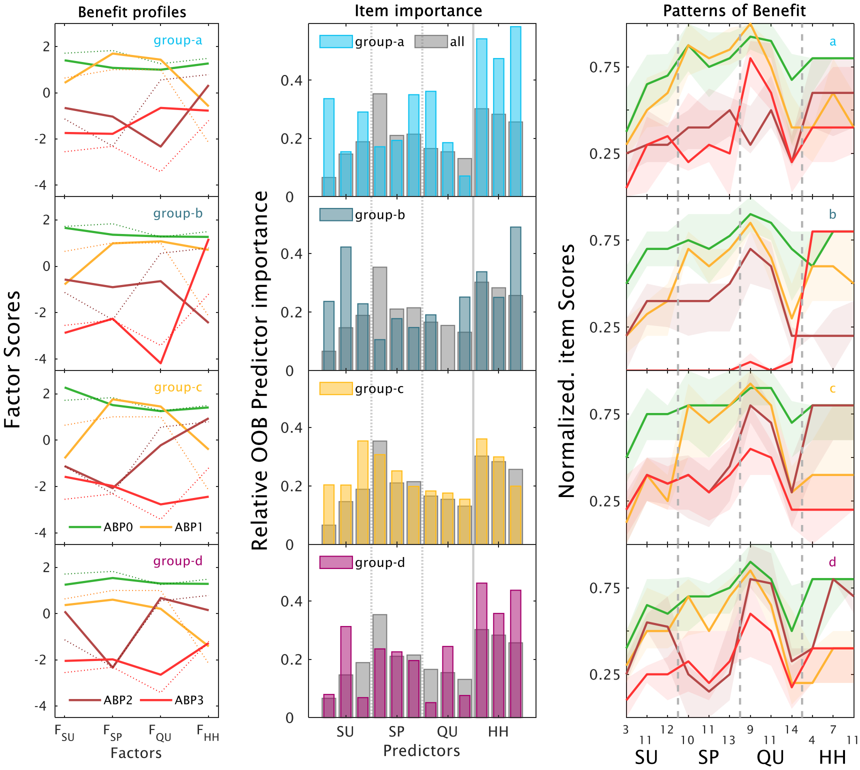


Figure 6.2: Data-driven analysis of subjective data stratified into audiometric groups. Each row corresponds to an audiometric group (a-d). Left panel: Archetypal benefit patterns (ABP) resulting from step III of the method. Each row corresponds to an audiometric group (a-d). Four patterns are ranked and labelled between optimal (ABP0; green) to suboptimal (ABP3; red). The analysis of the overall data is shown by the dotted lines. Middle panel: Relative importance of the individual items estimated by the Out-of-the-bag permuted features delta error of a random forest. The three items with higher importance are shown for each subdomain (SU, SP, QH and HH). The grey bars show the results of using the same procedure on the entire dataset. Right panel: Normalized median scores and interquartile distances for each of the clusters (benefit profile) derived from the benefit patterns (BP0, BP1, BP2, and BP3) across the most important items of each subdomain. The SSQ and HHQ scores were normalized between 0 and 1, with '0' corresponding to poor and '1' to good performance.

Table 6.2: Items of the SSQ and HHQ questionnaires with the highest importance for all the stratified groups. Three questions were selected for each of the subscales based on the sum of the OOB predictor importance obtained in step V: priority estimation stage of the data-driven method. The pragmatic subscale for each question is taken from (Gatehouse and Akeroyd, 2006)

Subscale Item	Question	Pragmatic Subscale
Speech 3	You are in a group of about five people, sitting around a table. It is an otherwise quiet place. You can see everyone else in the group. Can you follow the conversation?	Speech in quiet
Speech 11	You are in conversation with one person in a room where there are many other people talking. Can you follow what the person you are talking to is saying?	Speech in Speech context
Speech 12	You are with a group and the conversation switches from one person to another. Can you easily follow the conversation without missing the start of what each new speaker is saying?	Multiple speech-streams
Spatial 10	Can you tell from the sound which direction a bus or truck is moving, e.g. from your left to your right or right to left?	Distance and movement
Spatial 11	Can you tell from the sound of their voice or footsteps which direction a person is moving, e.g. from your left to your right or right to left?	Distance and movement
Spatial 13	Can you tell from the sound whether a bus or truck is coming towards you or going away?	Distance and movement
Qualities 9	Do everyday sounds that you can hear easily seem clear to you (not blurred)?	Sounds clarity and naturalness
Qualities 11	Do everyday sounds that you hear seem to have an artificial or unnatural quality?	Sounds clarity and naturalness
Qualities 14	Do you have to concentrate very much when listening to someone or something?	Listening effort
Handicap 4	How often is your self-confidence affected by your hearing difficulty?	Emotional
Handicap 7	How often does your difficulty with your hearing affect the way you feel about yourself ?	Emotional
Handicap 11	How often does your hearing difficulty restrict your social or personal life?	Social

(Table 5.2). The predictor importance is shown for each of the subgroups (in color) and the overall data (in grey). The highest importance shown in the analysis of the overall data corresponded to the spatial hearing (SP) difficulties. In particular, the item related to lateral sound movement and the ones related to handicap-related (HH) were found to be important. In contrast, the questions related to speech understanding (SU), and hearing qualities (QH) questions were found to be less important. Group-a showed higher importance than the overall group for SU in questions related to conversations with multiple talkers in quiet, SP difficulties in terms of distance, QH difficulties in the clarity of sounds and the three HH related questions. Group-b showed higher importance for SP difficulties related to speech-in-noise perception, QH difficulties related to listening effort and HH related to social participation. Group-c showed similar importance for SP and HH as the overall data. However, the importance obtained for SP difficulties related to multi-talker scenarios, specifically the ability to get the start of the sentences during conversational turn-taking, was higher than overall, as was the importance of the HH question about affected self-confidence. Group-d showed higher importance for speech-in-noise in the SU domain and the naturalness of the voices in the QH domain compared to the analysis of the entire data, and the highest importance obtained for the handicap-related items. Overall, the stratified approach for the analysis of priorities for hearing rehabilitation revealed different patterns of importance for the different audiometric groups of listeners.

Stratified patterns of benefit across of the important items

The right panel of Figure 6.2 shows the normalized median results and interquartile distributions for each of the clusters derived from the benefit profiles (BP0, BP1, BP2, and BP3) across the important questions shown in Table 6.2 for each audiometric profile group. The optimal pattern (BP0) was similar in all subgroups with normalized scores between 0.5 and 0.8 for the speech-related items, and around 0.8 for most of the items in the other domains. The suboptimal pattern (BP3) corresponded to low scores close to 0.3 for SU, SP and HH, but slightly higher scores for QH. The suboptimal patterns (BP3) were similar for the participants in groups a, c and d. In contrast, the suboptimal pattern group-b corresponded to low scores close to 0 in the difficulty subdomains and to high scores in the handicap-related items. The other benefit patterns are described in comparison to BP0. The near-optimal pattern (BP1) of group-a showed a decreased score in HH items, whereas the near-suboptimal (BP2) showed scores around 0.4 in all the items. The BP1 of group-b showed reduced scores in SU items (< 0.4) but not in the SP and QH items. The BP2 showed decreased scores in SP and HH items, whereas the SU and QH related difficulties were in line with the BP1. The BP1 of group-c showed decreased scores in SU and HH items. The BP2 pattern showed higher scores than BP1 for the HH questions but lower for the SP items, whereas QH and HH were in line with BP0. The BP1 of group-d showed reduced scored in HH items. The BP2 indicated higher scores than in BP1 for the HH items but substantially lower scores for the SP related items, whereas SU and QH related items where similar to the BP0 scores.

6.4 Discussion

Interpretation of the patterns of benefit

The data-driven analysis revealed four patterns of benefit, labeled as BP0, BP1, BP2, BP3. BP0, or “optimal”, is a pattern shown by patients who do not require additional intervention and may only need periodic follow-up visits (e.g. once per year). The HCP should ensure that this optimal result does not change by evaluating the intervention periodically. BP1, or “near-optimal”, corresponds to an intervention that requires minor improvements. BP1 corresponds to patients who require some adjustments or instructions in regular follow-up visits to improve the treatment. The HCP should be aware of the limitations and allocate time to perform these improvements successfully. BP2, or “near-suboptimal”, corresponds to an intervention that requires major improvements. Patients who show this pattern might reflect problems that require substantial additional intervention through structured sessions that are focused on different difficulties and handicaps. BP3, or “suboptimal”, is a pattern associated with patients with low benefit. This suggests that the initial intervention (i.e. the type of device or initial diagnosis) should be reconsidered. In this case, the HCP should evaluate the possibilities of changing the device (e.g. from a hearing-aid to a bone-anchored hearing device in cases of a conductive hearing loss) or evaluate the need for a multi-disciplinary approach (in cases of central auditory disorders or other comorbidities).

Specific priorities for hearing rehabilitation in different audiometric groups

Hearing rehabilitation can involve a broad variety of interventions. The intervention is often prioritized in terms of sensory management, perceptual training and counselling in a “holistic approach” (Boothroyd, 2007). In contrast, a pre-assessment with the SSQ12 and HHQ based on the present findings, as well as tempering expectations (Whitmer et al., 2016), can effectively guide further rehabilitation. The intervention can then be focused on overcoming specific hearing difficulties or handicaps in a systematic approach with the help of the present stratification. Table 6.3 shows the priorities for hearing rehabilitation for each of the four audiometric groups. The priorities are set by the differences between the patterns of benefits and the important items (Figure 6.2, right and middle panels). For example, when a patient classified as audiometric group-c receives a new hearing device, priority I should be to test the patient's ability to follow a conversation (deviation between BP1 and BP0), priority II to evaluate how the hearing difficulties affect his/her self-confidence, and priority III to assess spatial hearing abilities (deviation between BP2 and BP1). If the result of the evaluation is not optimal, an intervention focused on overcoming the specific hearing difficulties or handicaps should then be planned by the HCP.

Table 6.3: Priorities of hearing rehabilitation inferred from the data-driven approach on subjective data of hearing difficulties and handicaps.

	Priority I	Priority II	Priority III	Intervention
a	Handicap (social)	Speech (speech-in-quiet)	Spatial (Distance)	Counseling and adjustment of HA advanced features
b	Speech (speech-in-speech)	Spatial (movement)	Handicap (social)	HA advanced features, perceptual training and counseling
c	Speech (multi-streams)	Handicap (emotional)	Spatial (movement)	HA advanced features, counseling and perceptual training
d	Handicap (emotional)	Spatial (distance)	Speech (speech-in-speech)	Counseling and HA style

Audiometric groups and auditory profiles

The analysis of the subjective self-reported scores of hearing difficulties and handicaps identified patterns of benefit and priorities for hearing rehabilitation. The participants were divided into audiometric groups based on the auditory profiles from Sanchez-Lopez et al. (2020b). The auditory profiles were the result of a data-driven analysis of multidimensional data that involved measures of audibility, loudness perception, speech perception, binaural processing abilities and spectro-temporal resolution. The audiometric groups used in the present study to stratify the participants (a, b, c and d) cannot be considered equivalent to the auditory profiles (A, B, C and D) derived in Sanchez-Lopez et al. (2020b). However, there are similarities and discrepancies between the objective hearing deficits observed in the auditory profiles and the subjective difficulties and handicaps associated with the audiometric groups and their benefit profiles. In Sanchez-Lopez et al. (2020b), auditory profile A listeners showed, on average, a low degree of perceptual deficits and a close-to-normal speech intelligibility, whereas the results for audiometric group-a listeners of the present study indicated a high importance of speech communication in quiet, spatial perception difficulties and a high priority of rehabilitating hearing handicaps. Profile B listeners showed reduced speech-in-noise perception performance, which is consistent with the priorities of rehabilitation showed by group-b listeners (i.e. difficulties in speech understanding followed by spatial perception and hearing handicaps). Profile C listeners showed a high degree of perceptual deficits, consistent with the results of the audiometric group-c listeners that indicated a priority for rehabilitating difficulties in speech-in-noise perception and social participation. Finally, Profile D listeners showed near-normal suprathreshold perception, except for their abnormal loudness perception. However, the results of the audiometric group-d listeners showed a priority for rehabilitating hearing handicaps followed by spatial hearing difficulties.

Insights for hearing-aid evaluation and validation

Different shorter versions of the SSQ have been proposed (Demeester et al., 2012; Gablenz et al., 2018; Moulin et al., 2019; Noble et al., 2013). However, in the present study, the subset of questions shown in Table 6.2 was not intended to create a new short version of the questionnaire, but to better understand the differences among the audiometric groups. The SSQ12 (Noble et al., 2013) is a twelve-item questionnaire that was the result of an item selection process between three parties and based on different criteria, involving the scores reported in a factor analysis (Akeroyd et al., 2014) and including all ten pragmatic subscales (Gatehouse and Akeroyd, 2006). Noble et al. (2013) concluded that SSQ12 should be accompanied by the HHQ to provide a complete evaluation of the level of hearing disability and handicap. In the present study, the patterns of benefit revealed that disabilities and handicaps were in many cases independent, and a minor degree of difficulties do not always imply minor participation restrictions.

However, different patterns of difficulties and handicaps were observed in different audiometric groups, suggesting that the results of outcome measures used for assessing a clinical practice might be divided into meaningful groups to minimize the confounds of the sensory hearing deficits.

6.5 Conclusion

The data-driven approach for inferring patterns of benefit and priorities for hearing rehabilitation revealed different benefit profiles for the four audiometric groups of listeners considered in the present study. The observed patterns of benefit and priorities for hearing rehabilitation together with the four clinical subpopulations showing significant differences in perceptual deficits presented in Sanchez-Lopez et al. (2020) could help to guide the hearing rehabilitation based on perceptual deficits beyond the audiogram. The patterns of benefit and the use of stratification might improve the clinical intervention of the hearing loss and the efficiency and quality of service in the hearing care clinic.

Acknowledgments

The authors thank A. Ahrens, O. Cañete for their comments in an earlier version of the manuscript. The authors acknowledge the valuable feedback from the members of the Hearing Sciences Scottish Section during the realization of the present study. This work was supported by Innovation Fund Denmark Grand Solutions 5164-00011B (Better hEAring Rehabilitation project) Oticon, GN Hearing, WSAudiology and other partners (Aalborg University, University of Southern Denmark, the Technical University of Denmark, Force, Aalborg, Odense and Copenhagen University Hospitals). The funding and collaboration of all partners are sincerely acknowledged. WW was supported by the Medical Research Council [grant number MR/S003576/1]; and the Chief Scientist Office of the Scottish Government

7

Overall discussion

In this thesis, the basis for precision medicine in the field of rehabilitative audiology was explored. It was hypothesized that the complexity of sensorineural hearing loss can be characterized by the perceptual consequences of the hearingimpairments. The hearing deficits are assumed to be represented by two independent auditory perceptual distortions, distortion type-I and distortion type-II. Along these two dimensions, four archetypal auditory profiles can be identified: Profile A reflecting a low degree of distortions type-I and type-II; Profile B reflecting a high degree of distortion type-I and a low degree of distortion type-II; Profile C associated with a high degree of distortions type-I and type-II; and Profile D representing a low degree of distortion type-I and a high degree of distortion type-II. The following research questions were addressed in this work:

1. Can data-driven profiling identify these four auditory profiles?
2. Which auditory tests have the potential to be used in a clinical auditory profiling?
3. What are the characteristics that define the sub-populations of patients based on the clinical auditory profiles?
4. Do the listeners in the four auditory profiles have different needs in terms of hearing-aid compensation and hearing-rehabilitation?

7.1 Summary of main results

In *Chapter 2*, a data-driven method for the analysis of behavioral auditory outcomes was proposed and evaluated by analyzing two existing data sets from the literature (Johannesen et al., 2016; Thorup et al., 2016). The method was based on the hypothesis that the hearing deficits can be characterized by two independent perceptual distortions. It was also hypothesized that distortion type-I might be associated with audibility-related deficits and reduced spectral processing abilities, whereas distortion type-II might be associated with non-audibility-related deficits and reduced temporal processing abilities not reflected by the audiogram. The results of the analyses showed that four meaningful groups of listeners could be effectivelly identified based on the similarity of the listeners with "archetypal" auditory profiles derived from the data-driven profiling method. However, the analysis of the two considered data sets showed mixed results. The analysis of a data set obtained from a clinical setup and with listeners with either near-normal hearing or mild-to-moderate high-frequency hearing loss (Thorup et al., 2016) showed that distortion type-I was mainly associated with high-frequency hearing loss and reduced speech intelligibility in multi-talker scenarios, whereas distortion type-II was associated with reduced binaural processing abilities and elevated most comfortable levels at low frequencies. In contrast, the analysis of the other data set from a study performed in a research unit and with hearing-impaired listeners with moderate-to-severe hearing losses (Johannesen et al., 2016) showed that distortion type-I was mainly associated with high-frequency hearing loss and basilar membrane compression loss at high frequencies, whereas distortion type-II was associated with low-frequency hearing loss and reduced temporal fine structure sensitivity. Although the distortion type-I was in both studies associated with high-frequency hearing loss, the binary decision resulting from a decision tree trained with the corresponding data was different for the two studies. Furthermore, distortion type-II was associated with non-audibility related deficits in the analysis of one dataset (Thorup et al., 2016) but with low-frequency hearing loss in the analysis of the other dataset (Johannesen et al., 2016). Overall, the study provided a novel view on hearing loss characterization and provided insights into the capabilities and limitations of the chosen data-driven auditory profiling approach.

In *Chapter 2*, It was concluded that the correct characterization of the two hypothesized perceptual distortions, and the definition of the four auditory profiles, required a new dataset consisting of outcome measures from various aspects of auditory processing obtained from listeners with a variability of hearing abilities. In *Chapter 3*, an auditory test battery was proposed and tested with a group of seventy-five listeners. The criteria for inclusion were based on a thorough literature review. The selected tests explored different aspects of auditory processing and had shown potential for clinical implementation. The tests focused on six domains: audibility, loudness perception, speech perception, binaural processing abilities, spectro-temporal modulation sensitivity, and spectro-temporal resolution. The listeners were recruited from a clinical population of hearing-aid users and a control group of near-normal hearing listeners. In this chapter, each of the tests in the test battery was introduced and the motivation for their inclusion was explained. The normative data obtained with the normal-hearing and hearing-impaired listeners, as well as the test-retest reliability of each of the tests, were presented and discussed. Furthermore, an explorative analysis of the outcome measures of the tests was provided. A correlation analysis showed that the outcome variables were mainly divided into two correlated clusters. First, the variables of the outcomes obtained at high frequencies (e.g. hearing thresholds or temporal masking release and spectral masking release) were found to be interrelated. In this cluster, the variables corresponding to speech-in-noise perception were also found to be correlated to the outcomes measured at high-frequencies. Second, the outcomes measured at low frequencies (e.g. hearing threshold or loudness functions) were interrelated and correlated to the speech reception thresholds in quiet. A factor analysis revealed four principal factors corresponding to: 1) low-frequency processing and speech perception in quiet; 2) high-frequency processing; 3) most comfortable levels; and 4) speech-in-noise perception. The results were discussed in terms of the relationships across different aspects of auditory processing and the clinical feasibility of the tests.

In *Chapter 4*, the new dataset generated in *Chapter 3* was then analyzed with an iterative data-driven profiling method based on the method presented in *Chapter 2*. The robust data-driven auditory profiling approach yielded four clinically relevant subgroups of listeners. The archetypal patterns (“prototypes”) showed that Profiles[a] reflecting a high degree of distortion type-I (B and C) presented associated deficits in terms of high-frequency processing, binaural processing, speech intelligibility in noise and tone-in-noise detection at low frequencies. In contrast, profiles with a high degree of distortion type-II (C and D) showed reduced low-frequency processing, loudness perception, and spectral masking release. Overall, the results reported in *Chapter 4* strongly suggested that the listeners’ hearing deficits can be characterized by two independent auditory distortions. Distortion type-I was associated with “speech intelligibility related deficits” that affected listeners with audiometric thresholds above 50 dB HL at high-frequencies. Distortion type-II was associated with “loudness perception related distortion” that affected listeners with audiometric thresholds above 30 dB HL at low frequencies. Furthermore, the four profiles (A-B-C-D) showed similarities to the audiometric phenotypes provided in Dubno et al. (2013) suggesting that Profile B might be considered a sensory loss and Profile D might represent a metabolic loss.

Chapter 5 and *Chapter 6* explored the differences in terms of preference and likely priorities for hearing rehabilitation of the four auditory profiles identified in *Chapter 4*. *Chapter 5* represented a proof-of-concept study conducted with a few listeners who also participated in the study of *Chapter 4*. The listeners evaluated different compensation strategies using a realistic hearing-aid simulator in different realistic scenarios. Four hearing-aid strategies were implemented based on the conclusions from *Chapter 4*. Signal-to-noise (SNR) improvement

---

[a] Throughout the discussion the term "Profile" in capital letters corresponds to the group of listeners that share the same hearing deficits as the "archetypal pattern". Therefore, Profile A corresponds to "the listeners belonging to the group that showed high similarity to the archetypal profile A".

was selected as the strategy to overcome the "speech intelligibility related" deficits exhibited by the listeners in Profiles B and C. Loudness normalization was assumed to be the best solution for the loudness perception related deficits exhibited by Profiles C and D. The results showed that listeners belonging to the different auditory profiles might have different patterns of preference for the tested hearing-aid settings. As expected, listeners in Profiles C and D favored the hearing-aid settings aiming for loudness normalization in quiet scenarios. Profile C listeners preferred SNR improvement even for positive SNRs in noisy scenarios, whereas Profile B listeners preferred fast-acting compression but did not show a significant preference for SNR improvement suggesting that they might be susceptible to the distortions introduced by noise suppression algorithms. The findings of the proof-of-concept study support further investigation of the present approach in a larger clinical population and using clinically fitted hearing aids.

Chapter 6 focused on the data-driven analysis of subjective data from questionnaires of hearing disabilities and handicaps. The goal was to identify patterns of benefit associated with the four auditory profiles. Since this was a retrospective study, an approximation of the auditory profiles was done by dividing the participants into four audiometric groups based on the previous findings from Chapter 4. The groups were labeled as groups a-d (in lowercase) to distinguish them from the auditory profiles (A-D). The method applied for the subjective data yielded four archetypal patterns of benefit along with four latent factors: speech understanding, spatial perception, qualities of hearing and hearing handicaps. The patterns of benefit were ranked in terms of their priorities for hearing rehabilitation from "optimal", which corresponds to patients who do not require additional intervention, to "suboptimal", which corresponds to patients with low benefit. Based on the four benefit patterns and the importance of the individual items, the priorities for the improvements in hearing rehabilitation in the four audiometric groups (a-d) were estimated. Group-a showed a priority to overcome the handicaps, improve their speech in quiet and the perception of distance. Group-b showed a priority for improvements of speech in noisy scenarios, the perception of movement and an increase of social participation. Group-c showed a priority for improvements in speech understanding in multi-talker scenarios, better self-confidence, and an improved perception of movement. Group-d showed a priority for improving their self-confidence, spatial hearing and speech understanding in speech-in-speech scenarios. These findings were discussed in terms of their implications for planning hearing rehabilitation in a clinical context. The stratification of the listeners in audiometric groups revealed differences in the patterns of benefit across groups. The distinct patterns of benefit for each of the representative groups suggested that follow-up visits, focusing on obtaining specific improvements that can be measured in the hearing-care clinic, might be beneficial.

7.2 Hearing loss characterization

The presented data-driven approach for hearing loss characterization was limited to: 1) the use of psychoacoustical measures; 2) the use of auditory tests with potential for clinical implementation; and 3) the use of a clinical population of older adults with bilateral hearing loss.

Physiological measures, such as otoacoustic emissions and auditory evoked potentials, were not considered in the current approach. This was a decision in the interest of the characterization of the perceptual consequences derived of the hearing deficits rather than the “sources” of the hearing loss. Several paths can be taken from here to include physiological measures in the auditory profiling approach. For example, the auditory profiles defined in *Chapter 4* may serve as a starting point for a hearing loss characterization in terms of physiology and physiological measurements can be carried out in listeners previously classified into the four groups. A similar approach has been taken in Vaden et al. (2018) where the audiometric phenotypes from Dubno et al. (2013) were further characterized using otoacoustic emissions. Alternatively, otoacoustic emissions (OAE) and auditory evoke potentials (AEP) could be obtained in a new study and analyzed using a similar data-driven profiling method. The clinical devices for measuring OAE and AEP are usually focused on estimating the hearing thresholds, since they are typically used as a form of "objective" audiometry. Therefore, the physiological measures used in such an investigation might include responses to supra-threshold stimuli, especially designed to pinpoint certain aspects of auditory processing (Vasilkov and Verhulst, 2019). Furthermore, the auditory profiling approach could be used as part of genetic studies for a better understanding of the deficits presented in certain genotypes associated with specific impaired mechanisms (Bruce et al., 2019).

Cognitive factors are also important for characterizing the overall “listening profile” of individuals with hearing loss, as suggested in several studies (e.g., Humes, 2007; Rönnberg et al., 2016). In the present thesis, cognitive factors where considered as a confound rather than a missing part of the auditory profiling approach. A better understanding of the sensory dysfunction is needed to provide an efficient compensation of the hearing deficits rather than a compensation for the audibility loss. Therefore, the thesis focused only on the “auditory” aspects assuming that the perceptual measurements would be only partially influenced by the cognitive factors (Rönnberg et al., 2016). However, it would be of great interest to explore the cognitive factors from a bidirectional point of view. Cognition can affect the perception of the auditory stimuli presented in the test battery and the listener’s cognitive resources can also be affected by the “distortions” reflected by the auditory profiles and lead to an effortful listening experience (Peelle, 2018; Pichora-Fuller et al., 2016).

Another limitation of the present thesis is that the selected auditory measures were required to have potential for clinical implementation. Therefore, important aspects of auditory processing were left out of the test battery. For example, classic measurements of frequency selectivity and behavioral estimates of cochlear compression (Glasberg and Moore, 1990; Nelson et al., 2001) were not considered because they are time-consuming and often show training effects. However, some efforts to optimize these measures for clinical implementation have recently been developed (Fereczkowski et al., 2020; Hyvärinen et al., 2020). Furthermore, other aspects reflecting auditory perception abilities such as auditory stream segregation (Madsen et al., 2018), localization (Stecker and Gallun, 2012), modulation discrimination (Wiinberg et al., 2019), frequency modulation detection (Johannesen et al., 2016) or spectral modulation (Davies-Venn et al., 2015) were in fact considered during the process of designing the test battery. These aspects are relevant for hearing research but were eventually discarded for similar reasons, since they did not seem to be feasible for clinical setups yet. However these measures might be useful to understand the auditory distortions reflected in the four auditory profiles.

Besides the potential for clinical implementation, the tests that were language independent were prioritized. However, a test battery representing speech intelligibility deficits would be of great relevance. Such a test battery could be tested on a population of people with different hearing abilities and analyzed using a similar data-driven profiling method as the one presented here. This test battery might involve speech intelligibility tests in the presence of different interferers and spatial configurations (e.g. Lőcsei et al., 2016; Rosen et al., 2013). Besides, it might contain tests where speech intelligibility is affected by reverberation (Reinhart and Souza, 2016), distortions (Pichora-Fuller et al., 2007), or the use of amplification. In such a study, phenomena such as masking release or binaural unmasking could be further investigated using a data-driven approach.

Overall, the hearing loss characterization presented in this thesis provided four robust listener subpopulations that can be further characterized in terms of physiological aspects, cognitive factors, auditory processing and speech intelligibility in different conditions.

7.3 Implications for hearing technology

Hearing-aid candidacy

In rehabilitative audiology, there exists a stratification of the listeners depending on the degree and type of hearing loss. The hearing device selected for the hearing loss compensation can be: 1) Bone-anchored hearing devices indicated in cases of chronic middle-ear dysfunction. These devices stimulate the inner-ear bypassing the middle ear using bone-conduction; 2) Cochlear implants indicated in cases of severe-to-profound hearing losses. These devices stimulate the auditory nerve by electrical current applied by electrodes implanted by surgery (Vickers et al., 2016); 3) Electro-acoustic stimulation (EAS) indicated in steeply sloping high-frequency hearing losses and a residual hearing at low frequencies. These devices are a special type of hearing devices, which apply electric stimulation to the high frequencies and acoustic stimulation to the low frequencies. These devices are indicated in severe-to-profound hearing losses with residual hearing at low frequencies. 4) Hearing aids are indicated for the rest of cases and are the most common solution for hearing loss compensation.

The four auditory profiles defined in *Chapter 4* differed in terms of two perceptual distortions “speech intelligibility related deficits” and “loudness perception-related deficits”. One of the profiles (Profile C) showed a poor performance in most of the considered auditory measures. Although there was no systematic assessment of the benefit obtained with real hearing aids by this group, the same listeners participated in a study exploring their aided performance with a hearing aid simulator (*appendix B*). The present findings support the consideration of a different candidature for this group. Since Profile C still showed elevated speech reception thresholds, it is possible that they benefit from other hearing devices, such as electro-acoustic stimulation. This is consistent with the idea that the listeners with a high degree of distortion type-I might have an associated IHC loss. Therefore, acoustic stimulation might not elicit the "desired" sensation since the organ of Corti is damaged, and electric stimulation might be beneficial in such conditions. However, no studies have been reported assessing the speech-in-noise perception benefit in patients with hearing aids (pre-implanted) and with EAS (post-implanted).

Assuming that listeners belonging to the four auditory profiles would be candidates for the use of bilateral hearing aids, further considerations should be made in terms of the hearing aid selection, gain prescription and advanced features.

Hearing-aid selection

The selection of the adequate hearing aid style and acoustic coupling (i.e. the eartips) is crucial for hearing rehabilitation. In *Chapters 5* and *Chaper 6*, preferences for hearing-aid settings and priorities for hearing rehabilitation were explored in the context of the auditory profiling. Two findings were observed here that can be useful for hearing-aid selection: 1) The indication of in-the-ear hearing aids for listeners in profile D. It has been shown that the placement of the microphone at the entrance of the ear canal provides a more acoustically transparent response and maintain the binaural cues (Denk et al., 2018). Therefore, this hearing style would fulfill the priority for spatial hearing observed in Profile D. 2) The use of custom earpieces in profiles B, C and D. The reason for this is that profiles C and D showed preference for settings with low-frequency amplification. In real hearing aids, this would require a "close fitting", to enable the correct amplification at low frequencies by occluding the ear canal. However, other burdens are typically associated to occlusive fittings (Winkler et al., 2016) such as the abnormal perception of the listener’s own voice. Besides, Profiles B and D indicated a preference for SNR improvement. However, an effective SNR improvement will only be possible if the signal is processed by the hearing aid, which is difficult when using large vents or “instant” eartips (Keidser et al., 2007).

Gain prescription

Currently, the gain prescription is based on the audiogram and some factors like gender or experience with hearing aids. However, the gain prescription could be prescribed by different formulas in different subpopulations based on their hearing deficits. In *Chapter 5*, a fitting formula that applied different offsets to the nonlinear gain prescription was evaluated. The results suggested that Profiles C and D might benefit from a prescription aiming to normalize the loudness perception. A more accurate loudness normalization can be implemented in the clinic by using the results of loudness scaling tests (ACALOS) and the prescription for loudness restoration (Oetting et al., 2018). Another approach could be to set certain weights in the model-based prescription of the formula of the National Acoustics Laboratory (NAL-NL2 Keidser et al., 2011). NAL-NL2 prescribes the hearing aid gain based on an optimization process and a trade-off between two models: a speech intelligibility model and a loudness model. Regarding the findings from *Chapters 4* and *5*, the models used in the NAL-NL2 prescription may be modified and include the speech intelligibility related deficits and the loudness perception related deficits observed in this thesis. Consequently, the prescription would apply different criteria depending on the auditory profile of the listener and the optimization process would provide a weighted solution where either speech or loudness are prioritized.

Advanced features

The advanced features of the hearing aids are particularly useful for providing listening comfort and increasing the satisfaction of the listener in complex situations. Often, the hearing-aid functionalities are modified depending on the sound scene, providing an optimized set of parameters for specific sound environments (Keidser et al., 2005). It can be argued that the current hearing-aid technology focuses in the patient's ecology such that the hearing-aid parameters are automatically modified depending on the listening condition. For example, the noise management algorithms are often activated in challenging hearing situations to improve the listening comfort, and deactivated in less challenging situations to maintain the sound quality. However, the individualization of the hearing-aid parameters based on the hearing deficits might provide an additional benefit. In *Chapter 5*, different hearing-aid settings were tested in realistic scenarios. The listeners with a higher degree of speech intelligibility related distortions showed a preference for hearing-aid settings with SNR improvement. Importantly, this preference was not only maintained but even more strongly preferred in scenarios with positive SNRs, i.e. in less challenging sound scenarios. This finding suggests that aggressive parameters of noise suppressors might still be relevant in more favorable conditions, especially in listeners with reduced speech intelligibility in noise. Hearing-aid advanced features are beneficial for noise reduction and listening comfort. However, these algorithms also provide audible distortions. For example, aggressive single-channel noise reduction leads to audible "musical noise" (Kates, 2017), whereas aggressive beamforming can produce distortions of binaural cues (Neher et al., 2016). *Appendix B* was concerned on the performance of listeners divided into auditory profiles that were also participants of the study in *Chapter 4*. All listeners experienced an improvement when the target speech was in front and aggressive noise reduction and beamforming were applied. In contrast, speech intelligibility dropped to 0% when the target was presented from one side. These findings suggest that the trade-off between SNR improvement and audible distortions introduced by the hearing aids might be guided by the hearing deficits. Listeners with larger difficulties in terms of speech understanding might profit from more aggresive parameters. However, hearing tests to evaluate the acceptable limits of these aggressive parameters might be necessary in the future audiological practice.

7.4 Modelling hearing deficits

Computational auditory models are useful for testing hypothesis about the origin of certain hearing phenomena and for explaining perceptual behaviors. Overall, the existing auditory models are either based on an accurate reproduction of the biological and physiological stages of the auditory system (e.g. Carney, 1993), or in the reproduction of a perceptual behavior by efficient signal processing, not necessarily inspired in physiological findings (e.g., Dau et al., 1996; Taal et al., 2010). With the findings of the present thesis, two approaches can help to better characterize the hearing deficits: 1) To explore the hypothesis of the association between auditory profiles and the audiometric phenotypes (metabolic and sensory types); 2) To reproduce the data obtained in *Chapter 3* (test battery) with a "hearing-impaired" model.

Existing physiological models are typically inspired by animal studies and are useful for reproducing human electrophysiological data. In contrast, these models are challenged when reproducing perceptual data. This is in part because there are no accepted and established criteria about what response should be used for such an evaluation. The physiological aspects associated with the auditory profiles might be investigated using an exploratory approach rather than aiming to reproduce the perceptual data. Physiological auditory models usually have parameters that can also simulate the effects of hearing-impairments, such as OHC loss, IHC loss or neurodegeneration (Bruce et al., 2018; Carney, 1993; Meddis et al., 2010; Panda et al., 2014). However, only few of them allow the direct alteration of the endocochlear potential loss associated with the metabolic type of presbycusis (Verhulst et al., 2018). The exploratory approach could involve the assessment of four models, one of each representing an auditory profile. Each model would provide an internal representation of the incoming signal, which has been affected by the hearing impairments associated with each profile. The differences in the neural coding “acuity” of relevant stimuli, for example the ones used in some of the tests of *Chapter 3*, might provide interesting insights about the connection between impaired mechanisms and auditory perception.

Perceptual models can be efficient but still biologically inspired (Meddis et al., 2010). The main point of these type of models is that the output can be interpreted as a metric of perceptual sensitivity, which is useful for applications in signal detection theory. An example of these types of models is the perceptual auditory model of Dau et al. (1997a,b). This model is inspired by both, findings from physiological studies (e.g. the adaptation stage; Dau et al., 1996) and from perceptual studies (modulation filter-bank; Dau et al., 1997a). This model has been able to reproduce different auditory tasks and it has been further developed for auditory perception and for speech intelligibility. Therefore, the last realization of the model (CASP; Jepsen et al., 2008; Kowalewski et al., 2020 and sCASP;Relaño-Iborra et al., 2019), which can account for behavioral data from hearing-impaired listeners, would be the interesting candidates for reproducing the data from *Chapter 3*. However, the model need to include a binaural extension (Chabot-Leclerc et al., 2016). The model's behaviour should be able to represent categorical scaling data (Trevino et al., 2016). The individualization process need to be reconsider and not be based on physiological estimates (Jepsen and Dau, 2011) and some stages of the model should be revised as suggested in Relaño-Iborra (2019). Overall, CASP can exploit the dataset generated here and shed light to the aspects of auditory signal processing that can be affected in connection to the auditory profiles.

The implementation of new efficient computational models usually starts with the goal of reproducing a specific auditory behavior by a normal hearing "model" that can be generalized to reproduce several auditory tasks. Perceptual models can be based on the reproduction of data from different tests performed by people with different hearing abilities (e.g. Schädler et al., 2020). However, the challenge is to create a model that is efficient, accurate and still based on biological foundations that can provide insights about the mechanisms producing the hearing deficits. The current data, and the increasing knowledge in auditory modelling and hearing loss characterization might promote the implementation of new "impaired models". An impaired model would be based on reproducing data from hearing-impaired listeners, but could still be able to reproduce normal-hearing behaviour. The model can be individualized by modifying "impairing factors" in a single task with data from a variety of listeners and then be generalized towards other auditory tasks (*Chapter 3*). This might provide a better understanding of the hearing deficits, their dimensions and interactions, and whether signal processing might be able to compensate for them. Overall, the use of computational auditory models combined with the outcomes of the present thesis might help to simplify the exploration of complex physiological models and to guide the implementation of efficient, biologically inspired, models of auditory perception.

7.5 Perspectives for "precision audiology"

In the last two decades, the audiological practice has experienced substantial changes. Newborn screening protocols using objective measures of the auditory function has been successfully implemented in several countries (Wroblewska-Seniuk et al., 2017). Therapies for protecting and most likely regenerating impaired mechanisms in the auditory systems are being tested showing promising results in animal studies (Kujawa and Liberman, 2019; Wang and Puel, 2018). The candidacy criteria for implanted devices have evolved rapidly (Vickers et al., 2016). Besides, hearing-aid technology has developed tremendously and the benefit reported by the people with hearing loss has increased (Kochkin, 2010). New technologies are also providing important insights into "how" hearing-aid fitting parameters might change based on real-life behavioural data (Johansen et al., 2017; Pasta et al., 2019). However, a most recent paradigm shift has been towards new forms of services, especially the establishment of over-the-counter (OTC) hearing aids and hearable technologies. These devices allow the consumer to acquire a hearing solution with no need to consult a hearing care professional, which improves the accessibility of hearing devices for people with hearing difficulties (Edwards, 2020). Nevertheless, this new scenario might compromise the quality of the hearing rehabilitation of the patients that decide to “self-fit” their hearing aids without professional advice.

In terms of technology, some advanced OTC hearing aids do not differ substantially from clinically fitted hearing aids (Callaway and Punch, 2008). Therefore, the audiological clinical practice might need to be upgraded and include new elements that support the importance of hearing-care practice in the audiology clinic. In this context, the current opportunity of implementing “precision audiology” appears to be attractive for audiologists, manufacturers, and hearing-aid users. Although “precision audiology” can be associated with current advances in genetics and therapeutics (Rudman et al., 2018), there exist other approaches related to hearing rehabilitation with hearing aids that can be implemented in the clinical practice by applying current knowledge about the hearing deficits. This new approach might increase the quality of hearing rehabilitation of the patients by improving the sensory compensation.

The conditions for applying precision medicine (Trusheim et al., 2007) to the field of audiology are: 1) A heterogeneous disease associated with different biological mechanisms, shown by the variability of the perceptual deficits observed in hearing-impaired listeners; 2) Multiple treatment options, shown by the numerous parameter spaces and the signal processing available in the current and future hearing-aid technology; 3) Clinical markers, which can be based on diagnostic or auditory tests that are associated with specific treatments or can be used to predict the optimal treatment.

Four approaches can be explored that may satisfy these three conditions.

Prediction of optimal parameters

In this approach, the hearing-aid parameters can be fitted based on individual audiological tests. If an audiological test can predict the benefit and/or preference for certain sets of parameters, the hearing-aid fitting can be individualized (Schädler et al., 2018; Zaar et al., 2019). This approach could easily be implemented and would be the most conservative approach. Likely, some manufacturers have already identified some correlations between the acceptance of some aggressive parameters and specific hearing deficits (Jensen et al., 2019; Jepsen and Soli, 2016; Krueger et al., 2019; Serman, M., Fischer, R-L., Herbig, R., & Hannemann, 2017; Theill, 2018; Zaar et al., 2019). Therefore, the development of this type of precision audiology would depend on the hearing-aid manufacturers, which would most likely explore the possibilities available in their specific products for such individualization. Overall, this perspective represents the most plausible scenario since there would be a direct connection between the new tests to be adopted in the clinic and the hearing aid fitting. However, the individualization can be too specific and lead to different tests for each manufacturer's products.

Interactive hearing-aid fitting

In this approach, some hearing-aid parameters could be adjusted based on the responses of the listeners to specific tests as it was proposed by Carhart (1946). Interactive hearing-aid fitting does not imply self-fitting by the user but represents a more elaborated approach (Alphons Marie Franck et al., 2007; Kiessling et al., 2006). This idea has also been termed "trainable", so the hearing-aid learns the optimal settings (Dillon et al., 2006) from the patient. The first step would be to identify the tasks and parameters that are susceptible to be used in an interactive approach. For example, in the field of optometry, the prescription of the power of the ophtalmic lenses (diopters), is estimated by a test of character discrimination and visual acuity. In audiology, the tasks could involve speech intelligibility tests, loudness perception or acceptable noise level with a “master hearing aid” able to modify the parameters of interest during the interactive process (Neher and Wagener, 2016; Oetting et al., 2018). This approach would require exhaustive research efforts since the aided performance tasks and parameters that can be generalized are still unclear. Furthermore, the procedure can make use of recent research for hearing-aid personalization using machine learning techniques (Nielsen et al., 2014). Besides, this “personalized hearing-aid fitting” does not often consider the hearing deficits of the patient and it might be only driven by personal preferences. Overall, this perspective represents an uncertain scenario that requires more evidence and a limited complexity to be adopted in the audiological practice.

Susceptibility to audible distortions

In this approach, some hearing-aid algorithms such as noise reduction, compression or directionality could be adjusted depending on “traits” or “profile cues”. This approach has already been explored by identifying listeners who are more susceptible to noise or distortions associated with certain algorithms (Neher, 2014; Neher et al., 2016). Also, the sensitivity of the listeners to distortions associated with temporal or spectral cues can be used for identifying groups of listeners that obtain a better response with certain sets of parameters (Souza et al., 2015; Souza et al., 2020). This can be useful for adjusting the aggressiveness of advanced algorithms that can improve the SNR at the cost of introducing audible distortions (see appendix B). This approach has shown promising results for its clinical implementation. However, the susceptibility to the audible distortions has not been found to be directly associated with hearing deficits but with cognitive factors (Souza et al., 2019). Furthermore, the tests used for identifying the traits (Völker et al., 2018) or profile cue (Souza et al., 2020) are not feasible for their clinical implementation yet. Overall, this perspective represents a relevant scenario since it provides additional information of the patient's auditory processing that can be related to specific distortions associated to hearing-aid processing.

Stratified hearing-aid fitting

In this approach, sub-populations of listeners with hearing loss are likely to respond better to specific hearing-aid settings. When the "profile" of the listener has been identified, the hearing-aid parameters are adjusted to promote an optimal response associated with that specific subgroup. This approach relies on the possibility of using "auditory markers" that are common to the listeners belonging to the same group allowing the classification of the listeners in the different sub-groups through audiological tests. This is the approach proposed by the auditory profiling discussed in the present thesis. Despite the possibility of implementing this approach with the current technology, the real potential would be in the hearing aid development. If a manufacturer designs different technologies for each of the sub-groups, which are tailored towards overcoming their specific hearing deficits, both the benefit obtained from a more appropiate sensory management as well as the avoidance of the undesired side effects of signal processing algorithms may improve their satisfaction. Furthermore, the use of auditory models might be particularly useful to create new forms of signal processing able to overcome the deficits reflected by specific auditory impairments (Bondy et al., 2004; Chen et al., 2005). Therefore, the investigation of the optimal rehabilitation for each sub-population can be guided by previous research in stratified medicine (Lonergan et al., 2017). Overall, this perspective represents a consistent scenario since it provides additional information of the patient's hearing deficits supported by evidence. However, the development of new forms of signal processing is still to be investigated.

The two main critical uncertainties for these four scenarios are the 1) the complexity of the tests to be included in the future clinical practice and 2) the evidence supporting the additional benefit of the new approach. However, these perspectives are not mutually exclusive and can be complementary to each other leading to a scenario where these four approaches are combined. For example, the use of more advanced diagnostics to classify patients can lead to a better "first fit" for relevant subpopulations. The susceptibility to certain distortions that are associated to generic hearing-aid signal processing can guide the selection of the most adequate hearing aid technology. Furthermore, the use of specific tests might predict the optimal parameters of certain signal processing algorithms used by specific hearing aid models. Finally, interactive iterations can improve the hearing aid adjustments for the individual needs and preferences. The strategies suggested in these approaches might themselves be individualized. This means that that some perspectives might have different efficacy in different people so the use of these future tools can be reduced to specific target populations.

Personalization beyond sensory management

The current holistic approach (Boothroyd, 2007) used for hearing rehabilitation depends on the experience and knowledge of the hearing care professional, who has to assess the hearing difficulties in a non-systematic way. This means that an "experienced" and "smart" HCP can provide an excellent service by individualizing the rehabilitation. However, the decisions taken are based on experience and not on evidence supporting those decisions which compromises the overall quality of service. The outcomes of *Chapter 6* (i.e., the patterns of benefit and priorities for hearing rehabilitation) and *Chapter 4* (i.e., four clinical subpopulations with significant differences in perceptual deficits) could help to optimize the hearing rehabilitation based on perceptual deficits beyond the audiogram. More personalized pathways would be possible by applying a decision-making based on audiometric data and short questionnaires. These decisions can then be evaluated and improved based on evidence.

Overall, the verification and validation of the hearing intervention at the individual level in the hearing-care clinic may include; 1) An improved hearing-aid evaluation exploring aided performance measures based on auditory tests rather than self-reported questionnaires; 2) Personalized auditory training considering the use of auditory training in listeners showing low HA benefit and reflecting the specific hearing difficulties (Bees et al., 2019; Lawrence et al., 2018); and 3) Interactive ecological intervention, which might help to guide the follow-up visits. For example, the trial-and-error approach for fine tuning can be substituted by a self-fine tuning, involving the patient in the process and also compiling their hearing experiences in real life (Convery et al., 2019; Lund et al., 2020).

This thesis aimed at enabling "precision audiology" and provide new avenues for developing auditory-profile based compensation strategies for hearing rehabilitation. The success of any form of "precision audiology" would imply a better understanding perceptual consequences of different hearing impairments and new methods for their compensation.

> "Our greatest hopes could become reality in the future. With the technology at our disposal, the possibilities are unbounded. All we need to do is make sure we keep talking.”
>
> *Stephen Hawking (1942-2018)*

Bibliography

Abrams, H., T. H. Chisolm, and R. McArdle (2002). "A cost-utility analysis of adult group audiologic rehabilitation: are the benefits worth the cost?" In: *Journal of rehabilitation research and development* 39.5, pp. 549–58.

Ahroon, W. A., R. I. Davis, and R. P. Hamemik (1993). "The Role of Tuning Curve Variables and Threshold Measures in the Estimation of Sensory Cell Loss". In: *International Journal of Audiology* 32.4, pp. 244–259.

Akeroyd, M. A., F. H. Guy, D. L. Harrison, and S. L. Suller (2014). "A factor analysis of the SSQ (Speech, spatial, and qualities of hearing scale)". In: *International Journal of Audiology* 53.2, pp. 101–114.

Al-Salim, S. C., J. G. Kopun, S. T. Neely, W. Jesteadt, B. Stiegemann, and M. P. Gorga (2010). "Reliability of Categorical Loudness Scaling and Its Relation to Threshold". In: *Ear and Hearing* 31.4, pp. 567–578.

Alphons Marie Franck, B., M. Boymans, and W. Albert Dreschler (2007). "Interactive fitting of multiple algorithms implemented in the same digital hearing aid". In: *International Journal of Audiology* 46.7, pp. 388–397.

Anderson, M. C., K. H. Arehart, and P. E. Souza (2018). "Survey of Current Practice in the Fitting and Fine-Tuning of Common Signal-Processing Features in Hearing Aids for Adults". eng. In: *Journal of the American Academy of Audiology* 29.2, pp. 118–124.

Baker, R. and V. Hazan (2011). "DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs". In: *Behavior Research Methods* 43.3, pp. 761–770.

Baumgärtel, R. M. et al. (2015). "Comparing Binaural Pre-processing Strategies I: Instrumental Evaluation". In: *Trends in hearing* 19, pp. 1–16.

Bees, K. L., D. Guan, N. Alsarrage, and G. D. Searchfield (2019). “The effects of auditory object identification and localization (AOIL) training on noise acceptance and loudness discomfort in persons with normal hearing”. In: *Speech, Language and Hearing* 22.2, pp. 71–78.

Bernstein, J. G. W., H. Danielsson, M. Hällgren, S. Stenfelt, J. Rönnberg, and T. Lunner (2016). “Spectrotemporal Modulation Sensitivity as a Predictor of Speech-Reception Performance in Noise With Hearing Aids.” In: *Trends in hearing* 20, pp. 1–17.

Bernstein, J. G., G. Mehraei, S. Shamma, F. J. Gallun, S. M. Theodoroff, and M. R. Leek (2013). “Spectrotemporal Modulation Sensitivity as a Predictor of Speech Intelligibility for Hearing-Impaired Listeners”. In: *Journal of the American Academy of Audiology* 24.4, pp. 293–306.

Bianchi, F., L. H. Carney, T. Dau, and S. Santurette (2019). “Effects of Musical Training and Hearing Loss on Fundamental Frequency Discrimination and Temporal Fine Structure Processing: Psychophysics and Modeling”. In: *Journal of the Association for Research in Otolaryngology* 20.3, pp. 263–277.

Bisgaard, N., M. S. M. G. Vlaming, and M. Dahlquist (2010). “Standard Audiograms for the IEC 60118-15 Measurement Procedure”. In: *Trends in Amplification* 14.2, pp. 113–120.

Boldt, J., U. Kjems, M. S. Pedersen, T. Lunner, and D. Wang (2008). “Estimation of the Ideal Binary Mask using Directional Systems”. In: *Proceedings of the 11th International Workshop on Acoustic Echo and Noise Control.*

Bondy, J., S. Becker, I. Bruce, L. Trainor, and S. Haykin (2004). “A novel signal-processing strategy for hearing-aid design: neurocompensation”. In: *Signal Processing* 84.7, pp. 1239–1253.

Boothroyd, A. (2007). “Adult Aural Rehabilitation: What Is It and Does It Work?” In: *Trends in Amplification* 11.2, pp. 63–71.

Bramhall, N. et al. (2019). “The search for noise-induced cochlear synaptopathy in humans: Mission impossible?” In: *Hearing Research* 377, pp. 88–103.

Brand, T. and V. Hohmann (2002). “An adaptive procedure for categorical loudness scaling”. In: *The Journal of the Acoustical Society of America* 112.4, pp. 1597–1604.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone (2017). *Classification and regression trees.*

Brons, I., R. Houben, and W. A. Dreschler (2014). “Effects of Noise Reduction on Speech Intelligibility, Perceived Listening Effort, and Personal Preference in Hearing-Impaired Listeners”. In: *Trends in Hearing* 18, p. 233121651455392.

Brown, M. and F. Musiek (2013). “Pathways: The Fundamentals of Masking Level Differences for Assessing Auditory Function”. In: *The Hearing Journal* 66.1, p. 16.

Bruce, I. C., Y. Erfani, and M. S. Zilany (2018). “A phenomenological model of the synapse between the inner hair cell and auditory nerve: Implications of limited neurotransmitter release sites”. In: *Hearing Research* 360, pp. 40–54.

Bruce, I. C. et al. (2019). “Phenotyping and computational modeling of diverse forms of genetic hearing loss”. In: *23rd International Congress on Acoustics , ICA 2019*. September. Aachen , Germany.: Deutsche Gesellschaft für Akustik, pp. 686–687.

Brungart, D. S., B. M. Sheffield, and L. R. Kubli (2014). “Development of a test battery for evaluating speech perception in complex listening environments”. In: *The Journal of the Acoustical Society of America* 136.2, pp. 777–790.

Callaway, S. L. and J. L. Punch (2008). “An Electroacoustic Analysis of Over-the-Counter Hearing Aids”. In: *American Journal of Audiology* 17.1, pp. 14–24.

Carhart, R. (1946). “Tests for selection of hearing aids”. In: *The Laryngoscope* 56.12, pp. 780–794.

Carney, L. H. (1993). “A model for the responses of low-frequency auditory-nerve fibers in cat”. In: *The Journal of the Acoustical Society of America* 93.1, pp. 401–417.

Chabot-Leclerc, A., E. N. MacDonald, and T. Dau (2016). “Predicting binaural speech intelligibility using the signal-to-noise ratio in the envelope power spectrum domain”. In: *The Journal of the Acoustical Society of America* 140.1, pp. 192–205.

Chang, Y.-S. et al. (2019). “Standard Audiograms for Koreans Derived through Hierarchical Clustering Using Data from the Korean National Health and Nutrition Examination Survey 2009–2012”. In: *Scientific Reports* 9.1, p. 3675.

Chen, Z., S. Becker, J. Bondy, I. C. Bruce, and S. Haykin (2005). “A Novel Model-Based Hearing Compensation Design Using a Gradient-Free Optimization Method”. In: *Neural Computation* 17.12, pp. 2648–2671.

Ching, T. Y. C., H. Dillon, R. Katsch, and D. Byrne (2001). “Maximizing Effective Audibility in Hearing Aid Fitting”. In: *Ear and Hearing* 22.3, pp. 212–224.

Choi, J. E. et al. (2016). “Evaluation of Cochlear Implant Candidates using a Non-linguistic Spectrotemporal Modulation Detection Test”. In: *Scientific Reports* 6.April, p. 35235.

Chung, K. (2004). “Challenges and recent developments in hearing aids. Part I. Speech understanding in noise, microphone technologies and noise reduction algorithms.” In: *Trends in amplification* 8, pp. 83–124.

Convery, E., G. Keidser, M. McLelland, and J. Groth (2019). “A Smartphone App to Facilitate Remote Patient-Provider Communication in Hearing Health Care: Usability and Effect on Hearing Aid Outcomes”. In: *Telemedicine and e-Health*, tmj.2019.0109.

Cox, R. M. and G. C. Alexander (1999). “Measuring Satisfaction with Amplification in Daily Life: The SADL Scale”. In: *Ear and Hearing* 20.4, pp. 306–320.

Cox, R. M. (2003). “Assessment of subjective outcome of hearing aid fitting: getting the client’s point of view”. In: *International Journal of Audiology* 42.sup1, pp. 90–96.

Cox, R. M. and G. C. Alexander (1995). “The Abbreviated Profile of Hearing Aid Benefit”. In: *Ear and Hearing* 16.2, pp. 176–186.

Cox, R. M., J. A. Johnson, and J. Xu (2016). “Impact of Hearing Aid Technology on Outcomes in Daily Life I”. In: *Ear and Hearing* 37.4, e224–e237.

Cox, R. et al. (2000). “Optimal Outcome Measures, Research Priorities, and International Cooperation”. In: *Ear and Hearing* 21.Supplement, 106S–115S.

Cramer, E. M. and W. H. Huggins (1958). “Creation of Pitch through Binaural Interaction”. In: *The Journal of the Acoustical Society of America* 30.5, pp. 413–417.

Cunningham, L. L. and D. L. Tucci (2017). “Hearing Loss in Adults”. In: *New England Journal of Medicine* 377.25. Ed. by A. H. Ropper, pp. 2465–2473.

Cutler, A. and L. Breiman (1994). “Archetypal Analysis”. In: *Technometrics* 36.4, pp. 338–347.

Darwin, C. J. (1997). “Auditory grouping”. In: *Trends in Cognitive Sciences* 1.9, pp. 327–333.

Dau, T., B. Kollmeier, and A. Kohlrausch (1997a). “Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers”. In: *The Journal of the Acoustical Society of America* 102.5, pp. 2892–2905.

Dau, T., B. Kollmeier, and A. Kohlrausch (1997b). “Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration”. In: *The Journal of the Acoustical Society of America* 102.5, pp. 2906–2919.

Dau, T., D. Püschel, and A. Kohlrausch (1996). “A quantitative model of the “effective” signal processing in the auditory system. I. Model structure”. In: *The Journal of the Acoustical Society of America* 99.6, pp. 3615–3622.

Davies-Venn, E., P. Nelson, and P. Souza (2015). “Comparing auditory filter bandwidths, spectral ripple modulation detection, spectral ripple discrimination, and speech recognition: Normal and impaired hearinga)”. In: *The Journal of the Acoustical Society of America* 138.1, pp. 492–503.

Davies-Venn, E., P. Souza, M. Brennan, and G. C. Stecker (2009). “Effects of Audibility and Multichannel Wide Dynamic Range Compression on Consonant Recognition for Listeners with Severe Hearing Loss”. In: *Ear & Hearing* 30, pp. 494–504.

Dawes, P, A Maharani, J Nazroo, G Tampubolon, and N Pendleton (2019). “Evidence that hearing aids could slow cognitive decline in later life”. In: *Hearing Review* 26.1, pp. 10–11.

Demeester, K. et al. (2012). “Hearing Disability Measured by the Speech, Spatial, and Qualities of Hearing Scale in Clinically Normal-Hearing and Hearing-Impaired Middle-Aged Persons, and Disability Screening by Means of a Reduced SSQ (the SSQ5)”. In: *Ear and Hearing* 33.5, pp. 615–616.

Denk, F., S. M. A. Ernst, S. D. Ewert, and B. Kollmeier (2018). “Adapting Hearing Devices to the Individual Ear Acoustics: Database and Target Response Correction Functions for Various Device Styles”. In: *Trends in Hearing* 22, p. 233121651877931.

Desloge, J. G., C. M. Reed, L. D. Braida, Z. D. Perez, and L. A. D’Aquila (2017). “Masking release for hearing-impaired listeners: The effect of increased audibility through reduction of amplitude variability". In: *The Journal of the Acoustical Society of America* 141.6, pp. 4452–4465.

Dillon, H., J. A. Zakis, H. McDermott, G. Keidser, W. Dreschler, and E. Convery (2006). "The trainable hearing aid". In: *The Hearing Journal* 59.4, p. 30.

Dreschler, W. A., H Verschuure, C Ludvigsen, and S Westermann (2001). "ICRA noises: artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment." In: *Audiology: Official Organ of the International Society of Audiology* 40.3, pp. 148–157.

Dubno, J. R., M. A. Eckert, F.-S. Lee, L. J. Matthews, and R. A. Schmiedt (2013). "Classifying Human Audiometric Phenotypes of Age-Related Hearing Loss from Animal Models". In: *Journal of the Association for Research in Otolaryngology* 14.5, pp. 687–701.

Dudoit, S. and J. Fridlyand (2003). "Bagging to improve the accuracy of a clustering procedure". In: *Bioinformatics* 19.9, pp. 1090–1099.

Durlach, N. I. (1963). "Equalization and Cancellation Theory of Binaural Masking-Level Differences". In: *The Journal of the Acoustical Society of America* 35.8, pp. 1206–1218.

Edwards, B. (2020). "Emerging Technologies, Market Segments, and MarkeTrak 10 Insights in Hearing Health Technology". In: *Seminars in Hearing* 41.01, pp. 037–054.

Elberling, C., C. Ludvigsen, and P. E. Lyregaard (1989). "Dantale: A New Danish Speech Material". In: *Scandinavian Audiology* 18.3, pp. 169–175.

Esch, T. E. M. van et al. (2013). "Evaluation of the preliminary auditory profile test battery in an international multi-centre study". In: *International Journal of Audiology* 52.5, pp. 305–321.

Esch, T. E. M. van and W. A. Dreschler (2011). "Measuring spectral and temporal resolution simultaneously: a comparison between two tests." In: *International journal of audiology* 50.7, pp. 477–90.

Eustaquio-Martín, A. and E. A. Lopez-Poveda (2011). "Isoresponse Versus Isoinput Estimates of Cochlear Filter Tuning". eng. In: *Journal of the Association for Research in Otolaryngology* 12.3, pp. 281–299.

Ewert, S. (2013). “AFC - A modular framework for running psychoacoustic experiments and computational perception models”. In: *Proceedings of the International Conference on Acoustics AIA-DAGA 2013*. Merano, Italy, pp. 1326–1329.

Fereczkowski, M, T Dau, and E. N. MacDonald (2020). “Comparison of clinical feasibility of behavioural and physiological estimates of peripheral compression”. In: *Trends in Hearing* Submitted.

Festen, J. M. and R. Plomp (1990). “Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing”. In: *The Journal of the Acoustical Society of America* 88.4, pp. 1725–1736.

Fleming, S. and S. Docs (2016). *Commissioning Services for People with Hearing Loss: A Framework for Clinical Commissioning Groups*. Tech. rep.

Franklin, J. (2005). “The elements of statistical learning: data mining, inference and prediction”. In: *The Mathematical Intelligencer* 27.2, pp. 83–85.

Frawley, W. J., G. Piatetsky-Shapiro, and C. J. Matheus (1992). “Knowledge Discovery in Databases: An Overview”. In: *AI Magazine* 13.3, p. 57.

Füllgrabe, C., A. J. Harland, A. P. Sęk, and B. C. J. Moore (2017). “Development of a method for determining binaural sensitivity to temporal fine structure”. In: *International Journal of Audiology* 56.12, pp. 926–935.

Füllgrabe, C. and B. C. J. Moore (2017). “Evaluation of a Method for Determining Binaural Sensitivity to Temporal Fine Structure (TFS-AF Test) for Older Listeners With Normal and Impaired Low-Frequency Hearing”. In: *Trends in Hearing* 21, p. 233121651773723.

Gablenz, P. von, F. Otto-Sobotka, and I. Holube (2018). “Adjusting Expectations: Hearing Abilities in a Population-Based Sample Using an SSQ Short Form”. In: *Trends in Hearing* 22, p. 2331216518784837.

Gallun, F. J. et al. (2018). “Development and validation of Portable Automated Rapid Testing (PART) measures for auditory research”. In: *The Journal of the Acoustical Society of America*. Vol. 143. 3, p. 050002.

Gatehouse, S. (1993). “Role of perceptual acclimatization in the selection of frequency responses for hearing aids.” In: *Journal of the American Academy of Audiology*.

Gatehouse, S. (1999). “Glasgow Hearing Aid Benefit Profile : Derivation and Validation of a Client-centered Outcome Measure for Hearing Aid Services”. In: *J Am Acad Audiol* 10, pp. 80–103.

Gatehouse, S. and M. Akeroyd (2006). “Two-eared listening in dynamic situations”. In: *International Journal of Audiology* 45.sup1, pp. 120–124.

Gatehouse, S., G. Naylor, and C. Elberling (1999). “Aspects of auditory ecology and psychoacoustic function as determinants of benefit from and candidature for non-linear processing in hearing aids”. In: *18th DanavoxSymposium: Auditory Models and Non-linear Hearing Instruments.* Ed. by A. N. Rasmussen, P. A. Osterhammel, T Anderson, and T Poulsen. Copenhagen, Denmark: Holmens Trykkeri, pp. 221–233.

Gatehouse, S., G. Naylor, and C. Elberling (2006a). “Linear and nonlinear hearing aid fittings – 1. Patterns of benefit”. In: *International Journal of Audiology* 45.3, pp. 130–152.

Gatehouse, S., G. Naylor, and C. Elberling (2006b). “Linear and nonlinear hearing aid fittings – 2. Patterns of candidature”. In: *International Journal of Audiology* 45.3, pp. 153–171.

Gatehouse, S. and W. Noble (2004). “The Speech, Spatial and Qualities of Hearing Scale (SSQ)”. In: *International Journal of Audiology* 43.2, pp. 85–99.

Geetha, C. and P. Manjula (2014). “Effect of compression, digital noise reduction and directionality on envelope difference index, log-likelihood ratio and perceived quality”. In: *Audiology Research* 4.1.

Gelfand, S. A. (2009). *Essentials of Audiology.*

Gerlach, M., B. Farb, W. Revelle, and L. A. Nunes Amaral (2018). “A robust data-driven approach identifies four personality types across four large data sets”. In: *Nature Human Behaviour* 2.10, pp. 735–742.

Gieseler, A., M. A. S. Tahden, C. M. Thiel, K. C. Wagener, M. Meis, and H. Colonius (2017). “Auditory and Non-Auditory Contributions for Unaided Speech Recognition in Noise as a Function of Hearing Aid Use”. In: *Frontiers in Psychology* 8.

Glasberg, B. R. and B. J. Moore (1989). “Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech.” In: *Scandinavian audiology. Supplementum* 32, pp. 1–25.

Glasberg, B. R. and B. C. Moore (1990). “Derivation of auditory filter shapes from notched-noise data”. In: *Hearing Research* 47.1-2, pp. 103–138.

Gligorijević, V., N. Malod-Dognin, and N. Pržulj (2016). “Integrative methods for analyzing big data in precision medicine”. In: *PROTEOMICS* 16.5, pp. 741–758.

Goldstein, D. P. and S. D. G. Stephens (1981). “Audiological Rehabilitation: Management Model I”. In: *International Journal of Audiology* 20.5, pp. 432–452.

Gorga, M. P. and S. T. Neely (2003). “Cost-effectiveness and test-performance factors in relation to universal newborn hearing screening”. In: *Mental Retardation and Developmental Disabilities Research Reviews* 9.2, pp. 103–108.

Graydon, K, C Waterworth, H Miller, and H Gunasekera (2019). “Global burden of hearing impairment and ear disease”. In: *The Journal of Laryngology & Otology* 133.1, pp. 18–25.

Hagerman, B. and Å. Olofsson (2004). “A method to measure the effect of noise reduction algorithms using simultaneous speech and noise”. In: *Acta Acustica united with Acustica*.

Hallberg, L. R. (1998). “Evaluation of a Swedish version of the Hearing Disabilities and Handicaps Scale, based on a clinical sample of 101 men with noise-induced hearing loss”. In: *Scandinavian Audiology* 27.1, pp. 21–29.

Hamernik, R. P., J. H. Patterson, G. A. Turrentine, and W. A. Ahroon (1989). “The quantitative relation between sensory cell loss and hearing thresholds”. In: *Hearing Research* 38.3, pp. 199–211.

Henry, K. S., M. Sayles, A. E. Hickox, and M. G. Heinz (2019). “Divergent Auditory Nerve Encoding Deficits Between Two Common Etiologies of Sensorineural Hearing Loss”. In: *The Journal of Neuroscience* 39.35, pp. 6879–6887.

Hétu, R, L. Getty, L. Philibert, F. Desilets, W. Noble, and D. Stephens (1994). “Mise au point d’un outil Clinique pour la mesure d’incapacités auditives et de handicaps”. In: *J Speech-Lang Path Audiol* 18, pp. 83–95.

Hinrich, J. L., S. E. Bardenfleth, R. E. Roge, N. W. Churchill, K. H. Madsen, and M. Morup (2016). “Archetypal Analysis for Modeling Multisubject fMRI Data”. In: *IEEE Journal of Selected Topics in Signal Processing* 10.7, pp. 1160–1171.

Holube, I., S. Fredelake, M. Vlaming, and B. Kollmeier (2010). “Development and analysis of an International Speech Test Signal (ISTS)”. In: *International Journal of Audiology* 49.12, pp. 891–903.

Houtgast, T. and J. M. Festen (2008). “On the auditory and cognitive functions that may explain an individual’s elevation of the speech reception threshold in noise”. In: *International Journal of Audiology* 47.6, pp. 287–295.

Hu, Y. and P. C. Loizou (2008). “Evaluation of Objective Quality Measures for Speech Enhancement”. In: *IEEE Transactions on Audio, Speech, and Language Processing* 16.1, pp. 229–238.

Humes, L. E. (1994). “Response to Comments on ’Evaluating a speech-reception threshold model for hearing impaired listeners’ [J. Acoust. Soc. Am. 96, 586–587 (1994)]”. In: *The Journal of the Acoustical Society of America* 96.1994, p. 588.

Humes, L. E. (2007). “The Contributions of Audibility and Cognitive Factors to the Benefit Provided by Amplified Speech to Older Adults”. In: *Journal of the American Academy of Audiology* 18.7, pp. 590–603.

Hyvärinen, P, M Fereczkowski, and E MacDonald (2020). “Evaluation of a notched-noise test on a mobile phone”. In: *Trends in Hearing* (u.rev.

ISO 16832 (2006). “Acoustics - Loudness scaling by means of categories”. In: *International Organization for Standardization* 2006, pp. 1–12.

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

ISO 8253-3 (2012). “Acoustics. Audiometric test methods - Part 3: Speech audiometry”. In: *International Organization for Standardization.*

Imsiecke, M., B. Krüger, A. Büchner, T. Lenarz, and W. Nogueira (2019). “Interaction Between Electric and Acoustic Stimulation Influences Speech Perception in Ipsilateral EAS Users”. In: *Ear and Hearing*, p. 1.

Irving, S., L. Gillespie, R. Richardson, D. Rowe, J. B. Fallon, and A. K. Wise (2014). “Electroacoustic Stimulation: Now and into the Future”. In: *BioMed Research International* 2014, pp. 1–17.

Jameson, J. L. and D. L. Longo (2015). “Precision Medicine — Personalized, Problematic, and Promising”. In: *New England Journal of Medicine* 372.23, pp. 2229–2234.

Jensen, N., A. Pedersen, E. Petersen, and J. Theill (2019). “When one size does not fit all: Special hearing solutions for special occasions”. In: *Hearing Review* 26.8, pp. 22–26.

Jenstad, L. M. and P. E. Souza (2005). “Quantifying the Effect of Compression Hearing Aid Release Time on Speech Acoustics and Intelligibility”. In: *Journal of Speech Language and Hearing Research* 48.3, p. 651.

Jepsen, M. L. and T. Dau (2011). “Characterizing auditory processing and perception in individual listeners with sensorineural hearing loss.” In: *The Journal of the Acoustical Society of America* 129.1999, pp. 262–281.

Jepsen, M. L., S. D. Ewert, and T. Dau (2008). “A computational model of human auditory signal processing and perception”. In: *The Journal of the Acoustical Society of America* 124.1, pp. 422–438.

Jepsen, M. L. and S. D. Soli (2016). *Method of fitting a hearing aid system and a hearing aid fitting system. U.S. Patent No. 9,363,614.*

Jerger, J., L. Anthony, S. Jerger, and L. Mauldin (1974). “Studies in Impedance Audiometry: III. Middle Ear Disorders”. In: *Archives of Otolaryngology - Head and Neck Surgery* 99.3, pp. 165–171.

Jerger, J. and S. Jerger (1967). “Psychoacoustic Comparison of Cochlear and VIIIth Nerve Disorders”. In: *Journal of Speech and Hearing Research* 10.4, pp. 659–688.

Jerger, J. and S. Jerger (1974). “Audiological Comparison of Cochlear & Eighth Nerve Disorders”. In: *Annals of Otology, Rhinology & Laryngology* 83.3, pp. 275–285.

Johannesen, P. T., P. Pérez-González, S. Kalluri, J. L. Blanco, and E. A. Lopez-Poveda (2016). “The Influence of Cochlear Mechanical Dysfunction, Temporal Processing Deficits, and Age on the Intelligibility of Audible Speech in Noise for Hearing-Impaired Listeners”. In: *Trends in Hearing* 20, p. 233121651664105.

Johansen, B., M. K. Petersen, N. H. Pontoppidan, P. Sandholm, and J. E. Larsen (2017). “Rethinking Hearing Aid Fitting by Learning From Behavioral Patterns”. In: *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors* *in Computing Systems - CHI EA '17*. New York, New York, USA: ACM Press, pp. 1733–1739.

Jorgensen, L. E. (2016). “Verification and validation of hearing aids: Opportunity not an obstacle”. In: *Journal of Otology* 11.2, pp. 57–62.

Jürgens, T., B. Kollmeier, T. Brand, and S. D. Ewert (2011). “Assessment of auditory nonlinearity for listeners with different hearing losses using temporal masking and categorical loudness scaling”. In: *Hearing Research* 280.1-2, pp. 177–191.

Kaernbach, C. (1990). “A single-interval adjustment-matrix (SIAM) procedure for unbiased adaptive testing”. In: *The Journal of the Acoustical Society of America* 88.6, pp. 2645–2655.

Kamerer, A. M., J. G. Kopun, S. E. Fultz, S. T. Neely, and D. M. Rasetshwane (2019). “Reliability of Measures Intended to Assess Threshold-Independent Hearing Disorders”. In: *Ear and Hearing* 40.6, pp. 1267–1279.

Kates, J. M. (2005). “Principles of Digital Dynamic-Range Compression”. In: *Trends In Amplification* 9.2, pp. 45–76.

Kates, J. M. (2017). “Modeling the effects of single-microphone noise-suppression”. In: *Speech Communication* 90, pp. 15–25.

Keidser, G., H. Dillon, M. Flax, T. Ching, and S. Brewer (2011). “The NAL-NL2 prescription procedure”. In: *Audiology Research* 1.1S, pp. 1–3.

Keidser, G., C. Brew, S. Brewer, H. Dillon, F. Grant, and L. Storey (2005). “The preferred response slopes and two-channel compression ratios in twenty listening conditions by hearing-impaired and normal-hearing listeners and their relationship to the acoustic input”. In: *International Journal of Audiology* 44.11, pp. 656–670.

Keidser, G., L. Carter, J. Chalupper, and H. Dillon (2007). “Effect of low-frequency gain and venting effects on the benefit derived from directionality and noise reduction in hearing aids”. In: *International Journal of Audiology* 46.10, pp. 554–568.

Keidser, G., H. Dillon, and D. Byrne (1995). “Candidates for Multiple Frequency Response Characteristics”. In: *Ear and Hearing* 16.6, pp. 562–574.

Keidser, G., H. Dillon, L. Carter, and A. O'Brien (2012). “NAL-NL2 Empirical Adjustments”. In: *Trends in Amplification* 16.4, pp. 211–223.

Keidser, G. and F. Grant (2001). “Comparing Loudness Normalization (IHAFF) with Speech Intelligibility Maximization (NAL-NL1) when Implemented in a Two-Channel Device”. In: *Ear and Hearing* 22.6, pp. 501–515.

Kiessling, J. (2001). “Hearing aid fitting procedures - state-of-the-art and current issues”. In: *Scandinavian Audiology* 30.1, pp. 57–59.

Kiessling, J., M. Müller, and M. Latzel (2006). “Fitting strategies and candidature criteria for unilateral and bilateral hearing aid fittings”. In: *International Journal of Audiology* 45.sup1, pp. 53–62.

Killion, M. C. and P. A. Niquette (2000). “What can the pure-tone audiogram tell us about a patients SNR loss?” In: *The Hearing Journal* 53.3, pp. 46–48.

Killion, M. C., P. A. Niquette, G. I. Gudmundsen, L. J. Revit, and S. Banerjee (2004). “Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners”. In: *The Journal of the Acoustical Society of America* 116.4, pp. 2395–2405.

Kirkby-Strachan, G and C Que-Hee (2016). “Implantable hearing devices – An update”. In: *Australian Family Physician* 45, pp. 370–373.

Kochkin, S. (2010). “MarkeTrak VIII : Consumer satisfaction”. In: *The Hearing Journal* 63.1, p. 19.

Kollmeier, B. and J. Kiessling (2018). “Functionality of hearing aids: state-of-the-art and future model-based solutions”. In: *International Journal of Audiology* 57.sup3, S3–S28.

Koo, T. K. and M. Y. Li (2016). “A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research.” In: *Journal of chiropractic medicine.*

Kowalewski, B., M. Fereczkowski, O. Strelcyk, E. MacDonald, and T. Dau (2020). “Assessing the effects of hearing-aid compression on auditory spectral and temporal resolution using an auditory modeling framework”. In: *Acoustical Science and Technology* 41.1, pp. 214–222.

Krohne, L. G., Y. Wang, J. L. Hinrich, M. Moerup, R. C. K. Chan, and K. H. Madsen (2019). “Classification of social anhedonia using temporal and spatial network features from a social cognition fMRI task”. In: *Human Brain Mapping* 40.17, pp. 4965–4981.

Krueger, H., J. Raether, S. Allegro-Baumann, J. Rehmann, and T. Boeld (2019). *Method and apparatus for fitting a hearing device. U.S. Patent No. 10,219,727.*

Kujawa, S. G. and M. C. Liberman (2009). “Adding Insult to Injury: Cochlear Nerve Degeneration after "Temporary" Noise-Induced Hearing Loss”. In: *Journal of Neuroscience* 29.45, pp. 14077–14085.

Kujawa, S. G. and M. C. Liberman (2019). “Translating animal models to human therapeutics in noise-induced and age-related hearing loss”. In: *Hearing Research* 377, pp. 44–52.

Lansbergen, S. and W. Dreschler (2020). “Classification of Hearing Aids Into Feature Profiles Using Hierarchical Latent Class Analysis Applied to a Large Dataset of Hearing Aids”. In: *Ear & Hearing* (under rev.

Laplante-Lévesque, A., L. Hickson, and L. Worrall (2010). “Rehabilitation of Older Adults With Hearing Impairment: A Critical Review”. In: *Journal of Aging and Health* 22.2, pp. 143–153.

Larsby, B and S Arlinger (1998). “A method for evaluating temporal, spectral and combined temporal-spectral resolution of hearing.” In: *Scandinavian audiology* 27, pp. 3–12.

Larson, V. D. et al. (2002). “A Multi-Center, Double Blind Clinical Trial Comparing Benefit from Three Commonly Used Hearing Aid Circuits*”. In: *Ear and Hearing* 23.4, pp. 269–276.

Lawrence, B. J. et al. (2018). “Auditory and Cognitive Training for Cognition in Adults With Hearing Loss: A Systematic Review and Meta-Analysis”. In: *Trends in Hearing* 22, p. 233121651879209.

Lecluyse, W., C. M. Tan, D. McFerran, and R. Meddis (2013). “Acquisition of auditory profiles for good and impaired hearing”. In: *International Journal of Audiology* 52.9, pp. 596–605.

Li, X., S. Rao, Y. Wang, and B. Gong (May 2004). “Gene mining: a novel and powerful ensemble decision approach to hunting for disease genes using microarray expression profiling”. In: *Nucleic Acids Research* 32.9, pp. 2685–2694.

Liberman, M. C., M. J. Epstein, S. S. Cleveland, H. Wang, and S. F. Maison (2016). “Toward a Differential Diagnosis of Hidden Hearing Loss in Humans”. In: *PLOS ONE* 11.9. Ed. by M. S. Malmierca, e0162726.

Lidén, G. and M. Korsan-Bengtsen (1973). “Audiometric Manifestations of Retrocochlear Lesions”. In: *Scandinavian Audiology* 2.1, pp. 29–40.

Lin, F. R. et al. (2013). “Hearing Loss and Cognitive Decline in Older Adults”. In: *JAMA Internal Medicine* 173.4, p. 293.

Lobarinas, E., R. Salvi, and D. Ding (2013). “Insensitivity of the audiogram to carboplatin induced inner hair cell loss in chinchillas”. In: *Hearing Research* 302, pp. 113–120.

Lonergan, M. et al. (2017). “Defining drug response for stratified medicine”. In: *Drug Discovery Today* 22.1, pp. 173–179.

Lopez-Poveda, E. A. (2014). “Why do I hear but not understand? Stochastic undersampling as a model of degraded neural encoding of speech”. In: *Frontiers in Neuroscience* 8.October, pp. 1–7.

Lopez-Poveda, E. A., P. T. Johannesen, P. Pérez-González, J. L. Blanco, S. Kalluri, and B. Edwards (2017). “Predictors of Hearing-Aid Outcomes”. In: *Trends in Hearing* 21, p. 233121651773052.

Lopez-Poveda, E. a. and P. T. Johannesen (2012). “Behavioral Estimates of the Contribution of Inner and Outer Hair Cell Dysfunction to Individualized Audiometric Loss”. In: *Journal of the Association for Research in Otolaryngology* 13.4, pp. 485–504.

Lund, K, R. P. Ordoñez, J. B. Nielsen, and D Hammershøi (2020). “Sentence-based experience-logging in new hearing aid users”. In: *American Journal of Audiology* Special is.[under review].

Lőcsei, G., J. H. Pedersen, S. Laugesen, S. Santurette, T. Dau, and E. N. MacDonald (2016). “Temporal Fine-Structure Coding and Lateralized Speech Perception in Normal-Hearing and Hearing-Impaired Listeners”. In: *Trends in Hearing* 20.0, p. 233121651666096.

Madsen, S. M., T. Dau, and B. C. Moore (2018). “Effect of harmonic rank on sequential sound segregation”. In: *Hearing Research* 367, pp. 161–168.

Manchaiah, V., D. Stephens, and R. Meredith (2011). “The patient journey of adults with hearing impairment: the patients’ views”. In: *Clinical Otolaryngology* 36.3, pp. 227–234.

Mclean, A. (2008). *Development and Implementation of a National Funding and Service System for Hearing Aids Stage One Report March 2008*. Tech. rep.

Meddis, R., W. Lecluyse, C. M. Tan, M. R. Panda, and R. Ferry (2010). "Beyond the Audiogram: Identifying and Modeling Patterns of Hearing Deficits". In: *The Neurophysiological Bases of Auditory Perception*. New York, NY: Springer New York, pp. 631–640.

Mehraei, G., F. J. Gallun, M. R. Leek, and J. G. W. Bernstein (2014). "Spectrotemporal modulation sensitivity for hearing-impaired listeners: Dependence on carrier center frequency and the relationship to speech intelligibility". In: *The Journal of the Acoustical Society of America* 136.1, pp. 301–316.

Mellor, J. C., M. A. Stone, and J. Keane (2018). "Application of Data Mining to "Big Data" Acquired in Audiology: Principles and Potential". In: *Trends in Hearing* 22, p. 233121651877681.

Miller, C. W., R. A. Bentler, Y.-H. Wu, J. Lewis, and K. Tremblay (2017). "Output signal-to-noise ratio and speech perception in noise: effects of algorithm". In: *International Journal of Audiology* 56.8, pp. 568–579.

Mills, J., R. Schmiedt, B. Schulte, and J. Dubno (2006). "Age-Related Hearing Loss: A Loss of Voltage, Not Hair Cells". In: *Seminars in Hearing* 27.4, pp. 228–236.

Moore, B. C. J., M Huss, D. A. Vickers, B. R. Glasberg, and J. I. Alcántara (2000). "A test for the diagnosis of dead regions in the cochlea". In: *British journal of audiology* 34.4, pp. 205–224.

Moore, B. C. J. B. J., D. A. Vickers, C. J. Plack, and A. J. Oxenham (1999). "Inter-relationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism". In: *The Journal of the Acoustical Society of America* 106.5, pp. 2761–2778.

Moore, B. C. J. (2001). "Dead Regions in the Cochlea: Diagnosis, Perceptual Consequences, and Implications for the Fitting of Hearing Aids". In: *Trends in Amplification* 5.1, pp. 1–34.

Moore, B. C. J. (2007). *Cochlear Hearing Loss*. Ed. by B. C. Moore. West Sussex, England: John Wiley & Sons, Ltd, pp. 1–332.

Moore, B. C. J. (2016). "A review of the perceptual effects of hearing loss for frequencies above 3 kHz". In: *International Journal of Audiology* 55.July, pp. 707–714.

Mørup, M. and L. K. Hansen (2012). “Archetypal analysis for machine learning and data mining”. In: *Neurocomputing* 80, pp. 54–63.

Motallebzadeh, H., J. A. M. Soons, and S. Puria (2018). “Cochlear amplification and tuning depend on the cellular arrangement within the organ of Corti”. In: *Proceedings of the National Academy of Sciences* 115.22, pp. 5762–5767.

Moulin, A., J. Vergne, S. Gallego, and C. Micheyl (2019). “A New Speech, Spatial, and Qualities of Hearing Scale Short-Form: Factor, Cluster, and Comparative Analyses”. In: *Ear and Hearing* 40.4.

Neher, T. (2014). “Relating hearing loss and executive functions to hearing aid users’ preference for, and speech recognition with, different combinations of binaural noise reduction and microphone directionality”. In: *Frontiers in Neuroscience* 8.

Neher, T. (2017). “Characterizing the binaural contribution to speech-in-noise reception in elderly hearing-impaired listeners”. In: *The Journal of the Acoustical Society of America.*

Neher, T., S. Laugesen, N. Søgaard Jensen, and L. Kragelund (2011). “Can basic auditory and cognitive measures predict hearing-impaired listeners’ localization and spatial speech recognition abilities?” In: *The Journal of the Acoustical Society of America* 130.3, pp. 1542–1558.

Neher, T. and K. C. Wagener (2016). “Investigating Differences in Preferred Noise Reduction Strength Among Hearing Aid Users”. In: *Trends in Hearing* 20, p. 233121651665579.

Neher, T., K. C. Wagener, and R.-L. Fischer (2016). “Directional Processing and Noise Reduction in Hearing Aids: Individual and Situational Influences on Preferred Setting”. In: *Journal of the American Academy of Audiology* 27.8, pp. 628–646.

Nelson, D. A., A. C. Schroder, and M. Wojtczak (2001). “A new procedure for measuring peripheral compression in normal-hearing and hearing-impaired listeners”. In: *The Journal of the Acoustical Society of America* 110.4, pp. 2045–2064.

Newman, C. W., B. E. Weinstein, G. P. Jacobson, and G. A. Hug (1990). “The Hearing Handicap Inventory for Adults”. In: *Ear and Hearing* 11.6, pp. 430–433.

Nielsen, J. B. and T. Dau (2011). “The Danish hearing in noise test”. In: *International Journal of Audiology* 50.3, pp. 202–208.

Nielsen, J., J. Nielsen, and J. Larsen (2014). "Perception-based Personalization of Hearing Aids using Gaussian Processes and Active Learning". In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, pp. 1–1.

Nilsson, M., S. D. Soli, and J. a. Sullivan (1994). "Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise". In: *The Journal of the Acoustical Society of America* 95.2, pp. 1085–1099.

Noble, W., N. S. Jensen, G. Naylor, N. Bhullar, and M. A. Akeroyd (2013). "A short form of the Speech, Spatial and Qualities of Hearing scale suitable for clinical use: The SSQ12". In: *International Journal of Audiology* 52.6, pp. 409–412.

Oetting, D., T. Brand, and S. D. Ewert (2014). "Optimized loudness-function estimation for categorical loudness scaling data". In: *Hearing Research* 316, pp. 16–27.

Oetting, D., V. Hohmann, J. E. Appell, B. Kollmeier, and S. D. Ewert (2018). "Restoring Perceived Loudness for Listeners With Hearing Loss". In: *Ear and hearing*.

Oxenham, A. J. and A. M. Simonson (2009). "Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference". In: *The Journal of the Acoustical Society of America* 125.1, pp. 457–468.

Panda, M. R., W. Lecluyse, C. M. Tan, T. Jürgens, and R. Meddis (2014). "Hearing dummies: Individualized computer models of hearing impairment". In: *International Journal of Audiology* 53.10, pp. 699–709.

Papakonstantinou, A., O. Strelcyk, and T. Dau (2011). "Relations between perceptual measures of temporal processing, auditory-evoked brainstem responses and speech intelligibility in noise". In: *Hearing Research* 280.1-2, pp. 30–37.

Parthasarathy, A., S. Romero Pinto, R. M. Lewis, W. Goedicke, and D. B. Polley (2020). "Data-driven segmentation of audiometric phenotypes across a large clinical cohort". eng. In: *Scientific Reports* 10.1, pp. 6704, 6704.

Pasta, A., M. K. Petersen, K. J. Jensen, and J. E. Larsen (2019). "Rethinking hearing aids as recommender systems". In: *Proceedings of the 13th ACM Conference on Recommender Systems. HealthRecSys'19*. Copenhagen, Denmark.

Patuzzi, R. (1993). "Otoacoustic emissions and the categorization of cochlear and retro-cochlear lesions". In: *British Journal of Audiology* 27.2, pp. 91–95.

Pauler, M., H. F. Schuknecht, and A. R. Thornton (1986). “Correlative studies of cochlear neuronal loss with speech discrimination and pure-tone thresholds”. In: *Archives of Oto-Rhino-Laryngology* 243.3, pp. 200–206.

Peelle, J. E. (2018). “Listening Effort”. In: *Ear and Hearing* 39.2, pp. 204–214.

Pichora-Fuller, M. K., B. a. Schneider, E. MacDonald, H. E. Pass, and S. Brown (2007). “Temporal jitter disrupts speech intelligibility: A simulation of auditory aging”. In: *Hearing Research* 223.1-2, pp. 114–121.

Pichora-Fuller, M. K. et al. (2016). “Hearing Impairment and Cognitive Energy”. In: *Ear and Hearing* 37, 5S–27S.

Picou, E. M., S. C. Marcrum, and T. A. Ricketts (2015). “Evaluation of the effects of nonlinear frequency compression on speech recognition and sound quality for adults with mild to moderate hearing loss”. In: *International Journal of Audiology* 54.3, pp. 162–169.

Plomp, R. and A. M. Mimpen (1979). “Speech-reception threshold for sentences as a function of age and noise level.” In: *The Journal of the Acoustical Society of America* 66.5, pp. 1333–42.

Plomp, R. (1978). “Auditory handicap of hearing impairment and the limited benefit of hearing aids”. In: *The Journal of the Acoustical Society of America* 63.2, pp. 533–549.

Plomp, R. (1994). “Comments on “Evaluating a speech-reception threshold model for hearing- impaired listeners” [J. Acoust. Soc. Am. 93 , 2879–2885 (1993)]”. In: *The Journal of the Acoustical Society of America* 96.1, pp. 586–587.

Plomp, R. (2001). *The Intelligent Ear.*

Prosser, S, M Turrini, and E Arslan (1991). “Effects of different noises on speech discrimination by the elderly”. In: *Acta Oto-Laryngologica* 111.sup476, pp. 136–142.

Ragozini, G., F. Palumbo, and M. R. D’Esposito (2017). “Archetypal analysis for data-driven prototype identification”. In: *Statistical Analysis and Data Mining: The ASA Data Science Journal* 10.1, pp. 6–20.

Raufer, S., J. J. Guinan, and H. H. Nakajima (2019). “Cochlear partition anatomy and motion in humans differ from the classic view of mammals”. In: *Proceedings of the National Academy of Sciences* 116.28, pp. 13977–13982.

Reinhart, P. N. and P. E. Souza (2016). “Intelligibility and Clarity of Reverberant Speech: Effects of Wide Dynamic Range Compression Release Time and Working Memory”. In: *Journal of Speech, Language, and Hearing Research* 59.6, pp. 1543–1554.

Relaño-Iborra, H. (2019). “Predicting speech perception of normal-hearing and hearing-impaired listeners”. PhD thesis. Technical University of Denmark, p. 209.

Relaño-Iborra, H., J. Zaar, and T. Dau (2019). “A speech-based computational auditory signal processing and perception model”. In: *The Journal of the Acoustical Society of America* 146.5, pp. 3306–3317.

Remus, J. J. and L. M. Collins (2008). “Comparison of adaptive psychometric procedures motivated by the Theory of Optimal Experiments: Simulated and experimental results”. In: *The Journal of the Acoustical Society of America* 123.1, pp. 315–326.

Rhebergen, K. S., N. J. Versfeld, and W. a. Dreschler (2006). “Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise”. In: *The Journal of the Acoustical Society of America* 120.6, pp. 3988–3997.

Rieke, C. C. et al. (2017). “Fixed-Level Frequency Threshold Testing for Ototoxicity Monitoring”. In: *Ear and Hearing* 38.6, e369–e375.

Rix, A., J. Beerends, M. Hollier, and A. Hekstra (2001). “Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs”. In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* Vol. 2. IEEE, pp. 749–752.

Ronde-Brons, I. de, W. Soede, and W. Dreschler (2019). “Systematic Evaluation of Self-Reported Hearing Ability in Six Dimensions Before and After a Hearing Aid Trial”. In: *Journal of Speech, Language, and Hearing Research*, pp. 1–15.

Rönnberg, J. et al. (2016). “Hearing impairment, cognition and speech understanding: exploratory factor analyses of a comprehensive test battery for a group of hearing aid users, the n200 study”. In: *International Journal of Audiology* 55.11, pp. 623–642.

Rosen, S., P. Souza, C. Ekelund, and A. A. Majeed (2013). “Listening to speech in a background of other talkers: Effects of talker number and noise vocoding”. In: *The Journal of the Acoustical Society of America* 133.4, pp. 2431–2443.

Rosowski, J. J., S. Stenfelt, and D. Lilly (2013). “An Overview of Wideband Immittance Measurements Techniques and Terminology”. In: *Ear and Hearing* 34, 9s–16s.

Ross, B, K. L. Tremblay, and T. W. Picton (2007). “Physiological detection of interaural phase differences.” In: *The Journal of the Acoustical Society of America* 121.2, pp. 1017–1027.

Rudman, J. R., C. Mei, S. E. Bressler, S. H. Blanton, and X.-Z. Liu (2018). “Precision medicine in hearing loss”. In: *Journal of Genetics and Genomics* 45.2, pp. 99–109.

Sanchez-Lopez, R., F. Bianchi, M. Fereczkowski, S. Santurette, and T. Dau (2018a). “Data-Driven Approach for Auditory Profiling and Characterization of Individual Hearing Loss”. In: *Trends in Hearing* 22.1, p. 233121651880740.

Sanchez-Lopez, R., T. Dau, and W. M. Whitmer (2020a). “Audiometric profiles and patterns of benefit. A data-driven analysis of subjective hearing difficulties and handicaps”. In: *medRxiv* 20045690.

Sanchez-Lopez, R., M. Fereczkowski, T. Neher, S. Santurette, and T. Dau (2020b). “Robust data-driven auditory profiling for precision audiology”. In: *medRxiv* 20036442.

Sanchez-Lopez, R., M. Fereczkowski, S. Santurette, T. Dau, and T. Neher (2020c). “Auditory Profile-based Hearing-aid Fitting: A Proof-of-concept study”. In: *medRxiv* 20036459.

Sanchez-Lopez, R. et al. (2018b). “Technical evaluation of hearing-aid fitting parameters for different auditory profiles”. In: *Euronoise 2018*. Heraklion, Crete: 11th European Congress and Exposition on Noise Control Engineering, pp. 381–388.

Sanchez-Lopez, R. et al. (2019). *Data from "Auditory tests for characterizing hearing deficits: The BEAR test battery"*.

Sanchez-Lopez, R. et al. (2020d). “Auditory tests for characterizing hearing deficits: The BEAR test battery”. In: *medRxiv* 20021949.

Sanchez, R., F. Bianchi, M. Fereczkowski, S. Santurette, and T. Dau (2017). “Data-driven approach for auditory profiling”. In: *Proceedings of the International*

Symposium on Auditory and Audiological Research 6.0 SE - 2017/4. Assessment of specific auditory functions and hearing ability.

Sanders, N. C. and S. B. Chin (2009). “Phonological Distance Measures”. In: *Journal of Quantitative Linguistics* 16.1, pp. 96–114.

Santurette, S. and T. Dau (2012). “Relating binaural pitch perception to the individual listener’s auditory profile”. In: *The Journal of the Acoustical Society of America* 131.4, pp. 2968–2986.

Saunders, G. H., D. L. Field, and M. P. Haggard (1992). “A clinical test battery for obscure auditory dysfunction (OAD): Development, selection and use of tests”. In: *British Journal of Audiology* 26.1, pp. 33–42.

Schädler, M. R., A. Warzybok, and B. Kollmeier (2018). “Objective Prediction of Hearing Aid Benefit Across Listener Groups Using Machine Learning: Speech Recognition Performance With Binaural Noise-Reduction Algorithms”. In: *Trends in Hearing* 22, p. 233121651876895.

Schädler, M. R., D. Hülsmeier, A. Warzybok, and B. Kollmeier (2020). “Individual aided speech recognition performance and benefit predictions for listeners with impaired hearing”. In: *Trends in hearing* under revi.

Schmiedt, R. A., H. Lang, H. O. Okamura, and B. A. Schulte (2002). “Effects of furosemide applied chronically to the round window: A model of metabolic presbyacusis”. In: *Journal of Neuroscience* 22.21, pp. 9643–9650.

Schorn, K and E Zwicker (1990). “Frequency selectivity and temporal resolution in patients with various inner ear disorders.” In: *Audiology : official organ of the International Society of Audiology* 29.1, pp. 8–20.

Schuknecht, H. F. (1964). “Further Observations On the Pathology of Presbycusis”. In: *Archives of Otolaryngology - Head and Neck Surgery* 80.4, pp. 369–382.

Schuknecht, H. and M. R. Gacek (1993). “Cochlear Pathology in Presbycusis”. In: *Annals of Otology, Rhinology & Laryngology* 102.1_suppl, pp. 1–16.

Schütt, H. H., S. Harmeling, J. H. Macke, and F. A. Wichmann (2016). “Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data”. In: *Vision Research* 122, pp. 105–123.

SenseLab, dept. (2017). *SenseLabOnline (version 4.0.2). Listening test software.* Hørsholm, Denmark.

Serman, M., Fischer, R-L., Herbig, R., & Hannemann, R (2017). "Moving beyond the audiogram towards a patient-centered, interactive first fit". In: *AudiologyOnline*, p. 20493.

Shah, P. et al. (2019). "Artificial intelligence and machine learning in clinical development: a translational perspective". In: *npj Digital Medicine* 2.1, p. 69.

Sharma, M., S. C. Purdy, and P. Humburg (2019). "Cluster Analyses Reveals Subgroups of Children With Suspected Auditory Processing Disorders". In: *Frontiers in Psychology* 10.

Simeonsson, R (2000). "Revision of the International Classification of Impairments, Disabilities, and Handicaps Developmental issues". In: *Journal of Clinical Epidemiology* 53.2, pp. 113–124.

Sørensen, A. J., M. Fereczkowski, and E. N. MacDonald (2018). *Task dialog by native-Danish talkers in Danish and English in both quiet and noise.*

Souza, P. E., R. A. Wright, M. C. Blackburn, R. Tatman, and F. J. Gallun (2015). "Individual Sensitivity to Spectral and Temporal Cues in Listeners With Hearing Impairment". In: *Journal of Speech, Language, and Hearing Research* 58.2, pp. 520–534.

Souza, P., K. Arehart, T. Schoof, M. Anderson, D. Strori, and L. Balmert (2019). "Understanding Variability in Individual Response to Hearing Aid Signal Processing in Wearable Hearing Aids". In: *Ear and Hearing* 40.6, pp. 1280–1292.

Souza, P., F. Gallun, and R. Wright (2020). "Contributions to Speech-Cue Weighting in Older Adults With Impaired Hearing". In: *Journal of Speech, Language, and Hearing Research* 63.1, pp. 334–344.

Stebbins, W. C., J. E. Hawkins, L.-G. Johnsson, and D. B. Moody (1979). "Hearing thresholds with outer and inner hair cell loss". In: *American Journal of Otolaryngology* 1.1, pp. 15–27.

Stecker, G. C. and F. Gallun (2012). "Binaural Hearing, Sound Localization, and Spatial Hearing". In: *Translational Perspectives in Auditory Neuroscience: Normal Aspects of Hearing.* Ed. by K. Tremblay and R. F. Burkard. Plural Publishing Inc. Chap. 14, pp. 383–433.

Stratford, P. W. and C. H. Goldsmith (1997). "Use of the Standard Error as a Reliability Index of Interest : An Applied Example Using Elbow Flexor Strength Data". In: *Physical Therapy* 77.7, pp. 745–750.

Strelcyk, O. and T. Dau (2009). "Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing". In: *The Journal of the Acoustical Society of America* 125.5, p. 3328.

Strelcyk, O., P. Zahorik, J. Shehorn, C. Patro, and R. P. Derleth (2019). "Sensitivity to Interaural Phase in Older Hearing-Impaired Listeners Correlates With Nonauditory Trail Making Scores and With a Spatial Auditory Task of Unrelated Peripheral Origin". In: *Trends in Hearing* 23, p. 233121651986449.

Summers, V., M. J. Makashay, S. M. Theodoroff, and M. R. Leek (2013). "Suprathreshold Auditory Processing and Speech Perception in Noise: Hearing-Impaired and Normal-Hearing Listeners". In: *Journal of the American Academy of Audiology* 24.4, pp. 274–292.

Sweets, J. A. and R. M. Pickett (1982). *Evaluation of Diagnostic Systems.* Elsevier.

Taal, C. H., R. C. Hendriks, R. Heusdens, and J. Jensen (2010). "A short-time objective intelligibility measure for time-frequency weighted noisy speech". In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* IEEE, pp. 4214–4217.

Tecca, J. (2018). "Are post-fitting follow-up visits not hearing aid best practices?" In: *Hearing Review.* 25(4).4, pp. 12–22.

Theill, J. (2018). *Method of fitting a hearing aid system, a hearing aid fitting system and a computerized device. U.S. Patent No. 10,158,956.*

Thorup, N., S. Santurette, S. Jørgensen, E. Kjærbøl, T. Dau, and M. Friis (2016). "Auditory profiling and hearing-aid satisfaction in hearing-aid candidates." In: *Danish medical journal* 63.10, pp. 1–5.

Trevino, A. C., W. Jesteadt, and S. T. Neely (2016). "Development of a multi-category psychometric function to model categorical loudness measurements". In: *The Journal of the Acoustical Society of America* 140.4, pp. 2571–2583.

Trusheim, M. R., E. R. Berndt, and F. L. Douglas (2007). "Stratified medicine: strategic and economic implications of combining drugs and clinical biomarkers". In: *Nature Reviews Drug Discovery* 6.4, pp. 287–293.

Uus, K, J Bamford, and R Taylor (2006). “An analysis of the costs of implementing the National Newborn Hearing Screening Programme in England”. In: *Journal of Medical Screening* 13.1, pp. 14–19.

Vaden, K. I., L. J. Matthews, and J. R. Dubno (2018). “Transient-Evoked Otoacoustic Emissions Reflect Audiometric Patterns of Age-Related Hearing Loss”. In: *Trends in Hearing* 22, p. 233121651879784.

Van Esch, T. E. M. and W. A. Dreschler (2015). “Relations Between the Intelligibility of Speech in Noise and Psychophysical Measures of Hearing Measured in Four Languages Using the Auditory Profile Test Battery”. In: *Trends in Hearing* 19, p. 233121651561890.

Van Veen, B. D. and K. M. Buckley (2009). “Beamforming Techniques for Spatial Filtering”. In: *The Digital Signal Processing Handbook*. Vol. 5. 2, pp. 1 –22.

Vasilkov, V. and S. Verhulst (2019). “Towards a differential diagnosis of cochlear synaptopathy and outer-hair-cell deficits in mixed sensorineural hearing loss pathologies”. In: *medRxiv*, p. 19008680.

Verhulst, S., A. Altoè, and V. Vasilkov (2018). “Computational modeling of the human auditory periphery: Auditory-nerve responses, evoked potentials and hearing loss”. In: *Hearing Research* 360, pp. 55–75.

Vestergaard Knudsen, L., M. Öberg, C. Nielsen, G. Naylor, and S. E. Kramer (2010). “Factors Influencing Help Seeking, Hearing Aid Uptake, Hearing Aid Use and Satisfaction With Hearing Aids: A Review of the Literature”. In: *Trends in Amplification* 14.3, pp. 127–154.

Vickers, D., L. De Raeve, and J. Graham (2016). “International survey of cochlear implant candidacy”. In: *Cochlear Implants International* 17.sup1, pp. 36–41.

Vinay, A. S. Hansen, Ø. Raen, and B. C. J. Moore (2017). “Reference thresholds for the TEN(HL) test for people with normal hearing”. In: *International Journal of Audiology* 56.9, pp. 672–676.

Vlaming, M. S. M. G. et al. (2011). “HearCom: Hearing in the Communication Society”. In: *Acta Acustica united with Acustica* 97.2, pp. 175–192.

Völker, C., T. Bisitz, R. Huber, B. Kollmeier, and S. M. A. Ernst (2018). “Modifications of the MUlti stimulus test with Hidden Reference and Anchor (MUSHRA) for use in audiology”. In: *International Journal of Audiology* 57.sup3, S92–S104.

WHO (2018). *Global estimates on prevalence of hearing loss.* Tech. rep. World Health Organization, pp. 1–16.

Wagener, K., J. L. Josvassen, and R. Ardenkjær (2003). “Design, optimization and evaluation of a Danish sentence test in noise: Diseño, optimización y evaluación de la prueba Danesa de frases en ruido”. In: *International Journal of Audiology* 42.1, pp. 10–17.

Wang, J. and J.-L. Puel (2018). “Toward Cochlear Therapies”. In: *Physiological Reviews* 98.4, pp. 2477–2522.

Whitmer, W. M., P. Howell, and M. A. Akeroyd (2014). “Proposed norms for the Glasgow hearing-aid benefit profile (Ghabp) questionnaire”. In: *International Journal of Audiology* 53.5, pp. 345–351.

Whitmer, W. M., K. F. Wright-Whyte, J. A. Holman, and M. A. Akeroyd (2016). “Hearing Aid Validation”. In: ed. by G. R. Popelka, B. C. J. Moore, R. R. Fay, and A. N. Popper. Vol. 56. Springer Handbook of Auditory Research September 2016. Cham: Springer International Publishing, pp. 291–321.

Wiinberg, A., M. L. Jepsen, B. Epp, and T. Dau (2019). “Effects of Hearing Loss and Fast-Acting Compression on Amplitude Modulation Perception and Speech Intelligibility”. In: *Ear and Hearing* 40.1, pp. 45–54.

Wilson, B. S., D. L. Tucci, M. H. Merson, and G. M. O’Donoghue (2017). “Global hearing health care: new findings and perspectives”. In: *The Lancet* 390.10111, pp. 2503–2515.

Winkler, A., M. Latzel, and I. Holube (2016). “Open Versus Closed Hearing-Aid Fittings: A Literature Review of Both Fitting Approaches”. In: *Trends in Hearing* 20, p. 233121651663174.

Wolak, T., K. Cieśla, A. Pluta, E. Włodarczyk, B. Biswal, and H. Skarżyński (2019). “Altered Functional Connectivity in Patients With Sloping Sensorineural Hearing Loss”. In: *Frontiers in Human Neuroscience* 13.

Wolff, A., S. S. Houmøller, D. Hougaard, M. Gaihede, D. Hammershøi, and J. Schmidt (2020). “Health-related Quality of Life in Hearing Impaired Danish Adults before and after Hearing Aid Rehabilitation”. In: *International Journal of Audiology* (u. rev.)

Wong, A. C. Y. and A. F. Ryan (2015). “Mechanisms of sensorineural cell damage, death and survival in the cochlea”. In: *Frontiers in Aging Neuroscience* 7.

Wroblewska-Seniuk, K. E., P. Dabrowski, W. Szyfter, and J. Mazela (2017). “Universal newborn hearing screening: methods and results, obstacles, and benefits”. In: *Pediatric Research* 81.3, pp. 415–422.

Wu, P., L. Liberman, K. Bennett, V. de Gruttola, J. O’Malley, and M. Liberman (2019). “Primary Neural Degeneration in the Human Cochlea: Evidence for Hidden Hearing Loss in the Aging Ear”. In: *Neuroscience* 407, pp. 8–20.

Wu, P., W. Wen, J. T. O’Malley, and M. C. Liberman (2020). “Assessing fractional hair cell survival in archival human temporal bones”. In: *The Laryngoscope* 130.2, pp. 487–495.

Zaar, J., L. Simonsen, T. Bherens, T. Dau, and S. Laugesen (2019). “Investigating the relationship between spectro-temporal modulation detection, aided speech perception, and noise reduction preference”. In: *Proceedings of ISAAR 2019: Auditory Learning in Biological and Artificial Systems. Vol 7.* Nyborg (Denmark).

Zeng, F.-G., S. Oba, S. Garde, Y. Sininger, and A. Starr (1999). “Temporal and speech processing deficits in auditory neuropathy”. In: *NeuroReport* 10.16, pp. 3429–3435.

A

Technical evaluation of hearing-aid fitting parameters for different auditory profiles [a]

Abstract

Hearing-aid users have reported increased satisfaction since digital technology and advanced signal processing became available in hearing aids. However, many users still experience difficulties in noisy environments and complex listening scenarios. Although numerous parameters can be adjusted to provide an individualized hearing solution, hearing-aid fitting currently consists of: 1) the gain prescription and adjustment based on the pure-tone audiogram, 2) the activation of advanced features on-demand, such as beamforming and noise reduction. In a previous study (Sanchez-Lopez et al., 2018a, *Chapter 2*), a novel approach for auditory profiling was suggested, where the hearing deficits were characterized according to two types of distortion. This allowed the classification of listeners into four auditory profiles according to a high/low degree of hearing distortions along the two dimensions. The present study aimed to evaluate different hearing-aid compensation strategies that may fit the needs of different auditory

---

[a] This chapter is based on Sanchez-Lopez, R., Fereczkowski, M., Bianchi, F., Piechowiak, T., Hau, O., Pedersen, M.S., Behrens, T., Neher, T., Dau, T., Santurette, S. (2018). Technical evaluation of hearing-aid fitting parameters for different auditory profiles. In Euronoise 2018. (pp. 381–388). Heraklion, Crete: 11th European Congress and Exposition on Noise Control Engineering.

profiles via technical measures. A hearing-aid simulator, consisting of beamforming, noise reduction, and dynamic range compression, was used to test which parameter spaces and outcome measures may be of interest for a “profile-based hearing-aid fitting”. The simulator consists of two dummy behind-the-ear hearing aids and off-line sound processing performed on a personal computer. Technical measures, such as signal-to-noise ratio (SNR) improvement, envelope degradation, and a metric of spectral distortions, were used to evaluate the effects of different signal processing strategies on the signal at the output of the simulator. Several parameter settings were evaluated using speech in the presence of various interferers at different SNRs. Here, the results of this technical evaluation are presented and discussed, with a view towards identifying the effective compensation strategies for different auditory profiles.

A.1 Introduction

Satisfaction reported by hearing-aid users has increased significantly since digital technology became available (Kochkin, 2010). This can be attributed to the ability of modern hearing aids (HAs) to deliver non-linear amplification as well as advanced signal processing features, such as beamforming and noise reduction. However, many HA users still experience difficulties in understanding speech in noisy environments and other complex listening scenarios.

While numerous parameters can be adjusted to provide an individualized hearing solution, current hearing-aid fitting procedures are relatively simple. Usually, frequency and level-dependent gain is first determined based on the listener’s pure-tone sensitivity, i.e., the audiogram. Subsequently, advanced features, including algorithms like beamforming and noise reduction, may be activated depending on personal preferences. Importantly, the fitting procedure does not take supra-threshold performance, e.g., measures of the listener’s performance at moderate sound levels and in complex environments, into account. Therefore, listeners with similar audiograms receive similar fitting solutions. The individual listener's needs are addressed during fine-tuning, which depends solely on the audiologist's skills and experience. Given the nonlinear nature of many hearing-aid algorithms and their interactions, the design of individualized compensation strategies can be a complex task. This complexity is further increased by a broad range of sound scenarios encountered by individual HA users as well as inherent variability in a given individual's responses.

Evaluating a listener's supra-threshold performance requires tools beyond the pure-tone audiogram. The listener's performance may be estimated using a test battery and individual data can then be used to quantify the degree of perceptual distortions perceived by each listener. Recently, a data-driven approach to characterize individual listeners' hearing along two dimensions has been proposed (Sanchez-Lopez et al., 2018a), where each dimension represented an independent type of supra-threshold distortions. Each listener was assigned to one of the four possible auditory profiles defined by their degree of perceptual distortions in the two dimensions. It is reasonable to assume that the most efficient compensation of a given hearing loss depends on the type of auditory distortions present, such as the ability to perceive the temporal and spectral features of sounds. Hence, a "profile-based" HA fitting would ideally activate algorithms that can compensate for the specific types of distortions present in each listener. To approach this ideal scenario, a technical characterization of how modern HA features can affect specific distortions in the physical signal should be obtained first. Such a characterization was the aim of the present study to help define feature combinations that are adapted to different auditory profiles.

A profile-based HA parameter space may require different directionality, noise reduction, and compression settings. Although the two first types of strategies aim for signal-to-noise ratio (SNR) improvement, directionality applies spatial filtering that keeps the signal in front unaltered while noise reduction applies spectral filtering on the noisy mixture. The effects of noise reduction and directionality on speech-in-noise perception have been a topic of interest in previous studies

(Brons et al., 2014; Neher and Wagener, 2016). Furthermore, the influence of the parameters used in dynamic range compression has been broadly studied (Davies-Venn et al., 2009; Jenstad and Souza, 2005). The characteristics of these processing algorithms in isolation have also been assessed through technical measures, such as speech intelligibility prediction or physical measures of the acoustic signal (Baumgärtel et al., 2015; Hu and Loizou, 2008), which do not require the participation of a listener. The present study is inspired by the approaches used in these previous studies and focuses on characterizing the effects of the HA algorithms on established metrics reflecting distortions in the physical signal.

In the literature, the SNR improvement and other physical measures at the output of real hearing aids have been explored in connection to speech-in-noise perception (Miller et al., 2017) as well as perceived quality measurements (Geetha and Manjula, 2014). In this context, speech intelligibility prediction models and speech quality models are commonly used to quantify the expected performance of specific algorithms (Baumgärtel et al., 2015; Geetha and Manjula, 2014). While these objective measures may correlate with the observed perceptual performance of normal-hearing listeners, there is no guarantee that hearing-impaired listeners would exhibit the same behavior. Therefore, in the present study, such model-based objective performance measures were complemented with technical metrics related to SNR, spectral, and temporal signal distortions. The idea was to characterize how the combination of parameters in HA algorithms affects such metrics rather than predicting HA user performance.

For this purpose, a hearing-aid simulator (HASIM) was designed and evaluated with a set of five objective metrics. The chosen physical measures were the segmental SNR and objective measures of temporal-envelope and spectral distortions. The objective speech intelligibility and quality measures used here were the short-time objective intelligibility (STOI; Taal et al., 2010) and the perceptual evaluation of speech quality (PESQ; Rix et al., 2001). The main goal was to characterize the performance of each algorithm in isolation as well as their interaction in several sound scenarios. Additionally, it was of interest to identify the combinations of parameters that lead to the best/worst performance in terms of the five chosen metrics.

A.2 Hearing-aid simulator (HASIM)

The HASIM was implemented in MATLAB via the combination of three processing algorithms. As shown in Figure A.1, the signal recorded from the frontal and rear microphones of a hearing aid was processed by a beamformer, a noise reduction algorithm and a wide-dynamic range compressor.

Beamformer

The Beamformer (BF) provides an omnidirectional sum of both microphones and two polar patterns, a fixed unilateral BF and a binaural BF. To obtain the optimized beam-patterns for the two BF types, a head and torso simulator (HATS) was placed in the center of an anechoic room facing a speaker at 0 degrees (distance 1.5 m). The impulse responses from the speaker to each of the four microphones were measured with a 5-s maximum length sequence (MLS) with a code length of 11 bit at a sound pressure level (SPL) of about 65 dB. This was repeated for loudspeakers situated in the horizontal plane for angles from 0 to 360 degrees with a resolution of 5 degrees. After the impulse responses were obtained, a linear filter was built for each microphone (front, rear) and optimized in a least-square sense to a predefined beampattern (Van Veen and Buckley, 2009). Optimization was performed only in the frequency region between 1 and 5 kHz. Below 1 kHz, the front microphone signal alone was used as the output, and above 1 kHz unilateral beamforming was applied. For the binaural BF, the four outputs of the left (L) and right (R) ear devices were processed similarly. This resulted in a diotic signal. However, the use of a diotic signal removes spatial cues that are important for localization and spatial separation in real environments. Therefore, to improve the acceptance of the binaural beamformer, a portion of the signal from the front microphone was added to each device. In this case, 85% of the processed signal

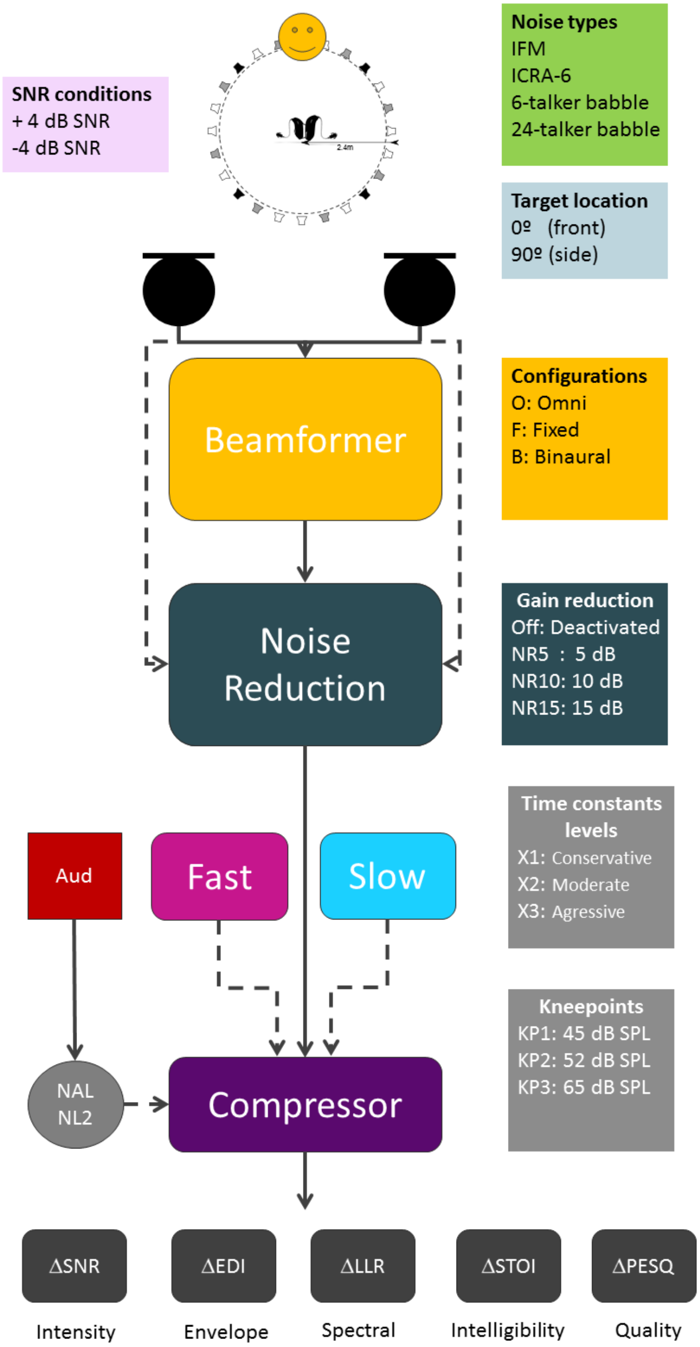


Figure A.1: Diagram of the hearing-aid simulator (HASIM) including the sound scenarios tested and the objective measures considered in the study. Panels placed on the right briefly explain the different levels used for each stage.

and 15% of the front microphone were considered in the simulations.

Noise Reduction

The noise reduction system (NR) was based on the spatial properties of two closely-spaced microphones and the assumption that the sounds of interest would be primarily located in front of the listener. From the two microphone signals, two first-order differential arrays (cardioids), pointing in opposite directions (towards the front and back), were created as described in (Boldt et al., 2008). Hereby, the front-facing cardioid primarily captured sounds in front of the listener (sound of interest) while the rear-facing cardioid primarily captured sounds behind the listener (noise). By comparing the power spectral density estimates of the two cardioids in each time-frequency frame, a binary mask was created which determines if a given time-frequency tile mainly contains energy from the front or the back of the listener. The time-frequency mask was converted into a binary gain, which attenuates time-frequency tiles with more energy in the rear-facing cardioid compared to the front-facing cardioid with a fixed amount of attenuation (Boldt et al., 2008).

Dynamic range compressor

The wide dynamic range compressor (WDRC) consisted of a 15-band filterbank (0.1-10 kHz), a percentile estimator, and an amplifier with non-linear gain. The bandwidth of the filterbank was approximately one-third octaves for the eleven mid-frequency bands and half an octave for the four remaining upper and lower bands. The envelope of the individual bands was estimated based on the low-pass filtered squared signal. The envelope was then transformed to the logarithmic domain and passed through the percentile estimator that effectively controlled the time constants of the compression system. The output of the percentile was increased with a fixed-rate attack time if the envelope was greater than the output. Similarly, the output of the percentile was decreased with a fixed-rate release time if the envelope was smaller than the output. The percentile estimator calculated the desired gain in the compressor input-gain function and was set for each of the compressor conditions. The amplifier's gain function was a broken-stick nonlinearity with a single kneepoint used to set the insertion gain for conversational speech level (65 dB SPL). The upper and lower slopes of the function were calculated to match the target gains for soft (50 dB SPL) and loud (80 dB SPL) speech targets. The calculated gain was applied to the individual frequency bands based on the prescription rule corresponding to the individual pure-tone audiometric thresholds. The compressed output signal was formed by the sum of all bands (Kates, 2005).

A.3 Method

Sound scenarios

The sound scenarios used in the technical evaluation were recorded in an anechoic chamber with 24 loudspeakers placed in the horizontal plane, in steps of 15°, around a chair located in the middle of the chamber. A HATS was placed on the chair while wearing HA satellites consisting of a HA housing with a front and a rear microphone. The international speech test signal (ISTS; Holube et al., 2010) was used as the target signal, which was recorded when played from the loudspeakers located at 0° and 90° degrees at 65 dB SPL. Two noises were used; the international female noise (IFN), a stationary noise with the same long-term average spectra (LTAS) as the ISTS (Holube et al., 2010), and (ICRA-6; Dreschler et al., 2001), a fluctuating noise composed of the envelope of six talkers and the fine structure of a random noise. The two noise maskers were recorded from the two loudspeakers located at ±45. Additionally, two multi-talker noise environments were constructed using recordings of real conversations (Sørensen et al., 2018). A 6-talker babble was recorded from loudspeakers located at ±15°, ±30°, and ±45°. A 24-talker babble was recorded by playing the speech of one independent talker from each of the 24 loudspeakers.

The sound scenes were prepared by combining the signal from each of the microphones of the target signal and each of the sound environments. The conditions considered for each of the noise environments were:

1. Target at 0° and +4 dB SNR.
2. Target at 90° and +4 dB SNR[b]
3. Target at 0° and -4 dB SNR.
4. Target at 90° and -4 dB SNR[b].

Besides, each of the sound scenes was constructed either with the target in phase ($S0N0$) or in antiphase ($S\pi S0$). This was done to enable the extraction of the target and the noise signals in each stage of the HASIM using the Hagerman-Olofsson separation technique (Hagerman and Olofsson, 2004).

Hearing-aid parameter spaces

Each of the three HASIM stages was tested in a number of conditions. The BF was tested in three modes: Omni (O), Fixed (F), and Binaural (B). The NR algorithm was tested with attenuations of 5 (NR5), 10 (NR10), and 15 dB (NR15), as well as when the algorithm was deactivated (Off). The parameters of the WDRC adjusted in the simulations were the kneepoint (KP) and the time constants (TC). The KP was set at either 45, 52, or 65 dB SPL. The TC were divided into 'fast' and 'slow' options and tested with three levels in each case:

1. Fast1: Attack = 15 ms; Release = 50 ms.
2. Fast2: Attack = 10 ms; Release = 10 ms.
3. Fast3: Attack = 5 ms; Release = 10 ms.
4. Slow1: Attack = 40 ms; Release = 400ms.

---

[b] The SNR is referred to as the tested device (left) only.

5. Slow2: Attack = 100 ms; Release = 800ms.
6. Slow3: Attack = 250 ms; Release = 1250ms.

The compression ratio was determined by applying the NAL-NL2 (Keidser et al., 2011) prescription rule to different audiometric profiles based on the proposed standard audiograms (Bisgaard et al., 2010). The audiometric thresholds of the audiograms N1, N2, N3, N4, S1, S2, and S3 were entered into the NAL-NL2 software and the target gains at 50, 65, and 80 dB SPL were transferred to the compressor algorithm. Additionally, the 0-dB linear-gain condition was tested in order to explore the processing algorithms (BF and NR) in isolation. In total, 216 different parameter combinations (3 BF x 4 NR x 3 KP x 6 TC) were tested per audiometric profile.

Procedure

The simulations were carried out in the same way for each of the sound scenarios and set of HA parameters. Once the sound scenario at the input of the frontal and rear microphones was constructed, the resulting signals were used as the input to the BF. As mentioned above, this was done for both the S_0N_0 and $S_\pi N_0$ versions of each sound scenario. After the BF stage, the resulting signal as well as the original signals from the frontal and rear microphones were input to the NR algorithm. The last stage was the WDRC which was fed with the signal obtained at the output of the NR. Once this was done, the reference signal for the evaluation, corresponding to the omni-directional and linear condition (OmLin), was obtained by performing a simulation in which the prescribed gain per frequency band corresponded to the long-term spectrum of the output signal. This was done in order to 1) minimize the effect of the spectral shape of the output signal for each audiogram and more clearly observe the effects of the WDRC parameters, and 2) reduce the effect of the input SNR. Moreover, using OmLin as a reference yielded a reference output signal that had been processed by the whole HASIM but was not influenced by the distortions and enhancements created by each algorithm.

Objective measures

The technical evaluation involved three physical measures of the acoustic signal. These were the segmental signal-to-noise ratio (segSNR) at the output of the HASIM, the log-likelihood ratio (LLR) between the unprocessed and processed signals, and the envelope distortion index (EDI) between the unprocessed target and the isolated target at the output of HASIM as defined in Jenstad and Souza (2005) and Hu and Loizou (2008). Additionally, two performance measures were considered: STOI and PESQ. In both performance measures, the reference signal was the clean target from the OmLin condition and the test signal was the noisy speech at the output of HASIM.

A.4 Results & Discussion

The simulations were first carried out for the processing algorithms (BF & NR) and the fitting algorithm (WDRC) in isolation. A multi-way ANOVA for all the sound scenarios showed a significant effect of NR [F(3,191)=3.23, p=0.02] and BF [F(2,191)=9.73, p<0.01] on the segSNR but not their interaction, which was only significant on the LLR [F(6,191)=3.29, p<0.01]. In contrast, only BF had a significant influence on EDI [F(2,191)=171.8, p<0.001]. When comparing the different sound scenarios, NR had no effect when the noise was located in front, due to the inefficiency of the SNR estimation algorithm in such a setting. In contrast, for the 24-talker babble, BF had a significant effect on the three physical measures and NR affected segSNR and LLR significantly. In the following, only the results for the 24-talker babble scenario in its four conditions are reported and discussed.

Figure A.2 shows the changes in segSNR, EDI, and LLR scores, relative to the OmLin condition. The left panel shows the performance of BF and NR for different SNR and target location conditions. While the segSNR scores increased when BF and NR were activated and the target was located at 0^{o}, the scores of the binaural BF (B) were 2.5 dB lower when the target was located at 90°. This was also

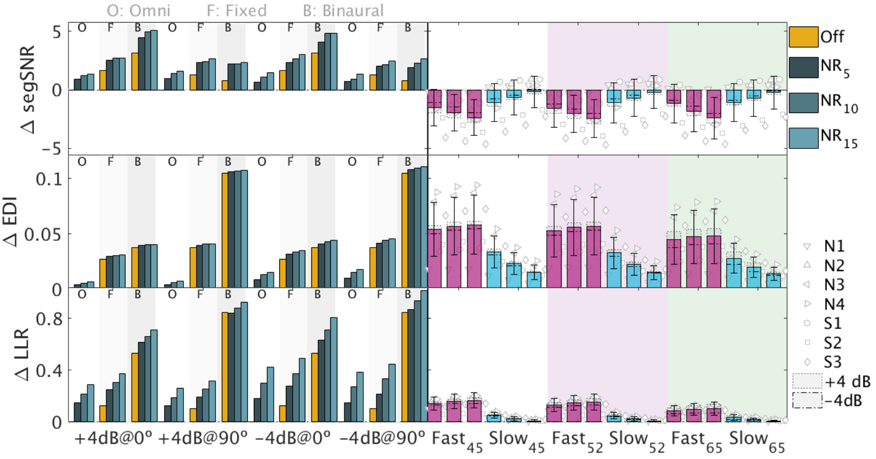


Figure A.2: SNR improvement (ΔsegSNR), envelope (ΔEDI), and spectral distortions (ΔLLR), of the the HASIM algorithms in isolation. Left panel shows the beamformer (BF) and noise reduction (NR) for the different sound scenarios and Right panel the fitting algorithm (WDRC) for the different combinations of parameter kneepoint (KP=45,52 or 65 dB SPL) and time constants (TC=Fast or Slow in their three levels) for positive and negative SNRs as well as the mean of each audiometry.

observed in the EDI scores, which increased dramatically when the binaural BF was activated and the target was located at 90°. Furthermore, the LLR increased when the algorithms were more aggressive, regardless of the target location. Moreover, the condition B and NR15 yielded the largest change in segSNR (5 dB) but also the highest amount of spectral distortion (>0.8).

The right panel of Figure A.2 shows the mean results for the different WDRC conditions (colored bars), the average results for positive and negative SNRs (shadowed bars), as well as the mean of each audiometric configuration (markers). In contrast to the results shown in the left panel, WDRC reduced the segSNR, particularly for the fast-acting compression settings. While the influence of KP on the EDI scores was small but significant [F(2,495)=18.9 p<0.001], there were large differences between the slow and fast-acting configurations [F(1,495)=1033.8, p<0.001]. The effective compression became more linear with increasing time constants (slow-acting) showing a reduced amount of distortions and a smaller SNR reduction. Therefore, the selection of fast-acting compression may counteract the SNR enhancement provided by the processing stages and can introduce additional distortions in the temporal envelope (i.e. higher EDI scores). When comparing the results for the individual audiometric configurations, the audiograms with a higher degrees of hearing loss, particularly at low frequencies (N3 and N4), led to even larger envelope distortions. On the other hand, the spectral distortions (LLR) introduced by the WDRC were much lower than the ones introduced by the processing algorithms (BF & NR).

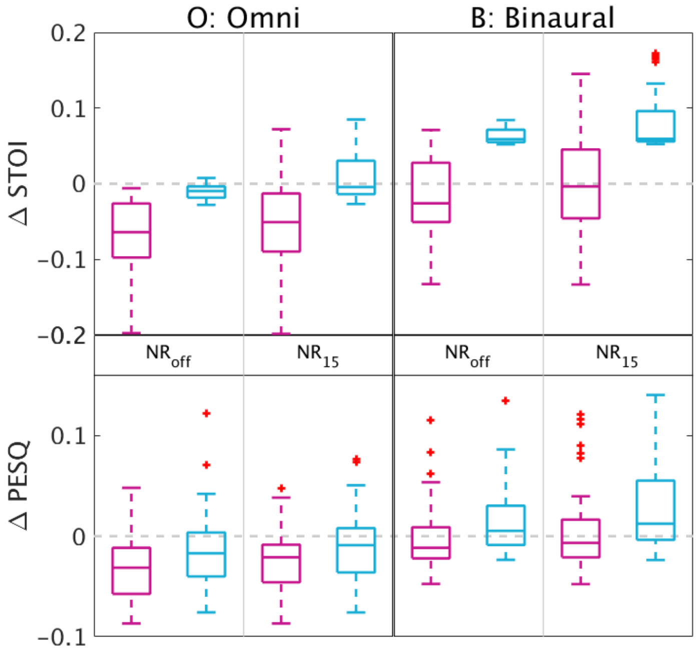


Figure A.3: Objective performance measures (STOI and PESQ) of the HASIM with the target direction at 0°. Each result corresponds to a combination of the three algorithms. The magenta boxplots show results of the fast-acting compressor and cyan the slow-acting compressor.

Figure A.3 illustrates the differences in performance due to the combined effects of the BF, NR, and WDRC. Only the condition with the target in front was considered here, and the boxplots show the scores relative to the OmLin condition for all noise types and audiometric configurations. The results suggested a clear improvement of the STOI scores when BF was binaural and an additional improvement when NR was activated with the highest attenuation (NR15). One should note that the variance of the results of the fast-acting compression was higher than for the slow-acting HA configuration. This is mainly due to the slow-acting compression linearizing the long-term response and acting as a gain reduction that does not affect the spectro-temporal features of the signal. However, fast-acting compression has different effects depending on the compression ratio applied, which depends on the audiometric thresholds. In contrast, the results for the PESQ metric did not show significant differences neither in terms of the mean values nor the variance.

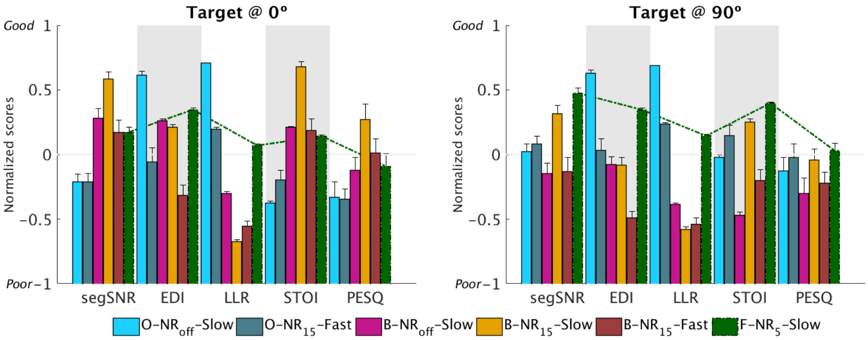


Figure A.4: Normalized scores in the five chosen objective measures for six HA parameter settings. Results are divided in low (mild) and high (sev) degree of hearing loss and by target direction. The results were normalized between the 10th and 90th percentiles. The normalized scores of EDI and LLE were multiplied by a factor (-1) so -1 always corresponds to a poor performance and 1 to a good performance.

To test different profile-based compensation strategies, it is of interest to explore HA parameter spaces that differ widely from one another, not only in terms of performance, but also in terms of spectral and temporal distortions. Therefore, six HA parameter settings were chosen for that purpose. Figure A.4 shows the average results across noise types, SNR conditions, and hearing profiles for these six settings. The normalized results of the five objective measures are shown for the

0°(left panel) and 90°(right panel) target condition. As expected, the HA setting with no processing activated and slow-acting compression (O-NRoff-Slow) provided good scores for the distortion measures (i.e., EDI, and LLR), but slightly negative scores for the segSNR, STOI, and PESQ metrics. In contrast to the unprocessed HA setting, a HA setting with all the algorithms activated at their most aggressive level (B-NR15-fast) showed clear spectral and temporal distortions. In addition, B-NR15-fast showed an improvement in SNR and STOI when the target was located in front but poorer scores when it was located at 90°. For the fourth HA setting (B-NR15-slow), this improvement was even higher and exceeded the unprocessed HA setting in all cases. The HA setting with moderate processing parameters (F-NR5) and slow-acting compression showed positive scores for both target directions, suggesting an improvement in speech intelligibility compared to most of the other HA settings considered here.

A.5 Conclusions

Several HA parameter spaces were characterized by using objective physical measures at the output of a HA simulator. While the processing algorithms (BF and NR) tended to enhance the SNR and introduce spectral distortions, fast-acting compression had a detrimental effect on SNR improvement and temporal distortion. Parameter spaces towards a profile-based HA fitting were proposed by choosing combinations of parameters that provided different results in terms of SNR benefit, physical distortions and performance predictors. Overall, a perceptual evaluation using these identified parameters spaces should provide meaningful differences among the different HA settings and may help in the implementation of a profile-based compensation of the hearing deficits.

A.6 Acknowledgements

We thank M. L. Jepsen, E. Schmidt, N. Bisgaard and the rest of BEAR partners for their input during the realization of this study. This work was supported by Innovation Fund Denmark Grand Solutions 5164-00011B (Better hEAring Rehabilitation project).

B

Perceptual evaluation of six hearing-aid processing strategies from the perspective of auditory profiling: Insights from the BEAR project [a]

Abstract

The current study forms part of the Better hEAring Rehabilitation (BEAR) project, which aims at developing new clinical tools for characterizing individual hearing loss and for assessing hearing-aid (HA) benefit. Its purpose was to investigate potential interactions between four auditory profiles and three measures of HA outcome obtained for six HA processing strategies. Measurements were carried out in a realistic noise environment at signal-to-noise ratios that were set based on individual aided speech reception thresholds (SRT_{50}). Speech recognition scores and ratings of overall quality and noise annoyance were collected in two spatial conditions. The stimuli were generated with the help of a HA simulator and presented via headphones to 60 older, habitual HA users who had previously been profiled based on a data-driven approach (Sanchez-Lopez et al., 2020b, *Chapter 4*). The four auditory profiles differed significantly in terms of mean aided

---

[a] This chapter is based on Wu, M., Sanchez-Lopez, R., El-Haj-Ali, M., Fereczkowski, M., Dau, T., Santurette, S., Neher, T. (2020). Perceptual evaluation of six hearing-aid processing strategies from the perspective of auditory profiling: Insights from the BEAR project. In Proceedings of ISAAR 2019: Auditory Learning in Biological and Artificial Systems.

SRT_{50} and interacted significantly with the HA processing strategies for speech recognition in one spatial condition. Moreover, the correlation-pattern between the speech recognition scores and subjective ratings differed among the auditory profiles.

B.1 Introduction

Hearing-aid (HA) benefit in noisy environments is known to vary substantially among users, and several researchers have investigated ways to improve individual HA outcome (e.g., Lopez-Poveda et al., 2017; Neher and Wagener, 2016; Souza et al., 2019). Additionally, modern HA technology offers various features to improve speech intelligibility, e.g. directional microphones (Keidser et al., 2011), noise reduction (Brons et al., 2014), and dynamic range compression (Picou et al., 2015). Despite these efforts, clinical HA fittings are still mainly based on the audiogram, even though pure-tone hearing thresholds are unable to capture all the supra-threshold deficits induced by a hearing loss (Johannesen et al., 2016; Plomp, 1978). Moreover, the advanced features are not utilized in a systematic way. The Better hEAring Rehabilitation (BEAR) project aims at developing new clinical tools for individual hearing loss characterization and HA benefit assessment. For that purpose, an auditory test battery and a data-driven approach for classifying listeners into four distinct auditory profiles were proposed in an earlier study (Sanchez-Lopez et al., 2020b). In that study, 75 participants from four auditory profiles differed in terms of their performance on various auditory measurements as shown in Table B.1. In the present study, 60 of the subjects tested by Sanchez-Lopez et al. (2020b) participated and evaluated six processing strategies for HA treatment in three perceptual tasks. The main purpose of the current study was to evaluate the perceptual HA outcomes of these six HA processing strategies in relation to the four auditory profiles. Furthermore, correlations between aided speech-in-noise intelligibility and the subjective ratings of overall quality and noise annoyance were analysed. Since a better speech recognition score with a given HA setting does not necessarily correspond to high preference for that HA setting (Cox et al., 2016), we hypothesized that the four auditory profiles may help explain this inconsistency.

Auditory Profile	Audibility		Binaural processing	Loudness	Speech perception	Spectro-temporal
A (n=14)	☺	😐	☺	☺	☺	☺
B (n=13)	☺	☹	😐	😐	☹	☹
C (n=20)	☹	☹	☹	☹	☹	☹
D (n=8)	☹	😐	☺	☹	☺	😐

Table B.1: . Overall relative performance on the main measures from the BEAR auditory test battery. LF = low frequencies, HF = high frequencies. ☺: better performance, ☹: poorer performance and neutal smile: average performance.

B.2 Methods

The perceptual evaluation was carried out in a simulated speech-in-noise environment and consisted of a speech recognition task and a subjective rating task. To achieve high face validity, the testing conditions were chosen to reflect the difficulties that older HA users often encounter in complex noisy scenarios (Neher et al., 2011; Prosser et al., 1991).

Participants

Sixty subjects aged 60-80 years (mean = 70.8 years) were recruited for the study. Twenty-nine of them were tested at Odense University Hospital, Odense, while the other ones were tested at Bispebjerg Hospital, Copenhagen. All participants had bilateral symmetrical sensorineural hearing loss and were experienced HA users.

The range of hearing loss configurations was chosen to lie in-between the N1 and N4 standard audiograms (Bisgaard et al., 2010). Prior to this study, all participants completed a comprehensive auditory test battery developed by Sanchez-Lopez et al. (2020d). Based on these measurements, the participants were classified into one of the four auditory profiles using a data-driven approach (Sanchez-Lopez et al., 2020b). Five of the participants tested here could not be reliably allocated to any of these profiles and were thus not included in the data analysis described here. The distribution of the remaining 55 participants was as shown in the first column of Table B.1.

Test setup

The measurements were performed either in an anechoic chamber or a soundproof booth. Audio playback was via an RME Fireface UC soundcard, an SPL Phonitor Mini amplifier and a pair of Sennheiser HDA200 headphones. All stimuli were generated with the help of a hearing-aid simulator (HASIM) implemented in Matlab (Sanchez-Lopez et al., 2018b).

Stimuli

The target speech stimuli were DANTALE-II sentences spoken by a female native Danish speaker (Wagener et al., 2003). The target speech was presented from either 0º (front) or 90º (the side of the 'better' ear according to previously conducted unaided speech-in-noise measurements). The background noise was a spatially diffuse cafeteria noise recorded in a university canteen with a pair of HA satellites. In addition, the International Speech Test Signal (Holube et al., 2010) was used as a directional distractor from either 90º (target speech from 0º) or 0º (target speech from 90º). The directional distractor was presented at a signal-to-noise ratio (SNR) of +2 dB relative to the diffuse cafeteria noise.

Hearing-aid simulator (HASIM)

The HASIM included directional processing (omnidirectional, fixed cardioid or fixed binaural beamformer setting), noise reduction (maximal attenuation of 0, 5 or 15 dB) and amplitude compression (attack times of 5 or 250 ms and release times of 10 or 1250 ms for ‘fast’ and ‘slow’, respectively). For each listener, gains were set according to the NAL-NL2 fitting rule (Keidser et al., 2011). Four HA processing strategies (Table B.2) were selected to maximize differences in the sound processing. HA1 corresponded to very basic processing and served as a reference. HA6 resembled typical ‘commercial’ HA processing. For further details about the HASIM, see Sanchez-Lopez et al. (2018b).

	Directional processing	Noise reduction	Amplitude compression
HA1	Omnidirectional	Off	Slow
HA2	Omnidirectional	Strong	Fast
HA3	Binaural beamformer	Off	Slow
HA4	Binaural beamformer	Strong	Slow
HA5	Binaural beamformer	Strong	Fast
HA6	Cardioid	Mild	Slow

Table B.2: Description of the six tested HA processing strategies

Procedure

Each participant completed two visits. At the first visit, aided speech reception thresholds (SRT_{50}) were measured in an adaptive procedure (1-down 1-up procedure with a step size of 4 dB for the first five trials and 2 dB afterwards) to establish a baseline performance level for each participant. For the aided SRT_{50} measurements, the baselines of the stimuli were amplified according to individual gains (NAL-NL2 prescription for an input level of 65 dB SPL) and the target was amplified linearly during measurements. Aided SRT_{50} was only tested in the 0°condition. The six HA processing strategies were then evaluated for both spatial conditions using a speech recognition task at a fixed SNR that corresponded to the individual aided SRT_{50}. The speech recognition measurements were repeated at the second visit. The subjective assessment included ratings of overall quality and noise annoyance for the six HA in two spatial conditions. A multi-stimulus comparison method with a hidden anchor ('MUSHA') was implemented in the SenseLabOnline 4.0.2 software (SenseLab, 2017). The anchor stimulus used for the subjective ratings was a speech-in-noise stimulus that had been heavily distorted using random binary mask processing to approximate undesired spectral distortion of the tested noise reduction scheme. On a given trial, participants were presented with a graphical user interface containing seven playback buttons and sliders (6 HA settings + 1 anchor stimulus). Each stimulus was rated four times per spatial condition. The test SNR used for the subjective ratings corresponded to SRT_{50} + 4dB.

B.3 Results

Effect of auditory profile on SRT_{50}

On average, profile A had the lowest SRT_{50} (mean = 0.5 dB SNR, SD = 1.2 dB SNR) while profile C had the highest (mean = 5.1 dB SNR, SD = 3.6 dB SNR). According to a series of independent t-tests, profile B (mean = 2.7 dB SNR, SD = 2.3 dB SNR) and profile C differed significantly from profile A and profile D (mean = 0.6 dB SNR, SD = 1.2 dB SNR), respectively (all $p < 0.01$).

Effects of auditory profile on HA outcomes

For both speech recognition (Figure B.1) and the subjective ratings, listeners from the four auditory profiles showed similar patterns of benefit from the six HA processing strategies. More specifically, all auditory profiles gained larger benefits from the same or similar HA processing strategies for each outcome measure. To assess the effect of auditory profile on the different HA outcomes, linear mixed effects models were implemented. The dependent variable was the individual standardized score. For speech recognition, due to the data being split based on spatial condition, the model included four components (HA, auditory profile (AP), HA*test SNR, HA*AP). The random effect was the individual intercept. For the subjective ratings, the model included nine parts (HA, spatial condition (spa), AP, HA*spa, HA*AP, AP*spa, HA*test SNR, spa*test SNR, HA*spa*AP). For all three outcomes, a significant effect of HA was found (all $p < 0.001$). For the subjective ratings, the effects of spa and HA*spa were also significant (all $p < 0.001$). Furthermore, for speech recognition assessed in the 90º spatial condition there was a significant interaction between AP and HA ($F_{9,201} = 4.3$, $p < 0.001$), which was driven by low-benefit HA strategies (HA2 and HA3, see Fig. B.1). Overall, there were no significant main effects of auditory profile or significant interaction with auditory profile (all $p > 0.05$).

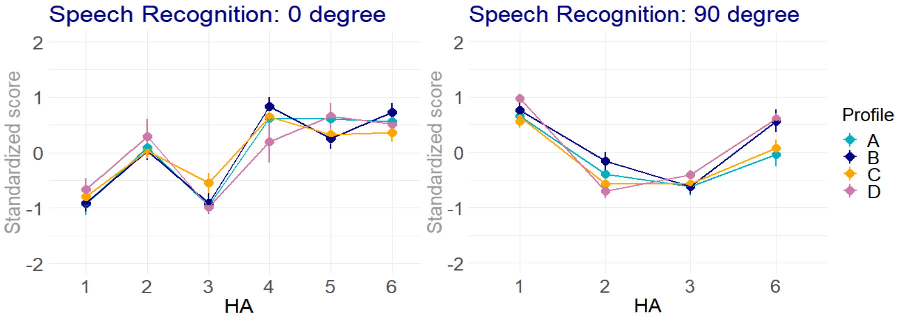


Figure B.1: Mean standardized speech recognition scores and standard errors for each test condition and auditory profile. Scores were averaged across test and retest. HA4 and HA5 were excluded in the 90° condition because of strong flooring effects.

Correlation analysis

Spearman's correlation coefficients were calculated to investigate potential relations between the three outcome measures across the four auditory profiles (Table B.3). In general, more correlations were found for the 90° spatial condition than for the 0° spatial condition. In particular, the overall quality ratings were positively correlated with the speech scores for all auditory profiles in the 90° (but not the 0°)

condition. Some differences among the four profiles were observed. Participants from profiles B showed relatively large, positive correlations between sentence recognition scores and both types of subjective ratings, while for profile A, which had a near-normal SRT_{50}, the different outcomes were not significantly correlated in most cases.

		OVERALL & SPEECH		NOISE & SPEECH	
Profile		0°	90°	0°	90°
A	*r*	0.07	**0.40**	0.02	0.17
	p	0.52	<0.01	0.88	0.22
B	*r*	**0.29**	**0.60**	**0.34**	**0.29**
	p	0.02	<0.001	<0.01	0.04
C	*r*	-0.01	**0.61**	**0.36**	**0.25**
	p	0.96	<0.001	<0.001	0.04
D	*r*	0.08	**0.71**	0.04	**0.57**
	p	0.60	<0.001	0.81	<0.01

Table B.3: Results of correlation analyses performed on the speech scores and subjective ratings for each auditory profile. OVERALL = overall quality, SPEECH = speech recognition, NOISE = noise annoyance.

B.4 Discussion

In the current study, speech recognition measurements and subjective ratings were applied to investigate potential links between four auditory profiles and response to six different HA processing strategies in a simulated speech-in-noise environment. Differences in aided SRT_{50} between four auditory profiles indicate different needs in terms of SNR improvement in HA processing. However, the four profiles barely differed in terms of their responses to the six tested HA processing strategies. One possible explanation could be that the participants were equated in terms of baseline performance level, which was based on their aided SRT_{50}. In other words, both the HASIM and the participants were exposed to different input signals. Another potential explanation for the lack of differences among the four profiles could be that the acoustic scene contained only one type of noise. It is possible that the use of a multi-talker scenario or more fluctuating noises would elicit more pronounced differences among the profiles in terms of their ability to utilize spatial and temporal cues in such scenarios. Moreover, in the present study, a limited set of HA settings were considered, with gains being prescribed according to the NAL-NL2 rule in all conditions. Previous research suggested that individuals with sloping audiograms obtain larger benefits from different HA amplification than individuals with flat audiograms (Keidser and Grant, 2001). Thus, it is possible that individuals from four auditory profiles obtain high HA benefit from different amplification rationales. Whether there is a three-way interaction between HA setting, amplification rationale and auditory profile in terms of perceptual HA outcome requires further study in the future.

The correlation analyses revealed that the four auditory profiles differed in terms of the extent to which speech recognition is related to overall quality and noise annoyance. For profile B, there were consistent positive correlations between the two types of measurements. This result might indicate that for profile B listeners HA preference is governed by the clarity or naturalness of the target speech. However, for profiles A and D, this was only the case in the 90° condition. Considering that these two groups were tested at lower SNRs, it is reasonable to think that the HA processing strategies rendered the speech more unclear or distorted in this condition.

It is well established that HA benefit in complex speech-in-noise environments depends on both auditory and non-auditory factors (Gatehouse and Akeroyd, 2006). Our study suggests that preference for HA processing can be broken down into different types of psychoacoustic function. Whether those auditory factors are indeed linked to a general preference for speech naturalness requires further research. More generally, the question of whether the auditory profiles tested here influence HA outcome still needs further investigation. Ideally, this work should use real HAs, various background noises and aided outcome measures, and should also provide the participants with the possibility to acclimatize to the tested HA settings.

Acknowledgements

This work was supported by Innovation Fund Denmark Grand Solutions 5164-00011B (BEAR project), Oticon, GN Resound, Widex and other partners (Aalborg University, University of Southern Denmark, the Technical University of Denmark, Force, Aalborg, Odense and Copenhagen University Hospitals). The funding and collaboration of all partners is sincerely acknowledged. The authors sincerely thank Rikke Skovhøj Sørensen (Technical University of Denmark) and Christer P. Volk (SenseLabOnline, FORCE Technology) for their support. We also want to thank the participants and student helpers in this study.

C

A clinical test battery for Better hEAring Rehabilitation (BEAR): Towards the prediction of individual auditory deficits and hearing-aid benefit [a]

Abstract

One aim of the Better hEAring Rehabilitation (BEAR) project is to define a new clinical profiling tool, a test battery, for individualized hearing loss characterization. Recently, (Sanchez-Lopez et al., 2020d, *Chapter 4*) proposed a test battery that includes six types of measures: audibility, speech perception, binaural-processing abilities, loudness perception, and spectro-temporal resolution. The results of 75 listeners were analyzed using a data-driven approach (Sanchez-Lopez et al., 2020b, *Chapter 4*), which provided evidence for the existence of two independent sources of auditory distortion and four different auditory profiles. The classification of the listeners into auditory profiles allows the prediction of the performance of the listeners on different psychoa-

---

[a] This chapter is based on:

Sanchez Lopez, R., Nielsen, S. G., Cañete, O., Fereczkowski, M., Wu, M., Neher, T., Dau, T., Santurette, S. (**2019**). "A clinical test battery for Better hEAring Rehabilitation (BEAR): Towards the prediction of individual auditory deficits and hearing-aid benefit," In Proceedings of the 23rd International Congress on Acoustics (pp. 3841-3848).

coustic tasks as well as their expected aided speech intelligibility. For clinical practice, a decision tree with a small set of highly predictive tests is desirable for an efficient classification of hearing-impaired individuals. The main aim of the present study was to investigate the optimal decision tree and to propose a clinically feasible test battery with a minimum number of tests for accurate listener classification. The clinical test battery will be used in a large-scale field study that will help implement a hearing-aid fitting protocol for better hearing rehabilitation.

C.1 Introduction

The Better hEAring Rehabilitation (BEAR) project pursues the development and implementation of new methods for the diagnosis of hearing deficits as well as new hearing-aid compensation strategies to improve hearing rehabilitation. Since digital hearing aids were introduced to the market, hearing-aid users have reported increased benefit (Kochkin, 2010), probably because of the advanced signal processing techniques (or features) that are now commonly available, such as directionality, noise reduction and dynamic range compression. However, the hearing-aid fitting is still based on the audiogram only which provides the basis for frequency-dependent gain prescription. The other features are adjusted based on preferences and not according to the individual auditory deficits of the user. Furthermore, in hearing care clinics it is common to “fine-tune” some hearing-aid parameters during follow-up visits (Tecca, 2018). If the initial fitting is near-optimal, the follow-up visit may focus on individualization of the fitting parameters according to the “life-style” of the patient. However, if the initial fitting is far from optimal, the audiologist needs to tailor-fit hearing-aid parameters to the hearing deficits of the listener by “trial-and-error”. The BEAR project attempts to improve this situation by identifying groups of listeners – or “auditory profiles” – with specific performance patterns on a range of threshold and supra-threshold tasks and by providing tailored solutions with proposed dedicated hearing-aid compensation strategies for each auditory profile.

In an attempt to identify the auditory profiles, Sanchez-Lopez et al. (2018a) hypothesized that the hearing deficits of a given listener can be described as the combination of two independent types of auditory distortions. The hypothesis was based on the idea that each type of distortion can cause both threshold and supra-threshold deficits and that these deficits are not necessarily independent. In Figure C.1(left panel), the two types of distortions create a two-dimensional space where a given listener's location is determined by the degree of severity of these distortions. As a result, the listener can be identified as belonging to a certain auditory profile. As shown in Figure C.1, normal-hearing listeners are located at the bottom left-hand corner, exhibiting no distortions. Profile A corresponds to a group with minor distortions and therefore good performance in general. Profile C exhibits a high degree of both types of distortions. Profile B exhibits a high degree of distortion type I. Profile D shows a high degree of distortion type II. Using a data-driven approach (Sanchez-Lopez et al., 2018a), four auditory profiles were identified by analyzing the data from two previous studies, providing evidence for the validity of this approach. However, the substantial differences in terms of listeners and tests applied in these two studies limited the overall conclusions that could be drawn from this work.

In order to overcome the aforementioned limitations discussed in Sanchez-Lopez et al. (2018a), a new test battery including a range of supra-threshold psychoacoustic tests was proposed and evaluated in 75 listeners with various types of audiometric configurations. Additionally, the test-retest reliability of the new test battery was investigated in a subset of 11 listeners (Sanchez-Lopez et al., 2020d). The dataset obtained in this manner will in the following be referred to as BEAR3. For the classification of the 75 listeners, unsupervised learning techniques were used to carry out iterative auditory profiling based on the data-driven approach (Sanchez-Lopez et al., 2020b). After this analysis, 70 of the 75 listeners were reliably identified as belonging to one of the four auditory profiles A-D and the remaining five listeners (shown in grey in Figure C.1) were left unclassified.

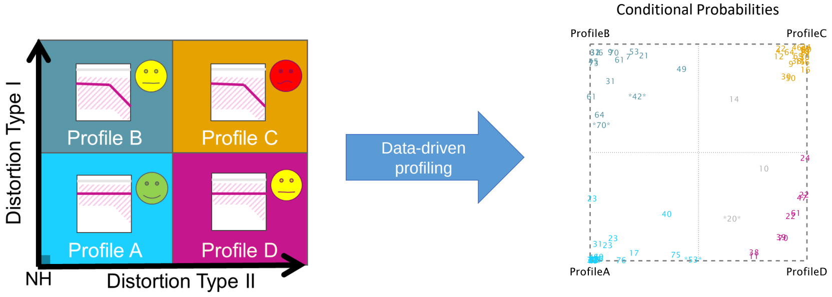


Figure C.1: Hypothesized auditory profiles together with the results of the data-driven profile identification. Left panel: The listeners are placed in a two-dimensional space along two dimensions of auditory distortion. Right panel: Using a data-driven analysis, listeners are placed in the two-dimensional space as a function of their probability of belonging to a specific profile.

However, this iterative unsupervised method requires the entire dataset to identify the four groups and is therefore not suitable for the classification of new listeners. Decision trees are a well-known simple classification tool that may prove useful for classifying unseen data, i.e. new listeners. The efficacy of decision trees can be explored by evaluating their classification performance (Sweets and Pickett, 1982).

When implementing a new protocol for diagnosing a specific disease in the clinic, it is crucial to evaluate its ability to correctly identify the patients who are affected by the health problem under consideration. In general, two types of errors can occur in this classification process: truly affected patients may be “missed” (false negatives) and healthy patients may be “misclassified” as being affected (false positives). Confusion matrices are typically used to quantify the test performance of a classifier. In addition to the classification performance, it is of interest to investigate the cost efficiency of a new clinical protocol (Gorga and Neely, 2003) by estimating the cost of having false negatives or false positives as well as the benefit that the correct classification would provide.

The goal of the current study was to develop a decision tree for a large field study to be conducted as part of the BEAR project, where listeners will have to be classified into the four hypothesized auditory profiles. It is also of interest to identify an additional group of unclassified listeners (Uc) who do not seem to belong to any of the four primary profiles. The BEAR3 dataset (Sanchez-Lopez et al., 2019) was used for investigating the accuracy and efficiency of different decision trees. Using supervised learning techniques, different classification strategies were tested and evaluated in terms of both test performance and cost effectiveness.

C.2 Methods

Decision tree classifiers were trained for predicting the identified auditory profiles from Sanchez-Lopez et al. (2020b) using supervised learning. The analysis of the cost efficiency was based on considerations made in the context of the aforementioned field study to be carried out in different hearing clinics. These considerations cover the recruitment of a random sample of 500 listeners where at least 60 listeners per profile are expected.

Classification methods

The classification of the listeners into the four auditory profiles was performed using supervised learning with tests that showed good to excellent reliability as input, and the labels of the four auditory profiles as well as the unclassified group (Uc) as output. The classification algorithm used here was a standard classification and regression tree (CART), which makes use of recursive binary partitions in order to fit the data to the best set of binary decisions or splits (Franklin, 2005). Four classification schemes were considered:

- DTA: A simple classification based only on the audiogram.
- DT10: A multi-label “fitted” classification. Decision tree based on all reliable tests and 10 binary decisions.
- DT7: A multi-label “pruned” classification with seven binary decisions.

- DT4: A multi-label “pruned” classification with four binary decisions.

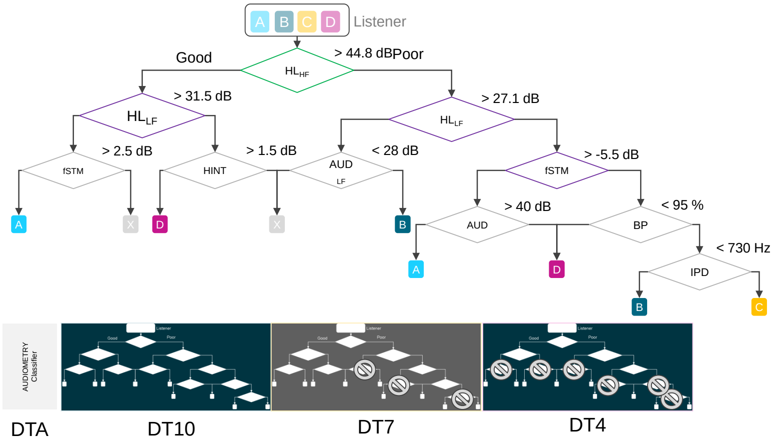


Figure C.2: Top panel: The complete decision tree (DT10). Bottom panel: The four decision trees considered in the present study. DTA: Audiometry-based classifier, DT10: Same as in top panel; DT7: Decision tree with seven binary decisions (DT10 with three pruned splits), and DT4: Decision tree with four binary decisions (DT10 with six pruned splits).

Figure C.2 illustrates the complete classification tree (DT10). Each diamond (split) corresponds to a logic rule related to a given variable, for example $\text{HL}_{\text{HF}} > 45$ dB HL. The right branch corresponds to poorer outcome and the left branch to better outcome. The decision trees DT7 and DT4 are the result of pruning the decision tree DT10 by discarding some of the nodes, as illustrated in the figure.

Test performance and cost-efficiency evaluation

In order to evaluate both the classification performance and the cost efficiency, a synthetic dataset was created for bootstrapping. The original dataset was copied seven times which resulted in 525 observations. Next, the specific standard error of the measurements (SEM) of each test (Sanchez-Lopez et al., 2020d) was used for introducing some uncertainty (additive Gaussian noise) in the outcomes to simulate the data from the aforementioned field study. Confusion matrixes were then calculated for 100 iterations and the cost-efficiency was estimated.

The cost-efficiency was calculated based on (Gorga and Neely, 2003) and adapted for the multi-label case. Consider a 2x2 matrix of costs **C**. Following the previous assumptions, C_{00} is the cost of a true negative, i.e. a participant to be excluded from the study or correctly "not-classified" as a given profile. The cost C_{00} would be equal to the session cost. C_{11} is the cost of a correct classification. C_{01} and C_{10} correspond to false positives and false negatives, respectively, which would introduce outliers in the final results. These would be equal to the cost of misclassification. Additionally, consider the matrix **P** with the probabilities of each of the previous cases, where P_{11} is the probability of correct classification, P_{00} that of correct rejection, and P_{01} and P_{10} those of the two types of misclassification. The expected cost is the Hadamard product of the **P** and **C** matrixes:

$$\text{Expected Cost} = \mathbf{P} \circ \mathbf{C} \tag{C.1}$$

This generic expression can then be simplified due to the fact that the probability of belonging to a given class class_i not truly belonging to that group $P(\text{class}_i|\text{class}_j)$ is equal to $1 - P(\text{class}_j|\text{class}_j)$. The index i denotes the predicted class and the index j the actual class. Therefore, the expression can be simplified to:

$$\text{Expected Cost} = \sum_i \sum_j P(\text{class}_i|\text{class}_j) - \beta P(\text{class}_j|\text{class}_j), \tag{C.2}$$

where β is defined as

$$\beta = \frac{C_{11} - C_{01}}{C_{00} - C_{10}} \frac{P(\text{class}_j)}{1 - P(\text{class}_j)}. \tag{C.3}$$

Given that the probabilities can be calculated in terms of the specificity and sensitivity, Equation 2 can be written as follows:

$$\text{Expected Cost} = \sum_{j} \text{Specificity}_j - \beta\,\text{Sensitivity}_j \tag{C.4}$$

Cost assumptions

The following assumptions were made:

- **Test cost:** Each additional test that is not part of current clinical practice incurs costs for the implementation, the training of the examiners and the documentation. Uus et al. (2006) analyzed the costs of implementing a newborn screening program. The average set-up cost across 16 sites for implementing two tests was £665 for 1000 infants. Therefore, in the present study, a hypothetical total cost of $600 was considered for a field study that involves 500 listeners. The cost per new test per session would therefore be $1.2.

  Session cost: The duration of the session has a cost that involves the salary of the examiner and the use of the facilities. Taking the average of the costs suggested in Abrams et al. (2002), Fleming and Docs (2016), and Mclean (2008) and assuming that one session lasts for one hour, this leads to a hypothetical cost of $60 per session or $1 per minute.

  Correct classification: The correct classification of a given listener increases the probability for the study to be successful. As suggested in Gorga and Neely (2003), this should involve the long-term benefits, including the future reduction of follow-up visits in the clinics if the project is a success. In this case, we limited the expected benefit to the reduction of follow-up visits. Tecca (2018) recently studied the number of visits and the incidence of hearing-aid fitting-related complaints during the first six weeks of hearing aid use. It was shown that the first and second visit involved changes in the gain and advanced features in more than 70% of the cases. Since the new

BEAR fitting rationale aims to provide a better first-fit solution, a cost of one follow-up visit ($60) was considered here.

Misclassification cost: The cost of classifying the listener as belonging to a different auditory profile would correspond to the extra efforts used for this listener to obtain their optimal fitting, i.e. follow-up visits. Here, it is assumed that these listeners would be unsatisfied with their fittings because that corresponds to any other auditory profile and that this would lead to two extra follow-up visits for fine-tuning and verification ($120).

Table C.1 shows the characteristics of each of the considered classifiers in terms of number of tests, the duration of the session and the total test cost and session cost.

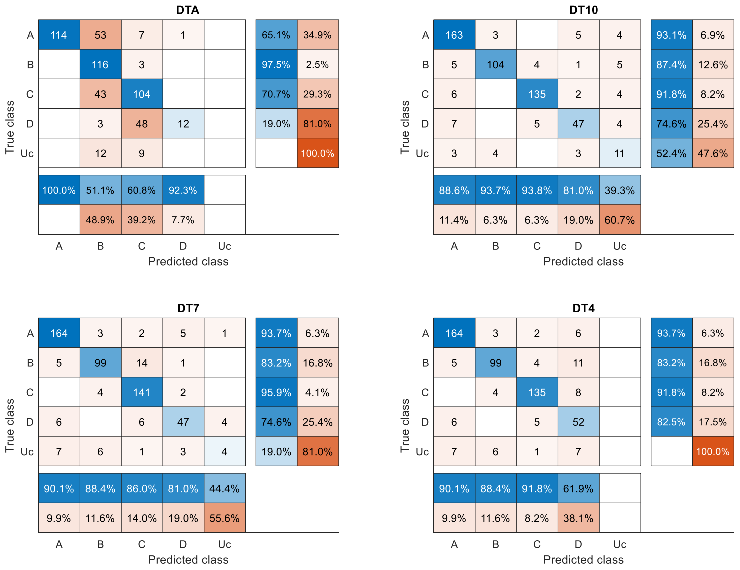
Table C.1: Description of the four classifiers in terms of number of tests, duration and costs. The number of tests includes the outcome measure HINT by default even in the case of DTA and DT4 where this test is not part of the decision trees.

Decision Trees	Description	Number of tests	Duration (min)	Test cost ($)	Session cost ($)
DTA	Audiometry classifier	1	27	1.2	28.2
DT10	Complete classifier	5	68	6.0	74.0
DT7	Pruned classifier I	4	41	4.8	45.8
DT4	Pruned classifier II	3	34	3.6	37.6

C.3 Results

Classification performance

The four classifiers were tested with a constructed data set based on the original data of the BEAR3 data after applying bootstrapping. Figure C.3 shows the confusion matrices of the four classifiers.



DTA

	A	B	C	D	Uc		
A	114	53	7	1		65.1%	34.9%
B		116	3			97.5%	2.5%
C		43	104			70.7%	29.3%
D		3	48	12		19.0%	81.0%
Uc		12	9				100.0%
	100.0%	51.1%	60.8%	92.3%			
		48.9%	39.2%	7.7%			

DT10

	A	B	C	D	Uc		
A	163	3		5	4	93.1%	6.9%
B	5	104	4	1	5	87.4%	12.6%
C	6		135	2	4	91.8%	8.2%
D	7		5	47	4	74.6%	25.4%
Uc	3	4		3	11	52.4%	47.6%
	88.6%	93.7%	93.8%	81.0%	39.3%		
	11.4%	6.3%	6.3%	19.0%	60.7%		

DT7

	A	B	C	D	Uc		
A	164	3	2	5	1	93.7%	6.3%
B	5	99	14	1		83.2%	16.8%
C		4	141	2		95.9%	4.1%
D	6		6	47	4	74.6%	25.4%
Uc	7	6	1	3	4	19.0%	81.0%
	90.1%	88.4%	86.0%	81.0%	44.4%		
	9.9%	11.6%	14.0%	19.0%	55.6%		

DT4

	A	B	C	D	Uc		
A	164	3	2	6		93.7%	6.3%
B	5	99	4	11		83.2%	16.8%
C		4	135	8		91.8%	8.2%
D	6		5	52		82.5%	17.5%
Uc	7	6	1	7			100.0%
	90.1%	88.4%	91.8%	61.9%			
	9.9%	11.6%	8.2%	38.1%			

Figure C.3: Confusion matrices corresponding to one single iteration of classification for each of the four tested decision trees.

The audiometry-based classifier (DTA) was able to correctly predict 67% of the data with a low sensitivity in the predictions of profiles B and C and a low specificity in the case of profile D (19%). The classifiers DT10, DT7 and DT4 had an overall accuracy of 85%. However, they differed in terms of the specificity and sensitivity for each of the profiles. DT10 and DT7 were able to identify some Uc listeners correctly. In contrast, while DT4 lead to a higher specificity in the classification of profile D, it had the disadvantage that all the Uc listeners would be missclassified as any of the four profiles. The main difference between DT10 and DT7 was the sensitivity, especially for the Uc listeners. DT10 was more accurate in the prediction of the Uc listeners, and it had also higher specificity for predicting the four auditory profiles. DT7 predicted less Uc listeners and misclassified more true B listeners, but had also higher sensitivity for profile C. Overall, the decision trees that contain binary decisions for identifying Uc listeners (DT10 and DT7) were both more accurate and more specific.

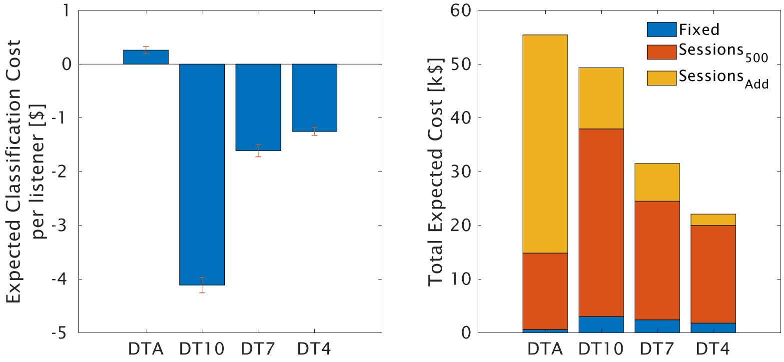


Figure C.4: Expected classification cost per session and total classification cost. The total classification cost involves the costs of the implementation of new tests (Fixed), the session costs for 500 listeners (Sessions$_{500}$) and the cost of additional sessions (Sessions$_{\text{Add}}$).

Expected cost

The expected cost was calculated according to Equation 4. Additionally, the total expected cost of the field study was estimated by the sum of the fixed costs, the planned sessions for 500 listeners recruited randomly, and the additional sessions needed for fulfilling the requirement of testing 60 listeners in each profile. Figure C.4 shows the expected costs. The left panel illustrates the differences among the four decision trees where DTA resulted in higher costs than the other three classifiers. DT10 provided higher cost-efficiency with $4 gained per listener, followed by DT7 with $1.5 and DT4 with $1.2 gained per listener. The right panel of Figure C.4 shows the total cost of the field study. The audiometry-based DTA classifier was the one with the lowest fixed and planned costs but the highest total cost. This is because of the risk of misclassification, which requires numerous additional listeners to get 60 subjects in profile D, with a total of 1992 listeners. DT10 and DT7 required a similar number of listeners ($\approx$ 675 listeners in total) but differed in the session cost, making DT7 a cheaper decision tree overall. The last decision tree DT4 provided the lowest total cost and required a lower number of additional measurements (564 listeners in total). However, the disadvantage is that DT4 cannot identify Uc listeners and would therefore classify them as belonging to one of the four auditory profiles. If the aim is to achieve a low number of misclassified listeners, this classifier would not be the optimal choice.

Overall, the results suggest that DT10 would be the best candidate for the considered field study, due to its higher sensitivity and specificity. Moreover, the clinical test battery that can help to better define the auditory profiles in a larger population by gathering information related to auditory spectro-temporal resolution, speech intelligibility, loudness perception and binaural processing abilities.

C.4 Discussion

The results of the present study speak in favor of a reduced test battery based on five tests included in DT10 for classifying listeners in clinical practice. These tests include adaptive categorical loudness scaling (ACALOS), the hearing in noise test (HINT), the binaural pitch (BP) test, the frequency threshold for identifying interaural phase differences (IPD), and a fast version of the spectro-temporal modulation sensitivity test (fSTM). As such, the proposed clinical version of the test battery covers four domains: loudness perception, speech-in-noise intelligibility, binaural processing abilities and spectro-temporal modulation sensitivity. Although the original BEAR test battery (Sanchez-Lopez et al., 2020d) also involved tests related to audibility and spectro-temporal resolution (as well as some additional tests in the four covered domains), the five tests were found to be the most informative and reliable for the classification of the listeners in auditory profiles.

The ACALOS test is able to estimate hearing thresholds, which are comparable to the ones provided by pure-tone audiometry (Al-Salim et al., 2010). In the present study, these estimates were used as the most informative predictors for the fitted classifier (DT10) instead of the audiometric thresholds. ACALOS is also able to provide supra-threshold information related to loudness perception, such as the slope of the loudness functions, the most comfortable level and the overall dynamic range. Therefore, the use of ACALOS could be of interest not only for the purpose of auditory profiling but also for hearing-aid fitting. For example, it would provide information about the growth of loudness of the patient that could guide fine-tuning of the gain at different input levels. Moreover, fitting formulas based on loudness normalization could be refined if loudness is measured with this technique (Brand and Hohmann, 2002).

The results of the fSTM test showed that profile C listeners have significantly poorer performance than listeners belonging to profiles A, B or D. This makes this new test quite interesting for classification. Additionally, the HINT results showed that profile B and C listeners had elevated speech reception thresholds in noise, suggesting that hearing-aid outcome will improve for these listeners if advanced processing is able to increase the signal-to-noise ratio. Therefore, it would be interesting to investigate whether the tests involved in DT4 only (ACALOS, fSTM) are sufficient for a short version of the clinical test battery. Moreover, these tests are not language-dependent, in contrast to HINT. Although DT4 could be more easily adopted by the public health centers due to the cost-efficiency and shorter duration of the tests (35 min), the use of more informative tests, including the complete decision tree (DT10), should be of higher priority for a field study with research as the main purpose.

The unclassified group can only be identified using DT10 or DT7. It is of interest to identify this group during the field study further explore their supra-threshold auditory performance, which could help better understand the consequences of hearing loss in those listeners.

C.5 Conclusion

The results of the present study support the implementation of new audiological tests in the clinic to achieve a more comprehensive definition of the hearing abilities of patients with hearing loss. Four decision trees were evaluated in terms of classification performance and cost efficiency. The most informative and reliable tests beyond the audiogram were found to be the evaluation of spectro-temporal modulation sensitivity, loudness perception and binaural processing abilities. The BEAR clinical test battery will be evaluated in a large-scale study together with the new profile-based hearing-aid fitting strategy. The BEAR clinical test battery based on DT10, and proposed for such a field study, is available in a public repository[b].

Acknowledgements

We want to thank Dorte Hammershøi, James Harte and the other BEAR partners for their input during the realization of this study. This work was supported by Innovation Fund Denmark Grand Solutions 5164-00011B (Better hEAring Rehabilitation project), Oticon A/S, GN Hearing, Widex A/S, Aalborg University, University of Southern Denmark, the Technical University of Denmark, Force Technology, and the university hospitals in Aalborg, Odense and Copenhagen. The funding and collaboration are sincerely appreciated.

---

[b] `https://bitbucket.org/hea-dtu/bear-test-battery/`

Contributions to Hearing Research

Vol. 1: *Gilles Pigasse*, Deriving cochlear delays in humans using otoacoustic emissions and auditory evoked potentials, 2008.
External examiners: Mark Lutman, Stefan Stenfeld

Vol. 2: *Olaf Strelcyk*, Peripheral auditory processing and speech reception in impaired hearing, 2009.
External examiners: Brian Moore, Kathrin Krumbholz

Vol. 3: *Eric R. Thompson*, Characterizing binaural processing of amplitude-modulated sounds, 2009.
External examiners: Michael Akeroyd, Armin Kohlrausch

Vol. 4: *Tobias Piechowiak*, Spectro-temporal analysis of complex sounds in the human auditory system, 2009.
External examiners: Jesko Verhey, Steven van de Par

Vol. 5: *Jens Bo Nielsen*, Assessment of speech intelligibility in background noise and reverberation, 2009.
External examiners: Björn Hagerman, Ejnar Laukli

Vol. 6: *Helen Connor*, Hearing aid amplification at soft input levels, 2010.
External examiners: Inga Holube, Birgitta Larsby

Vol. 7: *Morten Løve Jepsen*, Modeling auditory processing and speech perception in hearing-impaired listeners, 2010.
External examiners: Birger Kollmeier, Ray Meddis

Vol. 8: *Sarah Verhulst*, Characterizing and modeling dynamic processes in the cochlea using otoacoustic emissions, 2010.
External examiners: David Kemp, Stephen Neely

Vol. 9: *Sylvain Favrot*, A loudspeaker-based room auralization system for auditory research, 2010.
External examiners: Bernhard Seeber, Michael Vorländer

Vol. 10: *Sébastien Santurette*, Neural coding and perception of pitch in the normal and impaired human auditory system, 2011.
External examiners: Christopher Plack, Christian Lorenzi

Vol. 11: *Iris Arweiler*, Processing of spatial sounds in the impaired auditory system, 2011.
External examiners: Joost Festen, Jürgen Tchorz

Vol. 12: *Filip Munch Rønne*, Modeling auditory evoked potentials to complex stimuli, 2012.
External examiners: Bob Burkard, Stephen Neely

Vol. 13: *Claus Forup Corlin Jespersgaard*, Listening in adverse conditions: Masking release and effects of hearing loss, 2012.
External examiners: Stuart Rosen, Christian Lorenzi

Vol. 14: *Rémi Decorsière*, Spectrogram inversion and potential applications for hearing research, 2013.
External examiners: Michael Stone, Oded Ghitza

Vol. 15: *Søren Jørgensen*, Modeling speech intelligibility based on the signal-to-noise envelope power ration, 2014.
External examiners: John Culling, Martin Cooke

Vol. 16: *Kasper Eskelund*, Electrophysiological assessment of audiovisual integration in speech perception, 2014.
External examiners: Lawrence Rosenblum, Matthias Gondan

Vol. 17: *Simon Krogholt Christiansen,* The role of temporal coherence in auditory stream segregation, 2014.
External examiners: Shihab Shamma, Guy Brown

Vol. 18: *Márton Marschall,* Capturing and reproducing realistic acoustic scenes for hearing research, 2014.
External examiners: Sascha Spors, Ville Pulkki

Vol. 19: *Jasmina Catic,* Human sound externalization in reverberant environments, 2014.
External examiners: Bernhard Seeber, Steven van de Par

Vol. 20: *Michał Feręczkowski,* Design and evaluation of individualized hearing-aid signal processing and fitting, 2015.
External examiners: Christopher Plack, Enrique Lopez-Poveda

Vol. 21: *Alexandre Chabot-Leclerc,* Computational modeling of speech intelligibility in adverse conditions, 2015.
External examiners: Steven van de Par, John Culling

Vol. 22: *Federica Bianchi,* Pitch representations in the impaired auditory system and implications for music perception, 2016.
External examiners: Ingrid Johnsrude, Christian Lorenzi

Vol. 23: *Johannes Zaar,* Measures and computational models of microscopic speech perception, 2016.
External examiners: Judy Dubno, Martin Cooke

Vol. 24: *Johannes Käsbach,* Characterizing apparent source width perception, 2016.
External examiners: William Whitmer, Jürgen Tchorz

Vol. 25: *Gusztáv Löcsei,* Lateralized speech perception with normal and impaired hearing, 2016.
External examiners: Thomas Brand, Armin Kohlrausch

Vol. 26: *Suyash Narendra Joshi*, Modelling auditory nerve responses to electrical stimulation, 2017.
External examiners: Laurel Carney, Bob Carlyon

Vol. 27: *Henrik Gerd Hassager*, Characterizing perceptual externalization in listeners with normal, impaired and aided-impaired hearing, 2017.
External examiners: Volker Hohmann, Piotr Majdak

Vol. 28: *Richard Ian McWalter*, Analysis of the auditory system via synthesis of natural sounds, speech and music, 2017.
External examiners: Maria Chait, Christian Lorenzi

Vol. 29: *Jens Cubick*, Characterizing the auditory cues for the processing and perception of spatial sounds, 2017.
External examiners: Ville Pulkki, Pavel Zahorik

Vol. 30: *Gerard Encina-Llamas*, Characterizing cochlear hearing impairment using advanced electrophysiological methods, 2017.
External examiners: Roland Schaette, Ian Bruce

Vol. 31: *Christoph Scheidiger*, Assessing speech intelligibility in hearing-impaired listeners, 2018.
External examiners: Enrique Lopez-Poveda, Tim Jürgens

Vol. 32: *Alan Wiinberg*, Perceptual effects of non-linear hearing aid amplification strategies, 2018.
External examiners: Armin Kohlrausch, James Kates

Vol. 33: *Thomas Bentsen*, Computational speech segregation inspired by principles of auditory processing, 2018.
External examiners: Stefan Bleeck, Jürgen Tchorz

Vol. 34: *François Guérit*, Temporal charge interactions in cochlear implant listeners, 2018.
External examiners: Julie Arenberg, Olivier Macherey

Vol. 35: *Andreu Paredes Gallardo,* Behavioral and objective measures of stream segregation in cochlear implant users, 2018.
External examiners: Christophe Micheyl, Monita Chatterjee

Vol. 36: *Søren Fuglsang,* Characterizing neural mechanisms of attention-driven speech processing, 2019.
External examiners: Shihab Shamma, Maarten de Vos

Vol. 37: *Borys Kowalewski,* Assessing the effects of hearing-aid dynamic-range compression on auditory signal processing and perception, 2019.
External examiners: Brian Moore, Graham Naylor

Vol. 38: *Helia Relaño Iborra,* Predicting speech perception of normal-hearing and hearing-impaired listeners, 2019.
External examiners: Ian Bruce, Armin Kohlrausch

Vol. 39: *Axel Ahrens,* Characterizing auditory and audio-visual perception in virtual environments, 2019.
External examiners: Pavel Zahorik, Piotr Majdak

Vol. 40: *Niclas A. Janssen,* Binaural streaming in cochlear implant patients, 2019.
External examiners: Tim Jürgens, Hamish Innes-Brown

Vol. 41: *Wiebke Lamping,* Improving cochlear implant performance through psychophysical measures, 2019.
External examiners: Dan Gnasia, David Landsberger

Vol. 42: *Antoine Favre-Félix,* Controlling a hearing aid with electrically assessed eye gaze, 2020.
External examiners: Graham Naylor, Jürgen Tchorz

The end.

To be continued…

Communication is an important ability that shapes our social life. A critical factor that affects successful oral communication is hearing loss. If the verbal information is not properly perceived, the interpretation can be flawed leading to an unsatisfactory, tiring and uncomfortable conversation. The sources and consequences of a sensorineural hearing loss are diverse and the hearing devices, especially hearing aids, have multiple configurations that can be adjusted for specific needs. However, the hearing-aid fitting to the individual hearing loss is currently performed based mainly on the audiogram, which is not necessarily related to listening abilities such as the speech understanding. In this thesis, the basis for "precision audiology" was explored. The prerequisites for implementing precision treatments are 1) that the diseases must be heterogenous, 2) that there exist multiple options for treatment and 3) that there are "markers" that associate certain characteristics of the patient to specific treatments. The present work focused on the investigation of auditory biomarkers that allow the link between perceptual deficits and hearing-aid settings. Different approaches for precision audiology may be implemented in the near future. These can drive hearing-aid development, hearing loss characterization and the quality of service in the hearing-care clinic towards a better hearing rehabilitation and an evidence-based audiological practice.

DTU Health Tech
Department of Health Technology

Ørsteds Plads
Building 352
DK-2800 Kgs. Lyngby
Denmark
Tel: (+45) 45 25 39 50
www.hea.healthtech.dtu.dk