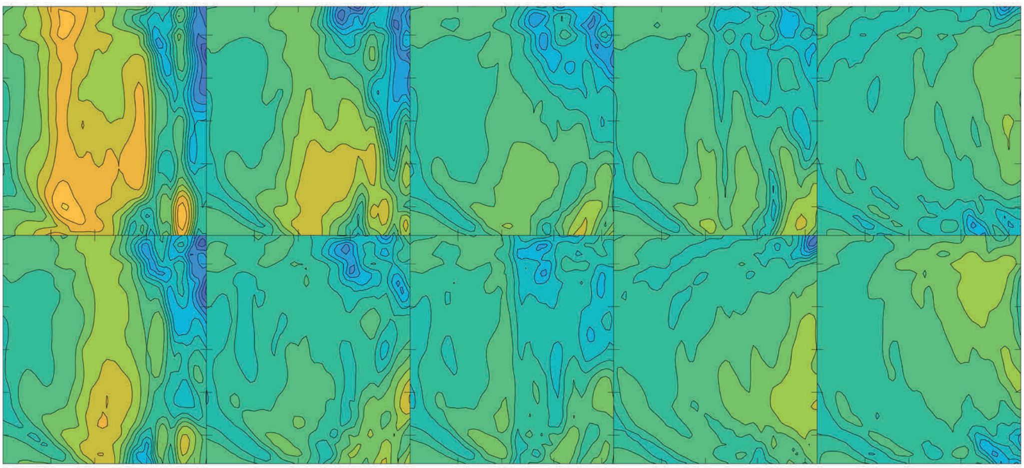
Florian Denk

Characterizing and conserving the transmission properties of the external ear with hearing devices



BIS-Verlag der Carl von Ossietzky Universität Oldenburg

Florian Denk

Characterizing and conserving the transmission properties of the external ear with hearing devices

BIS-Verlag der Carl von Ossietzky Universität Oldenburg

Oldenburg, 2020

Verlag / Druck / Vertrieb

BIS-Verlag
der Carl von Ossietzky Universität Oldenburg
Postfach 2541
26015 Oldenburg

E-Mail: bisverlag@uni-oldenburg.de

ISBN 978-3-8142-2387-2

Von der Fakultät V für Mathematik und Naturwissenschaften
der Carl von Ossietzky Universität Oldenburg
zur Erlangung des Grades und Titels

eines Doktors der Naturwissenschaften
(Dr. rer. nat.)

angenommene Dissertation

von Florian Denk
geboren am 18.04.1992 in Günzburg

Erstgutachter :	Prof. Dr. Dr. Birger Kollmeier
Zweitgutachterin:	Prof. Dr. Dorte Hammershøi
Drittprüferin:	Prof. Dr. Ulrike Feudel
Eingereicht:	30.09.2019
Tag der Disputation:	03.12.2019

Preface

„All good things come in threes“ holds for most dissertations: Three peer-reviewed publications within three years covering either one of three approaches (i.e. theory, experiments or applications) make up a good dissertation. However, Florian Denk somehow overstretched this rule: The three main chapters of his dissertation contain EACH three peer-reviewed papers covering ALL three approaches... and it was even finished within LESS than three years! Apparently, from the three times 25 PhD candidates that I was privileged to supervise so far (including Florian as the 75th candidate), he clearly sticks out: Starting with a rather application-oriented task (i. e., to construct hearing devices that provide acoustic transparency to the user, or " hearing aids without side effects" as Florian commonly explains it in short), he did not only apply for a patent for future use but decided to dig much deeper into experiments and even hearing theory. As a result, Florian established a freely available database of individual head-related impulse responses with a high spatial resolution not only for the "naked" ear, but also for all different kinds of ear-level hearing devices and form factors. Based on this, he could characterise both experimentally and with model predictions the localisation ability of good-localisation-performance subjects and poor performers as well as an exact account of how much localisation ability the user of a certain hearing device will loose—there is nothing as practical as a good theory! Please read yourself!

Based on his extensive work in psychoacoustics, auditory modelling and acoustic engineering which even led to a commercially available research prototype of his transparent earpiece (the "hearpiece"), it is not surprising that Florian has convinced the three most relevant institutions that he performs excellent work: First, the international scientific community who accepted, utilised and cited his publications and his free database. Second, the German research funding organisation DFG who recognised his great work in the context of our structured joint research projects Forschergruppe, Sonderforschungsbereich Hearing Acoustics and Cluster of Excellence, and even signalised that they would support Florian's scientific work much further. Third, last but not least, the dissertation committee (consisting of three professors) and the School of Mathematics and Sciences who agreed on accepting his dissertation with the highest possible distinction "summa cum laude". Please read yourself to get convinced that his dissertation justifies this award!

But Florian Denk would not be himself without providing a role model for future PhD students with respect to caring for others and bringing the research results from the dark dungeons of acoustics to the shining light of PR activities or even science slams: Among others, Florian has served as the spokesperson for all PhD students in our Cluster of Excellence "Hearing4all", has been invited as an entertaining science slammer in various scientific conferences and meetings, and has even been awarded with the Oldenburg "Golden Brain" (one of the highest science slam award available in the North-West region of Germany). Again: All good things come in threes!

Taken together, it has been a privilege, a great jump forward in science, and—most of all—much fun to work with Florian and to supervise his dissertation—up to a point where it became unclear who supervises whom! Please read yourself to get convinced, that his dissertation does not only cover great science but also appears like the starting point of a great scientific career!

Oldenburg, March 2020, Birger Kollmeier

Abstract

Hearing devices are electro-acoustic instruments worn in the ear to modify the heard environment. This comprises hearing aids, which amplify sound to compensate for consequences of hearing impairment, but also similar devices targeted at normal-hearing users, like hear-through headsets or electronic hearing protectors. An ideal hearing device would allow for natural sound perception just as with the open ear, except for the modification that is desired by design. This property is referred to as acoustic transparency. Typically, current hearing devices do not fulfill this requirement but introduce additional artifacts. These can be described as an alteration or destruction of the acoustic transmission properties of the open ear that a listener is trained to exploit for analyzing the acoustic environment. Such artifacts lead to a degraded auditory perception, characterized by a reduced ability to localize, separate and understand different sound sources, as well as an impaired sound quality. This thesis is a contribution to improving the conservation of external ear transmission properties in hearing devices by a comprehensive assessment of the hearing device related ear acoustics, an evaluation of perceptual consequences and the development of technical solutions.

The limitations of acoustic transparency imposed by the signal captured by different hearing device styles were assessed by measuring direction-dependent acoustic transfer functions of individual ears, including the transfer functions to both the eardrum of the open ear and to different hearing device microphone locations. This data also allowed to simulate the full processing chain of hearing devices while freely changing the individual processing steps, like the occurrence of a delay. Acoustic analyses, auditory model simulations and psychoacoustic experiments were conducted to quantify the artifacts occurring in linear hearing devices and their influence on timbre, sound quality, directional cues and sound localization. Furthermore, a real-time demonstrator hearing device was implemented and extended. This included the refinement of an electro-acoustic earpiece, and the development and implementation of algorithms for achieving acoustic transparency in integration with other building blocks on a PC-based real-time platform. The real-time device was used for evaluation purposes, as well as for exploring the possibility of integrating ear-level electroencephalography sensors for brain-computer-interface applications in hearing devices.

This thesis provides novel insights regarding the limiting factors for acoustic transparency, which led to the development of methods for establishing optimum acoustic transparency for a given device. To construct hearing devices with a perceptually convincing acoustic transparency, the results show that a) comb-filtering effects have to be avoided, b) the frequency response of the open ear should be conserved for diffuse-field rather than free-field incidence, and c) directional cues created by the pinna have to be conserved or replicated by the hearing device. These requirements translate into a low delay or high suppression of leakage components acoustically entering the ear canal, the need for appropriate individual equalization of the hearing device output, and devices that mechanically obstruct the pinna as little and as uniformly as possible. Exploiting these findings, methods for practical

implementation of acoustic transparency in hearing devices were proposed and demonstrated. This includes the definition of an optimum aided response target, as well as algorithms to achieve this response by means of an equalization filter design, which reduces comb-filtering effects as well. In essence, the fundamental research outcomes of this thesis regarding (aided) external ear acoustics, its implementation in different prototype systems, and the psychoacoustic verification performed here can contribute to better, more natural sounding hearing devices of the future.

Zusammenfassung

Hörsysteme sind im Ohr getragene elektro-akustische Geräte, welche die gehörte Umgebung gezielt verändern. Dazu zählen Hörgeräte, die Schall verstärken um Schwerhörigkeit abzumildern, andererseits aber auch Geräte für Normalhörende, wie Kopfhörer mit Durchhörfunktion und elektronische Gehörschützer. Ein ideales Hörsystem würde nur die gewollte Veränderung der akustischen Umgebung bewirken, ansonsten aber natürliches Hören wie mit dem offenen Ohr erhalten. Diese Eigenschaft bezeichnet man als Akustische Transparenz. Aktuelle Hörsysteme erfüllen diese Anforderung meist nicht, sondern führen zusätzlich Artefakte ein, die einer Veränderung oder Vernichtung der akustischen Übertragungseigenschaften des offenen Außenohrs entsprechen. Solche Artefakte führen zu einer gestörten Hörwahrnehmung, wie Schwierigkeiten bei der Schalllokalisation, der Unterscheidung von mehreren Quellen, dem Sprachverstehen und einer verschlechterten Klangqualität. Diese Dissertation soll einen Beitrag zur besseren Erhaltung der Außenohrübertragungseigenschaften in Hörsystemen leisten. Dazu wurden die ohrakustischen Effekte im Zusammenhang mit Hörsystemen durch ausgiebige Messungen erfasst und bewertet, deren Auswirkung auf die Hörwahrnehmung analysiert, und auf dieser Basis technische Lösungsansätze entwickelt.

Als Grundlage dienten Messungen der richtungsabhängigen Übertragungsfunktionen individueller Ohren, sowohl zum Trommelfell des offenen Ohres als auch zu unterschiedlich im Ohr positionierten Mikrophonen verschiedener Hörsystem-Bauformen. Auf Basis dieser Daten wurden prinzipielle Limitationen für die Erzeugung akustischer Transparenz abgeleitet, die sich aus den vom Hörsystem aufgenommenen (gestörten) akustischen Informationen ergeben. Diese Daten haben es auch ermöglicht, die Verarbeitungskette in einem Hörsystem realistisch zu simulieren, wobei einzelne Verarbeitungsschritte wie das Auftreten einer Latenz mit großer Freiheit vernachlässigt oder berücksichtigt werden konnten. So konnten mit akustischen Analysen, Simulationen mit auditorischen Modellen und in psychoakustischen Studien die in linearen Hörsystemen auftretenden Artefakte charakterisiert und quantifiziert werden. Außerdem wurde ein Echtzeit-Demonstrationshörsystem realisiert und erweitert. Dazu wurde ein elektroakustisches Ohrpassstück weiterentwickelt, Algorithmen für das Herstellen akustischer Transparenz abgeleitet und in Kombination mit weiteren Verarbeitungsschritten eines Hörsystems auf einer PC-basierten Echtzeitplattform implementiert. Dieses Demonstrationssystem wurde einerseits für Evaluationsstudien genutzt, andererseits wurde damit auch gezeigt, dass sich elektroenzephalographische Sensoren zur Realisierung von Hirn-Computer-Schnittstellen in ein aktives Hörsystem integrieren lassen.

Basierend auf den Ergebnissen dieser Dissertation konnten die Voraussetzungen und Limitationen für akustische Transparenz ermittelt, und daraus Methoden zur praktischen Umsetzung abgeleitet werden. Hierbei stellten sich als wichtigste Voraussetzungen für akustische Transparenz heraus: a) Kammfiltereffekte müssen bestmöglich vermieden werden, b) Die Hörsystem-Übertragungsfunktion muss gegenüber der des offenen Ohres entzerrt werden, vorzugsweise für ein diffuses Schallfeld, c) Richtungsinformationen, die durch die Ohrmuschel erzeugt werden, müssen

so gut wie möglich bewahrt oder nachgebildet werden. Erforderlich dafür sind eine möglichst kurze Verarbeitungslatenz und/oder eine starke Direktschallabdämpfung, eine individuelle Entzerrung der Hörsystemwiedergabe, und eine möglichst kleine Bauform, welche die Pinna so wenig und gleichmäßig wie möglich füllt. Diese Erkenntnisse führten zu direkt anwendbaren Methoden, um mit einem gegebenen Gerät die bestmögliche akustische Transparenz zu erzielen. Dies beinhaltet die Definition eines optimalen Ziel-Frequenzgangs, sowie Algorithmen zur individuellen Entzerrung der Wiedergabe bei gleichzeitiger Reduktion von Kammfiltereffekten. Die grundlegenden Erkenntnisse aus dieser Arbeit zur Außenohrakustik mit Hörsystemen und ihrer Wahrnehmung, sowie die Ergebnisse mit der Echtzeitplattform können so einen Beitrag zu besseren, natürlicher klingenden Hörsystemen der Zukunft leisten.

Contents

List of publications

Peer-reviewed publications this thesis is based on

A1 F. Denk, B. Kollmeier, and S. Ewert (2018f). "Removing reflections in semianechoic impulse responses by frequency-dependent truncation". *Journal of the Audio Engineering Society* 66(3), pp. 146–153.

A2 F. Denk, S. M. A. Ernst, S. D. Ewert, and B. Kollmeier (2018b). "Adapting Hearing Devices to the Individual Ear Acoustics: Database and Target Response Correction Functions for Various Device Styles". *Trends in Hearing* 22, p. 2331216518779313.

A3 F. Denk, S. D. Ewert, and B. Kollmeier (2018g). "Spectral directional cues captured by hearing device microphones in individual human ears". *The Journal of the Acoustical Society of America* 144(4), pp. 2072–2087.

A4 F. Denk, S. Vogl, H. Schepker, B. Kollmeier, M. Blau, and S. Doclo (2017d). "The Acoustically Transparent Hearing Device: Towards Integration of Individualized Sound Equalization, Electro-Acoustic Modeling and Feedback Cancellation". *Proc. 1st International Workshop on Challenges in Hearing Assistive Technology (CHAT-2017).* Stockholm, Sweden, pp. 89–94.

A5 F. Denk, M. Lettau, H. Schepker, S. Doclo, R. Roden, M. Blau, J. Bach, J. Wellmann, and B. Kollmeier (2019a). "A one-size-fits-all earpiece with multiple microphones and drivers for hearing device research". *Proc. AES Conference on Headphone Technology.* Paper 13. San Francisco, USA, pp. 1–9.[1]

A6 F. Denk, H. Schepker, S. Doclo, and B. Kollmeier (2018d). "Equalization filter design for achieving acoustic transparency in a semi-open fit hearing device". *Proc. 13. ITG Conference on Speech Communication.* Oldenburg, Germany, pp. 226–230.

A7 F. Denk, S. Ewert, and B. Kollmeier (2019b). "On the limitations of sound localization with hearing devices". *The Journal of the Acoustical Society of America* 146(3), pp. 1732–1744.

A8 F. Denk, M. Grzybowski, S. M. A. Ernst, B. Kollmeier, S. Debener, and M. G. Bleichner (2018e). "Event-Related Potentials Measured From In and Around the Ear Electrodes Integrated in a Live Hearing Device for Monitoring Sound Perception". *Trends in Hearing* 22, p. 233121651878821.

A9 H. Schepker, F. Denk, B. Kollmeier, and S. Doclo (2019b). "Subjective sound quality evaluation of an acoustically transparent hearing device". *Proc. 2nd AES Conference on Headphone Technology.* Paper 18. San Francisco, USA, pp. 1–10.[2]

[1] Honored with the Best Student Paper Award, AES International Conference on Headphone Technology 2019, San Francisco.

[2] HS and FD contributed equally to this paper.

Further publications associated with this thesis

The following conference proceedings were also published within the scope of this thesis but are not included in this document.

P1 F. Denk, J. Heeren, S. D. Ewert, B. Kollmeier, and S. M. A. Ernst (2017a). "Controlling the Head Position during individual HRTF Measurements and its Effect on Accuracy". *Fortschritte der Akustik - DAGA*. Kiel, Germany, pp. 1085–1088.

P2 F. Denk, M. Grzybowski, S. M. A. Ernst, B. Kollmeier, S. Debener, and M. G. Bleichner (2017c). "Measuring hearing instrument sound modification using integrated ear-EEG". *Proceedings of the International Symposium on Auditory and Audiological Research ISAAR, vol 6.* Nyborg, Denmark, pp. 351–358.

P3 H. Schepker, F. Denk, B. Kollmeier, and S. Doclo (2018). "Multi-loudspeaker equalization for acoustic transparency in a custom hearing device". *Proc. 13. ITG Conference on Speech Communication.* Oldenburg, Germany, pp. 36–40.

P4 F. Denk, F. Brinkmann, A. Stirnemann, and B. Kollmeier (2019c). "The PIRATE: an anthropometric earPlug with exchangeable microphones for Individual Reliable Acquisition of Transfer functions at the Ear canal entrance". *Fortschirtte der Akustik - DAGA*. Rostock, Germany, pp. 635–638.

P5 M. Buhl, F. Denk, M. Bodenstein, N. Wiedemann, S. Jacobsen, M. R. Schädler, and B. Kollmeier (2019). "Calibration of a mobile hearing aid prototype and its validation: Towards transparent listening with commodity hardware". *Fortschirtte der Akustik - DAGA*. Rostock, Germany, pp. 927–930.

Abbreviations

BTE Behind The Ear

FDT Frequency Dependent Truncation

HRTF Head-related Transfer Function

ITC In The Canal

ITE In The Ear

RTF Relative Transfer Function

TRCF Target Response Correction Function

Mathematical Notation

x, X	Scalar values or functions
$\mathbf{x}, \mathbf{X}$	Vectors
$\underline{\underline{\mathbf{x}}}, \underline{\underline{\mathbf{X}}}$	Matrices
$\hat{x}$	The estimate of a quantity x
$x(t)$	A function of a continuous variable t
$x[n]$	A function of a discrete variable n
$x(t) * y(t)$	The convolution of two functions $x(t)$ and $y(t)$
$\mathrm{var}_x[\cdot]$	The variance over a variable x

Fixed Symbols

α	Lateral angle in the interaural-polar coordinate system
β	Polar angle in the interaural-polar coordinate system
φ	Azimuthal angle in the spherical coordinate system
ϑ	Elevation angle in the spherical coordinate system
t	Time in seconds, s
f	Frequency in Hertz, $\mathrm{Hz} = \frac{1}{\mathrm{s}}$
ω	Radial frequency in $\frac{\mathrm{rad}}{\mathrm{s}}$, $\quad \omega = 2\pi f$
i	The imaginary unit, $i = \sqrt{-1}$
h	An impulse response
H	A transfer function
λ	Wavelength in meters, m

1 General introduction

Hearing devices are ear-worn electro-acoustic instruments that modify the heard acoustic environment in a way that depends on the application. This can include an amplification to compensate for a hearing loss as in hearing aids (Dillon 2012; Kollmeier and Kiessling 2018), an attenuation of high-level sounds in electronic hearing protectors (Albrecht et al. 2017; Brungart et al. 2007; Killion et al. 2011; Mazur and Voix 2013), or hearing the external sound environment electronically through a headphone while having the possibility to add additional virtual sound sources in multimedia or augmented reality applications (Brandenburg et al. 2018; Bujacz and Strumiłło 2016; Härmä et al. 2004; Lokki et al. 2004). Independent of the specific application, an ideal hearing device would not perceivably alter the acoustic environment, except for the exact modification that is desired. This property is referred to as *acoustic transparency*.

To achieve acoustic transparency, the transfer function of the wearer's open ear should be conserved up to the desired modification. Furthermore, the device should obviously not produce audible self-noise, non-linear distortions or other electronic artifacts, and should cover the full audible bandwidth (Härmä et al. 2004; Killion 1979; Moore and Tan 2003). Current hearing devices usually do not fulfill these requirements, and therefore lead to a distorted perception of the acoustic environment even if adjusted to produce no modification (Best et al. 2010; Brungart et al. 2007; Cubick et al. 2018; D'Angelo et al. 2001; Härmä et al. 2004; Hoffmann et al. 2014). If these distortions are greater than the benefit of the desired functionality of the device, they essentially render a device useless. Insufficient acoustic transparency therefore limits the performance, user acceptance and range of application for many classes of hearing devices. This thesis provides a comprehensive acoustic assessment of the limitations and technical realization regarding the conservation of the external ear transmission properties in hearing devices, as well as a psychoacoustic evaluation of the perceptual consequences of the occurring deviations.

External ear transmission effects and their importance

Sound transmission processes through the external ear create specific features or cues that humans rely on when making sense of their acoustic environment. Unlike in the visual system, the position and many other attributes of a sound source are not directly represented in the sensory input, but have to be retrieved from the sound pressure waveforms at the two eardrums by analyzing various acoustic cues in integration with the other senses (Blauert 1997; Majdak et al. 2019; Moore 2012). The correspondence between the external ear transmission features and sound source attributes, like the position, have been either learned over time or are 'hard-wired' neural circuits. Therefore, if the usual sound transmission through the external ear is altered as in most current hearing devices, the analysis of an acoustic scene is disturbed or even disrupted. As a result, listening becomes more difficult when wearing hearing devices, particularly in complex listening situations, like localizing and focusing on another person at a big table in a noisy restaurant. This difficulty can lead to a decreased sound quality, spatial awareness, speech understanding, and an increased listening effort (Brungart et al. 2007; Cubick and Dau 2016; Cubick et al. 2018; Deng et al. 2019; Hoffmann et al. 2014; Hofman et al. 1998).

Sound transmission to the eardrum can be described by a set of two linear time invariant transfer functions for the two ears, referred to as Head-related Transfer Function (HRTF). The pronounced dependence of the HRTF on frequency, incidence direction and distance is caused by diffraction and reflection processes at the torso, head and external ear (Blauert 1997; Genuit 1984; Mehrgardt and Mellert 1977; Shaw and Teranishi 1968; Wiener and Ross 1946). It should be noted that these processes can have different effects at different frequency regions, given the large audible frequency range, where the ratio between the largest and smallest wavelength of interest is about 1000. The resulting interactions make the HRTF a complex object that is not always easy to interpret. As a rule of thumb, structures affect the HRTF at a given frequency, if their characteristic size is at least one quarter of the corresponding wavelength.

The HRTF can be separated into direction-dependent and direction-independent parts (Genuit 1984; Hammershøi and Møller 1996; Mehrgardt and Mellert 1977; Møller 1992). The directional features of the HRTF include the Interaural Level Difference and the Interaural Time Difference, i.e., level and time differences between the left and right ear due to spatial separation and diffraction around the head (Rayleigh 1907). Furthermore, spectral directional cues are created by direction-dependent diffraction and reflection processes in the pinna, and reflections from the shoulders (Blauert 1997; Butler and Belendiuk 1977; Mehrgardt and Mellert 1977; Møller et al. 1995a). Another salient directional cue is the dynamic variation of interaural differences that is entangled with the rotation of a person's head (Wallach 1940). Non-directional features of the HRTF originate from cavity resonances of the ear canal and attached cavum conchae, and partly from sound diffraction around the head (Kuhn 1979b; Madaffari 1974; Shaw and Teranishi 1968). Thereby, the cavity resonance of the ear canal is influenced by the acoustic impedance of the eardrum, which is coupled to the middle ear (Hudde 1983). These resonances result in a broad band-pass like amplification with a peak around 3 kHz and 15-20 dB, and an amplification for most frequencies in the audible range above ca. 200 Hz. These resonances determine the timbre of an external ear. Since the ear of every person is differently shaped and the ear canals vary in length, both the directional as well as the non-directional parts of the HRTF differ considerably between persons (Genuit 1984; Hammershøi and Møller 1996; Møller et al. 1995a).

Current hearing device technology

The conservation of open-ear transmission characteristics in hearing devices is often motivated with achieving a high sound quality. In fact, acoustic transparency as introduced above complies with the definition for general audio quality (Fastl and

Zwicker 2007; Union 2003). However, for hearing devices sound quality is far more than a luxury problem, and cannot be compared to owning a good HiFi sound system for the living room. Their sound quality or grade of transparency rather determines the threshold for whom and in which situations such devices are useful and accepted (Jessen et al. 2014; Killion 1979, 2004b).

Hearing aid technology gains importance for the society as a whole in most industrial countries due to the increasing prevalence of hearing loss associated with the current demographic change and rising life expectation. Significant progress has been made in the last two decades, most importantly through advanced signal processing and fitting techniques facilitated by digital hearing aids. Modern hearing aids can compensate for many effects of conductive and sensory-neural hearing loss, by combining linear (frequency-dependent amplification, directional microphones) and non-linear (dynamic range compression, noise reduction) sound processing approaches that can be adapted to the needs of the individual patient (Dillon 2012; Kollmeier and Kiessling 2018). Still, the acceptance of hearing aids is limited, with only about 37 % of diagnosed hearing impaired people in Germany using a hearing aid (Anovum 2018; Bisgaard and Ruf 2018), while an early treatment would often result in better success (Kießling et al. 2018). Self-stated non-financial reasons for not using hearing aids are a persisting stigma related to hearing aids (David and Werner 2016), as well as a non-satisfactory performance.

Many of the performance issues of hearing aids can be ascribed to distortions that are not related to an insufficient compensation of the hearing loss (Jessen et al. 2014; Killion 2004a; Sockalingam et al. 2009). Current hearing aids lead to a degraded spatial perception of sound that can lead to poorer speech understanding in complex acoustic scenes, also when normal-hearing subjects wear them (Akeroyd and Whitmer 2016; Cubick and Dau 2016; Cubick et al. 2018). The timbre produced by hearing aids is often perceived as unnatural, partly of course due to the prescribed frequency-dependent amplification, but also due to suboptimal frequency response characteristics (Killion 1979; Killion and Monsor 1980) and a hollow or metallic sound originating from a processing delay and feedback (Groth and Søndergaard 2004; Spriet et al. 2008; Stone et al. 2008). Moreover, the presence of a hearing device changes how the own voice is perceived (Kuk et al. 2009; Stone and Moore 1999). If these undesired distortions are larger than the benefit from amplification and noise reduction, there is no use of wearing a hearing aid—leading to a minimum degree of hearing impairment that is required to benefit from a hearing aid, which could be referred to as 'minimum aidable hearing loss'. Acoustic transparency could therefore increase the range of impairment, where patients benefit from a hearing aid, and thus contribute to a more effective treatment of hearing loss.

For the acceptance of hearing devices targeted at normal-hearing users like electronic hearing protectors or hear-through headsets, acoustic transparency and sound quality is arguably even more crucial than in hearing aids. Since the user gains less benefit from wearing these devices than a hearing-impaired person from a hearing aid, they are less prone to accept undesired distortions. The problems to solve are very similar, except that the negative effects of non-linear processing as in hearing aids are usually not included (Hassager et al. 2017; Tikander 2009).

Open-fit hearing aids have been introduced about 15 years ago to solve some of these issues and quickly gained popularity (Kinkel 2016; Mueller 2006; Winkler et al. 2016). Similar approaches exist for augmented reality applications (Martin et al. 2009). Open-fit devices are mechanically designed to not attenuate external sounds reaching the ear, such that the open-ear cues are undisturbed and the hearing device output is an addition to the acoustically perceived environment. However, the open design results in only little control over the sound leaking into the ear canal, and bears further drawbacks that limit the acoustic performance (Blau et al. 2008; Gatehouse 1989; Keidser et al. 2007). Therefore, this work aims to generalize the results for devices ranging from completely blocking the ear canal to completely open fits.

Conserving the external ear acoustics in hearing devices

When acoustic transparency is desired in a hearing device, both directional and non-directional aspects of the HRTF should be conserved. As such, an ideal hearing device could be thought of as an infinitely small device that sits on the eardrum, which receives the exact sound field to conserve, only applies the desired modification to it, and plays the resulting signal to the eardrum at the same moment. It is clear that this scenario is not realistic, and in practice inaccuracies occur at various stages and have to be minimized.

First, with current technology a hearing device cannot capture the signal at the open eardrum, since the microphone is located somewhere in the ear canal, in the pinna, or behind the ear (Dillon 2012). However, an approximation of the signal reaching the eardrum of the open ear has to be explicitly or implicitly known. This approximation is in the following referred to as target, which can be either a signal at, or a transfer function to the eardrum. In hearing aid fitting, an aided response target is usually defined based on a measurement of the transfer function to the open ear. Another possibility to define the target is applying a frequency-dependent transformation to the signal/response observed at the hearing device microphone. This transformation resembles the ear canal resonances that are extinguished by occluding the ear with the hearing device. It was long ago pointed out by Killion (1979; 1980) that the transformation to be applied depends on the device style and microphone position. However, a detailed assessment on how this transformation should be obtained, and how it relates to the individual ear, device style and incidence direction is yet to be made and evaluated. This problem is assessed and evaluated in this thesis on the basis of extensive acoustic measurements in Sec. 2.2, and the perceptual similarity between the resulting target and the open-ear signal evaluated in Sec. 4.3.
Second, with most hearing device styles the microphone does not capture the full directionality of the HRTF, since it is not positioned in the ear canal, and the shape of the pinna is altered by filling. This fact is also known (Durin et al. 2014; Hoffmann et al. 2013b; Kuhn 1979a), however, only a small selection of devices

was considered in one study, and again the differences between individual ears have never been assessed. Furthermore, the perceptual consequences of these errors have never been studied in isolation. The size and nature of these effects are assessed in this thesis based on acoustic measurements in Sec. 2.3, and the impact on sound localization is studied in Sec. 4.1.
Third, even if the open-ear reference signal to conserve is perfectly known, another issue is to actually adjust this device to generate the target at the eardrum. This includes equalizing the transducer responses, which is not always possible with perfect accuracy. A more salient issue is that the sound field at the eardrum is not only determined by the output of the hearing device, but also includes a so-called leakage component that directly enters the ear canal acoustically. The superposition of the typically delayed hearing device output and the leakage can lead to disturbances referred to as comb-filtering effects (Groth and Søndergaard 2004; Stone et al. 2008). While even closed-fit devices cannot suppress a leakage component completely, a vent is used in many modern devices to improve the wearing comfort and the own-voice perception (Kuk et al. 2009; Winkler et al. 2016), making sound equalization more challenging and the occurrence of comb-filtering effects more likely. Finally, even if all transfer functions are known, a procedure for calculating an equalization filter for the hearing device output is required. A variety of approaches exists, yet it is not clear what the optimum approach is, and particularly, how leakage can be included in the filter design. In this thesis, a new algorithm for computing an individualized equalization filter that also reduces comb-filtering effects is proposed in Sec. 3.3, and the impact of a processing delay and the equalization algorithm on sound localization and perceived sound quality are assessed in Secs. 4.1 and 4.3, respectively.

Outline of thesis

In this thesis, general acoustic and psychoacoustic limitations and refinements of acoustic transparency in hearing devices as outlined above are assessed and evaluated. The results of this thesis are also put into practice by means of a real-time demonstrator device. It facilitates an evaluation of the developed approaches in a realistic setting, including interactions between different signal processing stages. The first version of the demonstrator, as described in (Denk et al. 2018c), marks the starting point of this thesis. The demonstrator served as an evaluation platform for newly developed and implemented techniques. The platform was also used to explore the integration of Electro-Encephalography sensors into live hearing devices (Popelka and Moore 2016). The work on the demonstrator also led to the development of an earpiece that is now available as a commercial product for other researchers (Denk et al. 2019a).
The main part of the thesis is presented in Chapters 2–4, each including three sections that are standalone peer-reviewed publications (see List of Publications). Furthermore, the results of the individual studies are discussed comprehensively in Chapter 5. Taken together, a broad set of research questions, experiments and technical implementations have been performed that provide principal insights

and technical solutions on the way to improve the conservation of transmission properties of the external ear in hearing devices of the future.

Chapter 2 describes extensive measurements and the assessment of external ear acoustics related to hearing devices. This also includes refinements of the state-of-the-art measurement techniques. The results reveal general limitations regarding acoustic transparency in dependence of the device style.
In Sec. 2.1, a novel part of the technology for measuring HRTFs in the available laboratory is proposed and evaluated. Specifically, this comprises a new method to window out undesired reflections in measured impulse responses without introducing low-frequency inaccuracies.
In Sec. 2.2, the measurement of a comprehensive database of hearing device HRTFs is described. This data is used to evaluate, how accurate the open-ear reference signal can be approximated based on the signal picked up by microphones at different locations in the ear. To this end, appropriate transformations termed Target Response Correction Function (TRCF) were derived and evaluated, which directly facilitate to define an accurate open-ear target without access to measurements at the eardrum.
In Sec. 2.3, spectral directional cues captured by the individual hearing device microphones are evaluated by means of common metrics and computational models of human sound localization using the data described in Sec. 2.2. These directional cues captured by the hearing device microphone constitute the upper boundary of how well spatial hearing can be conserved with hearing devices.

Chapter 3 describes research and developments on a real-time demonstrator platform with the aim to put an acoustically transparent hearing device into practice and making the underlying technology generally accessible.
In Sec. 3.1, the real-time demonstrator platform representing the state of the art at the starting point of this thesis is reviewed, including descriptions of the hardware and tailored algorithms for individual equalization to provide transparency, acoustic feedback cancellation and electro-acoustic models. Shortcomings of the presented state are discussed based on new evaluation measurements, and potential interactions from integrating the different algorithms are discussed.
In Sec. 3.2, the refinement of the earpiece from (Denk et al. 2018c) is described and evaluated. This includes a transformation from an earmould with removable electronics to a one-size-fits-all device where all transducers are included, while maintaining the custom layout and transducer placement. The revised earpiece has been made openly available.
In Sec. 3.3, a new algorithm to compute the equalization filter for adjusting the overall response of a hearing device to an open-ear target is proposed and evaluated. A least-squares based design is proposed that has a closed-form solution and incorporates a frequency-dependent regularization approach to reduce comb-filtering effects.

Chapter 4 describes a set of experiments that assess the perceptual impact of physical deviations against the acoustics of the open ear. The results reveal principal limitations and determine allowable physical inaccuracies imposed by the auditory system for solving the problem of achieving perceptually convincing

acoustic transparency.
In Sec. 4.1, the impact of different detrimental effects on the ability to localize single sound sources in a free-field environment are assessed. These aspects include the location of the microphone, the bandwidth, the equalization approach and the hearing device delay. By using the HRTFs from Sec. 2.2 and virtual stimuli, it was first possible to separately assess the influences of the different factors.
In Sec. 4.2, an experiment is described where ear-Electro-Encephalography sensors were included into the real-time demonstrator described in Sec. 3.1. On the one hand, the experiment demonstrated the possibilities of integrating such sensors with respect to neuro-controlled hearing devices (Popelka and Moore 2016). On the other hand, ear-EEG methods are tested as an objective tool for assessing hearing-device related cognitive processes.
In Sec. 4.3, a revised version of the real-time demonstrator is evaluated in a subjective listening quality experiment. As compared to the state documented in Sec. 3.1, the demonstrator included a new target response definition based on the results from Sec. 2.2, the sound equalization algorithm as described in Sec. 3.3, and different options for integrating the equalization algorithm with a feedback cancellation algorithm (Schepker et al. 2019a).

Chapter 5 includes a discussion and summary of the overarching results and insights of this thesis. Furthermore, suggestions for future research on questions that arise from this work are made.

2 External ear acoustics related to hearing devices

2.1 Removing reflections in semianechoic impulse responses by frequency-dependent truncation

Outline and context within the thesis

In this section, a novel method to remove undesired reflections from measured acoustic impulse responses is described. Specifically, it facilitates removal of acoustic reflections originating from equipment in otherwise anechoic conditions. Frequency-dependent truncation of the impulse response is proposed with short truncation windows in the high frequencies, while longer truncation windows are sufficient for the low frequencies, since the reflection from small pieces of equipment are inherently high-pass. The longer truncation windows in the low frequencies avoid truncation artefacts that occur when the broadband impulse response is windowed, resulting in an extension of the usable bandwidth down to the cut-off frequency of the loudspeaker, in the present case 60 Hz. The technique is tailored but by no means limited to measurements in the Oldenburg Virtual Reality lab, where most acoustic measurements in this thesis were conducted. It is thus a prerequisite for the measurement and evaluation of the hearing device HRTF data used in several of the following sections (Secs. 2.2, 2.3, 3.2 and 4.1).

This section is a formatted reprint of
F. Denk, B. Kollmeier, and S. Ewert (2018f). "Removing reflections in semianechoic impulse responses by frequency-dependent truncation". *Journal of the Audio Engineering Society* 66(3), pp. 146–153.

Author contributions: FD developed and implemented the approach to frequency dependent truncation, performed the measurements, prepared the figures and wrote the manuscript. SE and BK participated in writing the manuscript.

Abstract

Acoustic reflections in impulse responses can be eliminated by truncation to short ("pseudo-anechoic") observation times that exclude the reflections. However, truncating the response tail distorts the corresponding low-frequency transfer function. When reflections in "semi-anechoic" data originate from moderate-sized objects, e.g., equipment in anechoic chambers, their composition is high-pass. Consequently, truncation only must be performed in the mid-to-high frequencies where the information is contained in a brief time interval, while the impulse response tail is anechoic for the low frequencies and can be retained. We present a frequency-dependent truncation approach that exploits this observation by adaptation of the truncation length in each band, and therefore avoids low-frequency errors while disturbing reflections are windowed out. Among several tested formulations, a novel Short Time Fourier Transform based formulation generated least artifacts while the anechoic impulse response was retrieved well from both simulated and measured semi-anechoic data.

2.1.1 Introduction

Characterization measurements of acoustic systems like loudspeakers or transmission elements are ideally performed in anechoic chambers, providing an isolated environment that is free of disturbing reflections originating from walls, the ceiling and the floor. Nevertheless, even in this optimal setting reflections from equipment like (other) loudspeakers or mounts seem unavoidable in many situations. Such a condition is referred to as semi-anechoic measurement in the following. The common practise to reject reflections is to truncate the impulse response (IR) prior to the arrival of the first acoustic reflection. This *pseudo-anechoic* technique allows measurement of free-field responses above a certain boundary frequency dependent on the truncation length, even in reverberant environments (Müller and Massarani 2001; Rife and Vanderkooy 1989; Struck and Temme 1994). However, errors in the bass regime are introduced due to truncation of the IR tail containing low frequency information if the response is not flat (Benjamin 2004; Fincham 1985). Various approaches to overcome this difficulty have been introduced, mostly relying on replacing the low-frequency response by models or auxiliary near-field data and are only applicable to specific problems (Niedrist 1993; Struck and Temme 1994). Thereby, most authors concentrate on the amplitude response of the system under test, often in connection with loudspeaker characterization in normal rooms. A workaround would be to artificially flatten the low-frequency response and therefore shortening the IR prior to truncation (Benjamin 2004; Fincham 1985) – which is only acceptable if the exact determination of the low-frequency response is not required.

An alternative approach is to vary the truncation length with frequency, i.e., Frequency Dependent Truncation (FDT) of the IR to obtain an enhanced pseudo-anechoic transfer function. FDT is perfectly suited for semi-anechoic measurements: Reflections from moderate-sized objects generally have a high-pass characteristic, since the reflected energy decays for wavelengths larger than the object size (Neubauer and Dragonette 1967; Rayleigh 1896). In consequence, low-frequency bands are approximately anechoic. Since the desired IR segment containing the high frequencies is typically short, a sufficient pseudo-anechoic IR can be obtained by truncation only at high frequencies. Semi-anechoic conditions and therefore application scenarios for FDT are encountered in many current audio applications, such as directivity measurements of acoustic sources and receivers, or characterization of multi-channel loudspeaker arrays used in virtual acoustics.

Karjalainen and Paatero (2001) introduced a mathematical foundation for FDT, and applied it to enhance the Signal-to-Noise Ratio (SNR) in a measured IR by truncating out the noise floor depending on frequency after the IR has faded out. However, elimination of undesired acoustic components was not part of this study. Benjamin (2004) used a similar approach (termed as "multi-taper and multi-resolution analysis") in an attempt to extend pseudo-anechoic measurements in normal rooms to low frequencies. He showed that FDT bears limitations in a reverberant environment, since reflections from walls and room modes cannot be suppressed while at the same time truncation errors are avoided. One limitation of his method is that the variable frequency resolution across the spectrum leads to discontinuities and cannot be applied to retrieve the IR without further assumptions and processing steps. Therefore, his method can only be used to estimate the anechoic amplitude response of a system under test, not processing an impulse response without major extensions. To the authors' best knowledge, FDT has not been utilized to compute pseudo-anechoic IRs from semi-anechoic measurements. No previous work known to the authors evaluated an FDT scheme that is specifically tuned to suppress reflections and allows direct IR reconstruction, nor has an examination of artefacts created with this approach been conducted.

The current paper therefore extends the theory and application of FDT to the elimination of reflections in semi-anechoic IRs. Thereby, we focus on processing the IR waveform without introducing relevant artefacts, instead of estimating the anechoic magnitude frequency response only. We start with a theoretical part where the FDT formulation adapted from Karjalainen and Paatero (2001) is supplemented by a novel formulation based on the Short Time Fourier Transform (STFT). Both formulations of FDT are laid out and evaluated, and the connection to spectral smoothing is discussed. The second part of the paper covers two verification experiments including removing a reflection from simulated data, as well as a practical example. The simulations allow a precise evaluation of the signal processing principles and quantification of generated artefacts, since the anechoic reference IR to be retrieved is exactly known. In a practical loudspeaker IR measurement conducted in an anechoic chamber with multiple equipment installed, the validity of assumptions made regarding the acoustics of semi-anechoic measurements are evaluated. Finally, the procedure of restoring the anechoic IR from the practical semi-anechoic measurement by FDT is validated by comparison to the identical measurement performed in an empty anechoic chamber.

2.1.2 Methods

Several approaches to obtain a pseudo-anechoic IR $\tilde{h}$ from its semi-anechoic observation h are reviewed and presented. All calculations regard discrete-time signals, marked by square brackets around the time index n. Lower case and upper case symbols denote time-domain and frequency-domain quantities, respectively.

2.1.2.1 Broadband Truncation BT

Broadband truncation (BT) of semi-anechoic IR is the standard method to compute a pseudo-anechoic IR and can be written as sample-wise multiplication with a truncation window $w[n]$

$$\tilde{h}_{\mathrm{BT}}[n] = w[n]h[n], \tag{2.1.1}$$

where the maximum in the IR envelope shall be located at the time index $n = 0$. A common shape for $w[n]$ is a modified rectangular window with ramps, e.g. halves of a Hann window (Struck and Temme 1994). Such tempered design avoids temporal discontinuities while most of the IR waveform remains unaltered. The truncation window $w[n]$ is typically chosen to be asymmetric with respect to the peak, i.e. it extends further to positive times than negative times in order to minimize the proportion of noise floor in the utilized IR. In this context, $w[n]$ can be termed a bounded window, i.e. it is zero above a boundary index n_b in the positive time regime

$$w[n] = 0 \quad \forall \quad n \geq n_b, \tag{2.1.2}$$

and correspondingly for negative times. In the frequency domain, the operation in Eq. (2.1.1) equals the convolution of the frequency response with the fourier transform of $w[n]$, which corresponds to a smoothing of the spectrum with a constant absolute bandwidth. Note that for any bounded window $w[n]$, the corresponding fourier transform $W[k]$ is "leaky", i.e., it has non-zero amplitude at indices k larger than the spectral width k_b corresponding to the temporal duration of $w[n]$.

2.1.2.2 Frequency-Dependent Truncation FDT

FDT can be formulated in the time domain by applying truncation windows with variable length in each frequency bin k while performing a Discrete Fourier Transform DFT (c.f. Karjalainen and Paatero 2001), yielding the pseudo-anechoic frequency response

$$\tilde{H}_{\mathrm{FDT-DFT}}[k] = \sum_{n=0}^{K-1} w_k[n]h[n]e^{-\frac{2\pi ink}{K}} \tag{2.1.3}$$

There, the truncation length $n_b[k]$ of each truncation window $w_k[n]$ can be adjusted in each frequency bin k independently to match the anechoic time-frequency portion of the IR. The FDT-processed IR can be retrieved from $\tilde{H}_{FDT-DFT}$ by an inverse DFT

$$\tilde{h}_{\mathrm{FDT-DFT}}[n] = \frac{1}{K} \sum_{k=0}^{K-1} H_{\mathrm{FDT-DFT}}[k]e^{\frac{2\pi ink}{K}}. \tag{2.1.4}$$

Note that the same DFT length $K \geq \max(n_b)$ must be used to achieve spectrally uniform sampling for all frequencies, although the truncation window lengths vary. This means that a varying degree of zero-padding is employed, which potentially causes artefacts due to discontinuities of the frequency response (see also Benjamin 2004, Fig. 16).

2.1.2.3 Short-time Fourier Transform Formulation of FDT

FDT can alternatively be formulated in a Short-Time Fourier Transform (STFT) framework (Allen 1977). Consider the time-frequency representation of the semi-anechoic IR given by its STFT

$$H[k,l] = \sum_{n=0}^{K-1} v_{an}[n]h[n-lL]e^{-\frac{2\pi ink}{K}}. \tag{2.1.5}$$

$v_{an}[n]$ is an analysis window of length K, here equal to the DFT length, l the frame index and L the hop size between frames. Then, FDT can be written as

$$\tilde{H}_{\mathrm{FDT-STFT}}[k,l] = W_k[l]H[k,l], \tag{2.1.6}$$

where $W_k[l]$ denotes a set of bounded windows operating along the frames l, which may be chosen independently for each frequency bin k. The truncated IR is then computed from its time-frequency representation by Weighted Overlapp-Add (WOLA, Allen 1977)

$$\tilde{h}_{\mathrm{FDT-STFT}}[n] = \sum_l v_{syn}[n] \left(\frac{1}{K} \sum_{k=0}^{K-1} \tilde{H}_{\mathrm{FDT-STFT}}[k,l] e^{\frac{2\pi ink}{K}} \right) \tag{2.1.7}$$

with the synthesis window v_{syn}.

The STFT formulation leaves a set of open parameters, most prominently the analysis/synthesis window lengths K and their shapes. Here, a length of 3.75 ms was chosen to achieve sufficient resolution (267 Hz) of low-frequency bands while maintaining moderate temporal discrimination. Both as analysis and synthesis window, a square-root Hann function was employed, which is often used in speech processing (Martin and Cox 1999). The soft fade out in the synthesis process minimizes artefacts, and in frames where no modification is applied perfect reconstruction is achieved when $L \leq \frac{K}{2}$. This allows implementation of $W_k[l]$ as a binary time-frequency mask without introducing considerable errors.

The hop size L determines the temporal sampling of the STFT, as well as the ramp length of narrow-band truncation windows in time domain after reconstruction given by

$$N_{\mathrm{r}} = K(1 - \frac{L}{K}). \tag{2.1.8}$$

$L = K/2$ (=1.675 ms) was chosen to minimize the ramp length, which is still large as compared to common broadband truncation ramps. On the other hand, temporal sampling is limited. To minimize effects of the temporal sampling on truncation, zeros are appended at negative times prior to the IR such that the STFT windows are placed at points in time that are optimally compatible to the desired truncation lengths.

A benefit of the STFT against the DFT formulation is the possibility to adjust the truncation lengths directly based on the time-frequency representation the processing is applied to. Additional support in this task can be gained by joint inspection of the IR and frequency response, or alternative time-frequency representations with variable spectro-temporal resolution.

2.1.2.4 Relation to complex spectral smoothing

Reflections in the IR are noticeable as a ripple in the frequency response due to comb filtering effects. This error can be reduced by replacing the coefficient for each bin by the average over a certain spectral bandwidth, i.e. smoothing the complex spectrum. The main motivation behind this approach is usually to reduce perceptually irrelevant spectral detail, wherefore the averaging range is increased with ascending frequency to match the resolution of the auditory system (Hatziantoniou and Mourjopoulos 2000; Völk et al. 2011).

Mathematically, the operation can be expressed by a convolution of the frequency response with a window of ascending size. In the time domain, for each narrow-band portion of the IR this equals a multiplication with the inverse Fourier Transform of the corresponding averaging window. Spectral smoothing could therefore be tuned in a way that it is equivalent to FDT in the DFT formulation Eq. (2.1.3). However, in most realizations bounded (or asymptotically decreasing) frequency-domain windows are employed, and the corresponding time-domain window is always leaky (or asymptotically decreasing), i.e. there is no n_b above which $w_k[n]$ is zero. Also, the design of smoothing bandwidths is usually conducted in the frequency domain, making it hard to control the temporal effects of the operation. To summarize, spectral smoothing in the common utilization attenuates reflections by an amount dependent on the smoothing bandwidth, whereas the time-domain centred FDT approach aims to window out reflections without a loss of information about the anechoic portion of the IR.

2.1.3 Experiment I: Removing Reflections from a Simulated Semi-Anechoic IR

2.1.3.1 Procedure

All described methods are evaluated by means of their performance in removing a reflection in a simulated semi-anechoic IR. A bandpass IR modelling an ideal loudspeaker was created as an anechoic reference (4-th order butterworth, 65 Hz to 18 kHz, 48 kHz sampling rate). A high-pass reflection was simulated by filtering the anechoic IR with a lower cut-off frequency of 1 kHz (2nd order butterworth) and adding it with a delay of 6 ms and a gain of -10 dB. The high-pass order corresponds to a rigid rectangular plate (Neubauer and Dragonette 1967) while a spherical reflector would act as a 4-th order high pass (Rayleigh 1896); the cut-off frequency and delay correspond to an object with a characteristic size of the cut-off wavelength (30 cm) and 1.8 m additional sound propagation path. The resulting responses together with a time-frequency representation of the semi-anechoic IR are shown in Fig. 2.1.1.

The reflection was then removed from the simulated semi-anechoic IR using BT, FDT, and complex spectral smoothing. The FDT lengths were manually adapted to the time-frequency structure of the reflection and listed in Fig. 2.1.1. For FDT-DFT, the same truncation lengths and frequency resolution as with the FDT-STFT were used. BT was applied with the shortest length of the FDT bands (4.5 ms) and a 24 samples Hann ramp. Complex spectral smoothing (Hatziantoniou and Mourjopoulos 2000) was performed over bandwidths corresponding to the

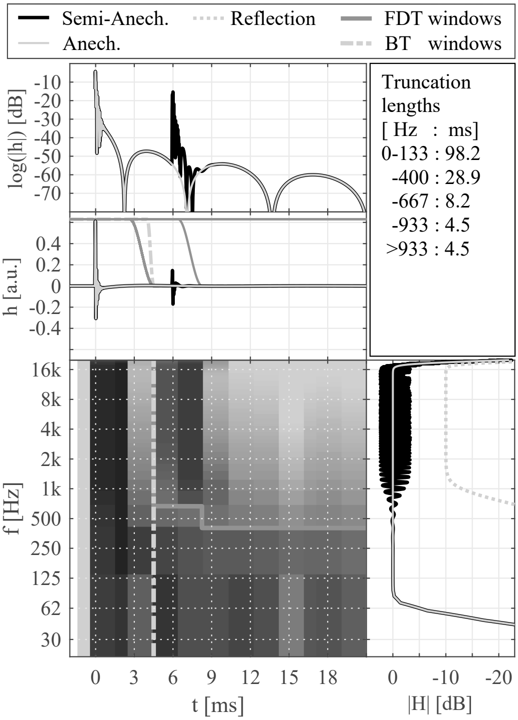


Figure 2.1.1: *Top two panels: Simulated anechoic bandpass IR with and without added high-pass reflection, with logarithmic (upper image) and linear (lower image) amplitude scale. Bottom left: Time-frequency representation of the simulated semianechoic IR, dark shades indicate high amplitude. Bottom right: Corresponding frequency responses of the anechoic IR, the simulated reflection and the resulting semianechoic response. Truncation windows for BT and FDT are plotted in the time-frequency and the linear IR depiction.*

inverse of the FDT truncation lengths in each frequency bin. Spectral analysis of the original and processed IRs was performed with a DFT length of 4096 samples.

In time domain, the deviation of any regarded IR $h[n]$ to the anechoic reference $h_{an}[n]$ can be described by a sample-wise Signal-to-Noise Ratio (SNR) in dB

$$\mathrm{SNR}[n] = 10\log_{10}\frac{h^2_{an}[n]}{(h_{an}[n]-h[n])^2}. \tag{2.1.9}$$

A high SNR indicates good match with the anechoic reference, infinite SNR means both are equal. The frequency domain error is expressed by the difference of responses in magnitude and phase.

2.1.3.2 Results

The IRs and corresponding frequency responses after processing are shown in Fig. 2.1.2 together with the deviation metrics from the anechoic reference in both domains.
With all truncation-based schemes, the reflection is well removed from the raw IR, whereas it is only partly attenuated when using complex spectral smoothing. This translates to a residual high-frequency ripple in the smoothed response (both in amplitude and phase), which is completely cancelled with the truncation approaches. The removal of the reflection can be quantified by the time-domain SNR around

6 ms, where the reflection occurs. As compared to the semi-anechoic data (-20 dB SNR), the SNR is most increased when using either FDT formulation (ca. 20 dB SNR), followed by BT (0 dB SNR) and spectral smoothing (ca. 0 dB SNR). Note that the SNR as computed in Eq. (2.1.9) was smoothed over 1 ms for better display.

Substantial low-frequency deviations are observed in the BT response below 200 Hz, and small ripples are apparent even beyond 1 kHz. In either FDT formulation, this error is substantially reduced, with deviations below 1 dB with the DFT realization and smaller than 0.3 dB with the STFT formulation. A similar observation can be made in the time-domain SNR. Minor artefacts (i.e. finite SNR values) are introduced through the FDT-DFT and spectral smoothing at all times, whereas with the FDT-STFT and BT, the processed and anechoic response only start deviating at times after the truncation starts. In the IR tail, the deviation from the anechoic response is very similar with both FDT versions, although some dips are visible on the FDT-DFT curve.

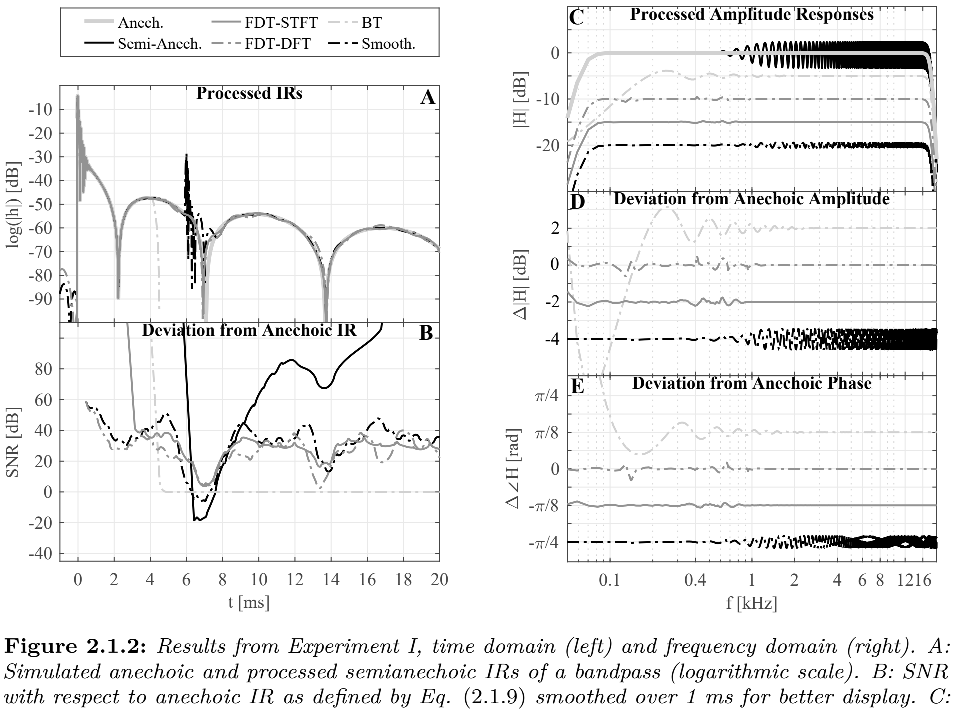


Figure 2.1.2: *Results from Experiment I, time domain (left) and frequency domain (right). A: Simulated anechoic and processed semianechoic IRs of a bandpass (logarithmic scale). B: SNR with respect to anechoic IR as defined by Eq.* (2.1.9) *smoothed over 1 ms for better display. C: Frequency response of the simulated semianechoic bandpass IR before and after processing. D: Amplitude deviation from the anechoic reference. E: Phase deviation from the anechoic reference. Note: Individual curves in C, D, and E were shifted along the y-axis for better display.*

2.1.3.3 Discussion

The FDT approach clearly outperforms standard BT as well as spectral smoothing. For the latter, it may be possible to better reject the response ripple by increasing smoothing bandwidths at high frequencies, however this would come at the cost of a decreased frequency resolution.

Within the FDT formulations, the STFT formulation appears to generate less artefacts than the DFT formulation. This can be accounted to discontinuities in the FDT-DFT spectrum at frequencies where the truncation length changes due to varying zero-padding. The issue may be resolved by additional smoothing of the spectrum, or using asymptotically decaying window functions (Völk et al. 2011), which on the other hand would reduce the performance in rejecting reflections. The possibly better frequency resolution in the DFT formulation would allow for gradual transition between truncation lengths, which potentially reduces the size of discontinuities but increases the number thereof. In conclusion, the results indicate that less artefacts are introduced by FDT when the STFT formulation is used.

2.1.4 Experiment II: Approximating the anechoic loudspeaker IR from semi-anechoic measurements

Acoustic reflections in a semi-anechoic measurement of a loudspeaker-microphone chain were removed using FDT and BT. The aim was to approximate the reference IR of the same loudspeaker-microphone chain that was measured in anechoic conditions.

2.1.4.1 Procedure

The IR of a *Genelec* 8030b loudspeaker was measured on axis at a distance of 1.8 m using a miniature electret microphone (*Knowles* FG-23329-P07, diameter 2.5 mm). The microphone was mounted on a thin wooden stick and a stand wrapped in absorber to have an omni-directional receiver that did not attenuate reflections from any incidence direction. The IR was measured using a 5 s long exponential sweep ranging from 10 Hz to 24 kHz (=half sampling rate of 48 kHz), played back from MATLAB through an *RME* Hammerfall II soundard. A median filter was then applied to the IRs obtained in 20 trials to compute the final IR.

The semi-anechoic IR was obtained in an anechoic chamber (chamber I, boundary frequency 120 Hz, 8.6 x 5.8 x 5.5 m^3, foam wedge absorbers with structure depth 0.6 m) shown in Fig. 2.1.3 with various equipment installed. Instalments included 86 loudspeakers (*Genelec* 8030b) in a 3D layout ca. 2.5 m from an acoustic centre, an acoustically optimized video screen in 180° of the horizontal plane and a grillage floor 1.7 m below the horizontal loudspeaker plane. The loudspeaker under test was positioned in front of the video screen; the microphone was located in the acoustic centre of the loudspeaker array. In this room it is impossible to avoid reflections from the instalments in an IR measurement. However, good approximations of the anechoic loudspeaker responses are required in many applications, e.g. for response inversion.

An anechoic IR serving as a reference was measured with the same equipment on the same day in a different anechoic chamber, which is also shown in Fig. 2.1.3 (chamber II, boundary frequency 50 Hz, 10 x 8.5 x 5.5 m^3, mineral wool absorbers with structure depth 1.5 m). Chamber II was empty except another experiment that was built up ca. 2 m behind the loudspeaker, which did not cause considerable reflections. Great care was taken in reproducing the distance between loudspeaker and microphone to minimize the need for further corrections.

An approximation of the anechoic IR was then computed from the measured response by performing BT and FDT. Prior to this step, minimal timing (shift by half sample, implemented by adding a linear phase) and level (+0.4 dB broadband gain) corrections were applied to the semi-anechoic IR to better match the anechoic IR. Given the results of experiment I, FDT was only utilized in the STFT formulation. The FDT truncation lengths were manually adjusted based on joint inspection of the IR, frequency response and the STFT of the semi-anechoic data, as shown in Fig. 2.1.4.

2.1.4.2 Results

Figure 2.1.4 shows the measured semi-anechoic and anechoic IRs, as well as corresponding frequency responses and a time-frequency depiction of the semi-anechoic IR in the same style as Fig. 2.1.1. Reflections occur in the semi-anechoic IR with various delays and frequency compositions, starting at a bit more than 3 ms after the first peak. The earliest reflection originates from the screen behind the loudspeaker and contains the most relevant energy above 1 kHz. The floor reflection is expected after about 6 ms, where it is clearly visible in the semi-anechoic response with the main energy concentrated around 1-2 kHz. Another temporally smeared reflection with large energy also at frequencies down to ca. 400 Hz occurs at about 13 ms, presumably from opposite loudspeakers and the mounting system of the loudspeaker array.

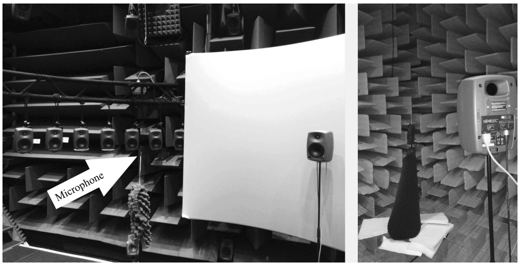


Figure 2.1.3: *Measurement setups in Experiment II. Left: Semianechoic chamber with multiple installments. A miniature electret microphone is mounted on a small stick in the center of a multichannel loudspeaker array. The loudspeaker under test is placed in front of an acoustically optimized video screen (right of image). Right: Setup of the anechoic reference measurement with the same equipment in a different anechoic chamber that was effectively empty.*

Also in the anechoic reference response, some minor reflections are observed. At about 6 ms, a reflection arrives that probably originates from the net floor. At later times, other disturbances are visible that make the IR appear noisy (best seen in logarithmic scale), although a smooth low-frequency oscillation would be expected.

In both measurements, the early impulse response was reproduced with very high accuracy. The IRs only start to considerably deviate after 3 ms, when the first acoustic reflections arrive. The frequency responses are very similar, with the only difference that the semi-anechoic response contains notable ripples that are not observed in the anechoic reference. Below 250 Hz, a small structural deviation between both curves becomes apparent. Most prominently, this comprises a peak at 125 Hz that is included in the semi-anechoic but not in the anechoic response.

Figure 2.1.5 shows the IRs and frequency responses after processing the semi-anechoic measurement. In the time domain, a notable reduction of reflections occurring after the main peak in the semi-anechoic response can be observed. A good correspondence between the FDT-processed IR and the anechoic IR becomes apparent throughout the depicted temporal range, whereas all information exceeding 3 ms delay are excluded in the BT-processed IR. In the frequency response, both BT and FDT of the semi-anechoic IR retrieve the anechoic frequency response almost perfectly for frequencies greater than about 300 Hz. At even lower frequencies, BT causes a large deviation to the anechoic response, whereas this error is significantly

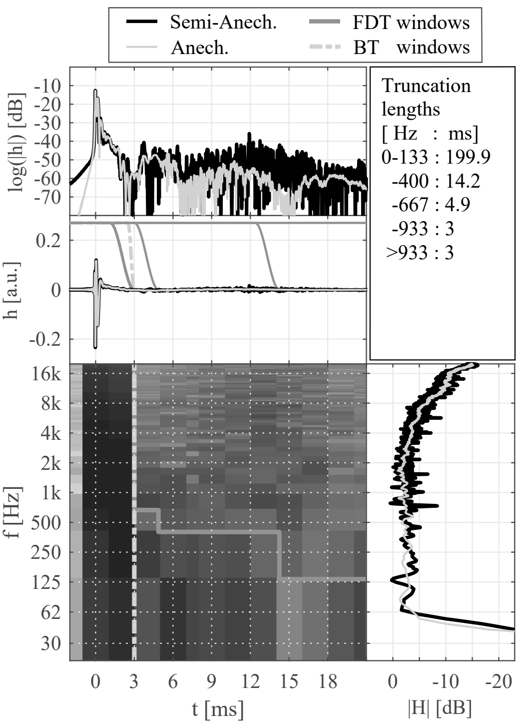


Figure 2.1.4: *Top two panels: Measured semianechoic and anechoic IRs, with logarithmic (upper image) and linear (lower image) amplitude scale. Bottom left: Time-frequency representation of the semianechoic IR, dark shades indicate high amplitude. Bottom right: Corresponding frequency responses of the measured IRs. Truncation windows for BT and FDT are plotted in the time-frequency and the linear IR depiction.*

reduced when using FDT. The FDT-processed response in this frequency region is very similar to the original semi-anechoic response.

2.1.4.3 Discussion

In this realistic application scenario with a significant amount of equipment installed in an anechoic chamber, the assumptions made for the case of a semi-anechoic measurement proved to be valid. Using FDT it was possible to retrieve the anechoic frequency response and IR in a wide frequency range with high accuracy from a semi-anechoic measurement, which is not possible using BT.

Below 300 Hz, differences observed between the semi-anechoic and anechoic frequency responses were not reduced by FDT. In this frequency regime, room modes in the chamber of the semi-anechoic measurement likely start to play a role. Given that the the cut-off frequency of the absorbers is approached and considering the dimensions of the chamber, modes at about 60 Hz and 120 Hz occur, which can be found in the response of the semi-anechoic IR. Similarly, room modes in the chamber where the anechoic measurement has been conducted (wedge length 1.5 m) may start playing a role near the cut-off frequency of the loudspeaker (55 Hz). FDT processing conserves the low-frequency response of the loudspeaker in anechoic chamber I very well, which due to non-negligible room acoustics differs slightly from the response in anechoic chamber II. The influence of such room modes cannot be eluded separately from modifying the measured low-frequency behaviour of the loudspeaker.

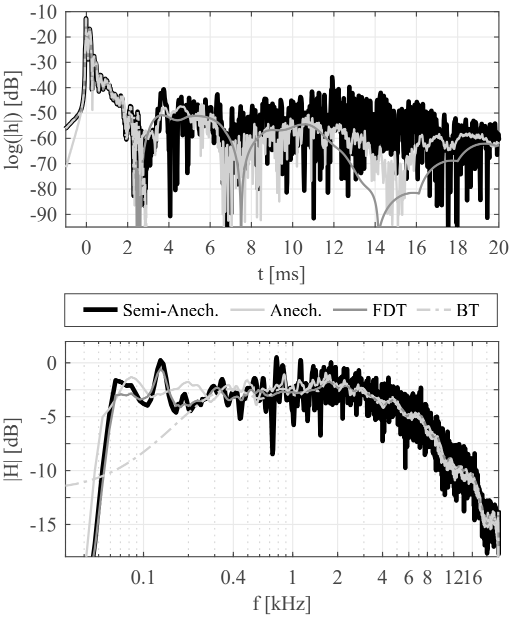


Figure 2.1.5: *Results of Experiment II after applying BT and FDT. Top: IRs from the anechoic and semianechoic measurement before and after processing, logarithmic scale. Bottom: Corresponding frequency responses.*

It was also observed that the anechoic IR is not perfectly free of acoustic reflections and other disturbances. Given the observed performance, FDT is a good alternative to spectral smoothing for enhancing anechoic IRs. Since in practice it is very hard to measure in perfect anechoic conditions, FDT provides a very data-oriented means to reduce imperfections of any kind. This also includes an SNR benefit for frequency bands that can be truncated to much shorter length than the longest IR observation, as already suggested by Karjalainen and Paatero (2001).

2.1.5 Summary and conclusion

To improve pseudo-anechoic measurements in semi-anechoic conditions, we presented a Frequency Dependent Truncation (FDT) approach for removing reflections in impulse responses (IRs). The method is acoustically justified when the reflections only contain relevant energy above a threshold frequency, which is the case if reflections are created by objects of moderate size and the recording room is approximately anechoic. Other than with standard broadband truncation (BT), the low frequency information contained in the impulse response tail can be conserved nearly perfect, while disturbing reflections are rejected with equal efficiency. The promising performance also suggests FDT for enhancing anechoic IRs, providing a more data-oriented alternative to common spectral smoothing. The source code and example data from this work are publicly available (*https://github.com/floriandenk/FDT*).

FDT has been mathematically formulated in conjunction with a DFT with variable truncation window length, as well as in an STFT framework. DFT and STFT formulations were compared in simulations. The applicability and performance of both formulations have been evaluated by processing simulated semi-anechoic data. For both formulations, results showed a clear benefit against standard BT of the impulse response prior to arrival of the first reflection, as well as the related approach of complex spectral smoothing. While both techniques achieved equal performance in removing a high-pass reflection, the STFT formulation produced less artefacts than the DFT realization. Future extensions could include more flexible time-frequency analyses, where the time-frequency resolution can be adjusted separately for each frequency (Karjalainen and Paatero 2001).

Tuning of truncation lengths to the particular data is an important non-trivial task that can be performed well based on the suggested joint inspection of the IR, frequency response, and time-frequency representation of the semi-anechoic data, as shown in Figs. 2.1.1 and 2.1.4. There, another benefit of the STFT formulation over the DFT formulation is the possibility to directly tune the truncation lengths based on the time-frequency representation that is used for processing.

Also in application to measured data, the acoustic assumptions for applicability of FDT proved to be valid. Using FDT, a loudspeaker IR measured in a an anechoic chamber with multiple instalments, the anechoic response measured in a different empty anechoic chamber was retrieved with high accuracy. Only in the bass regime where room modes in both chambers came into play, residual deviations were observed.

We conclude that FDT, particularly using the proposed STFT formulation, is very suitable for processing semi-anechoic measurements to retrieve the anechoic response in a wide frequency range. It thus overcomes problems that occur when standard broadband truncation of the impulse response is utilized. The acoustic conditions that are prerequisite for applying FDT are encountered in various current audio applications, such as multi-channel loudspeaker setups used in virtual acoustics or directivity measurements.

Acknowledgments

This work was funded by DFG through Research Unit 1732 *Individualized Hearing Acoustics*. Jan Heeren is acknowledged for assistance in conducting the measurements and fruitful discussions. The utilized implementations of FDT and the example data are freely available for download (*https://github.com/floriandenk/FDT*).

2.2 Adapting hearing devices to the individual ear acoustics: Database and target response correction functions for various device styles

Outline and context within the thesis

In this section, the external ear acoustics captured by hearing devices is comprehensively assessed and compared to the acoustics of the open ear up to the eardrum. To this end, HRTFs were measured at the eardrum and at microphone locations contained in a comprehensive selection of hearing devices in 16 human subjects and 3 dummy heads for 91 incidence directions. This database was made available to the public and is the basis for evaluation of different aspects of acoustic transparency in hearing devices within this thesis. In this section, relative transfer functions between the hearing device microphone locations and the open eardrum are assessed. The characteristics of these transfer functions were compared between acoustic field conditions, as well as between individual and non-individual acquisition. The evaluation conducted here therefore gives insights on how well the open-ear transmission characteristics can be generally approximated depending on the device style and microphone position. Furthermore, the results yield a target response definition approach for achieving acoustic transparency.

This section is a formatted reprint of
F. Denk, S. M. A. Ernst, S. D. Ewert, and B. Kollmeier (2018b). "Adapting Hearing Devices to the Individual Ear Acoustics: Database and Target Response Correction Functions for Various Device Styles". *Trends in Hearing* 22, p. 2331216518779313.

Author contributions: FD planned, implemented and conducted the measurements, performed the data evaluation, prepared the figures and wrote the manuscript. BK participated in planning the experiments, data evaluation and writing of the manuscript. SMAE participated in the planning of the measurements; SMAE and SDE participated in writing the manuscript.

Abstract

To achieve a natural sound quality when listening through hearing devices, the sound pressure at the eardrum should replicate that of the open ear, modified only by an insertion gain if desired. A target approximating this reference condition can be computed by applying an appropriate correction function to the pressure observed at the device microphone. Such Target Response Correction Functions (TRCF) can be defined based on the directionally dependent relative transfer function between the location of the hearing device microphone and the eardrum of the open ear. However, it is unclear how exactly the TRCF should be derived, and how large the benefit of individual, versus generic, correction is. We present measurements of Head-Related Transfer Functions (HRTF) at the eardrum and at 9 microphone locations of a comprehensive set of 5 hearing device styles, including 91 incidence directions, and recorded in 16 subjects and 2 dummy heads. Based on these HRTFs, individualized and generic TRCF were computed for frontal (referred to as free-field) and diffuse-field sound incidence. Spectral deviations between the computed target and listening with the open ear were evaluated using an auditory model and virtual acoustic scenes. Results indicate that a correction for diffuse-field incidence should be preferred over the free field, and individual correction functions result in notably reduced spectral deviations to open-ear listening, as compared with generic correction functions. These outcomes depend substantially on the specific device style. The HRTF database and derived TRCFs are publicly available.

2.2.1 Introduction

Frequency-response characteristics of hearing devices are ideally designed to replicate the individual transfer function to the eardrum of the open ear. In this case, the sound pressure generated at the eardrum of the subject approximates the pressure that would be present at the open eardrum, modified only by an insertion gain if desired. This approach to make hearing devices acoustic ally transparent has been applied to both hearing aids and consumer devices (Denk et al. 2018c; Dillon 2012; Härmä et al. 2004; Hoffmann et al. 2013a; Killion 1979; Rämö and Välimäki 2012). In the present work, we concentrate solely on the definition of a suitable target that approaches the open-ear reference, which is independent of the challenge of adjusting the hearing device to create that target at the eardrum of an individual subject. The current article provides an extensive database and analyzes the underlying assumptions and possibilities for deriving correction functions that can be employed to define a suitable target and in a comprehensive set of hearing device styles.

The target can either be defined as fixed transfer characteristics of the device as a frequency response, or as a time-varying signal depending on the current input signal. In either case, it is practical to compute the target by transforming the pressure (response) observed at the device's microphone by a frequency-dependent gain. The best possible transformation of the pressure observed at the microphone location of the hearing device to the open eardrum would be the Relative Transfer Function (RTF) between both locations (see Fig. 2.2.1). For most device styles, the microphone location deviates from an ideal location in the ear canal or at its entrance, which makes the RTF dependent on the direction of incidence (Durin et al. 2014; Hammershøi and Møller 1996; Hoffmann et al. 2013b; Killion and Monsor 1980). However, without assuming knowledge of the current sound field, only one direction-independent transfer function can be applied. Typical choices are a transformation that is correct for the frontal incidence direction, referred to as free-field correction, or for a random sound incidence, referred to as diffuse-field correction (Bentler and Pavlovic 1989, 1992; Dillon 2012; Killion 1979). In this

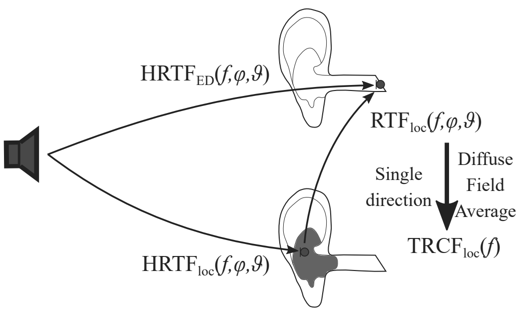


Figure 2.2.1: *Illustration of the transfer functions used. To equalize the head-related transfer function (HRTF) to a hearing device microphone (at location loc) to the eardrum of the open ear (ED), the directionally dependent relative transfer function (RTF) between both locations can be applied. φ and ϑ denote the azimuth and elevation angles, respectively. However, in operation, only a directionally independent correction function can be applied, here referred to as target response correction function (TRCF), which can be defined based on the measured directionally resolved RTFs.*

work, such a transformation will be referred to as the Target Response Correction Function (TRCF). The TRCF is the transformation that corrects the transfer function from the hearing device's microphone to the response to the open eardrum for a given acoustic field and a particular ear. This correction restores acoustic transmission features that are observed at the open eardrum but not at the hearing device microphone location. The TRCF is a generalization of the CORFIG principle (Coupler Response for Flat Insertion Gain; Killion and Monsor 1980), which is the TRCF with an additional correction for the hearing device response measured in a 2cc coupler. In contrast to the TRCF, the CORFIG is only applicable in devices with a fully occluding fit, since hear-through sound components are neglected, or it depends greatly on the coupling to the ear. It is well known that the TRCF depends significantly on the hearing device style and microphone location.

Although the TRCF is individual to each ear, in many applications, generic transfer functions must be utilized. The term generic for TRCFs describes non-individualized corrections that are used for any ear. Whereas in clinical hearing aid fitting the individual TRCF can be measured implicitly) using probe tube microphone techniques (Dillon 2012; Mueller 2001); in many applications, individual measurements at the eardrum are not available. This is the case in self-fit devices, in consumer products, or simply due to cost constraints. Generic transfer functions can be derived from average data of human subjects or dummy head measurements. Bentler and Pavlovic (1989, 1992) have compiled responses from the free and diffuse field to the eardrum and microphone locations of three standard hearing device styles (In-The-Canal, ITC; In-The-Ear, ITE; and Behind-The-Ear, BTE) that were pooled from a large number of separate measurements reported in the literature over several decades (Killion et al. 1987; Kuhn 1979b; Kuhn and Burnett 1977; Madaffari 1974; Shaw 1974, 1980; Shaw and Vaillancourt 1985; Wiener and Ross 1946). More recently, directionally resolved Head-Related Transfer Function (HRTF) measurements on a dummy head with an ear simulator that included a pair of three-channel BTE hearing aids were presented by Kayser et al. (2009). Durin et al. (2014) provided HRTF measurements on a dummy head and five hearing aid styles with high directional resolution, but excluding transfer functions to the eardrum. However, the existing datasets are limited in terms of device styles and microphone positions in the ear and do not capture differences between individual human ears.

We present HRTF measurements for 91 incidence directions to the eardrum and 9 microphone locations in a comprehensive set of 5 hearing device styles, obtained in both ears of 16 human subjects and 2 commercial dummy heads. The database is publicly available.[1] The data allow extensive analysis of (relative) transfer functions, as well as derivation and evaluation of corresponding TRCFs. Different possible ways to compute the TRCF from the RTFs are evaluated, including individual and average data from human subjects, from dummy head measurements, as well as with free- and diffuse-field corrections. Besides descriptive analyses of the transfer functions, the expected spectral distortion when listening through TRCF-corrected hearing device HRTFs compared with the open eardrum HRTF was evaluated by means of a psychoacoustic model for linear spectral distortions (Moore and Tan

2004). Using this approach, we tackled the following research questions relevant to the design and evalu tion of hearing devices:

— How can the features of the TRCFs and their dependence on the hearing device style and microphone location be related to known external ear acoustics?

— Is it more appropriate to apply an equalization to the free or to the diffuse field?

— How large is the difference between individual and average correction functions?

— How well can a dummy head-generated TRCF approximate the desired TRCF for the average human listener or a specific individual? Is it beneficial to employ a structural, instead of an arithmetic average?

— What is the putative perceptual relevance of these differences?

— What is the influence of the hearing device style on the TRCF, its difference between individuals and the best directional weighting of RTFs in defining the TRCF?

The article is structured as follows: In Sec. 2.2.2, the measurement routine as well as the hearing device styles and microphone locations are described. In Sec. 2.2.3, the further processing of the HRTF data (as published), the RTF extraction, and different possible ways to compute the TRCFs are described. In addition, the method of evaluating the spectral distortion after correction using the different TRCF definitions is described. Sec. 2.2.4 shows the measured transfer functions and computed TRCFs, as well as results of the TRCF evaluation. The results are comprehensively interpreted in Sec. 2.2.5 and the outcomes are summarized in the Sec. 2.2.6.

2.2.2 HRTF measurements

HRTFs for 91 directions were measured in both ears of 16 human subjects (10 male, 6 female, age 27.3±5.1) and 2 dummy heads (*Brüel&Kjær*HATS type 4128C and *G.R.A.S.* KEMAR type 45BM). All experiments were conducted according to the World Medical Association declaration of Helsinki. The subjects were provided with written information material and gave written consent about participation. The dataset contains transfer functions to the eardrum and the microphone locations of a comprehensive set of five hearing device styles. The dataset is publicly available.[1]

2.2.2.1 Pressure at the eardrum

The pressure at the eardrum of the open ear was measured by inserting an audiological probe tube micro phone until the subject reported contact with the tympanic membrane, and then pulled back by a minimal amount (see Fig. 2.2.2, panel Eardrum). This procedure (performed by a trained hearing aid acoustician) provided reliable probe tube positioning close to the eardrum, thus minimizing

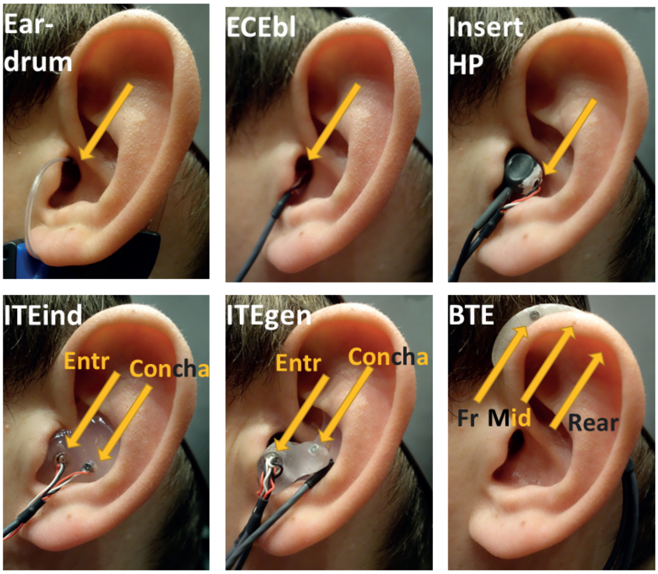


Figure 2.2.2: *Photograph of all hearing devices and microphone locations in the ear of a subject. Arrows mark the individual microphone locations. Note that two of the sound inlets of the behind-the-ear (BTE) device are behind the pinna. See text for further details, the abbreviations are listed in Tab. 2.2.1.*

errors due to standing waves in the ear canal in the frequency range of interest (Hellstrom and Axelsson 1993; Mueller 2001).

2.2.2.2 Hearing device styles and microphone locations

Wide frequency range miniature electret microphones *Knowles* FG-23329 and GA-38) were used in all hearing devices. The electret microphones were connected to a custom supply and amplifier box providing the operating voltage and 20 dB gain. The number of microphones used was minimized by removable insertion into the different devices whenever possible. All hearing devices with annotated microphone locations are shown in Fig. 2.2.2. Explanations of the abbreviations for the individual microphone locations are provided in Tab. 2.2.1.

The pressure at the blocked ear canal entrance ECEbl) was measured with a miniature microphone inserted flush into anthropometric earplugs available in three sizes (Lindau and Brinkmann 2012), which provide firm and reproducible fit in the ear canal. In a hearing systems context, the blocked ear canal entrance can be regarded as mostly equivalent to small hearing devices fitted into the ear canal of a subject, such as ITC, Completely-In-Canal, or even smaller hearing aids (Bentler and Pavlovic 1989; Durin et al. 2014).

Another microphone location was on a small insert headphone (InsertHP, *Sennheiser* CX200), as often used in augmented reality audio applications (Härmä et al. 2004;

Table 2.2.1: *Glossary of Abbreviations.*

HRTF	Head-related transfer function
HRIR	Head-related impulse response
DF	Diffuse field
FF	Free field, denoting frontal sound incidence
RTF	Relative transfer function
TRCF	Target response correction function
iDF	Individual, diffuse field
iFF	Individual, free field
mDF	Arithmetic subject average, diffuse field
mFF	Arithmetic subject average, free field
smDF	Structural subject average, diffuse field
dhDF	Arithmetic dummy head average, diffuse field
Hearing device styles/microphone locations	
ECEbl	Blocked ear canal entrance
InsertHP	Insert headphone with attached external microphone
ITEind, ITEgen	In-the-ear device, *individual* or *generic (non-individualized form)* earmold
Entr, Concha	Entrance and Concha microphone
BTE	Behind-the-ear hearing device
fr, mid, rear	*Frontal, Middle*, and *Rear* microphone

Hoffmann et al. 2013a; Rämö and Välimäki 2012). A minimal portion of flexible material was attached to the surface, and a miniature microphone inserted flush into a drilled hole. Depending on the subject's ear size, the headphone filled up to half of the cavum conchae. The microphone was placed near the bottom of the concha and pointed toward the rear concha wall (see Fig. 2.2.2). To realize an ITE-type hearing instrument, two microphones were inserted flush into an individual earmold that completely filled the bottom of the concha, one near the ear canal entrance (*Entrance Microphone*, Entr) and one in the rear part of the cavum conchae (*Concha Microphone*). Entrance and concha microphone were approximately 8 to 12 mm apart in a—preferably horizontal—orientation in the individual ears (the distances for the individual subjects are provided with the database). The hardware configuration, referred to as individual ITE device (ITEind), is equal to the outer microphones of the prototype hearing device presented by Denk et al. (2018c).

In a generic ITE device (ITEgen), the microphones were placed in a nonindividualized earplug with a layout comparable to ITEind. A custom adaptor piece was produced that was the same for all subjects and accommodated the transducers that fit into a generic headphone earplug with concha hook (*Bose* StayHear+, one of three sizes selected for each subject). The microphones were 1.1 cm apart and protruded further from the ear than the ITEind earpiece, and the cavum conchae

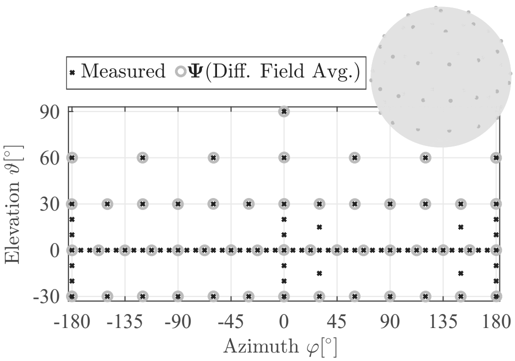


Figure 2.2.3: *Sound source positions in HRTF measurements in navigational coordinates: All positions measured (black crosses) and those used for diffuse-field averaging* $\mathbf{\Psi}$ *(blue circles). Inset: Sound source positions used for diffuse-field averaging* $\mathbf{\Psi}$ *projected on a sphere.*

was less uniformly filled. The ITEgen earpiece can also be understood as external microphones contained in a larger insert headphone.

A BTE dummy device with three microphones was produced based on a 3D scan of a commercial hearing aid (the same as used by Kayser et al. 2009). Miniature electret microphones were then placed at the locations of the original sound inlets.

2.2.2.3 Procedure

The measurements were conducted in an anechoic chamber featuring a 91-channel 3D loudspeaker setup; 48 speakers were located in the horizontal plane leading to a spatial resolution of 7.5°. The rest of the sphere was sampled with a spacing of approximately 30°, but with extra speakers in the median plane and another cone of confusion located at a 30° lateral angle. The spatial sampling grid is shown in Fig. 2.2.3. Two-way loudspeakers *Genelec* 8030b/8020b) were mounted upright between 2.5 and 3 m from the acoustic center. The effects of having separate sound sources for low- and high-frequency reproduction (spacing ca. 1.3°) as well as of the varying distance on the HRTF can be neglected (Brungart and Rabinowitz 1999). In this configuration, it is possible to measure the transfer functions from all different directions simultaneously using overlapping exponential sweeps (Majdak et al. 2007). The individual sweeps covered the frequency range between 50 and 20000 Hz with a duration of 4.1 s, leading to a total duration of 36 s for measuring the transfer functions from all loudspeakers. The recordings were made with a sampling rate of 48 kHz. Both the order of directions in each measured HRTF set, as well as the order of device styles for each subject, were randomized. The measurements included one further device style that is included in the public database[1] but not regarded in this work due to close similarities to ITEind. For each hearing device, the measurement was repeated four times without reinsertion of the device. Altogether, the experiment lasted between 60 and 90 min for each subject.

Assessing HRTFs at various points in the ear requires repeated measurements and exchanging the devices between measurements. A particular source of error in this situation is movement of the subject, which would result in HRTF deviations due to an effective shift of the sound incidence direction (Hirahara et al. 2010). To control this source of inaccuracy, a small headrest in combination with an interactive positioning method was employed (Denk et al. 2017a). The head position was monitored using a headtracker, and necessary corrections to restore a reference position and orientation were displayed to the subject on a graphical interface. This allowed the subject to correct and stabilize their head position with an accuracy of better than 0.5° source shift throughout the experiment, which eliminated the bias caused by variable head orientation. To further reduce the positioning errors, the trial with the best head position was selected for further evaluation, independently for each incident direction. The subjects and dummy heads were initially positioned using a pendulum marking the acoustic room center and a laser distance measurement device. The dummy heads were rotated such that the broadband interaural time difference in the HRTFs to the eardrum for frontal incidence was less than 20 μs (=1 sample @ 48 kHz).

2.2.2.4 Data processing

The raw impulse responses contained reflections from equipment, for example, the grating platform the subjects were seated on or the loudspeaker system. These distortions were removed from the data using frequency-dependent truncation (Denk et al. 2018f). The impulse response was truncated to 4 ms length for frequencies above 1 kHz, but, to avoid truncation errors, not truncated in lower frequency bins, where the reflections did not contain significant energy.

Spectral colorations introduced by the electroacoustic measurement system were compensated by regularized spectral division of the raw frequency responses by the free-field response of every individual loudspeaker (measured with a *Brüel&Kjær* type 4189 microphone), and the individually determined microphone sensitivities. In frequencies exceeding the lower boundary of the measurement (<60 Hz), the responses were extrapolated. To counteract temperature-dependent sensitivity changes of the electret microphones, a broadband gain was applied to each set of HRTFs, to adjust the directionally averaged low-frequency response (average below 150 Hz) to the expected 0 dB. Finally, the resulting impulse responses were truncated to a length of 256 samples, including 16 samples Hann window ramps.

2.2.3 Analysis methods

2.2.3.1 Preprocessing and incidence directions

For further analysis, HRTFs were calculated from the stored impulse responses with a spectral sampling of 5.9 Hz (8128-point fast fourier transform at 48 kHz sampling rate). Perceptually irrelevant spectral detail was removed by applying complex smoothing of the spectral power and phase separately, with averaging windows shaped like the responses of 4th-order gammatone filters with 1 ERB bandwidth (Breebart and Kohlrausch 2001).

A directionally balanced subset $\mathbf{\Psi}$ of the measured directions was considered for further evaluation, as indicated in Fig. 2.2.3. The spacing is 22.5° in the horizontal plane (elevation 0°), 30° in planes with an elevation of 30°, 60° in the plane with 60° elevation, and there is one direction at an elevation of 90° (i.e., directly above). This set of N_Ψ incidence directions is approximately equally spaced on the sphere (for elevations $\vartheta \geq$ -30°, see Fig. 2.2.3), and an average of transfer functions across these incidence directions can be regarded as an approximation of the corresponding diffuse-field transfer function. Since in the diffuse field, all incidence directions superimpose incoherently, the directional averaging operation was performed on the spectral power values. Phase coefficients were unwrapped and averaged independently.

2.2.3.2 Relative transfer functions

RTF between any given microphone location loc and the eardrum ED in the Free Field (FF) were computed for each subject X and incidence direction separately by complex division of the appropriate HRTFs

$$\mathrm{RTF}_{\mathrm{loc}}^{(\mathrm{FF},X)}(f,\varphi,\vartheta) = \frac{\mathrm{HRTF}_{\mathrm{ED}}^{(X)}(f,\varphi,\vartheta)}{\mathrm{HRTF}_{\mathrm{loc}}^{(X)}(f,\varphi,\vartheta)}. \tag{2.2.1}$$

As given in the formula, the RTF is, by definition, dependent on the incidence direction; φ and ϑ denote the azimuth and elevation angles, respectively. In the Diffuse Field (DF), the RTF magnitude is given by the quotient of the diffuse field-to-ear transfer functions, approximated by directionally averaged spectral densities

$$\mathrm{RTF}_{\mathrm{loc}}^{(\mathrm{DF},X)}(f,\varphi,\vartheta) = \frac{\sqrt{\frac{1}{N_\Psi}\sum_{\mathbf{\Psi}}\left|\mathrm{HRTF}_{\mathrm{ED}}^{(X)}(f,\varphi,\vartheta)\right|^2}}{\sqrt{\frac{1}{N_\Psi}\sum_{\mathbf{\Psi}}\left|\mathrm{HRTF}_{\mathrm{loc}}^{(X)}(f,\varphi,\vartheta)\right|^2}} \tag{2.2.2}$$

2.2.3.3 Target response correction functions

The RTF between a hearing device microphone location and the eardrum is the optimal correction function for each incidence direction that must be applied to the microphone signal to obtain the current signal at the eardrum of the open ear. However, despite potential spatial variability of the RTF, only one correction function can be applied to the input signal. In the following, this correction function is referred to as the TRCF. We define and evaluate different realizations, based on the observed RTFs using individual and averaged data in the human subjects, dummy head data, and different sound-incidence conditions. By averaging across subjects, a transfer function is desired that is correct on average on the dB scale as a rough estimate of the perception. Therefore, TRCF averaging across subjects is conducted in dB values. Note that TRCFs are expressed here by their magnitudes in dB only. The abbreviations for the individual TRCF definitions are summarized in Tab. 2.2.1.

— Individual Free-Field correction, iFF: Utilizing the RTF for a specific microphone location loc for the frontal incident direction in a specific subject's ear X. This is equal to the difference between free-field-to-eardrum and free-field-to-microphone location responses

$$\mathrm{TRCF}_{\mathrm{loc}}^{(\mathrm{iFF},X)}(f) = 20\log_{10}\left|\mathrm{RTF}_{\mathrm{loc}}^{(\mathrm{FF},X)}(f,\varphi=0,\vartheta=0)\right|. \tag{2.2.3}$$

— Individual Diffuse-Field correction, iDF: Utilizing the RTF for a specific microphone location loc for diffuse field incidence in a specific subject's ear X. This is equal to the difference between diffuse-field-to-eardrum and diffuse-field-to-microphone location responses

$$\mathrm{TRCF}_{\mathrm{loc}}^{(\mathrm{iDF},X)}(f) = 20\log_{10}\left|\mathrm{RTF}_{\mathrm{loc}}^{(\mathrm{DF},X)}(f)\right|. \tag{2.2.4}$$

— Mean Free-Field correction, mFF: Mean of iFF observed in all N_X human subjects $\mathbb{X}$. This is the free-field correction that is correct for the average of all human subjects

$$\mathrm{TRCF}_{\mathrm{loc}}^{(\mathrm{mFF})}(f) = \frac{1}{N_\Psi}\sum_{\mathbb{X}}\mathrm{TRCF}_{\mathrm{loc}}^{(\mathrm{iFF},X)}(f). \tag{2.2.5}$$

— Mean Diffuse-Field correction, mDF: Mean of iDF observed in all N_X human subjects $\mathbb{X}$. This is the diffuse-field correction that is correct for the average of all human subjects

$$\mathrm{TRCF}_{\mathrm{loc}}^{(\mathrm{mDF})}(f) = \frac{1}{N_\Psi}\sum_{\mathbb{X}}\mathrm{TRCF}_{\mathrm{loc}}^{(\mathrm{iDF},X)}(f). \tag{2.2.6}$$

— Structural mean Diffuse-Field correction, smDF: Structural mean of the iDF observed in all subjects. When individual transfer functions are averaged, the resonances tend to be smoothed out, due to variable peak frequencies across subjects. This can be avoided by finding structural correlates out of which the transfer functions can be computed (ear size etc.), averaging the structural parameters across subjects and computing a structural average (Genuit 1984; Hammershøi and Møller 1996; Mehrgardt and Mellert 1977). As a simple model, two second-order parametric bandpass filters with three parameters each (resonance frequency, resonance gain, and Q-factor), implemented as in (Orfanidis 1997) representing the cavity resonances in the ear canal and the cavum conchae were fitted to each iDF. The resonance frequencies were, however, constrained to sensible boundaries. The structural average transfer function was then computed by averaging the filter parameters across subjects and calculating the corresponding response.

— Dummy-head, Diffuse-Field correction, dhDF: As mDF, but averaged over all dummy head ears, here the KEMAR and *Brüel&Kjær*HATS.

2.2.3.4 Evaluation of spectral distortion in corrected HRTFs

Linear spectral distortions compared between listening through the usual HRTF to the eardrum and the hearing device HRTF corrected by a TRCF was evaluated using the model of Moore and Tan (2004). The model calculates the excitation patterns on the basilar membrane and predicts the perceived spectral distortion between a reference and test stimulus (quantified by the metric $D \in [0, 5]$) by evaluating the difference between excitation patterns and the standard deviation along the auditory filters (i.e., spectral ripple). This evaluation was carried out for each ear of the 16 human subjects.

Seven acoustic scenes as introduced by Grimm et al. (2016) were used as stimuli. They are spatially rendered speech-in-noise scenes with diverse spectral and spatial energy distribution, and reflect a comprehensive selection of acoustic communication scenarios. The segment between seconds 5 and 10 in each scene was evaluated by the model, with the sound pressure levels adjusted to 75 dB SPL. The same sound files as used by Grimm et al. (2016) were reused, which were rendered for playback on 48 loudspeakers in the horizontal plane. Stimuli referenced to the eardrum were created by convolving the loudspeaker signals with the appropriate Head-Related Impulse Responses (HRIRs). Spectral distortion was evaluated for two cases: For a stimulus with a full audio bandwidth and for a condition in which the acoustic scenes were low-pass filtered at 4 kHz. As a result, only distortions in this low-pass frequency band were captured. To create the test HRIRs, the smoothed HRTFs to the eardrum and the hearing device microphones were converted back to the time domain and truncated to a length of 400 samples (without loss of information). The TRCFs were then applied by convolving the 200 first samples of their corresponding minimum-phase impulse responses with the HRIRs to the hearing device microphones. Note that during the calculation of the subject-average TRCFs (mFr, mDF, and smDF) used in this evaluation, both ears of a given subject were excluded. The reference HRIRs to the eardrum were created by convolving the hearing device HRIR with the minimum-phase representation of the directionally resolved relative transfer function to the eardrum. This was to circumvent a possible bias in the results through the minimum-phase approximation of the TRCF.

In addition to the perceptive spectral distortion metric D of Moore and Tan (2004), a simpler physical measure based on the spectral difference between the eardrum HRTF- and TRCF-corrected hearing device HRTF was calculated for each condition. The root-mean-square average of the difference spectrum between both HRTFs was computed based on dB magnitudes evaluated in 1/2-ERB spaced auditory filter bands starting at 200 Hz. The RMS was then averaged over all 48 horizontal incidence directions used in the virtual acoustic scenes. The outcome is an easily interpretable spectral difference in dB, which can be compared with the spectral distortion metric for interpretation of the resulting values.

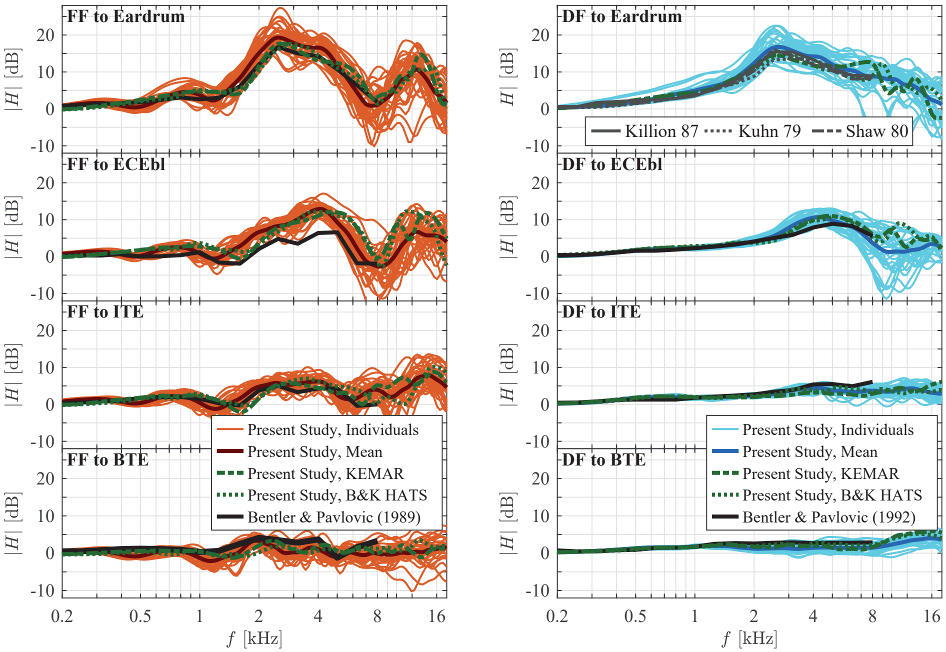


Figure 2.2.4: *Comparison of free (FF, red curves) and diffuse field (DF, blue curves) to ear transfer functions to literature values (black curves, taken from Bentler and Pavlovic 1989, 1992). ECEbl denotes the blocked Ear Canal Entrance, which is compared with the In-The-Canal (ITC) data of Bentler and Pavlovic (1989, 1992), ITE is the Entrance microphone of the individual In-The-Ear device (ITEind_Entr), and for the Behind-The-Ear (BTE) location, the middle microphone (BTE_mid) was selected.*

2.2.4 Results

2.2.4.1 Free and diffuse field to ear transfer functions

Figure 2.2.4 shows transfer functions from the free field and diffuse field to various points in the ear, including the individual curves for the human subjects, as well as their average. For comparison, comparable data compiled from previous studies (Bentler and Pavlovic 1989, 1992) are shown.

The current free-field transfer functions are generally in good agreement with literature data. For the eardrum and blocked ear canal entrance (ECEbl), the current data show a systematically higher amplitude in the region of the main resonance (up to 1.5 dB at the eardrum, up to 5 dB at the ECEbl), but the shapes of the curves are very comparable. The free field to ITE and BTE transfer functions from the current study are in excellent agreement with the curves given by Bentler and Pavlovic (1989). At all microphone locations, a slight spectral ripple below 1 kHz is visible in the current data, but not in the curves given by Bentler and Pavlovic (1989). As further assessed in Sec. 2.2.5, the ripple most probably originates from a reflection from the legs of the subjects.

The diffuse-field transfer functions approximations obtained in the current study are in excellent agreement with the diffuse-field curves provided by Bentler and Pavlovic (1992), which were all measured in reverberation chambers. The current diffuse-field-to-eardrum transfer functions are closer to the values reported by Killion et al. (1987) and Shaw (1980) than to those of Kuhn (1977). At the locations ECEbl and ITE, slight systematic differences (< 2 dB) to published values are seen in the higher frequencies.

The difference between subjects is generally larger in the free-field transfer functions. Also, a reduction of the interindividual variations with increasing distance from the eardrum is noted, which is more pronounced in the diffuse-field transfer functions.

2.2.4.2 Relative transfer functions

RTFs for all incidence directions from all microphone locations to the eardrum in the left ears of two exemplary subjects are depicted in Fig. 2.2.5. For the eardrum, the HRTF for all incidence directions is shown. VP_E1 is a man with comparatively large ears, whereas VP_N6 is a woman with one of the smallest ears of all subjects included in this study. The connection of the RTF data to acoustic transmission mechanisms in the external ear is addressed in detail in the Sec. 2.2.5.2.

At the ECEbl, the difference between RTFs observed at varying incidence directions is small. The RTFs at this location have one common resonance near 2.2 to 3 kHz and further characteristics in the high frequencies.

With increasing distance from the ECEbl, the difference between incidence directions becomes larger—the directional information at all other locations is biased (starting at frequencies >2-4 kHz, depending on the location). Furthermore, the main resonance increases in level and bandwidth, particularly in the frequencies above the peak. A very slight increase of the peak frequency can also be seen. At the BTE microphone locations, the average RTF approaches the average of the eardrum HRTF.

Besides the shape of the main resonance, the RTFs from the different microphone locations also differ in other aspects. Further peaks appear at higher frequencies, being most pronounced in the ECEbl, InsertHP, and ITEgen. Except for the ECEbl, those structures depend greatly on the direction of incidence, and are therefore differently represented in the diffuse-field and the free-field RTF. Between the two curves, the most prominent difference is a dip in the free-field RTF that is not observed in the diffuse-field RTF—and visible in all RTFs except at the ECEbl. Also, a spectral ripple at low frequencies (<1 kHz) is noted for the free-field RTF, but not in the diffuse field. For both subjects, this ripple is clearly visible in the HRTF to the eardrum and attenuated in the RTF of only some locations, which are not coherent across the subjects. Generally, this feature is more apparent in the RTF data of subject VP_N6.

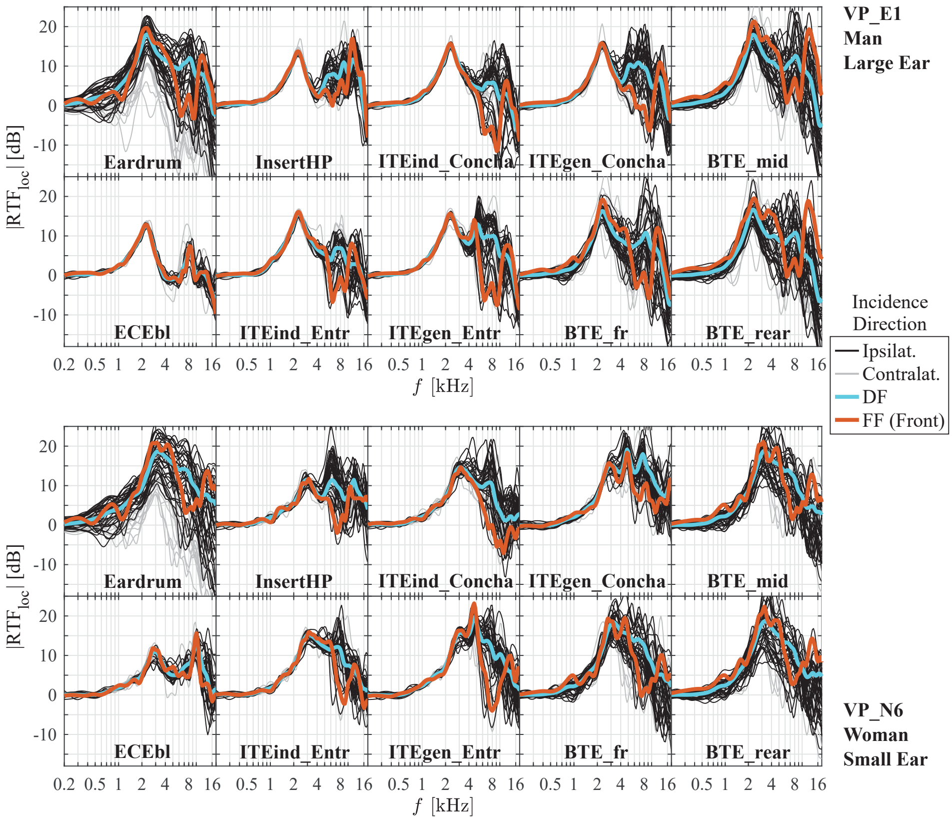


Figure 2.2.5: *Relative transfer functions (RTF) between all hearing device microphone locations and the eardrum of the open ear for all incidence directions, as well as free-field (FF, i.e., frontal incidence) and diffuse-field (DF) incidence, for two individual subjects. For the eardrum, the corresponding HRTF is shown. VP_E1 is a man with large ears and VP_N6 a woman with small ears.*

In the HRTF to the eardrum, a systematic level difference is observed between ipsi- and contralateral incidence directions that is not present in the RTFs. Only at the BTE locations, a slight variation becomes apparent for some directions, particularly toward higher frequencies.

Clear differences between the two subjects are apparent. The most obvious difference between both RTF sets is a shift in the frequency of the first resonance. Also, relative amplitudes of features at higher frequencies, as well as the spread between different incidence directions, vary notably between the ears shown.

Differences are observed between the RTFs for different microphone locations in the same device. This is particularly true for the free-field RTF (and other individual directions) but also for the diffuse-field average. The difference between the three BTE microphones is rather small, whereas it is largest between the microphones on the ITEgen.

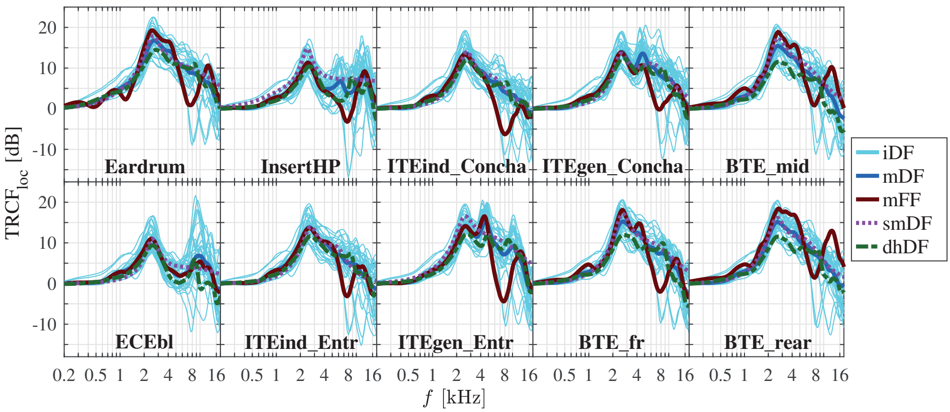


Figure 2.2.6: *Target response transfer function (TRCF) for each hearing device microphone location, showing the data of individual subjects (only diffuse field, iDF), together with the generic curves obtained from averages of subject and dummy head data. For the eardrum, free or diffuse-field transfer function is shown. See Tab. 2.2.1 for an explanation of the abbreviations.*

2.2.4.3 Target response correction functions

Figure 2.2.6 shows the TRCF derived from the RTF as defined by Eqs. (2.2.3) to (2.2.6) and listed in Tab. 2.2.1. The depiction includes all iDF TRCFs, as well as averages over subjects or dummy heads. Again, prominent variability between subjects is apparent in the iDF curves. The main resonance varies in a range of almost 2 kHz in frequency, as well as about 10 dB in level. At the higher frequencies, differences as large as 20 to 30 dB are observed between the individual iDF-TRCFs. The average Free-Field correction mFF is well in line with the DF curves up to about the first resonance, but clearly deviates from them above ca. 6 kHz. Around 7 kHz, the mFF-TRCF (except at the ECEbl) includes a spectral dip that is not observed in the diffuse-field curves. Only at the BTE microphones, the main resonance is higher in amplitude by a few dB in the mFF than in the mDF correction function. Whereas at most microphone locations, one prominent (broadened) peak is noted in the mid frequency range, and two separate peaks are observed for both ITEgen microphone locations in almost all TRCF versions.

The structurally averaged correction function (smDF) differs from the arithmetic average between subjects mDF), showing a main resonance that is slightly higher in level (1 to 4 dB, depending on the microphone location) and a generally smaller amount of spectral detail as a consequence of the two-resonator model used to calculate the structural average. The diffuse- field correction observed in the dummy heads (dhDF) is generally similar to the average of human subjects, but is typically lower in level (2 to 5 dB), especially around the main resonance. This level difference increases in the TRCFs with increasing distance from the eardrum.

In Fig. 2.2.7, all TRCFs are shown for subject VP_E1 and the ITEind_Entr location. Discrepancies between individual and average curves, as well as between FF and DF correction functions become apparent. First, the frequency of the first

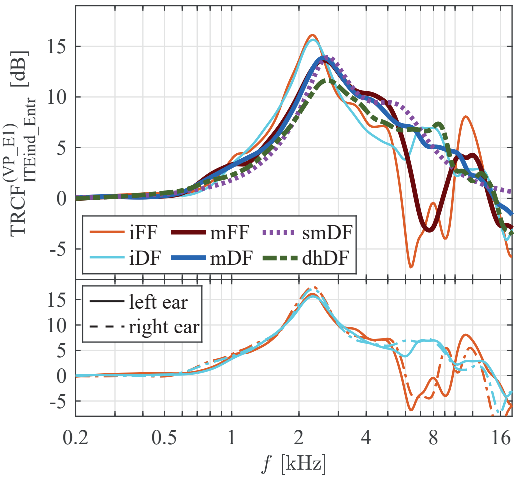


Figure 2.2.7: *Top Panel: Individual and generic target response transfer functions (TRCF) for one sample ear (left ear of subject VP_E1, as in top panel of Fig. 2.2.5). Bottom Panel: Individual TRCFs for both ears of the same subject, shown for free- (FF) and diffuse-field (DF) incidence. The line colors in the lower panel are specified in the legend of the top panel. See Tab. 2.2.1 for an explanation of the abbreviations.*

resonance is different in the individual (indicated by an i) and generic TRCFs—it is lower in this individual ear than for the average (mFF, mDF, smDF) and dummy head (dhDF) curves. Therefore, in this subject, using a generic response from a dummy head or a subject average) would introduce a clear error due to the shifted resonance. Second, in the dummy head and human-average curves, the main peak is broader than in the individual data. Compared with this subject, the structural average (smDF) better conserves the shape and bandwidth. Up to about 4 kHz, individual free- and diffuse-field TRCFs are virtually identical. At higher frequencies, more pronounced features appear in the free-field TRCF, most prominently a dip around 7 kHz. This and other directionally dependent cues are averaged out in the iDF-TRCF, which at the high frequencies (> 4 kHz) is rather similar to the averaged responses.

The individual TRCFs (iFF and iDF) are very similar in both ears, as shown in the lower panel of Fig. 2.2.7. An offset at the main resonance frequency of 2 dB between the sides is observed in both ears, and very similarly in the iFF and iDF correction function. At the higher frequencies, the correction functions between ears are very similar in structure, but deviate from each other in the fine details. The difference between ears is larger for the iFF TRCF. Similar observations were made for the other human subjects.

Figure 2.2.8 displays the distribution of characteristic parameters of the main resonance in the TRCFs. Distributions of the resonance frequency, the gain, and the bandwidth (± 3 dB around the maximum) are shown for the iDF TRCF in comparison to the appropriate parameters of the average transfer functions. The increase in gain, bandwidth, and frequency with increasing distance to the eardrum that was already noted in Fig. 2.2.6 is now clearly visible in the distribution of individual data and to a lesser extent in the generic curves as well. High interindividual variances (spread of iDF data), particularly for resonance frequency and gain, are observed in both ITEgen microphone locations, which are connected to the observed double peak in the correction functions (cf. Fig. 2.2.6). Otherwise,

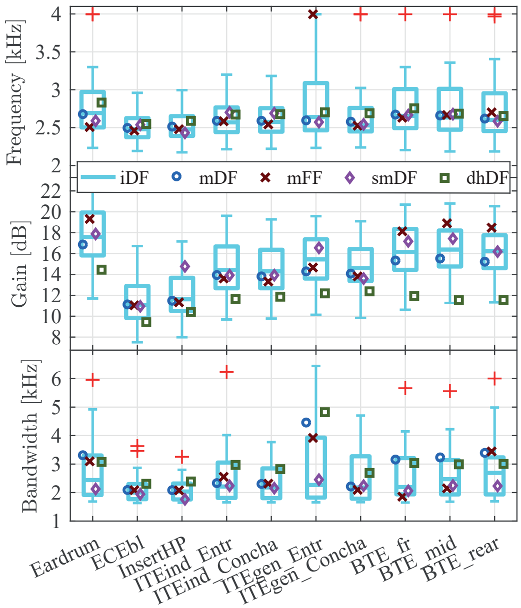


Figure 2.2.8: *Structural parameters of the TRCF main resonance for each hearing device microphone location, distribution over individual diffuse field (iDF) TRCF and results obtained with the generic curves. For the eardrum, the feature distribution of the diffuse field transfer function as shown in Fig. 2.2.4 is depicted. See Tab. 2.2.1 for an explanation of the abbreviations.*

the interindividual differences in the parameters increase with increasing distance to the eardrum.

The median resonance frequency seen in the iDF TRCF data is well conserved by the generic average functions, except for the average of dummy heads, which is systematically higher. The smallest deviation is observed in the TRCF of the ECEbl. The median gain at the peak is also generally best conserved in the mDF iDF arithmetically averaged over subjects). With increasing distance to the eardrum, the differences between individual and average curves becomes larger, without always showing a clear tendency. At the ITEind, ITEgen, and BTE microphone locations, the gain as compared with the median of individual data is lower with the dhDF TRCF, whereas the structurally averaged DF-TRCF (smDF) yields a resonance gain that is systematically too high. The mFF (arithmetic average of individual free field TRCF) main resonance has a gain that is very comparable to the median of iDF data, except in the BTE microphone locations where it is too high (as discussed for the TRCF curves, Fig. 2.2.6). The bandwidth of the main resonance is well conserved by the subject averages, as well as the structural average. At the ITEgen_Entr location, the double peak in the TRCF leads to a pattern that is not reasonably usable as a correction. The bandwidth of the dhDF curve is structurally larger than the median of the iDF data.

2.2.4.4 Spectral distortion with corrected hearing device HRTFs

Figure 2.2.9 shows spectral distortion metrics D between acoustic scenes auralized with the open-eardrum HRTFs and the hearing device HRTFs corrected by the different TRCFs. The spectral distortion metric is thus an estimate of perceived

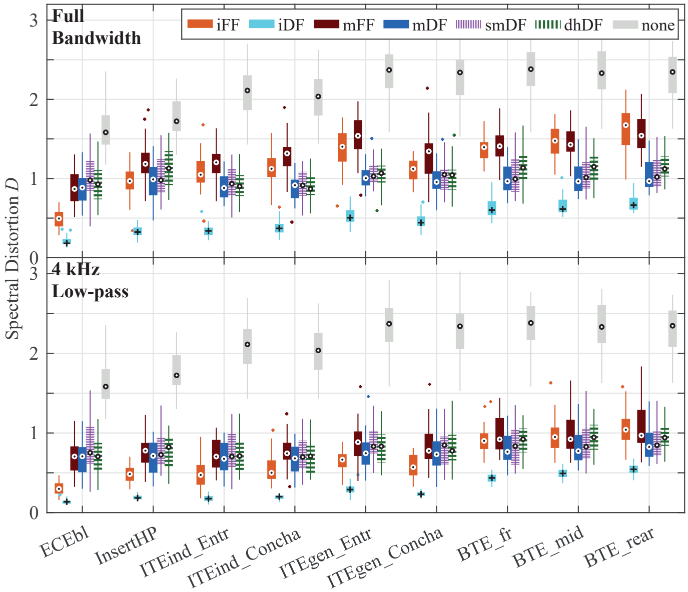


Figure 2.2.9: *Spectral distortion D calculated using the model of Moore and Tan (2004), between the scene auralized with the individual HRTF to the eardrum of the open ear, and with the individual HRTFs to the hearing device microphone locations corrected by the TRCF denoted by the color. See Tab. 2.2.1 for an explanation of the abbreviations. The boxplots group the results in the individual subjects' ears and all seven acoustic scenes. Boxes indicate the 25%/75% quantiles, the mark in the box the median, and the vertical lines the range of results. Top panel: Full audio bandwidth. Bottom panel: Results with the stimulus low-pass filtered at 4 kHz.*

distortion of natural sound transmission characteristics when a hearing device is adjusted to a target response given by the TRCFs; low values indicate a small distortion. The perceptual relevance of these distortions is further discussed in the last paragraph of this section, as well as in Sec. 2.2.5.

First, all TRCFs provide a clear benefit compared with applying no correction. For all microphone locations, the best correction (i.e., lowest residual distortion) is provided by the iDF TRCF. The performance with the iFF TRCF depends strictly on the microphone location: Whereas at the ECEbl, the residual deviation to the optimum approaches the diffuse-field equalization performance, it rapidly increases with increasing distance from the eardrum in the other microphone locations. This increase is more pronounced with the full stimulus bandwidth, where the distortion with the iFF quickly increases above the levels of many of the generic curves. The benefit of the individual- against mean diffuse-field-TRCFs (iDF vs. mDF) decreases with increasing distance from the eardrum; the smallest difference is observed at the BTE microphone locations.

Among the generic TRCFs, the mDF (arithmetic average of iDF) correction generally yields the lowest distortion values. No substantial (and usually insignificant) difference is observed against the structural average of individual data (smDF) or the dummy head average dhDF). The averaged free-field (mFF) correction yields

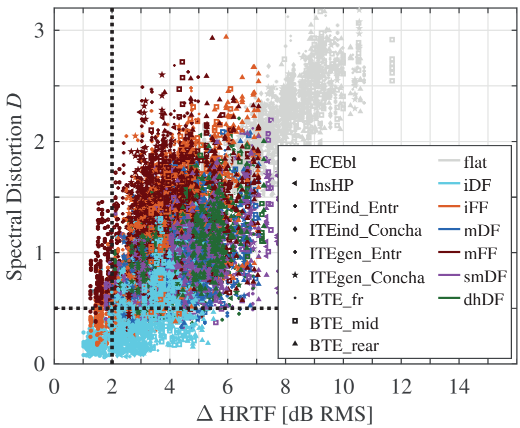


Figure 2.2.10: *Spectral distortion D according to the model of Moore and Tan (2004) plotted against the root-mean-square (RMS) difference calculated in ERB-bins between the HRTF to the ear drum and the TRCF-corrected HRTFs to different hearing device microphones (full bandwidth only). See Tab. 2.2.1 for an explanation of the abbreviations. Dotted black lines illustrate the estimated thresholds of audibility: Based on a 1-dB criterion extended by a 1 dB experimental uncertainty, the spectral deviation in data points with $D \geq 0.5$ can be expected to be noticeable (Δ HRTF $\geq$ 2 dB).*

very comparable values to the mDF correction when the stimuli are low-pass filtered, but in the full bandwidth condition, only at the ECEbl. At the other microphone locations, the distortion increases with the distance to the eardrum, comparable to the iFF TRCF. The modeled spectral distortion size generally decreases when applying a low-pass filter, which is more pronounced for microphone locations far away from the eardrum.

Figure 2.2.10 shows a comparison of the distortion values D from Fig. 2.2.9 with the appropriate RMS spectral difference (see Sec. 2.2.3 for details). The comparison to a more easily interpretable physical error aims to provide a better interpretation of the dimensionless D value. Generally, both metrics correlate well, indicating an almost linear connection of the modeled perceptual and physical errors. In neither metric does any correction completely eliminate all errors—the deviation is always larger than about 1 dB or $D = 0.1$. As an expected noticeable difference for the spectral distortions, we thus extend the common 1 dB criterion by 1 dB of measurement uncertainty. For the vast majority of data points where $D \geq 0.5$, the RMS spectral difference is larger than this 2 dB boundary. Therefore, we estimate the just noticeable spectral difference to be around $D = 0.5$.

2.2.5 Discussion

2.2.5.1 Raw data compared with previous studies

The free- and diffuse-field-to-ear transfer functions shown in Fig. 2.2.4 are generally in good agreement with previous reported studies. The most prominent differences were noted for both field conditions in the transfer functions to the ECEbl and ITE. These discrepancies are probably caused by slightly different shapes of the earplugs. The structurally lower gains in the literature curves support the assumption that the earplugs were less deeply inserted into the ear, leading to a higher attenuation of the cavum conchae resonance (Riederer 2004). Since the curves given by Bentler and Pavlovic (1989, 1992) arithmetic averages of individual transfer functions, average-out effects may also have played a role. Averaging transfer functions that include peaks that are shifted in frequency leads to more gentle and lower peaks (Genuit 1984; Mehrgardt and Mellert 1977).

We expected the transfer function to the eardrum to be reproduced with the highest accuracy, since the reference point in the ear is very well defined. The agreement with previous measurements is excellent for the case of the diffuse sound field, whereas some differences are noted for the free field. For the free field, details of the measurement setup, such as the subjects' posture, headrests, sound source distance, and so forth, play a larger role than in the diffuse field, where these factors usually average out. One example is a ripple in the current free-field responses below 1 kHz, which is most probably caused by a comb-filter effect due to a reflection from the knees of the seated subjects, and not observed in the dummy head data. Also, average effects and average methods (see Shaw 1974 for details of the literature free-field-to-eardrum curve) could have affected the final result. The fact that the current data fits well with previous measurements that have so far been used as the standard verifies both the quality of the measurement and data-processing procedures employed. It also validates the averaging procedure across directions to approximate diffuse-field responses. The data therefore enables us to utilize the current data, for example, to estimate the expected individual deviations from the average and the expected hearing device style-specific deviations. For consistency, only correction curves from data observed in this study are further evaluated, although similar results would be expected when using the curves given by Bentler and Pavlovic (1989, 1992).

The fact that the differences between subjects are larger in the free-field transfer functions can also be explained by the finer details in the free-field transfer functions, which are averaged out in the diffuse field. In both cases, individual differences increased with decreasing distance from the eardrum. This results from the simple fact that an increasing number of acoustic transmission elements that are individual to each person are captured, thus yielding more individual features in the transfer function.

2.2.5.2 Corrections functions related to external ear acoustics

The TRCF should artificially restore features that would occur during sound transmission from the respective microphone location to the eardrum of an open ear. This comprises directionally independent resonances, such as cavity resonances of the ear canal and the cavum conchae, as well as directionally dependent spatial features (Shaw and Teranishi 1968). These cues can best be inspected in the directionally resolved RTFs shown in Fig. 2.2.5, which also contain the individual TRCFs for the free field (frontal incidence, red curves) and diffuse field (blue curves).

The full HRTF is shown for the eardrum, where all acoustic cues generated by the external ear are included in the transfer functions. Starting at frequencies of several hundred Hz and more prominent for frequencies $>$ 1 kHz, there is a considerable spread over incidence directions, prominently between ipsi- and contralateral incidence. The physical correlate is the head-shadowing effect. This variation is, to a large extent, cancelled out in the RTFs in most of the hearing device microphone locations, with the exception of the BTE loctions. We conclude that in all hearing devices where the microphone is located in the concha, head-shadowing effects, and therefore interaural level cues, are conserved to a very high extent (consistent with Kayser et al. 2009).

The RTF from the ECEbl to the eardrum shows no considerable dependence on the incidence direction, and structurally consists of several peaks and notches that can be recognized in both subjects. The variance across incidence directions increases gradually with frequency and appears to be random, probably originating from experimental uncertainty. We conclude that sound transmission through the ear canal is not directionally dependent, verifying previous studies (Algazi et al. 1999; Hammershøi and Møller 1996; Mehrgardt and Mellert 1977). The lowest resonance in the ECEbl to eardrum RTF corresponds to the $\lambda/4$ resonance of the ear canal. Features at higher frequencies do not fit with a simple transmission-tube model and probably depend on the eardrum impedance and the shape of the ear canal (c.f. discussion of Hammershøi and Møller 1996).

Whereas at the ECEbl, the RTF decreases after the first peak, the peak extends further to higher frequencies in most other microphone locations. This can be understood as a superposition of the $\lambda/4$ resonance of the ear canal and broader peaks at higher frequencies. The physical correlate of this directionally independent feature is the cavity resonance of the cavum conchae (Shaw and Teranishi 1968), which is increasingly attenuated through filling by the hearing device. These resonances are best assessed for the diffuse-field RTFs. The InsertHP only partly fills the concha, and only higher frequencies of the resonance are affected. At low frequencies, a shape that is much comparable to the ECEbl is observed. By contrast, towards higher frequencies (> 4 kHz), a larger correction is necessary to restore the transfer function to the eardrum. The more the device fills up the cavum conchae or the further the microphone is positioned off the ear canal entrance, the less of the concha resonance is captured by the microphone. Consequently, more gain in the corresponding frequencies has to be applied in the TRCF. The ITEind fills the concha very uniformly, consequently the resonance is attenuated in all of its frequencies, leading to larger and smoother amplitude curves of the RTF in the diffuse field. For the ITEgen device, the concha is filled less uniformly and the microphone protrudes further, leading to more peaky RTFs. The reasons are probably cavity resonances in the residual cavum conchae segments.

In the InsertHP and ITE microphone locations, the variance across directions increases rapidly above approximately 4 kHz. At this frequency, the wavelength becomes comparable to the size of the pinna and concha. Therefore, spectral characteristics created by these structures are biased by the microphone placement and obstruction of the pinna in these conditions. The result excellently reproduces previous data on comparable hearing device microphones (Durin et al. 2014; Hoffmann et al. 2013b). In the band between 4 and 8 kHz in the RTFs for free-field (i.e., frontal) incidence, a distinct notch is observed. Referring to Fig. 2.2.4, this notch is present in the free-field-to-eardrum and ear canal entrance transfer function, but not in the ITE condition. This behavior can be related to the restoration of a destructive interference of a wave that directly enters the ear canal, and a component that is reflected by the concha back wall, often referred to as Concha Notch (Butler and Belendiuk 1977). This notch occurs in the median plane in frontal directions, but not for other incidence directions. The same observation was made in the RTFs of the BTE locations, but there, the concha notch is overlaid by other structures and is less prominent than in the ITE and InsertHP locations. Given the average over the free-field correction functions (TRCF-mFF, see Fig. 2.2.6), no large difference for this spectral region is observed between the InsertHP and the ITE locations, particularly the ITEind. Therefore, the spectral notch is not better included in the InsertHP, although it obstructs the cavum conchae less than the earmold of the ITEind. However, the conservation of this feature in the different microphone locations is subject to large differences between the individual ears, as can be observed in the individual RTFs given in Fig. 2.2.5.

2.2.5.3 Free-field versus diffuse-field correction

The findings with the spectral distortion model shown in Fig. 2.2.9 indicate a preference for correcting the target of a hearing device to the diffuse field rather than to the free field (i.e., frontal incidence). The results were obtained with a set of realistic acoustic scenes with several desired and distracting sound sources and reverberation in a complex spatial setup, and are therefore expected to hold in real-life acoustic environments.

The analysis in the previous section revealed that several directional cues for the prominent frontal direction like the Concha Notch) can only be included in the sound field if it is included in the TRCF. Such a procedure might, on the one hand, lead to a better spatial reproduction of the environment due to the inclusion of such features, but on the other hand, this will lead to larger errors for other directions than when applying a response correction that is correct for the directional average. In daily life, neither of the extreme cases, free field and diffuse field, occurs exactly. If the correction aims at equalizing for one prominent direction, the frontal direction is surely a well-grounded choice. However, due to reverberation, the pressure created by a sound source in the viewing direction is normally very different from what would be created in the free field. In rooms, fine structures, particularly notches, are flattened by reflections from objects and by reverberation (Shinn-Cunningham et al. 2005). This makes the transfer function to the eardrum generally more alike to that observed in an ideal diffuse field. Similar considerations and conclusions were drawn by Killion and Monsor (1980), but without a detailed assessment. The diffuse field may, however, not be the optimal reference for defining the TRCF. Whether a (weighted) average of the RTF over a constrained set of directions is a better option would be an interesting question for future psychoacoustic experiments.

Besides the directional effects originating from the pinna, a distinct ripple was seen in the free-field TRCF below 2 kHz, but not in the diffuse-field TRCF. The ripple most probably originates from a reflection from the knees of the subjects that only occurs for incidence directions in the frontal median plane. Given the size of the effect, large influences on the results regarding spectral distortions are not expected, besides the fact that this feature could also be regarded as a relevant directional cue.

Altogether, we recommend that the hearing device frequency response should be equalized by a correction that is correct to the diffuse field rather than to the free field. This holds for both when individual measurements are available, and for when generic correction functions are used. Future research should examine whether even better results can be achieved by correcting the response for the (weighted) average of different incidence directions, instead of the extreme cases of the free or diffuse field.

2.2.5.4 Individual versus generic correction and averaging methods

Given the notable variations of the TRCFs between individual ears shown in Fig. 2.2.6 and Fig. 2.2.7, it seems obvious to assume that using individual TRCFs is preferable against a generic correction derived from averaging subject data or dummy head measurements. This benefit was verified by the spectral distortion model results from Fig. 2.2.9 for the diffuse-field correction curves (iDF vs. mDF) in general, whereas for free-field corrections, the benefit depends on the microphone location in the ear. For the ECEbl, the directional information is valid over a wide frequency range, and the iFF TRCF provides a benefit against generic TRCFs. For the other microphone locations, the spectral distortion with the iFF TRCF was larger than with generic diffuse field (mDF, smDF, and dhDF) TRCFs. When the stimuli are low-pass filtered, the spectral distortion with the iFF was, for all microphone locations, at least as small as with the generic diffuse-field TRCF. Apparently, there is a trade-off between conserving the main resonance as the most characteristic individual feature and other errors that occur in different directions, mostly in the higher frequency regime above 4 kHz. Thus, whereas both the main resonance and the average high-frequency behavior of the iDF-TRCF fit all incidence directions on average, the notch in the higher frequencies of the iFF curve leads to higher errors for incidence directions other than from the front. This explanation fits well with the observed dependence on the stimulus bandwidth: When a 4-kHz low pass is applied, the correct spatial information is captured by all microphone locations except the BTE device. Consequently, an iFF correction is only beneficial with respect to the generic TRCFs with the low-pass stimulus.

As a general statement, the benefit of the individual TRCF compared with average values decreased with increasing distance to the eardrum, and almost vanishes at the BTE locations. This is rather surprising, because the further away from the eardrum the hearing device microphone is located, the more acoustic features have to be included in the TRCF—features that are in principle individual to each ear. Apparently, the superposition of many of such features leads to a decrease of individuality, particularly in the diffuse-field correction curves. This conclusion is even stronger when the bandwidth is restricted.

Among the generic DF correction functions (mDF, smDF, and dhDF), no large differences in spectral distortion were observed. Generally, the arithmetic average over individual TRCFs (mDF) produces the least spec tral distortions, and the ranking of structural average and dummy-head average depends on the microphone location. The difference between the mDF and dhDF-TRCFs is caused by the differences in geometry and eardrum impedance between the artificial and real ears. Both contain a comparable degree of spectral detail, but the dhDF lies 2 to 5 dB lower in level around the main resonance. The level offset might be explained by a mismatch between the impedances of the ear simulators (IEC711) and the real eardrum impedances in the subjects. Since the level gap increases with the distance from the eardrum, a slight misalignment in the dimensions of the pinna or the acoustic impedance of its material might also explain the mismatch. However, the spectral distortion observed with the correction function obtained from the dummy heads is only slightly higher than with the subject average (Fig. 2.2.9). We conclude that when no individual data are available, using the TRCF obtained from dummy head measurements is justified by the current results. This holds especially when considering the effort that is connected to measurements like the ones presented here, compared with the same experiments with dummy heads. Based on the current results, it might be beneficial to increase the gain of the main resonance of the dummy head TRCF as a heuristic correction.

The structural mean of the subjects' diffuse field TRCFs (smDF) avoids smearing effects due to arithmetic averaging over shifted peaks. The result should be a 'typical' curve, which would be represented in a better conservation of structural parameters of the TRCFs than with a standard averaging procedure. However, the structural parameters of the main resonance are not better represented by this structural average when compared with the arithmetic mean. Whereas the resonance frequency conserves well the subject median in all microphone locations, the gain is overestimated in the structural mean (see Figs. 2.2.6 and 2.2.8). Only the bandwidth of the main resonance is better conserved with the structural average. Given these results and the fact that the TRCF-smDF does not provide a benefit compared with the standard average in HRTF correction in terms of spectral distortion, we conclude that this simple structural model average should not be used. More sophisticated models may provide a better fit. This, however, makes averaging across the model parameters more challenging, and the model requires additional knowledge such as the geometric ear dimensions (Genuit 1984).

We conclude that an arithmetic average of individual human subjects' diffuse-field TRCF is the best option to obtain generic TRCFs. However, a correction curve based on dummy head data can be used with almost the same performance and could be improved by a heuristic correction.

2.2.5.5 Perceptual relevance of spectral distortions

The spectral distortion model used was based on physiologically motivated excitation patterns on the basilar membrane and yields a good prediction when a spectral difference is perceivable between two arbitrary stimuli (Moore and Tan

2004). This prediction worked very well for spectral naturalness judgments of electroacoustic transducers. However, in this context, it might be hypothesized that a spectral deviation from the individually correct external ear transfer function through which the subjects hear the world every day may be perceived as more disturbing than the same spectral distortion to the source signal. Final judgments about the benefit of individual response correction functions can thus only be made based on subjective listening experiments. Based on a common 1-dB criterion extended by the experimental uncertainty, the just noticeable spectral distortion was estimated to be around $D = 0.5$ (see Fig. 2.2.10 and Sec. 2.2.4).

This boundary is exceeded in the majority of conditions, and therefore the spectral distortions can, in general, be considered as relevant. Only when the iDF TRCF is applied did a considerable share of data points fall below this boundary. All these conditions belong to microphone locations inside the pinna, and include all devices except the BTE. The defined threshold was derived from the appropriate just-noticeable difference in psychoacoustic A-B comparisons, and therefore the lower boundary of detection sensitivity. In real-world environments with dynamically changing sound sources, the threshold for noticeable differences is likely to be much higher.

Altogether, we conclude that the construction of hearing devices that are acoustically transparent is possible in the perceptual sense with all device styles except the BTE. On the other hand, the result shows that acoustic transparency can only be achieved using individualized target-correction functions.

2.2.6 Conclusions

We studied external ear acoustics related to the task of equalizing hearing devices to the acoustics of the ear of the individual subject. In particular, correction functions that transform the pressure at the device's microphone to the reference observed at the open eardrum (here referred to as TRCF) were derived using different approaches and for various hearing device styles. For that purpose, we measured HRTFs to a comprehensive set of hearing device microphone locations and the eardrum in 16 human subjects and 2 dummy heads from 91 incidence directions. The HRTF dataset and the derived TRCFs are publicly available.[1]

TRCFs were calculated for each individual ear based on the relative transfer function between the microphone location and the eardrum of the open ear, for both a free and diffuse-sound field. Generic correction functions that do not include measurements on the specific subject were derived through arithmetic and structural averaging of the individual data, as well as using data from commercial dummy heads with standardized IEC711 ear simulators. Our main conclusions regarding the correction functions are:

— The TRCF depends greatly on the exact location of the microphone in the ear, as well as the degree to which the ear is filled by the device. It restores acoustic features that would normally occur during sound transmission between the microphone location and the eardrum of the open ear, most importantly a directionally independent resonance of the ear canal and cavum conchae between approximately 1 and 7 kHz.

— Diffuse-field equalization of the microphone signal to the open-ear reference should be preferred over free-field equalization.

— The TRCF differs between individuals, like each ear is different. Structural similarities between individuals in each microphone location exist, but these features are shifted in frequency and level.

— Individualized TRCFs provide a significantly better adaptation to the individual ear acoustics than generic correction functions. Using individual diffuse-field correction, it is possible to equalize the target of most hearing device styles down to an error compared with open-ear listening that is probably not noticeable in common acoustic scenes. Generic correction functions lead to spectral distortions that are probably noticeable.

— Among generic equalization functions, the best result is achieved when using an arithmetic average of the individual diffuse-field TRCFs. Using data obtained from dummy head measurements is a valid alternative that creates only marginally larger errors. Only minor differences to averaging over subjects were observed in the present data, most prominently a systematically lowered TRCF gain. A structural mean calculated using a rudimentary two-resonator model of the external ear resulted in a slightly poorer result than the arithmetic average.

— The benefit of diffuse-field correction and individual transfer functions interact with the device style: The less the hearing device microphone captures spatially dependent cues correctly, the more a diffuse-field TRCF should be preferred over a free-field TRCF. The benefit of individual TRCFs decreases with increasing distance of the hearing device microphone from the eardrum and increasing filling of the cavum conchae, particularly if it is filled uniformly. This means that individual correction functions are most beneficial in devices in the ear canal (such as CIC hearing aids) and least beneficial in BTE devices.

The results demonstrate the benefit of individualized hearing device fitting using probe tube measurements. According to our data, probe tube measurements should be performed in (approximated) diffuse-field conditions, for example, using several loudspeakers in a reverberant room. On the other hand, the generic TRCFs can be directly applied in devices that utilize a nonindividualized response target, such as consumer hear-through headsets (Hoffmann et al. 2013a).

For constructing acoustically transparent hearing devices, the optimum style would allow microphone positioning at the ear canal entrance capturing the full spatial HRTF information. Our data once more confirms that full spatial acoustic cues can only be observed with a microphone in the ear canal (Algazi et al. 1999; Hammershøi and Møller 1996), and thus acoustic transparency in the strict physical sense is only achievable with devices in the ear canal and individualized target responses. However, space constraints and nonindividualized shells require filling at least a part of the cavum conchae in many applications—in this case, the results indicate that it is beneficial to occlude the concha as uniformly as possible, even if it means that the device becomes larger. This holds especially if generic correction functions are utilized. With devices that have the microphone located inside the cavum conchae and using individualized TRCF, we conclude that it is possible to achieve acoustic transparency in a perceptual sense. With BTE microphones, it does not seem possible to construct acoustically transparent devices.

Future work should include a psychoacoustic validation of the findings. Open questions are to what extent noticeable differences to the usual transfer function to the eardrum translate to an impaired sound quality, and, specifically, what the subjective benefit of individual correction functions really is.

Acknowledgements

The authors thank Lena Haverkamp for help in placing the probe tube microphones, Jan Heeren and Felix Grossmann for assistance in preparing and conducting the HRTF measurements, and all voluntary subjects for patiently sitting in the anechoic chamber keeping their heads still. The authors also thank Andrew Oxenham and two anonymous reviewers for helpful remarks on an earlier version of the manuscript. The HRTF dataset and the derived TRCFs are publicly available.[1]

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Research Unit FOR1732 *Individualized Hearing Acoustics* and the Cluster of Excellence EXC1077 *Hearing4all*, both funded by the German Research Council DFG (Deutsche Forschungsgemeinschaft).

Note

[1] The HRTF database and the TRCFs derived are available for download at http://medi.uni-oldenburg.de/hearingdevicehrtfs/

2.3 Spectral directional cues captured by hearing device microphones in individual human ears

Outline and context within the thesis

In this section, the quality of spectral directional cues that are captured at different microphone locations of hearing devices in individual human ears is assessed. Such spectral cues are exploited for sound localization in many ways, and thus, how accurately these cues are captured by a hearing device microphone determines the boundary of how well spatial hearing can be conserved with hearing devices. To this end, the HRTF data obtained and described in Sec. 2.2 is utilized and the evaluation of spectral directional cues performed through computational models. A psychophysical evaluation of localization abilities based on the same data is made in Sec. 4.1.

Abstract

Spatial hearing abilities with hearing devices ultimately depend on how well acoustic directional cues are captured by the microphone(s) of the device. A comprehensive objective evaluation of monaural spectral directional cues captured at 9 microphone locations integrated in 5 hearing device styles is presented, utilizing a recent database of head-related transfer functions (HRTFs) that includes data from 16 human and 3 artificial ear pairs. Differences between HRTFs to the eardrum and hearing device microphones were assessed by descriptive analyses and quantitative metrics, and compared to differences between individual ears. Directional information exploited for vertical sound localization was evaluated by means of computational models. Directional information at microphone locations inside the pinna is significantly biased and qualitatively poorer compared to locations in the ear canal; behind-the-ear microphones capture almost no directional cues. These errors are expected to impair vertical sound localization, even if the new cues would be optimally mapped to locations. Differences between HRTFs to the eardrum and hearing device microphones are qualitatively different from between-subject differences and can be described as a partial destruction rather than an alteration of relevant cues, although spectral difference metrics produce similar results. Dummy heads do not fully reflect the results with individual subjects.

This section is a formatted reprint of
F. Denk, S. D. Ewert, and B. Kollmeier (2018g). “Spectral directional cues captured by hearing device microphones in individual human ears”. *The Journal of the Acoustical Society of America* 144(4), pp. 2072–2087.

Author contributions: FD designed, implemented and performed the simulations, created the figures and wrote the manuscript. BK and SE participated in writing the manuscript.

2.3.1 Introduction

Conservation of spatial hearing is one major unsolved issue for hearing aids hindering efficient communication in challenging situations (Kollmeier and Kiessling 2018). It also becomes highly relevant in hearing devices targeted at normal hearing users, such as augmented reality audio systems (Härmä et al. 2004; Rämö and Välimäki 2012). Spatial hearing abilities with any hearing device depend on how well the device conserves the acoustic directional cues utilized by the auditory system. These cues are created by sound transmission effects between the source and the eardrum and include interaural (level and time differences) and monaural spectral directional features that are all contained in the head-related transfer function (HRTF). Interaural cues are captured well in all common ear-level hearing device styles, and conservation of these features and connected localization in the lateral domain is mainly dependent on the signal processing and synchronization between the left and right device (Byrne and Noble 1998; Kollmeier et al. 1993; Van den Bogaert et al. 2006, 2011). In contrast, conservation of monaural spectral cues in the HRTF depends ultimately on how well the device microphone captures them. It is well known that the HRTF is distorted when the recording location is more than a few millimeters away from the (possibly blocked) ear canal entrance (Algazi et al. 1999; Hammershøi and Møller 1996), or the ear is (partially) filled up (Hofman et al. 1998; Riederer 2004). However, the characteristics and effects of these deviations have only been examined for a limited set of hearing device microphone locations and almost exclusively in artificial ears (Durin et al. 2014; Härmä et al. 2004; Hoffmann et al. 2013b; Rämö and Välimäki 2012). We recently presented HRTF measurements including microphone locations at the eardrum and a comprehensive set of hearing device styles, recorded in the ears of 16 human subjects and 3 dummy heads (Denk et al. 2018b). Based on these data, we here perform an objective evaluation of the directional information captured at a total of 9 microphone locations integrated in 5 different hearing device styles, in individual human and artificial ears.

Psychophysically, errors to the individual HRTF have the largest influence on localization performance in the vertical domain, as shown for modifications of the pinna shape (Gardner and Gardner 1973; Hofman et al. 1998) or listening through hearing devices (Best et al. 2010; Brungart et al. 2007; Byrne and Noble 1998; D'Angelo et al. 2001; Hoffmann et al. 2014; Van den Bogaert et al. 2011). For hearing devices, it is difficult to separate the influence of the HRTF at the hearing device microphone from other aspects of sound presentation, like processing or amplification settings, bandwidth restrictions or the influence of sound directly entering the ear canal (c.f. Van den Bogaert et al. 2011). Furthermore, only few connections between the subjective results and objective metrics have been made, and objective metrics that indicate the quality of directional information are hard to find. Durin et al. (2014) performed an objective evaluation of the HRTF information in five hearing device styles using auditory localization models. Their results also predict a decrease in vertical localization performance with the presence of errors in the HRTF. Their ranking of different microphone locations, however, depends much on the considered error metric.

Similar to the distortion of the individual HRTF in the hearing device microphone and the consequences for spatial perception, sound localization in a virtual acoustic environment is degraded when a sound field is reproduced over headphones using an HRTF from the ear of a different subject or dummy head (Minnaar et al. 2001; Møller et al. 1996; Wenzel et al. 1993). In this context, HRTF differences between individual subjects have been quantified and behaviorally studied in relation to subjective localization performance by Middlebrooks (1999a,b). The most prominent difference between individual HRTFs is a shift of features in the logarithmic frequency axis depending on the ear size, otherwise individual ears create structurally similar acoustic features. Vertical localization performance with another person's HRTF decreases with the physical spectral difference (SD) between the HRTFs of both persons. However, it is not clear whether the observed deviations to the individual HRTF in hearing devices is qualitatively comparable to between-subject differences, and thus whether these psychophysical results can be transferred.

Localization based on HRTF cues is a process that relies on learned spectral patterns that humans are capable of recalibrating to, at least to a certain extent (Carlile 2014; Hofman et al. 1998; Majdak et al. 2013; Mendonça 2014). Therefore, independent of the similarity of an HRTF (of another person or when wearing a hearing device) with the individual HRTF, the availability of any spatial information in the altered HRTF that could in principle be learned is of great interest. Durin et al. (2014) demonstrated that hearing device microphones still capture spectral directional cues with a reasonable spatial resolution.

To the authors' best knowledge, no evaluation of hearing device HRTFs has been performed that is comprehensive over device styles and human ears. In particular, previous investigations included HRTFs for either hearing aid styles (behind-the-ear or individualized shells) or non-individualized in-ear devices as usually utilized in consumer applications, but never both (Durin et al. 2014; Hoffmann et al. 2013b; Kayser et al. 2009). Also, it is by no means clear how HRTF information in the same device style differs between individual ears, since virtually all measurements on the matter have been made in dummy heads. Furthermore, the qualitative characteristics of the HRTF errors in hearing devices is not thoroughly clear, and no relation to between-subject differences in individual human HRTFs has been established.

We present a comprehensive evaluation of directional information captured by hearing device microphones. The analyses utilize a recent publicly available HRTF dataset including a comprehensive set of hearing device styles (Denk et al. 2018b). By means of descriptive analyses as well as a combination of previously established quantitative metrics and sound localization models (Durin et al. 2014; Langendijk and Bronkhorst 2002; Middlebrooks 1999a) we tackle the following research questions:

- What are the qualitative and quantitative errors to the individual HRTF depending on the microphone location and the device style, and how large is the spread of these errors over individual subjects' ears?

— What is the expected acute vertical localization performance with the HRTFs observed at the individual microphone locations?

— What is the quality of spectral directional information in the hearing device HRTFs, i.e., what potential localization cues in the hearing device HRTF are available to be exploited, e.g., after learning?

— Is the evaluation of hearing device HRTFs measured on dummy heads representative for human subjects?

The paper is structured as follows: In Sec. 2.3.2, the HRTF data including the hearing device styles used with microphone positions and data processing are outlined. Section 2.3.3 describes the qualitative inspection and quantitative metrics that are utilized to evaluate the HRTF information. The results are presented in Sec. 2.3.4, comprehensively discussed in Sec. 2.3.5, and the conclusions are given in Sec. 2.3.6.

2.3.2 HRTF data

We utilized head-related impulse responses (HRIRs) recorded at the eardrum and microphone positions of a comprehensive selection of hearing device styles (OlHeaD-HRTF Database Denk et al. 2018a, 2018b). The HRIRs were recorded in the ears of 16 human subjects (10 male, 6 female, age 27.3±5.1 years) and 3 dummy heads: KEMAR type 45BM, *Brüel&Kjær*HATS type 4128C, and a custom Dummy Head with Exchangeable Ear Canals (DADEC, Hiipakka et al. 2010), equipped with *G.R.A.S.* KB 1065/1066 Pinnae. As a reference and for comparisons with between-subject differences, HRIRs from the CIPIC HRTF database (Algazi et al. 2001) were used.

2.3.2.1 Hearing device styles and microphone locations

In this work, the HRTFs measured at nine hearing device microphone locations contained in five hearing device styles as well as at the eardrum as shown in Fig. 2.3.1 are considered. The following list outlines the device styles and the exact positioning. More details on the measurement technique, microphones and construction of the devices can be found in (Denk 2018a,b).

Eardrum: Measured using an audiological probe tube microphone.

ECEbl: Blocked ear canal entrance. Location at the terminating plane of the ear canal, which was occluded by a suitable earplug (as in Lindau and Brinkmann 2012), providing a firm and reproducible fit in the ear canal. In a hearing systems context, the blocked ear canal entrance can be regarded as mostly equivalent to small hearing devices fitted into the ear canal of a subject, such as completely-in-canal CIC) devices (Durin et al. 2014).

InsertHP: Location on a small insert headphone *Sennheiser* CX200), as has been used in augmented reality audio applications (Härmä et al. 2004; Hoffmann et al.

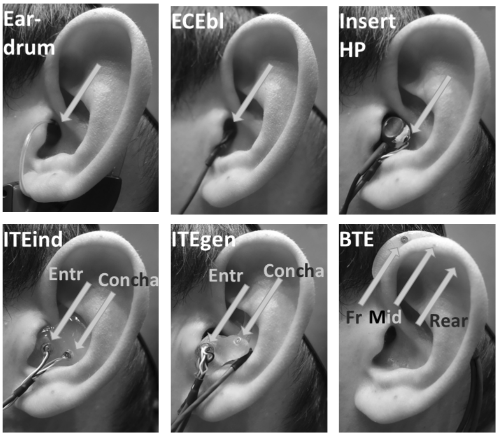


Figure 2.3.1: *Photograph of all hearing devices and microphone locations in the ear of a subject and utilized names, reprinted with permission from (Denk et al. 2018b).*

2013a). The microphone was placed near the concha bottom and points towards the rear concha wall (see Fig. 2.3.1). Regarding hearing aids, this device is comparable to an in-the-canal (ITC) device, although it might typically fill up a larger part of the concha than an individualized shell.

ITEind: Individual in-the-ear (ITE) type hearing instrument, implemented by two microphones flush inserted into an individual standard earmould that fills the concha bottom completely. One location is near the ear canal entrance (entrance microphone," Entr), and one in the rear part of the cavum concha ("Concha microphone"). Entrance and Concha microphones are approximately 8–12 mm apart from each other in a preferably horizontal orientation in the individual ears (the distances for the individual subjects are provided with the database). The hardware configuration is equal to the outer microphones of the prototype hearing device presented in (Denk et al. 2018c).

ITEgen: Generic ITE hearing device, i.e., a larger generic earplug that houses two external microphones (Entr and Concha microphones, as in ITEind). The microphones are 1.1 cm apart and stick further out of the ear than with the ITEind earpiece, and the cavum concha is filled less uniformly. The ITEgen earpiece can also be viewed as a larger insert headphone with integrated microphones.

BTE: Behind-the-ear hearing aid dummy with three microphones referred to as frontal (fr), middle (mid), and rear. The device has the same layout as the one used by Kayser et al. (2009).

2.3.2.2 Data processing and coordinate system

HRTFs were computed from impulse responses using a discrete Fourier transform with a length of 8192 samples. Perceptually irrelevant spectral detail was removed by smoothing the spectral amplitude with 1 ERB bandwidth (Breebart and Kohlrausch 2001). Directional transfer functions (DTFs) were then computed by dividing by the root-mean-square average over the HRTF magnitudes of a directionally balanced set of incidence directions $\mathbf{\Psi}$ (see below and Fig. 2.3.2). The differences between the DTF sets are thus equivalent to the residual differences after optimal equalization of the hearing device HRTFs using a single filter, i.e., a diffuse-field correction against the response at the eardrum (Denk et al. 2018b). Finally, the spectral sampling was reduced and weighted to approximate auditory importance by picking 1/2-ERB-spaced amplitude values spanning the full audio bandwidth (0.2–18 kHz).

HRTFs had been recorded for 91 sound incidence directions shown in Fig. 2.3.2. The horizontal plane is sampled with a resolution of 7.5°; otherwise the space is sampled with 30° spacing in azimuth and elevation. In the median plane and a sagittal plane displaced 30° to the right, the spatial resolution in the vertical domain in a range of ±30° around the horizontal plane was increased to 10° and 15°, respectively.

For the present evaluation, an ear-polar or interaural-polar coordinate system is considered, as shown in the inset of Fig. 2.3.2. There, the lateral angle $\alpha \in [-90°, 90°]$ denotes the lateral displacement from the median plane and determines the sagittal plane (median plane and dotted line in Fig. 2.3.2); negative values indicate the left-hand hemisphere. The polar angle $\beta \in [-90°, 270°[$ describes the position inside a sagittal plane; for the median plane 0° denotes frontal, 90° above, and 180° rear incidence. The ear-polar coordinate system reflects human sound localization: whereas the lateral angle of a sound source can be determined solely by interaural cues, the polar angle is ambiguous given interaural information, leading to a "cone of confusion." Thus, β is resolved by evaluation of monaural spectral directional cues. The subset of 47 measured incidence directions $\mathbf{\Psi}$ is approximately uniformly distributed on the sphere, and was utilized whenever averaging over incidence directions was applied.

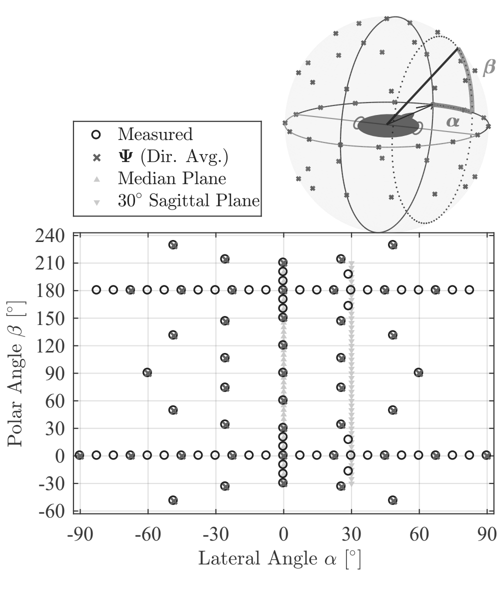


Figure 2.3.2: *Utilized incidence directions. Black circles indicate incidence directions for which HRTFs were measured; grey crosses the uniformly spaced subset W of the measured incidence directions that is included for directional averaging operations. Light grey triangles indicate incidence directions in the median plane and a sagittal plane 30° to the right hemisphere where the spatial sampling was interpolated to 5° resolution. Top right: visualization of the ear-polar coordinate system with lateral angle α and polar angle β (see text for detailed explanation). Head and arrow mark the frontal direction, the black bar an incidence direction determined by α, β. The crosses on the sphere denote the uniformly spaced subset of incidence directions $\mathbf{\Psi}$.*

Interpolation of the DTFs to a vertical resolution of 5° in the median plane and the additional cone of confusion ($\alpha = +30°$, right hemisphere) as shown in Fig. 2.3.2 was performed for each frequency bin separately using the spherical thin-plate splines method (Wahba 1981, 1982) as implemented by Brinkmann and Weinzierl (2017). The maximum angular distance between an interpolated and measured incidence direction was 15°. Around the horizontal plane, where the spatial variation in HRTFs is known to be larger (Møller et al. 1995a), the distance is 5° in the median plane and 7.5° in the 30° cone of confusion. Comparison of our data (ECEbl location) to the CIPIC database (Algazi et al. 2001) that was recorded with a polar sampling of 5.625° verified that the spatial resolution in the present data did not bias the results (see Sec. 2.3.5.1).

2.3.3 Analysis methods

The difference between DTF sets as well as the quality of the directional information was evaluated using a descriptive analysis as well as quantitative metrics and models compiled from previous literature. By means of a joint interpretation, we attempt to provide a comprehensive evaluation of the high-dimensional concept of spatial directional cues.

For quantitative metrics, the selected frequency range is very critical (Baumgartner et al. 2013; Durin et al. 2014; Langendijk and Bronkhorst 2002; Middlebrooks 1999a). For all metrics consistently, we selected the range between 2 and 13.5 kHz, sampled by 34 half-ERB spaced auditory filter channels. The lower boundary frequency reflects the range where spectral cues start differing between the regarded DTF sets. The upper cut-off frequency was chosen as low as possible but where previous data indicate that vertical sound localization is not impaired (King and Oldfield 1997; Langendijk and Bronkhorst 2002). Rationales for a low upper cut-off frequency were the relevance for hearing devices, to minimize the influence of possible measurement uncertainties and to not restrict frequency scaling operations (see Sec. 2.3.3.3, in consistency with Middlebrooks 1999a).

2.3.3.1 Descriptive analysis

The observed DTFs are qualitatively analyzed by a descriptive inspection of the DTFs observed in two representative subjects. The subjects were carefully selected to represent the span of results, and include a man with large ears (VP_E1) and a woman with small ears (VP_N6); the same subjects were selected for showing sample data in (Denk et al. 2018b).[1] Sound incidence in the median plane was chosen for this analysis.

2.3.3.2 Spectral differences

The difference between two DTFs from sets a and b for the same incidence direction (α, β) can be expressed by the SD as introduced by Middlebrooks (1999a). It is defined as the variance across frequency f of the difference spectrum calculated on the dB magnitudes of the DTFs using the ERB-scale frequency sampling

$$\mathrm{SD}_{a,b}(\alpha,\beta) = \mathrm{var}_f\left[\mathrm{DTF}_a(\alpha,\beta,f) - \mathrm{DTF}_b(\alpha,\beta,f)\right]. \tag{2.3.1}$$

The unit of the SD is dB^2. The SD is directly used to assess the directionally-dependent SD between two DTF sets.

To evaluate the overall difference between two DTF sets a and b, the SD is averaged over all relevant incidence directions, yielding the inter-set SD (ISSD) as introduced by Middlebrooks (1999a)

$$\mathrm{ISSD}(a,b) = \frac{1}{N_{\Psi_{ips}}}\sum_{\boldsymbol{\Psi}_{ips}} \mathrm{SD}_{a,b}(\alpha,\beta). \tag{2.3.2}$$

Here, spatial averaging was conducted over all incidence directions inside $\boldsymbol{\Psi}$, excluding all incidence directions with a lateral angle greater than 30° towards the contralateral side denoted by $\boldsymbol{\Psi}_{ips}$) that are not exploited in sound localization (Morimoto 2001).

2.3.3.3 Scaling of DTFs in frequency and level

Individual ears produce structurally similar spectral patterns, which are shifted against each other between listeners (Mehrgardt and Mellert 1977; Møller et al. 1995a). In the simplest case, an identical difference in all ear dimensions results in a scaling in frequency that can be compensated by a shift of the spectra on a logarithmic frequency axis. Middlebrooks has shown that such scaling in frequency can substantially reduce the ISSD between the DTFs of two listeners (Middlebrooks 1999a) and improves virtual sound source localization with non-individualized DTFs (Middlebrooks 1999b). Middlebrooks also showed that such frequency shifts are the main between-subject variation of DTFs (Middlebrooks 1999a).

We here apply an extended scaling approach to explore the qualitative aspects of the hearing device DTFs as compared to the eardrum DTF. Thereby, we utilize frequency scaling analogous to (Middlebrooks 1999a), as well as expansion/compression of the spectral profile, i.e., scaling the DTFs in level. Reduction of differences by frequency scaling would indicate that in the hearing device DTFs many cues are still contained, but shifted in frequency, corresponding to a transformed ear size. Such errors are expected mostly in the InsertHP, where the modified shape of the pinna is still comparable to an ear, but with a reduced size of the cavum concha. Reduction of differences by level scaling would indicate that normal cues are still contained in the hearing device DTF, but with a reduced spectral contrast. Such errors are expected to be most prominent in the ITEind, which conserves the characteristic dimensions of the cavum concha, but decreases the acoustic resonance quality of reflecting structures. Frequency and level scaling are applied both separately and jointly. No reduction of the ISSD between eardrum and hearing device DTF by either scaling approach can be interpreted as a complete destruction of regular cues in the hearing device DTF. The reduction in ISSD was assessed between individuals with the eardrum DTF, as well as between the eardrum and hearing device DTFs in the same ear.

For frequency scaling, a factor > 0 was applied to the original frequency vector of the smoothed DTF magnitude in dB prior to ERB-bin extraction (uniform frequency resolution of 5.86 Hz). Then, the ERB-spaced frequency bins corresponding to the frequencies of interest in the scaled frequency vector were extracted. This results in an effective shift of the DTF on the logarithmic frequency axis that is equivalent to the method in (Middlebrooks 1999a), although the implementation is different. Level scaling was implemented as follows: The average of the DTF (dB values, ERB bins) in the frequency range of interest was subtracted, a level scaling factor applied to the result, and the average added again. A level scaling factor < 1 thus results in a compression of the spectral profile, whereas a factor > 1 causes an expansion. As in (Middlebrooks 1999a), scaling operations between two DTF sets are applied symmetrically, i.e., scaling with inverse values is applied to both DTF sets. If frequency and level scaling are applied jointly, frequency scaling is performed prior to level scaling. The utilized scaling factors were between 0.7 and 1.4.

2.3.3.4 Modelling vertical sound localization

Sound localization within sagittal planes can be understood as a template-matching process that can be modelled computationally with reasonable accuracy. The probability that a stimulus is localized at a certain direction within a sagittal plane can be computed by means of a similarity metric between the stimulus spectrum (denoted by subscript "s") and the "stored" template DTF (denoted by subscript "t"). Following (Baumgartner et al. 2013; Langendijk and Bronkhorst 2002; Majdak et al. 2014), we utilize the similarity index (SI), which is the SD mapped to a value between 0 and 1 through a Gaussian function

$$\mathrm{SI}(\beta_s, \beta_t) = \exp\left(-\frac{\mathrm{SD}_{s,t}(\beta_s, \beta_t)}{2\sigma^2}\right). \qquad (2.3.3)$$

As suggested by Langendijk and Bronkhorst (2002) and verified by Majdak et al. (2014), a standard deviation $\sigma = 2$ (representing the sensitivity of a listener to SDs) was chosen to reflect an average listener. Note that the SI is used to compare DTFs from the same sagittal plane, and the dependence on α is dropped in the notation. Alternative sagittal plane localization models include positive spectral gradient features instead of the SD (Baumgartner et al. 2014) and showed a higher robustness against certain DTF modifications. We repeated our simulations using an adapted model of Baumgartner et al. (2014) and gained comparable results that did not change the main outcomes of this work. These results are provided as a supplement.[1]

We here compute the SI directly between DTFs in the representation of magnitudes in auditory bands, reflecting the estimated perceptual similarity between two identical stimuli presented from different or identical incidence directions. The probability that a stimulus from a certain direction is localized at a certain incidence direction is obtained by normalization

$$p_l(\beta_s, \beta_t) = \frac{\mathrm{SI}(\beta_s, \beta_t)}{\sum\limits_{\{\beta_s\}} \mathrm{SI}(\beta_s, \beta_t)}. \tag{2.3.4}$$

Left and right ears are modelled independently without including a binaural weighting stage in the model. According to previous studies, in the considered lateral angle range up to ±30° both ears contribute to vertical sound localization and results are very similar for monaural and binaural modelling (Baumgartner et al. 2014; Morimoto 2001).

For evaluation of the vertical localization performance, we calculate the local error (LE) and quadrant error (QE) as introduced by Middlebrooks (1999b). The LE is the root-mean-square error in sound localization around a target incidence angle, where large errors >90° are disregarded. We directly use the probability distribution of the localization estimates and calculate the LE from p_l as an expectancy value

$$\mathrm{LE}(\beta_s) = \sqrt{\frac{\sum\limits_{|\beta_s - \beta_t| < 90°} (\beta_{s'} - \beta_t)^2 p_l(\beta_{s'}, \beta_t)}{\sum\limits_{|\beta_s - \beta_t| < 90°} p_l(\beta_{s'}, \beta_t)}}. \tag{2.3.5}$$

For further evaluation, the LE is averaged over all stimulus incidence directions β_s, for each ear and sagittal plane independently. The LE is a metric for the local spatial resolution in sound localization. Contrarily, the QE is a metric for the occurrence of large localization errors, and is defined as the percentage of localizations >90° or further away from the stimulus incidence

$$\mathrm{QE}(\beta_s) = \sum_{|\beta_s - \beta_t| \geq 90°} p_l(\beta_{s'}, \beta_t). \tag{2.3.6}$$

The localization performance with the hearing device DTFs is modelled for 2 cases: *Acute localization* and *template-matched localization*. Acute localization simulates a subject trained to the individual eardrum DTF who listens with the hearing device DTF. There, the SI is calculated between the DTFs of the hearing device (as the virtual stimulus) and the eardrum-DTF of the individual person (as the stored template). The approach evaluates how well the hearing device DTF conserves the spectral cues of the unobstructed ear. Fortemplate-matched localization, the same DTF set of each microphone location is used as stimulus and template. The condition evaluates the quality of the spatial acoustic information in the DTF of the hearing device, independent of how similar it is to the usual DTF to the eardrum. This is equivalent to assuming perfect adaptation to the new DTF cues, irrespective of whether this is possible (Mendonça 2014).

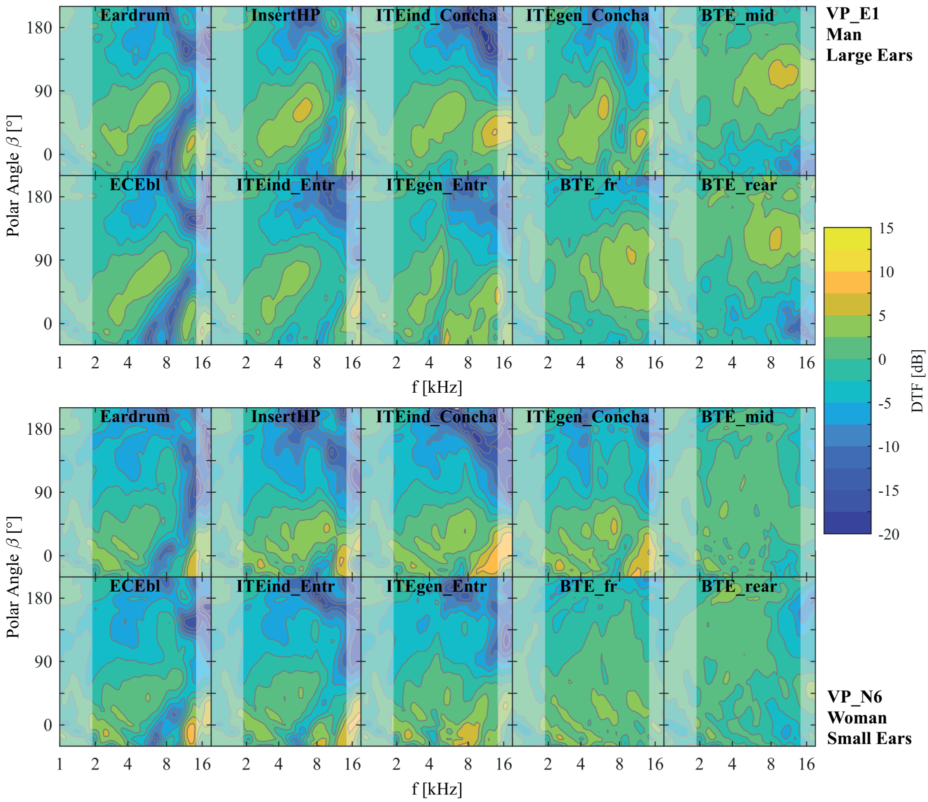


Figure 2.3.3: *DTFs in the median plane for all microphone locations in the left ears of two representative human subjects.*[1]*. VP_E1 is a man with large ears, VP_N6 a woman with small ears. Each panel displays the DTF for one microphone location in the ear, the x-axis denotes the frequency (logarithmic scaling), the y-axis the polar angle in the median plane (0° indicates sound incidence from the front, 90° from above, 180° from behind). Shaded areas mark the frequency regions that were excluded for computing the quantitative metrics.*

2.3.4 Results and analysis

2.3.4.1 Descriptive analysis

Figure 2.3.3 shows the DTFs for all microphone locations obtained in the median plane in the left ears of the two representative subjects.[1] The locations are sorted from top left to right bottom according to their distance from the eardrum. Frequency regions not considered in the quantitative metrics are shaded out.

The DTFs at the eardrum and the ECEbl are almost identical for both subjects, verifying that the differences between the two locations are not direction dependent (Hammershøi and Møller 1996; Mehrgardt and Mellert 1977; Møller 1992). Generally, the DTFs at the eardrum and the BTE microphones start deviating at frequencies above about 2 kHz, and above 4 kHz for the InsertHP, ITEind, and

ITEgen microphones, which is consistent with previous studies (Denk et al. 2018b; Hammershøi and Møller 1996; Hoffmann et al. 2013b; Kayser et al. 2009).

Typical spectral structures that were previously described can be identified in the eardrum DTF and to some extent at the hearing device microphone locations. One prominent example is a spectral notch occurring for frontal incidence directions that changes its frequency with the polar angle (Butler and Belendiuk 1977). The cue is crucial for perception of elevation in the frontal hemisphere. It originates from a destructive interference between a transmission path directly into the ear canal and a reflection from the concha back wall, where the delay is dependent on the elevation (Butler and Belendiuk 1977). The notch is still well visible at the InsertHP and the ITE microphone locations, but first deviates and then disappears with increasing distance from the eardrum. At the BTE microphone locations, the notch is not observed at all. When the notch is still present, deviations include a loss of depth (e.g., ITEind_Entr, more prominent in VP_E1), or a shift towards higher frequencies (InsertHP, both subjects).

Another well-visible directional feature is a higher amplitude for frontal incidence directions than for rear incidence, most prominent in frequencies above 10 kHz. These are shadowing effects of the pinna, a very important cue for resolution of the front vs. back hemisphere (Langendijk and Bronkhorst 2002). For both subjects, it lies mostly within the frequency range considered for the quantitative metrics, although it further extends into the upper excluded range. The cue is conserved quite well for microphone locations inside the pinna, although in the ITEgen microphone locations it becomes more ambiguous. This is intuitively explained, since pinna shadow effects should be independent of any obstructions of the cavum concha, but be dependent on how far the device sticks out of the pinna (as in the ITEgen). In the BTE microphone locations, a structurally similar dependence of the amplitude that is rather uniform across frequencies on the incidence directions can be seen. However, the attenuated and amplified incidence directions are rotated with respect to the eardrum DTF with the rotation dependent on the microphone (better seen in VP_E1), which is explained by altered or inverted pinna shadowing effects.

Deviations against eardrum DTF observed in the various microphone locations are notably different between the two subjects. This difference is not only the deviation between the DTFs observed at the eardrums of the different ears, which mostly comprises a shift of structurally similar cues towards higher frequencies from VP_E1 to VP_N6. For instance, the spectral notch is differently affected by the devices in the two subjects. For the InsertHP in VP_E1, the notch is conserved very well in shape but shifted upwards in frequency, while the notch almost disappears in VP_N6. With ITEind_Entr, the notch is better conserved in VP_N6 than in VP_E1. The complete structure of the DTF in VP_N6 is heavily distorted at both ITEgen microphone locations, whereas it is somewhat conserved in VP_E1. These differences might be related to the fact that the one-size earplugs stick further out of the smaller ear of VP_N6 than the larger ear of VP_E1. Likewise, the pinna shadow effect for the BTE is more pronounced in

VP_E1, probably because the BTE microphones are more occluded by the larger ears of VP_E1.

2.3.4.2 Spectral differences

Directional SD dependence
Figure 2.3.4 shows the directionally resolved SD, both for between-subject differences as well as differences between the eardrum and the other microphone locations in one ear. In both cases, the SD was averaged for each incidence direction independently over the 16 human subjects (i.e., 120 between-subject comparisons). Only the left ears were considered. The spatial distribution of SD for between-subject differences is very broad. The largest errors are observed for contralateral incidence directions. Apart from that, slightly larger SDs are observed from incidence directions in the frontal hemisphere than the rear hemisphere. We verified that the result for the current data is very similar to an equivalent evaluation of the CIPIC database.

For the hearing device microphone locations, the distribution of SD against the eardrum DTF differs from the between-subject SD. The SD is generally very small at the ECEbl, with no considerable directional spatial dependence. In the InsertHP and all ITE microphone locations, the largest errors occur for incidence directions around the front, or frontal incidence directions slightly displaced towards the contralateral side. This result is most pronounced at the ITEind_Entr location. For the other ITE microphone locations, considerable SD is also noted for rear sound incidence. The SD is larger on or below the horizontal plane than above. In

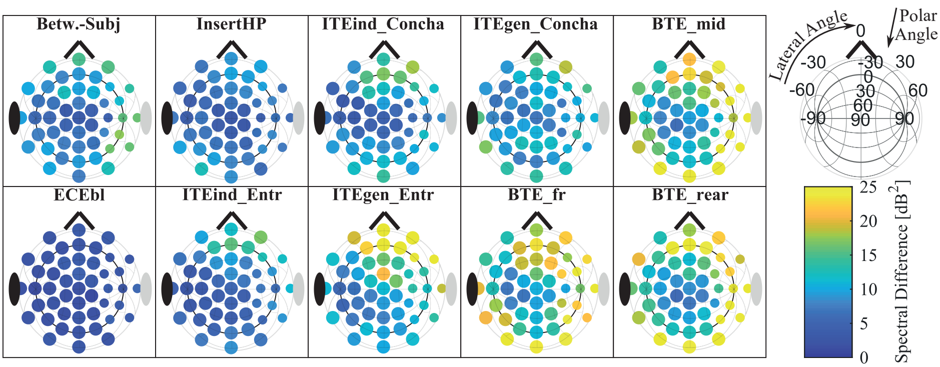


Figure 2.3.4: *Directionally resolved SD (see Sec. 2.3.3.2) for the left ear, averaged across subjects for each incidence direction independently. Upper left panel shows the average between-subject SD from comparing 120 ear pairs. The other panels show the SD between eardrum DTF and the denoted microphone location, averaged over 16 human subjects. The panel in the upper right corner denotes the spatial coordinate system of the plot: Incidence from above the head is located in the center, the horizontal plane is indicated by the thicker circle. Frontal incidence is from the top of the axis as marked by the nose. Grid and ticks denote the lateral and polar angle. Small color circles indicate contralateral incidence directions that are not relevant for sound localization.*

the BTE microphone locations, the SD is generally larger, but shows a comparable distribution.

Inter-set spectral differences
The resulting ISSD, i.e., the SD averaged over incidence directions, is shown in Fig. 2.3.5. The ISSD for the different hearing devices increases with increasing distance from the eardrum, and correlates well with the a priori sorting of the devices. The between-subject ISSD for each microphone location decreases with increasing distance from the eardrum. At the ECEbl, the between-subject ISSD in the current data is in very good agreement with the data from the CIPIC database. The intra-subject ISSD in the DTFs observed at the InsertHP and ITEind are well in the range of the between-subject ISSD for the eardrum and ECEbl. Thus, according to this metric the DTFs at these locations are about as different from the individual reference DTF as another person's DTF.

The ISSD as a function of the microphone location differs between individual ears. Especially, the correlation between ISSD at the InsertHP and the ITEind_Entr microphone locations are very low across subjects ($r = -0.28$, $p = 0.11$). On the other hand, the ISSD at both locations of the ITE devices correlate well (correlations between Entr and Concha locations, ITEind: $r = 0.67$, $p < 0.01$, ITEgen: $r = 0.67$, $p = 0.01$).

The results for the dummy heads are not always in the range of the subject data. The ISSD between dummy head and subjects' DTFs is increased as compared to the human between-subject data, comparably across all microphone locations. There, the results do not differ very much between the individual dummy heads. The relation of within-subject ISSD between human subjects and dummy heads depends much on the microphone location. Generally, the ISSD against the eardrum DTF

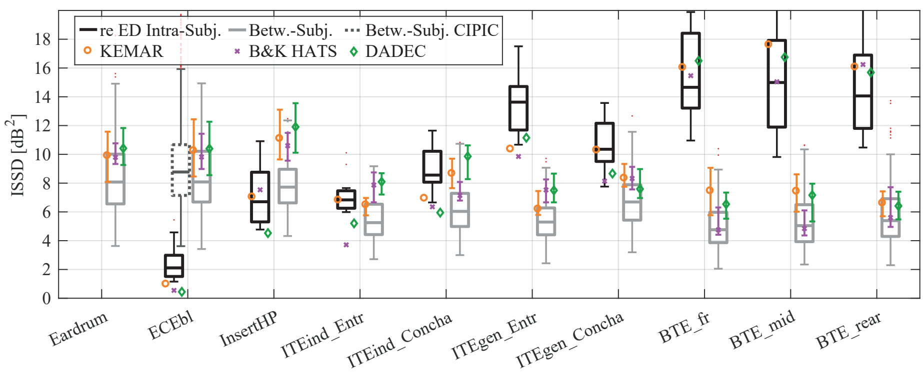


Figure 2.3.5: *ISSD (see Sec. 2.3.3.2) between the DTF at each microphone location and the DTF at the individual eardrum (black), and between subjects (grey, utilizing eardrum DTF). Boxplots show the distribution of results for the ears of the individual subjects. The horizontal line indicates the median, the box the 25%–75% quantiles and whiskers the whole data range excluding outliers, which are marked by dots above/below the whiskers. Individual symbols indicate the result in the dummy heads; in the between-subject condition the mean and standard deviation for the ISSD between the dummy head and all human subjects is shown.*

for the dummy heads is either lower than or equal to the human subjects. At the ECEbl, the ISSD is lower in all dummy heads. For the ITE, this is also the case except for the KEMAR result. For the InsertHP and the BTE, all dummy heads are in the range of the human data.

2.3.4.3 Scaling DTFs

Figure 2.3.6 shows the ISSD before and after the application of frequency, level, and joint frequency and level scaling, as well as the relative improvement by scaling in percent. Without showing figures it is worth noting that the between-subjects results from (Middlebrooks 1999a) were well reproduced. Results for scaling the DTFs observed at the ECEbl, ITEgen, and BTE microphone locations were discarded, since the random distribution of optimum scaling parameters showed that the improvement was merely of random nature and could not be attributed to a correction of physical deviations. This is especially understandable for the ECEbl, where only random measurement errors with respect to the eardrum DTF are expected. For the between-subject ISSD, a notable reduction is achieved by frequency scaling, but not level scaling. Joint application of both scaling methods does not result in a larger reduction than frequency scaling alone. For the hearing device DTFs, none of the individual scaling methods reduces the ISSD against the eardrum DTF in any way comparable to the between-subject result. In the InsertHP, virtually no difference in ISSD is observed after scaling. For the ITEind_Entr location a notable reduction through level scaling is observed. Joint frequency and level scaling between ITE and eardrum DTFs results in an ISSD reduction that is considerably higher than the separate scaling effects, and for the ITEind_Entr results in a relative improvement that is almost in the range of the between-subject differences. For this recording location, the optimal level scaling factors correspond to an expansion of spectral contrast in the DTF (1.21 ± 0.13, both for level-only and combined scaling), whereas the frequency scaling factors are evenly distributed around 0.

2.3.4.4 Modelled sound localization

SI for acute localization

Figure 2.3.7 shows the SI (see Sec. 2.3.3.4) for acute localization in the median plane for the two representative subjects,[1] i.e., the directionally resolved similarity between the eardrum DTF and the hearing device DTFs.

For VP_E1, the SI with the ECEbl is centered well around the diagonal, indicating very good localization performance. In this subject, the SI distribution still looks similar for the InsertHP, although especially for frontal sound incidence ($\beta = 0°$) the SI is very low around the diagonal but shows a distribution that predicts many front-back confusions. The same tendencies are more pronounced for both ITEind microphone locations. In the ITEgen microphone locations, the SI is generally

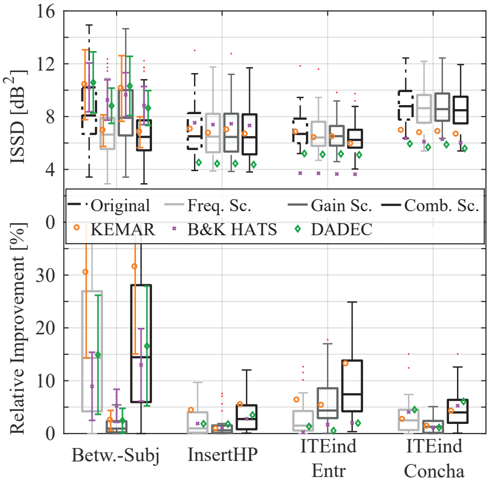


Figure 2.3.6: *ISSD (see Sec. 2.3.3.2) before and after scaling, absolute ISSD (upper panel), and relative improvement (lower panel). Boxplots show the distribution of results for the individual human ears. The horizontal line indicates the median, the box the 25%–75% quantiles and whiskers the whole data range excluding outliers, which are marked by dots above/below the whiskers. Symbols denote the results obtained with the individual dummy heads.*

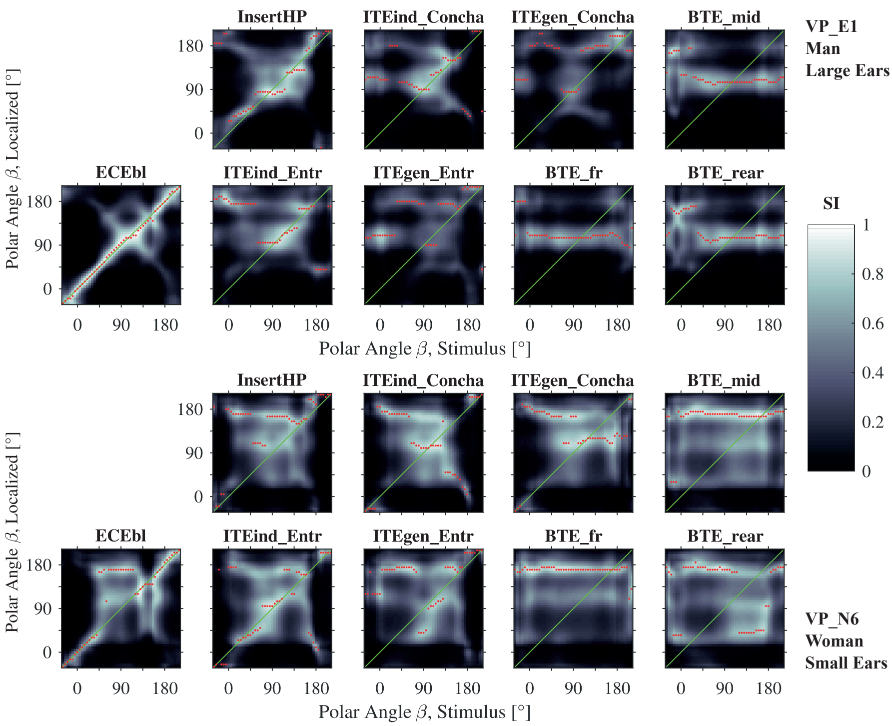


Figure 2.3.7: *SI (see Sec. 2.3.3.4) distributions for acute localization with the hearing device DTFs in the median plane for the two representative human subjects.*[1] *The distributions indicate the perceptual similarity between the DTF of the hearing device at angle on x-axis and the DTF at the eardrum at angle on y-axis, and thus correspond to the localization patterns when internal comparison to the eardrum DTF is assumed. The points indicate the angle with maximum SI per stimulus incidence angle, i.e., the direction a stimulus originating from the direction given by the x-axis is most probably localized at.*

reduced—the absolute similarity to the template DTF is very small. The SI pattern appears random, and no real correlation between actual and predicted angle is observed. In the BTE microphone locations, the SI is clustered mostly around a horizontal line at about 90° angle, which predicts that most incoming sounds would be localized above the head.

Some of these observations are different in VP_N6. First, the SI distributions are generally more widespread around the diagonal than in VP_E1. Second, the SI in the ECEbl is not as well clustered around the diagonal; especially for incidence from above (β around 90°). Third, the SI distribution for the InsertHP is much worse than in VP_E1, with very few SI values around the diagonal, especially around the horizontal plane ($\beta = 0°$ or 180°). On the other hand, the SI distribution at ITEind_Entr is better in VP_N6 than in VP_E1. For the ITEgen and BTE microphone locations, random SI distributions occur, which are structurally very similar between both subjects. However also at these microphone locations, the SI values are generally higher in VP_N6 than VP_E1.

SI for template-matched localization
Figure 2.3.8 shows the SI for template-matched localization in the median plane for the two representative subjects,[1] i.e., the directionally resolved similarity inside each DTF set. It can also be understood as a spatial autocorrelation function of the DTF.

In VP_E1, the SI distributions at the eardrum and ECEbl are very similar and centered around the diagonal, particularly for the frontal hemisphere ($\beta \leq 90°$). In all other microphone locations, the distribution is broadened, i.e., the spatial resolution is reduced. For the InsertHP, the general shape of the distribution is conserved, i.e., it is more concentrated around the diagonal for frontal incidence than incidence from above. In the ITEind microphone locations, a quite similar distribution is observed where the width is larger for frontal incidence than for rear incidence. In the ITEgen microphone locations, the distribution is comparable to the ITEind, but especially at the Concha microphone location the spread for frontal incidence directions is smaller. In all BTE microphone locations, the SI distribution is very wide, indicating poor spatial resolution of the DTF.

In subject VP_N6, some observations are again different to VP_E1. First, the SI distributions at the eardrum and ECEbl are notably broader than in VP_E1. Second, in VP_N6 the SI distributions look very similar between the InsertHP and the ITEind_Entr, with very similar widths. As in VP_E1, the SI distribution is narrower at the ITEind_Concha than at the ITEind_Entr microphone. For VP_N6, the distributions at all BTE microphones are very similar and the resolution is even worse than in VP_E1, with no significant resolution of directions in any of the three microphone locations.

Localization Performance
Figure 2.3.9 shows the results of the modelled localization, including the LE and QE as introduced in Sec. 2.3.3.4.[1] For comparison, the chance level obtained with a uniform SI was calculated for both metrics. The reference condition for localization performance (i.e., modelling free-field localization with unaided ears) is template-

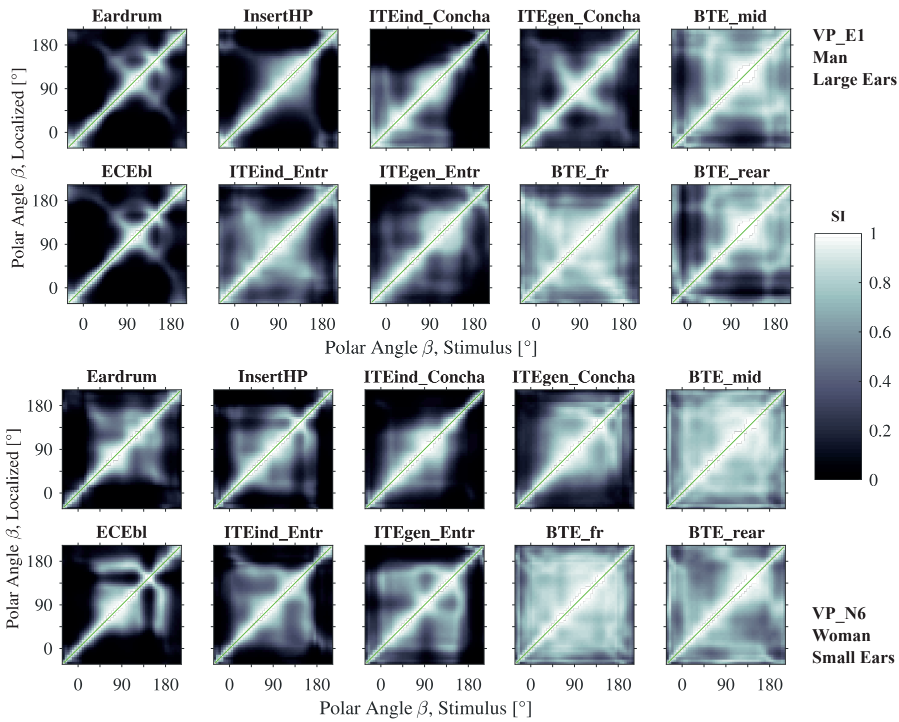


Figure 2.3.8: *SI (see Sec. 2.3.3.4) distribution for template-matched localization with the hearing device DTFs in the median plane for the two representative human subjects.*[1] *The distributions indicate the perceptual similarity between the DTF of the hearing device at angle on x-axis with a DTF from the same set at angle on y-axis, and thus correspond to the localization patterns when internal comparison to the same DTF set is assumed.*

matched localization with the eardrum DTF. Given that relevant localization cues might not be included in the frequency range used in the whole paper 2–13.5 kHz), we had repeated the simulations with a bandwidth between 1 and 16 kHz. This resulted in a general reduction of QEs, but had no noteworthy influence of the relative differences between conditions or conclusions of this work. For localization using the hearing device DTFs when trained to the individual eardrum DTF (acute localization, black boxes in Fig. 2.3.9), the error metrics for all microphone locations except the ECEbl are higher than for the reference condition. At the InsertHP, LE and QE are notably increased with respect to the reference condition, and are in the range of localization performance in the between-subject condition, i.e., localization with another person's DTF (leftmost boxes). The localization performance is very similar in both microphone locations of the ITEind, and notably worse as compared to the InsertHP. For the BTE microphone locations and the ITEgen_Entr, the localization performance is the poorest among the assessed conditions and around chance. The localization performance is better in the ITEgen_Concha than in the ITEgen_Entr, but not considerably different between the individual BTE microphone locations.

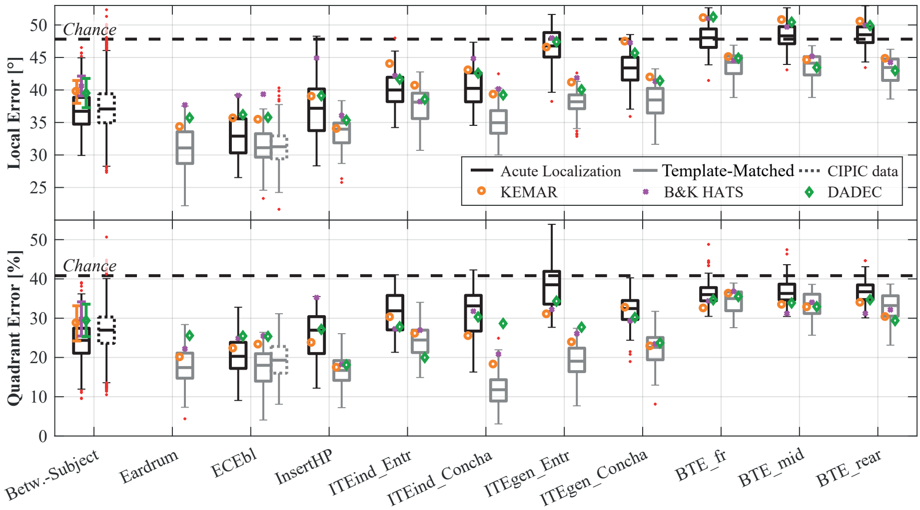


Figure 2.3.9: *Modelled localization performance using the DTF set denoted by the x-label in acute localization trained to the individual eardrum DTF (black lines) and template-matched localization using the respective DTF set (grey lines).*[1] *Top panel shows the LE and the bottom panel shows the percentage of QEs. Boxplots show the distribution of results in individual ears obtained in both regarded sagittal planes. The horizontal line indicates the median, the box the 25%–75% quantiles, and whiskers the whole data range excluding outliers, which are marked by crosses above/below the whiskers. Symbols denote the results for the dummy heads; the dashed lines indicate the chance levels (obtained with uniform SI).*

For localization when utilizing the appropriate DTF set as a template (template-matched localization, grey boxes in Fig. 2.3.9), all error metrics are notably reduced as compared to acute localization (black boxes). However, in terms of LE, reference performance is not achieved for any microphone location except the ECEb1. The smallest LE in template-matched localization is observed in the InsertHP, followed by ITEind_Concha, ITEind_Entr, and the ITEgen microphone locations. The largest LE is again observed in the BTE microphone locations, with no notable differences between the individual microphones. A QE that is very similar to the reference performance is observed at the ECEbl, the InsertHP and the ITEgen_Entr. In the ITEind_Concha, the QE is even lower than for the reference conditions, whereas in the ITEind_Entr and the ITEgen_Concha it remains slightly increased against the reference. The largest QE with template-matched localization is again observed in the BTE microphone locations, with no noteworthy differences between the microphones.

The correspondence of human and dummy head results depends on the condition and metric. The LE with the dummy head DTFs at the eardrum and the ECEbl is larger than in the human data. For the LE observed with the hearing device DTFs, the agreement is better. Also, the accordance is generally better for acute localization than for template-matched localization. In terms of QE, the data from human subjects and dummy heads are in good agreement. In the between-subject condition (here, the subjects' DTFs were used as the template, and the dummy head DTF as stimulus), the result with individual dummy heads are very similar, but all have a clear offset against the between-subject data.

2.3.5 Discussion

2.3.5.1 Data quality

The results observed with the current database and the CIPIC database are in all appropriate metrics very comparable. The between-subject ISSD (Fig. 2.3.5) is almost identical, the slightly larger values for the CIPIC database may be a consequence of the larger number of ears included. Also, modeled localization performance (Fig. 2.3.9) is almost identical between the two datasets. In particular, the equivalent LE results for the template-matched localization with the ECEbl-DTF demonstrate that the spatial sampling and interpolation method utilized here is sufficient to capture the acoustic resolution of the DTF in human ears.

2.3.5.2 SDs between DTFs: Quantitative and qualitative aspects

The smallest differences of the DTF with respect to the eardrum were observed at the blocked ear canal entrance ECEbl), with no notable deviations in the descriptive analysis, the directionally resolved SD (Fig. 2.3.4) or the ISSD (Fig. 2.3.5). The slightly larger errors in the human subjects compared to the dummy heads can be explained by larger random measurement errors, caused by small movements (also of the legs and body, which was not controlled) and lower SNR in the probe tube microphone than in the ear simulators. In accordance with previous studies we conclude that at the ECEbl the full directional information of the eardrum DTF is available (Algazi et al. 1999; Hammershøi and Møller 1996; Mehrgardt and Mellert 1977).

When the microphone location deviates from the ECEbl, the DTF is directionally biased, and the SDs and ISSDs (Figs. 2.3.4 and 2.3.5) increase quite monotonically with increasing distance to the eardrum, which confirms previous studies (Durin et al. 2014; Hoffmann et al. 2013b). Already at the three locations that are rather close to the ear canal entrance (InsertHP, ITEind_Entr, ITEind_Concha), an ISSD that is in the range of between-subject differences (observed at the eardrum, see Fig. 2.3.5) is observed. This corresponds well to the comparable alterations of the ear shapes, leading to differences in the physical processes that create the spectral directional cues (Shaw and Teranishi 1968). However, this result does not include any information about the qualitative characteristics of this error.

The directional distribution of the SD differs between between-subject and intra-ear evaluations (see Fig. 2.3.5). The SD for the between-subject case is rather widespread across incidence directions, whereas the SD for the hearing device DTFs occurs mostly around the frontal incidence direction. At frontal incidence, spectral profiles that have their origin in interferences of sound components reflected in the concha, are most distinct due to the orientation of the pinna (Butler and

Belendiuk 1977; Møller et al. 1995a). Whereas the DTFs are generally shifted against each other for between-subject differences (due to different sizes of the ears), recording with a hearing device microphone seems to affect mostly the cues at frontal incidence, by either altering the interference lengths or attenuating or even eliminating the reflections. This interpretation is consistent with the discussion of the notches made in the descriptive analysis of the DTFs.

Applying a frequency scaling approach to the present data (see Fig. 2.3.5, c.f. Middlebrooks 1999a) could not reduce the errors between eardrum and hearing device DTFs. Also, scaling in level did not notably reduce the deviations for either between-subject differences, nor the InsertHP or the ITEind_Concha. In the ITEind_Entr, however, level scaling results in an improvement that is larger than frequency scaling in this location. Moreover, combined frequency and level scaling results in an addition of both individual improvements.

We interpret these results as follows: The modification of the ear shape by the presence of a hearing device has a qualitatively different effect on the DTF than the variation of ear sizes and shapes between individuals. Given the results of the descriptive analysis, it is unlikely that the common cues in DTFs are destroyed completely, however, it is not necessary that the psychophysical results on listening with a non-individual HRTF are directly transferrable to hearing devices. The difference in improvement with the two scaling dimensions across the microphone locations gives an insight about how the DTF is distorted in the individual devices: In the InsertHP, DTF cues seem to be distorted in a way that they cannot be easily transformed back to the eardrum DTF, i.e., neither shifted in frequency nor compressed in level. One possible explanation is that the shape of the cavum concha is, also acoustically, very different with and without the device. On the other hand, the effectivity of gain scaling in the ITEind_Entr confirms that in this device (at least for some ears) the usual DTFs are mostly flattened out in level. Filling of the concha bottom conserves characteristic lengths that are relevant for the referred interference processes, but could reduce the size of the structure where sound is reflected at the concha back wall, leading to shallower spectral structures. Such flattening of the DTF has recently been connected with the perception of a decreased distance (Baumgartner et al. 2017).

2.3.5.3 Sagittal plane localization using the hearing device DTFs

The modelled localization performance verifies that in hearing devices, conservation of the individual eardrum DTF and capturing of directional information are two very different issues. When acute localization is considered, i.e., assuming internal comparison to the DTF observed at the eardrum, with all locations except the ECEbl the localization performance is significantly worse than in open-ear listening. The performance is also usually worse than in the case of listening with a different person's DTF. For microphone locations of the BTE and the ITEgen_Entr location, the chance level is approached or even exceeded—virtually no directional information that is consistent with the individual open-ear DTF is captured at these locations. The same can be observed in SI distributions that appear random (see

Fig. 2.3.7). Only with the InsertHP and ITEind microphone locations, localization seems to be to some extent possible using the unfamiliar DTFs. The results are in good agreement with those of Durin et al. (2014), however, there none of the rather large variations across subjects was captured since only data from a dummy head were analyzed.

For template-matched localization, i.e., assuming internal comparison of a stimulus with the appropriate DTF set, the localization performance is improved as compared to acute localization for all microphone locations and both metrics. This shows that significant directional information is still included in many hearing device DTFs, although it may not be consistent with the open-ear DTF. The BTE results are poorer than all other locations, indicating that the least usable directional information is captured there. This is consistent with the descriptive analysis, where it was shown that the least directional structure is observed in the corresponding DTFs. For the other microphone locations (InsertHP, ITEind, ITEgen) the performance in relation to open-ear listening depends on the metric. For the LE, the performance is always worse than with the unobstructed ear. All alterations of the pinna shape seem to reduce spatial resolution and information that encodes small shifts of the sound incidence direction, i.e., the natural shape appears to be to some extent optimal to create corresponding cues. On the other hand, the QE is with all these in-ear devices in the range of open-ear listening—apparently, cues originating from pinna shading effects that are included in all these devices, are sufficient to resolve the coarse incident direction. These results differ from comparable analyses in (Durin et al. 2014), who noted much smaller differences between microphones in five hearing aid styles, and better-than reference performance for some device styles. Possible explanations for the discrepancy are that individual subjects were included here (see also Sec. 2.3.5.5), but also a difference in the error metric.

To summarize, we expect vertical localization to be affected with DTFs from all microphone locations except the eardrum when the listener is trained to the eardrum DTF. If the hearing device DTFs could be used optimally, for example, through acclimatization, then coarse localization would be possible. However, the inherent spatial resolution of the hearing device DTFs is poorer than that of the open-ear DTFs, and thus localization accuracy would likely still be limited. Also, given previous results showing incomplete learning of new spectral cues (Carlile 2014; Majdak et al. 2013; Mendonça 2014), and particularly the results regarding qualitative aspects of the DTF distortions, it is by no means guaranteed that humans are able to fully accommodate to and exploit the spectral cues contained in hearing device DTFs.

2.3.5.4 Inter-individual differences

The broad distribution over individual ears and subjects in all quantitative metrics (Figs. 2.3.5, 2.3.6 and 2.3.9) demonstrates that the quality of spectral information encoded in hearing device microphones depends not only on the device style, but also on the individual ear. Due to the coherence of these differences across evaluation aspects, we expect that these differences are beyond random variations across individual DTFs (Møller et al. 1995a; Riederer 1998; Wightman and Kistler 1989).

Both the descriptive analysis (Fig. 2.3.3) and the inspection of SIs (Figs. 2.3.7 and 2.3.8) revealed that in different ears, the same hearing device style may capture certain directional features with different accuracy; or may provide a different quality of spatial information. This effect is most apparent when comparing the results at the InsertHP and ITEind_Entr in the two presented individual subjects. In VP_E1 (man, large ears), the individual eardrum DTF is better conserved with the InsertHP than in ITEind_Entr, whereas the ranking is opposite in VP_N6 (woman, small ears). This is not exclusive to the two shown subjects (see also supplementary material).[1] As indicated by the correlation values of ISSD between subjects, the ISSD ranking with respect to the eardrum DTF is quite consistent across all ITEgen and BTE, but very distinct between InsertHP and ITEind_Entr. Especially for the InsertHP and ITEind_Entr, the modification of the shape of the cavum concha is variable: The InsertHP is a device that has the same size in all ears, with the result that the cavum concha is obstructed to a larger degree in small ears than in large ears. This makes it understandable that more individual information is captured with this device style in VP_E1 than in VP_N6, and why the ISSD with respect to the eardrum DTF varies so largely across subjects at this location. The size of the ITEind, on the other hand, depends on the size of the individual ear, and the degree of obstruction is consistent across ear sizes.

In summary, our data show that to comprehensively characterize the directional information encoded in a hearing device microphone, it is crucial to measure in different ears to get a conclusive evaluation. The more embedded into structures of the external ear the device is, the larger are these individual differences.

2.3.5.5 Usability of dummy head data

Measuring the assessed transfer functions on dummy heads has many benefits: They are easier and cheaper to conduct and more reproducible than measurements in humans (Harder et al. 2015). In the present data, this is apparent by the reduced ISSD at ECEbl with respect to eardrum in the dummy heads as compared to the human subjects (see Fig. 2.3.5). However, some limitations of assessing hearing device DTFs on dummy heads that go beyond the inaccessible inter-subject variance became apparent in the present analysis. First, it appears that DTFs of dummy heads are more different from human DTFs than human DTFs differ from each other. This holds for the quantitative SDs at all regarded microphone locations (between-subject ISSD, grey boxes in Fig. 2.3.5), as well as for the localization performance (larger error in all metrics with the dummy head DTFs in the between-subject condition, leftmost boxes in Fig. 2.3.9). In this respect, our data are in line with previous results reporting that localization using recordings from dummy heads that are not replicating a carefully selected human ear is poorer than with recordings from the majority of other human subjects (Minnaar et al. 2001; Vorländer 2004).

Also, the directional resolution in dummy head DTFs appears to be poorer than in the average subject. As seen in the top panel of Fig. 2.3.9, the LE for template-matched localization at the reference locations (Eardrum, ECEbl) is increased in the dummy heads as compared to the subjects. The correspondence between dummy heads and subjects in terms of the LE with the hearing device DTFs depends strongly on the microphone location. Therefore, the relation of the LE between the reference and hearing device condition seen in the dummy heads does not always reflect the situation in the subjects. For instance, considering the case of template-matched localization, the dummy head data predict a decrease in LE for the InsertHP with respect to the reference (particularly for the *Brüel&Kjær*HATS, but also for the other heads), whereas an increase of LE is noted for the distribution of all subjects. This means that the directional resolution in the dummy head DTF is better at the InsertHP than at the eardrum or ECEbl. Interestingly, this particular observation exactly reproduces one result from Durin et al. (2014)—they observed a better directional resolution for the DTF of an ITC hearing aid than the ECEbl, also measured in a *Brüel&Kjær*HATS. Knowledge about acoustic origins of such effects would be very helpful for the design of both hearing devices as well as new artificial pinnae, but a detailed discussion would be beyond the scope of this investigation. One possible explanation, however, might be a larger symmetry in the artificial pinnae as compared to the human ears, which is reduced by filling a portion of the cavum concha as with an ITC hearing aid or the utilized InsertHP.

On the other hand, for the QE describing coarse localization errors (see Fig. 2.3.9), the dummy head data reflect well the distribution of subject data for the reference and hearing device locations. Also, the spectral deviations between eardrum and hearing device DTF are consistent between the dummy heads and human subjects. This is both true in the quantitative sense (ISSD re eardrum DTF, Fig. 2.3.5), considering a smaller experimental uncertainty) as well as in the qualitative sense, shown by the ISSD reduction through scaling (Fig. 2.3.6).

To summarize, the present results indicate that evaluation of directional information in hearing devices using dummy heads that have non-human pinnae results in a bias of the localization performance results as compared to measurements in human pinnae. If dummy heads are used for such tasks, our results suggest that a set of different pinnae that are replicated from carefully selected human subjects should be used, such as the systems in (Christensen et al. 2000; Harder et al. 2015; Lindau and Weinzierl 2006).

2.3.6 Conclusions

We analyzed the spectral directional information captured at different hearing device microphone locations as described by the DTF. Our observations confirm various findings from previous studies: The unbiased DTF observed at the eardrum is also obtained at the blocked ear canal entrance. For other microphone locations, a directionally dependent error with respect to the DTF at the eardrum is observed in frequencies >4 kHz for in-concha and >2 kHz for behind-the-ear locations. These errors can be expected to impair sound localization in the vertical domain.

In contrast to previous studies, we were here able to analyze these errors and predicted impact on sound localization in many individual human ears and dummy heads, and performed detailed analyses regarding the qualitative aspects of the errors in the DTFs, also in relation to between-subject differences in DTFs. Our findings can be summarized as follows:

— Considerable variations of the DTF errors and localization performance are observed between human ears and interact with the microphone location and device style. The ranking of device styles differs between ears.

— Errors in hearing device DTFs are centered around the frontal incidence direction.

— Prominent acoustic features in the DTF are rather destroyed than shifted or compressed. The error in hearing device DTFs is qualitatively different from the deviation of the DTFs between two subjects, where the features are mainly shifted in frequency.

— Acute sound localization in the vertical domain is predicted to be impaired by these errors. However, DTFs observed at most hearing device microphone locations except BTE) still contain directional information that could be learned, albeit poorer than in open-ear DTFs.

— Dummy head data do not always produce the same evaluation results as data from human subjects, irrespective of the inaccessible inter-subject variations.

Regarding microphone placement in hearing devices, the following practical conclusions can be drawn:

— Only when the microphone is placed in the ear canal or at its entrance, full spatial information can be conserved.

— For ITE type devices, it is important that the microphone is placed as shallow as possible in the cavum concha. Apart from this, if the shape of the ear is altered the exact shape of the device is less critical. That is, the results with a small insert headphone and an individualized ITE device are not obviously different. The size of the device is not necessarily a predictor of the captured directional information. Microphone positioning closer to the ear canal entrance did not result in better directional cues than in the rear part of the concha.

— BTE microphones capture virtually no spectral directional cues. There, the exact positioning of the microphones seems to be irrelevant, at least for spectral cues.

Reduced localization abilities remain a highly relevant issue in hearing support devices. The current contribution provides starting points for future research and development that should be made to overcome these difficulties. In future work, listening experiments are necessary to improve the understanding of how errors in spectral directional cues relate to localization abilities.

Acknowledgements

This research was funded by the German Research Council DFG through the Research Unit FOR1732 *Individualized Hearing Acoustics* and the Cluster of Excellence EXC1077 *Hearing4all*. We acknowledge Virginia Best and two anonymous reviewers for very helpful comments on an earlier version of the manuscript.

Notes

[1]See supplementary material at https://doi.org/10.1121/1.5056173 for a depiction of DTFs analogous to Fig. 2.3.3 for the left and right ears of all subjects and dummy heads for the median plane and sagittal plane at $\alpha = 30°$; a depiction of the SI for acute localization analogous to Fig. 2.3.7 for the left and right ears of all subjects and dummy heads for the median plane and sagittal plane at $\alpha = 30°$; a depiction of the SI for template-matched localization analogous Fig. 2.3.8 to for the left and right ears of all subjects and dummy heads for the median plane and sagittal plane at $\alpha = 30°$; and the sound localization results analogous to Fig. 2.3.9 obtained with a modified model of Baumgartner et al. (2014).

3 Research platforms and sound equalization strategies

3.1 The acoustically transparent hearing device: Towards integration of individualized sound equalization, electro-acoustic modeling and feedback cancellation

Outline and context within the thesis

In this section, the real-time demonstrator of an acoustically transparent hearing device, as of the beginning of this thesis, is presented and discussed. Apart from the approaches to achieving acoustic transparency, as previously described in (Denk et al. 2018c), it includes contributions by the coauthors that are to be integrated in the demonstrator platform. This includes a review of methods and data on sound equalization, electro-acoustic models for prediction of sound pressure at the aided eardrum, and feedback-cancellation. Regarding acoustic transparency and sound equalization, new evaluation data are shown that reveal shortcomings of the described state. Furthermore, possible interactions of a multi-channel feedback cancellation approach tailored to the multi-microphone earpiece from (Denk et al. 2018c) with the aim to achieve transparency are discussed. A revised version of the demonstrator that includes many of the results shown in this thesis is described and evaluated in Sec. 4.3.

This section is a formatted reprint of
F. Denk, S. Vogl, H. Schepker, B. Kollmeier, M. Blau, and S. Doclo (2017d). "The Acoustically Transparent Hearing Device: Towards Integration of Individualized Sound Equalization, Electro-Acoustic Modeling and Feedback Cancellation". *Proc. 1st International Workshop on Challenges in Hearing Assistive Technology (CHAT-2017)*. Stockholm, Sweden, pp. 89–94.

Note: In Sec. 3.1.5, the notation for the incidence angle was changed to φ here for consistency.

Author contributions: FD was responsible for the paper as a whole and Secs. 3.1.1, 3.1.2, 3.1.3 and 3.1.6. SV was responsible for Sec. 3.1.4; HS was responsible for Secs. 3.1.5. Responsibilities in the individual sections included development of the presented methods, data collection, evaluation, and creation of figures. All authors participated in writing the manuscript.

Abstract

Assistive hearing devices often suffer from a low acceptance by the end user due to poor sound quality. Recently, a novel acoustically transparent hearing device was developed that aims at increasing the acceptance and benefit, also for (near-to) normal-hearing people, by providing better sound quality. The hearing device integrates three microphones and two receivers and can be calibrated in-situ in an attempt to conserve the open-ear sound transmission characteristics of an individual person. To further improve the quality of acoustic transparency and extend the functionality of the hearing device, we outline the integration of further models and algorithms. Electro-acoustic models of the device can improve adjustment to transparency by providing a better estimate of the pressure at the eardrum with an in-ear microphone. In addition, the multi-microphone device layout allows the development of custom feedback cancellation algorithms by means of a beamformer in order to robustly steer a spatial null towards the hearing device receiver.

3.1.1 Introduction

Despite a great improvement in hearing technology in the past decades, the acceptance of assistive hearing devices is still limited, partially due to poor sound quality (Doclo et al. 2015; Killion 2004b; Kinkel 2016). This is particularly true for potential first-time users with a mild-to-moderate hearing loss or even (near-to) normal hearing. While they would benefit from features like speech enhancement or amplification in acoustically challenging situations, they are usually not willing to accept a general degradation of the sound quality. Therefore, an important challenge is to develop a device that is acoustically transparent, i.e., that allows hearing comparable to that of the open ear while being capable of providing a desired sound enhancement at the eardrum. These principles can be applied not only to hearing aids, but also to consumer products, e.g., hearables (Härmä et al. 2004; Hoffmann et al. 2013a).

We recently developed a prototype of an acoustically transparent hearing device that can be individually calibrated aiming to preserve the open-ear sound transmission characteristics of the particular user, even if the ear canal is partially occluded (Denk et al. 2018c). The used sound equalization approach exploits the microphone positions of a novel vented multi-microphone earpiece, including an in-ear microphone for monitoring the pressure in the ear canal. Acoustical transparency on the perceptual level was verified in a subjective listening experiment (Denk et al. 2018c), and convincing sound quality with the device was observed for normal hearing subjects (Denk et al. 2016). Nevertheless, the need for improving transparency in a physical sense was revealed in a recent technical evaluation (Denk et al. 2017b). Furthermore, other processing stages might interact with the desired goal of acoustic transparency.

After presenting the hardware of the device in Sec. 3.1.2, in this paper we first review the sound equalization approach to achieve acoustic transparency in Sec. 3.1.3, and then present two approaches that aim at improving and completing its functionality towards a full acoustically transparent hearing device. To improve the acoustic transparency feature, a promising approach is to include electro-acoustic models of the device (Vogl et al. 2016). These models provide an accurate estimate of the sound pressure at the eardrum with an in-ear microphone, which is key to precise sound equalization in a non-occluding fit. Principles and first results comparing the estimated and measured pressure at the eardrum are outlined in Sec. 3.1.4. In addition, the multi-microphone hardware layout facilitates feedback cancellation using a beamformer with a spatial null steered towards the receiver (Schepker et al. 2016a,b, 2017b) in addition to state-of-the-art adaptive feedback cancellation methods (Schepker et al. 2017a). The principle is briefly introduced and potential interactions of the null-steering approach with the aim of providing acoustic transparency are evaluated in Sec. 3.1.5. Challenges resulting from integrating all approaches are discussed in Sec. 3.1.6.

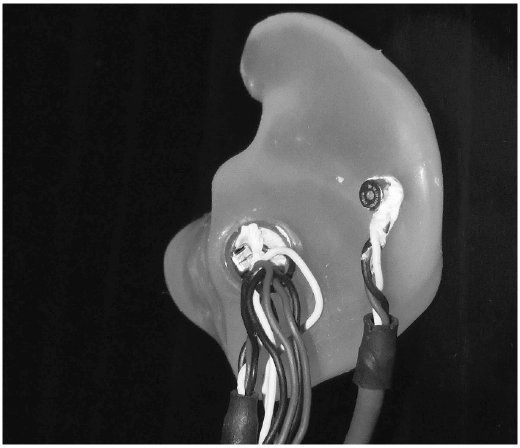


Figure 3.1.1: *Assembled earpiece, from (Denk et al. 2018c). 3 microphones and 2 receivers are fitted into an individual silicone earmould.*

3.1.2 Hardware

The custom in-the-ear type earpiece with relatively open acoustic properties is depicted in Fig. 3.1.1. A schematic drawing is shown in Fig. 3.1.2, together with the filter stages in sound equalization (see Sec. 3.1.3) and feedback cancellation (see Sec. 3.1.5), as well as references to the electro-acoustic model (see Sec. 3.1.4).

All electronic components are removably fitted into an individual silicone earmould that fills the concha bottom. In total, the device contains 3 microphones and 2 receivers. Two microphones (Type *Knowles GA-38*) and two balanced armature receivers are located in an acrylic tube referred to as the core, which is inserted into a bore through to the ear canal. The first microphone is located at the inner face of the core and points towards the eardrum ("in-ear microphone" with output voltage y_1 and pressure p_1) and serves to monitor the sound pressure in the ear canal. The second microphone is located at the outer face of the core and points outwards ("entrance microphone", with output voltage y_2 and pressure p_2). The third microphone ("concha microphone", Type *Knowles FG-23329*, with output voltage y_3 and pressure p_3) is placed in the back of the concha by flush insertion into a hole. The two independent receivers are positioned next to the microphones at both ends of the core, but both pointing towards the eardrum. The inner one is a tweeter (*Knowles WBFK-30019*, with input voltage u_1) and the outer one a woofer (*Knowles FK-26768*, with input voltage u_2). Note that although included in the electro-acoustic model, the woofer is not currently used in operation, i.e., it is not considered in sound equalization and feedback cancellation, which is indicated by a dashed line in Fig. 3.1.2. The hearing device is connected to a PC for real-time signal processing via a soundcard and a custom supply and amplifier box.

The residual space in the core between the microphones and receivers forms a vent, to increase wearing comfort by ventilation and reduction of the occlusion effect (Blau et al. 2008; Winkler et al. 2016). This also implies that sounds below 1 kHz reach the eardrum without considerable attenuation, and the frequency response of the receivers is restricted to above ca. 800 Hz (Denk et al. 2018c).

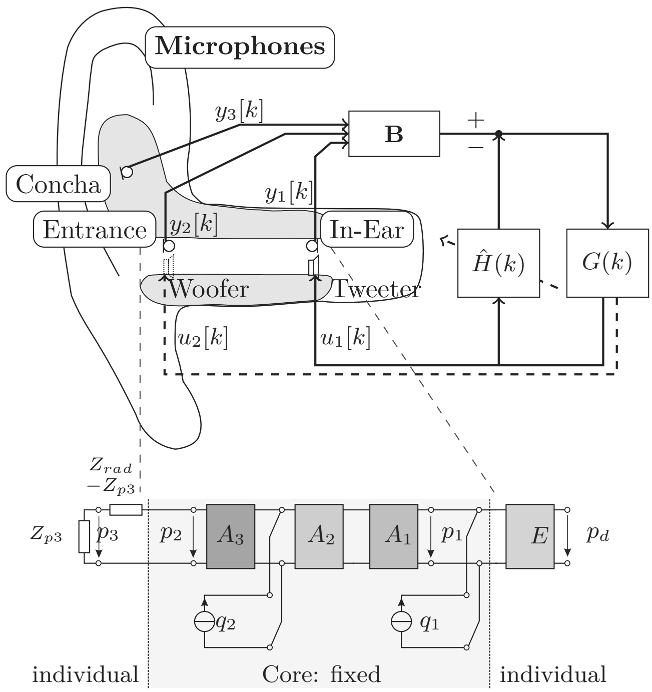


Figure 3.1.2: *Top part: Schematic drawing of the hearing device with filter stages for sound equalization and feedback cancellation. Lower part: Corresponding elements and circuit of the electro-acoustic model.*

3.1.3 Achieving acoustic transparency by individualized sound equalization

3.1.3.1 Principles

Acoustic transparency is achieved, when the superposition of direct sound leaking through the core and the electro-acoustically reproduced sound at the eardrum is physically or perceptually equal to the pressure that would be present with an open ear. Achieving acoustic transparency can be separated into two problems: First, the pressure at the eardrum with an open ear has to be estimated based on the available microphone signals to compute the so-called target pressure. Second, the device has to be adjusted such that the target pressure is generated at the eardrum of the individual subject when the device is in the ear, i.e., sound equalization is performed.

In (Denk et al. 2018c), the target pressure was defined as the pressure at the concha microphone, multiplied with an appropriate frequency-dependent gain function. This strategy is justified by observations from spatial audio technology showing that the relative transfer function between a recording point near the (blocked) ear canal entrance and the eardrum of an open ear is not direction-dependent

(Møller 1992). Thus, the concha microphone approximately contains the direction-dependent portion of the transfer function to the eardrum, and the optimal gain function is the relative transfer function between the concha microphone location and the eardrum in the individual ear. In (Denk et al. 2018c), a flat gain function was used, with the extension that the direct sound leaking through the individual core is considered.

To achieve sound equalization to the target pressure, the filter G of the hearing device is adjusted in a calibration routine conducted in-situ, i.e., when the device is inserted into the ear. The concha microphone is used to pick up external sound. Assuming that the pressure at the eardrum and the in-ear microphone are similar, the pressure at the eardrum generated by the external sound source and the active device is estimated using the in-ear microphone. Based on the observed deviation from the target pressure, the filter G is adapted until convergence is achieved.

3.1.3.2 Current limits and possible extensions

In psychoacoustic experiments with normal-hearing subjects, satisfactory results in terms of acoustic transparency on the perceptual level were observed (Denk et al. 2016, 2018c). However, physical evaluations still reveal some deficits with the current sound equalization approach. Figure 3.1.3 shows measurements of the Real-Ear Insertion Gain (REIG) of the transparent hearing device prototype as presented in (Denk et al. 2018c). The REIG is the difference between the sound pressure at the eardrum measured when the device is inserted and with an open ear. Acoustic transparency on a physical level is achieved if the REIG is 0 dB across all frequencies. The measurements were conducted in a free-field environment in both ears of 12 subjects, and include 3 incident directions in the horizontal plane (azimuth $\varphi = 0°, 90°, -135°$).

The measured REIGs deviate from 0 dB, particularly for frequencies above 2 kHz. The error is notably different between subjects and incidence directions, and the variation increases with frequency. This result shows that there is room for improvement in acoustic transparency, which may be tackled with various approaches.

Most of the observed error in the REIG can be explained by two factors: errors in the estimation of the target pressure, and inaccuracies in the sound equalization due to incorrect estimation of the pressure at the eardrum. To estimate the pressure at the eardrum, in (Denk et al. 2018c) the pressure at the in-ear microphone was used, which in most cases introduces an individual estimation error of up to ± 20 dB, which is highly variable across frequencies (Denk et al. 2017b). Thus, the sound equalization error could be reduced if a better estimate of the pressure at the eardrum were available. Electro-acoustic modeling approaches can be used for this purpose, which are treated in Sec. 3.1.4.

In addition, the occurrence of acoustic feedback due to the acoustic coupling between the receiver and the concha microphone has been neglected so far. While this is possible when only the concha microphone is utilized for sound pickup and

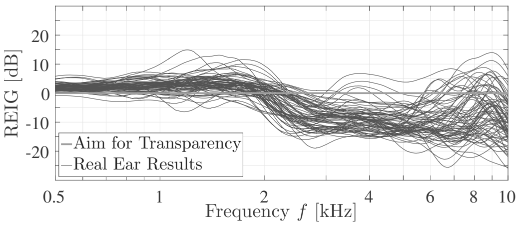


Figure 3.1.3: *Real-Ear Insertion Gain (REIG) after 1/6 octave smoothing measured with the hearing device prototype as presented in (Denk et al. 2018c). The data includes measurements in both ears of 12 subjects, with 3 incident directions in the horizontal plane.*

the applicable gain is limited, appropriate feedback management is a prerequisite when larger amplification than for acoustic transparency is required, or when both external microphones are used for sound pickup, e.g., when implementing a directional microphone. Feedback cancellation techniques tailored to the custom hardware layout are reviewed in Sec. 3.1.5, where possible interactions with acoustic transparency are examined.

3.1.4 Electro-acoustic model

In our previous work (Vogl et al. 2016), we proposed an electro-acoustic model, which serves to better understand the underlying physical principles of sound transmission in the hearing device, and to estimate quantities at locations where they cannot be directly measured, e.g., the sound pressure at the eardrum. The current focus is to predict the sound pressure at the eardrum p_d in *vivo*, based on measurements using the microphones of the hearing device only.

The model is made up of lumped elements and two-port networks, as depicted in Fig. 3.1.2. The middle part is the core, which can be regarded as fixed over individual subjects. On the other hand, both terminations, i.e. the external sound field and the residual ear canal, are individual to every ear. The complete model cannot be determined in one step, but is built up in a series of measurements and calculations that are described in the following.

3.1.4.1 Model of the core

First, the model of the core is obtained. It consists of:

— two microphones, characterized by their sensitivity measured prior to assembling the core, each converting its output voltage signal y_m to the corresponding pressure p_m.

— two receivers, which are modeled as ideal volume velocity sources, delivering the flux q_n. This source parameter was also measured prior to assembling the core, according to the technique described by Blau et al. (2010).

— the vent, represented by three acoustic transmission lines modeled as two-port networks $A_{1,2,3}$ according to Keefe (1984). The three parameters of each transmission line (length, radius and a loss factor) need to be fitted by referring to acoustic measurements. The microphones and receivers are coupled into the vent at locations depicted in Fig. 3.1.2.

To fit the free parameters of the transmission lines, the assembled core was coupled to a training setup with known termination impedances, and all four transfer functions between the two receivers and microphones 1 and 2 were measured. The medial termination was an IEC711 coupler, while at the lateral end the core was mounted in a baffle. The optimal parameters of the two-port networks were found by minimizing the differences between the measured and modeled transfer functions. Good agreement and computational effectiveness could be achieved with the Nelder-Mead-Simplex (Nelder and Mead 1965) algorithm, where the parameter values were constrained to realistic boundaries.

3.1.4.2 Model of the individual ear

In a second step, a model of the individual ear is estimated. It contains both terminations of the core, as shown in Fig. 3.1.2. The external sound field (outer termination of the core) is characterized by the radiation impedance Z_{rad}, which can be further split into the transfer impedance Z_{p3} between the outer core end and the concha microphone, and a remaining impedance $Z_{rad} - Z_{p3}$. Z_{rad} is approximated by the physical model of a piston in baffle. The model of the individual ear canal E (medial termination of the core) is individualized based on measurements in the ear of a subject. It is composed of four cascaded acoustic transmission lines with four radii and one total length as parameters, and two parallel load impedances Z_l and $Z_{l,residual}$ located medially (across them p_d is produced). Z_l is a purely resistive frequency-independent load to represent losses, $Z_{l,residual}$ is complex-valued and frequency dependent.

Assuming the core model and the outer termination are known and using the transfer function measurements from any of the two receivers to the in-ear microphone, the acoustic impedance Z_{ec} at the point of p_1 (i.e., the in-ear microphone) in the direction towards the ear canal can be calculated. Then, the parameters of the individual ear canal model E are fitted by minimizing both level and phase differences between measured and modeled impedances Z_{ec}, summed across the frequencies from approximately 1 to 15 kHz. Again, the Nelder-Mead-Simplex algorithm was applied with realistic boundaries. Since it was observed that results depended on initial values, 500 random initial values were used and the lowest cost result taken.

Several studies (e.g. Hudde et al. 1999, Sankowsky-Rothe et al. 2015) have shown that Z_{ec} - or the reflectance derived from it - can be used to estimate an ear canal model E and ultimately predict the sound pressure at the eardrum p_d.

3.1.4.3 Evaluation

The individual ear canal model E and the predicted pressure at the eardrum p_d were evaluated by means of probe tube measurements in 12 subjects.

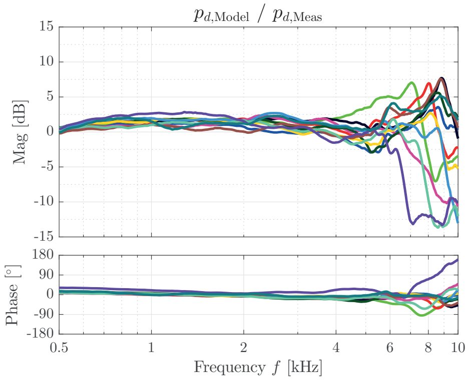


Figure 3.1.4: *Deviation between the pressure at the eardrum predicted by the model $p_{d,\text{Model}}$ and the one measured with a probe tube microphone $p_{d,\text{Meas}}$, for twelve subjects, with the woofer as sound source.*

The differences between the model predictions and the measurements of p_d created by the woofer are shown in Fig. 3.1.4. Below 6 kHz, the agreement in both magnitude and phase is very good. However, for higher frequencies, the differences increase. It should be noted that in this frequency range the probe tube measurements are more likely to be corrupted by errors, as the tube had to be in place together with the earmold which made visual inspection of its position impossible. Furthermore, around 8 kHz the core has a low source impedance, i.e., the impedance at the point where p_1 is measured towards the lateral direction is low compared to typical ear canal impedances Z_{ec}. This reduced measurement accuracy may additionally lead to deviations between the estimated and measured sound pressure.

3.1.5 Feedback cancellation

Acoustic feedback occurs when a signal is picked up by a microphone, amplified, played back by a receiver and picked up again by the microphone, creating a closed-loop system. In hearing devices, adaptive feedback cancellation (AFC) is typically used to reduce the detrimental effect of acoustic feedback, which is most often perceived as howling or whistling. In AFC, an adaptive filter is used to estimate the acoustic feedback path between the hearing device receiver and the microphone, theoretically allowing for perfect feedback cancellation (Waterschoot and Moonen 2011). However, due to the closed-loop electro-acoustic system, the estimate of the acoustic feedback path is generally biased (Siqueira and Alwan 2000; Spriet et al. 2005). Several algorithms have been proposed with the aim of reducing this bias, where the so-called prediction-error-method (Spriet et al. 2005) seems most promising. While an AFC algorithm can be applied for any hardware layout, the considered multi-microphone setup (c.f. Fig. 3.1.2) additionally allows for the use of multi-microphone feedback cancellation approaches. This includes a fixed null-steering beamformer that exploits the spatial diversity of the microphones to steer a spatial null towards the position of the hearing device receiver. Note that only the inner receiver of the device is considered here.

Several optimization approaches for calculating the null-steering beamformer coefficients have been proposed, including a robust least-squares design (Schepker et al. 2016a,b) and a robust min-max design (Schepker et al. 2017b) aiming at directly maximizing the maximum stable gain of the hearing device, i.e., the gain before the closed-loop system becomes unstable. Furthermore, the benefit of combining a fixed null-steering beamformer and an AFC algorithm based on the prediction-error-method to cancel residual feedback has recently been shown (Schepker et al. 2017a). However, in none of the presented null-steering beamformer optimization approaches (Schepker et al. 2016a,b, 2017b), the preservation of the pickup microphone directional response that is required for achieving acoustic transparency has been taken into account. This implies that the null-steering beamformer may alter spectral directional cues and bias spatial perception, e.g., sound localization. Therefore, after briefly introducing the optimization procedure, in the following we analyze the directional response of the fixed null-steering beamformer.

We assume time-invariance of the acoustic feedback paths $H_m(k) = H_m$, $m = 1, \ldots, M$ between the receiver and the mth microphone. Assuming the availability of I measurements of the acoustic feedback paths (e.g., obtained by prior measurement), the coefficients of the null-steering beamformer $\mathbf{B}$ are obtained by minimizing the following least-squares cost-function (Schepker et al. 2016a)

$$J_{LS}(\mathbf{b}) = \sum_{i=1}^{I} \|(\mathbf{H}^{(i)})^T \mathbf{b}\|_2^2, \tag{3.1.1}$$

where $\mathbf{b}$ is the ML_B-dimensional vector of the beamformer coefficients and $\mathbf{H}^{(i)}$ is the $ML_B \times (L_B + L_H - 1)$-dimensional matrix of concatenated convolution matrices of the acoustic feedback paths from the ith measurement, $i = 1, \ldots, I$, with L_B the number of beamformer coefficients for each microphone and L_H the length of the acoustic feedback path. To prevent the trivial solution of $\mathbf{b} = \mathbf{0}$, the beamformer coefficients in a reference microphone m_0 are constrained to correspond to a delay of L_d samples, i.e.,

$$\mathbf{b}_{m_0} = [\ \underbrace{0 \quad \ldots \quad 0}_{L_d} \quad 1 \quad 0 \quad \ldots \quad 0\ ]^T. \tag{3.1.2}$$

To obtain the beamformer coefficients, we first measured the acoustic feedback paths of the hearing device in the left ear of a dummy head with adjustable ear canals (Hiipakka et al. 2010), both in free-field and with a hand very close to the ear, using a sampling rate of 32 kHz. The beamformer coefficients were then computed by minimizing Eq. (3.1.1) subject to the constraint in Eq. (3.1.2) for $M = 3$ microphones (in-ear, entrance and concha microphone), $L_B = 32$, $m_0 = 2$ (entrance microphone), $L_d = 16$ and $I = 2$. The resulting added stable gain, i.e., the increase in gain margin compared to using only the entrance microphone ($m = 2$), was 18.3 dB and 22.6 dB for the free-field condition and the hand condition, respectively.

To compute the directional response of the null-steering beamformer for an incoming signal, the acoustic transfer functions to the microphones $D(\varphi_j)$, $j = 1, \ldots, J$ were measured for $J = 24$ equidistantly spaced angles φ_j surrounding the dummy head at a distance of approximately 2.5 m in the horizontal plane. Figure 3.1.5 shows the directional response of the beamformer $\tilde{D}(\varphi_j) = \mathbf{B}^T\mathbf{D}(\varphi_j)$ for multiple frequencies relative to the directivity $D_2(\varphi_j)$ of the entrance microphone. Ideally, the relative directional response would be equal to 0 dB for all frequencies and incidence angles. However, the response is different from 0 dB for most of the considered frequencies and directions. Nevertheless, for most frequencies and incident angles the null-steering beamformer alters the directivity only by approximately ±4 dB.

3.1.6 Discussion and summary

The principles of an acoustically transparent hearing device presented in (Denk et al. 2018c) and physical evaluation results have been reviewed in Sec. 3.1.3, and possible extensions towards improving and extending its functionality have been presented in Secs. 3.1.4 and 3.1.5. While the good performance of electro-acoustic modeling and customized feedback cancellation for the hearing device has been demonstrated for these approaches individually, a next challenge is the integration of the two approaches with the transparency feature of the hearing device in real-time operation.

Unbiased estimation of the pressure at the eardrum can improve acoustic transparency by improving sound equalization to a target pressure at the eardrum. The electro-acoustic model presented in Sec. 3.1.4 is able to predict the pressure at the eardrum that is generated by the hearing device receiver accurately up to approximately 6-7 kHz in magnitude and phase. However, when estimating the sound pressure at the eardrum in normal operation, the superposition with the direct sound leaking through the vent needs to be considered. Since the model is in principle also able to predict the pressure generated by the receiver at the in-ear microphone, this predicted pressure can be subtracted from the observed pressure at the in-ear microphone to obtain an estimate of the direct sound only. The pressure at the eardrum generated by the direct sound alone can then also be

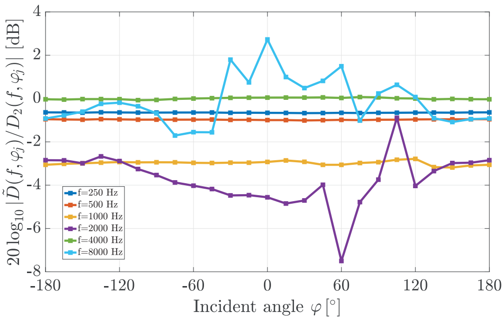


Figure 3.1.5: *Directional response of the beamformer output relative to the directional response of the entrance microphone ($m = 2$) as a function of the azimuth φ.*

predicted. It should be noted that for integrating the electro-acoustic model with the acoustically transparent hearing device, it is sufficient to extract all relevant transfer functions from the model after calibration measurements.

Although the null-steering beamformer presented and evaluated in Sec. 3.1.5 yields impressive results in terms of feedback cancellation, it was also noted that it introduces a direction-dependent bias compared to the reference microphone. Spectral directional cues contained in the reference microphone signal are thus altered, which may introduce perceptual errors regarding spatial hearing or other undesired artifacts. However, the deviations are in the range of about ± 4 dB, and their perceptual relevance is not yet clear. Another issue is the delay of L_d samples introduced by the beamformer, which should be considered when designing the equalization filter G. In principle, this can be achieved by performing the in-situ calibration (Denk et al. 2018c) with the beamformer output as hearing device input signal.

The electro-acoustic models of the individual ear, as well as the null-steering beamformer require knowledge of the transfer functions between the hearing device receivers and microphones. They can be measured in-situ in the individual ear using only the device as part of calibration measurements, which are also necessary to achieve transparency (Denk et al. 2018c). Gathering these data is therefore no practical obstacle to integrating the electro-acoustic model and the null-steering beamformer into a future version of the prototype.

In conclusion, there seem to be no principal problems hindering the integration of electro-acoustic models and customized feedback cancellation methods into our prototype hearing device. Both are promising approaches to improving the sound equalization to achieve acoustic transparency in our prototype hearing device, as well as increasing its functionality in more realistic application scenarios with higher gain settings and more than one pickup microphone in each side. Future work will hence focus on the implementation of the presented approaches to construct an improved version of our acoustically transparent hearing device.

Acknowledgement

This work was supported by the Research Unit FOR 1732 *Individualized Hearing Acoustics* and the the Cluster of Excellence *Hearing4all*, both funded by the German Research Council DFG.

3.2 A one-size-fits-all earpiece with multiple microphones and drivers for hearing device research

Outline and context within the thesis

In this section, the design and acoustic evaluation of a revised earpiece for the use in the real-time demonstrator, is described. The key novelty compared to the earpiece from (Denk et al. 2018c) was transferring the design to a generic shell, where all transducers are included, while the general layout and transducer placement were conserved as well as possible. The development was conducted to have a more stable hardware for long-term availability, dissemination and field-tests of technology for acoustically transparent hearing devices developed within this thesis and collaborating projects. The development was conducted together with InEar GmbH, Dieburg, Germany, and the resulting device was made available to the public as a commercial product.

Abstract

Earpieces that include one or more microphones and drivers are required in many research applications related to hearing devices, however suitable devices are often not readily available. In this contribution, we present the development and evaluation of an earpiece for research on assistive hearing devices and hearables. The earpiece includes two balanced armature drivers as well as four microphones, which are built into a one-size-fits-all acrylic shell. It features custom transducer positioning at different positions inside a vent, as well as a microphone inside the ear canal. We discuss details on the earpiece design, present acoustic measurements and discuss the eligibility for different applications. The earpiece is openly available both in a vented as well as an occluded version.

This section is a formatted reprint of
F. Denk, M. Lettau, H. Schepker, S. Doclo, R. Roden, M. Blau, J. Bach, J. Wellmann, and B. Kollmeier (2019a). "A one-size-fits-all earpiece with multiple microphones and drivers for hearing device research". *Proc. AES Conference on Headphone Technology.* Paper 13. San Francisco, USA, pp. 1–9.

Author contributions: FD led and managed the acoustic design, ML implemented the mechanic design and managed the production in close cooperation with FD. HS, SD, RR, MB, JB and BK participated in the acoustic design choices. JW developed the in-ear microphone pre-amplifiers. FD implemented and made the acoustic measurements, created the figures and wrote the manuscript, all authors participated in writing the manuscript.

3.2.1 Introduction

All research and development into better assistive hearing devices requires suitable electro-acoustic hardware that can be placed in the ear. Although the specific hardware requirements obviously depend on the application, they generally include the availability of at least one microphone to pick up sound at the ear level, at least one driver to play back sound, and a stable housing that allows testing in real-world settings. However, for many applications no suitable devices are available without larger efforts.

The inavailability of suitable electro-acoustic hardware constitutes a barrier for many researchers, or at least leads to time consuming and tedious work that has to be done in each laboratory independently. Researchers have pursued many different approaches to fulfill their own specific needs, like rewiring commercial hardware (Härmä et al. 2004; Liebich et al. 2016), manually attaching microphones to earphones (Albrecht et al. 2017; Hiipakka et al. 2010; Hoffmann et al. 2014), or building custom devices from scratch (Denk et al. 2018c). While manufacturers of headphones or hearing aids may provide special 'dummy' or 'satellite' devices with externally wired transducers, this typically happens inside specific projects and is thus often accessible for selected researchers only. Besides the excessive time spent on the hardware issue, the resulting zoo of custom approaches leads to poor comparability of results from different laboratories.

In this contribution, we present the design and acoustic evaluation of an in-the-ear type earpiece that is suitable for many current research topics related to hearing aids, hear-through headphones and active noise control. The earpiece is made openly available. It builds upon an earlier hand-made prototype presented in (Denk et al. 2018c), which had enabled a number of novel signal processing applications (Denk et al. 2017d, 2018d; Schepker et al. 2019a; Schepker et al. 2018; Vogl and Blau 2019). This first prototype comprised an earmould with a temporarily inserted hand-assembled transducer pack, which led to a large variance between devices and poor mechanical stability. Also, an individual earmould had to be made for each user. To overcome these issues, in the present work we transferred the key aspects of the first prototype into a new earpiece.

The new earpiece contains four microphones, including one in the ear canal, and two Balanced Armature (BA) drivers, which are positioned inside and around and a vent. All transducers are built into a stable one-size-fits-all acrylic shell that fits into about 90% of human ears and sits shallow in the cavum concha. The device can be connected to an arbitrary sound processing platform via a single flexible and sturdy cable. Two versions of the device are considered here: a vented version, and a closed version where the outer part of the vent is completely filled.

The design of the earpiece is described in detail in Sec. 3.2.2. Acoustic evaluation measurements and results are described and discussed in Sec. 3.2.3, followed by a summary and outlook in Sec. 3.2.4.

3.2.2 Earpiece design

In this section, the mechanical and electro-acoustic design of the earpiece is explained. In Sec. 3.2.2.1, the general layout is described, and the transducer placement is explained in Sec. 3.2.2.2. Details on the form of the shell are presented in Sec. 3.2.2.3, followed by details on the selected transducers and microphone preamplifiers in Secs. 3.2.2.4 and 3.2.2.5, respectively. Note that first, the design of the vented version is outlined, differences in the closed version are described in Sec. 3.2.2.6.

3.2.2.1 General layout

The general design of the earpiece is based on the first prototype presented in (Denk et al. 2017d, 2018c; Ernst et al. 2015). It enabled several novel approaches to individualized sound equalization (Denk et al. 2018d; Schepker et al. 2018), feedback cancellation (Schepker et al. 2019a) and electro-acoustic modeling (Vogl and Blau 2019). As discussed above, the first prototype had several issues regarding stability, usability and variation between devices. The aim of the new earpiece is therefore to overcome these issues while maintaining the properties of the first prototype.

To this end, in the new earpiece, all components are built into one acrylic shell that can be directly inserted into the ear. Figure 3.2.1 shows the CAD model of the new earpiece. The key features of the earpiece are:

- An in-the-ear fit comparable to an individual earmould.
- A vent containing several transducers, enabling electro-acoustic modelling.
- 3 microphones at the outer surface and the outer end of the vent to enable spatial sound processing.
- An in-ear microphone to monitor the sound pressure in the ear canal and at the eardrum.
- Two BA drivers with separate connections to enable multi-loudspeaker sound processing or two-way playback with a high bandwidth and quality.

The acrylic housing consists of 4 separate parts (see Fig. 3.2.1): 1) a shell (gray) 2) a faceplate (violet), which together with the shell forms the outer shape of the earpiece. 3) a vent tube (blue), which extends from the faceplate into the ear canal (some mm past the shell) and houses most transducers. 4) a ring (gray) at the inner end of the vent, which serves as a fixture for the silicone dome and includes the output port and a cerumen filter. All transducers of the earpiece can be connected to a sound processing platform by means of a 1.20 m long flexible 9-pin cable with a Sub-D 9 plug.

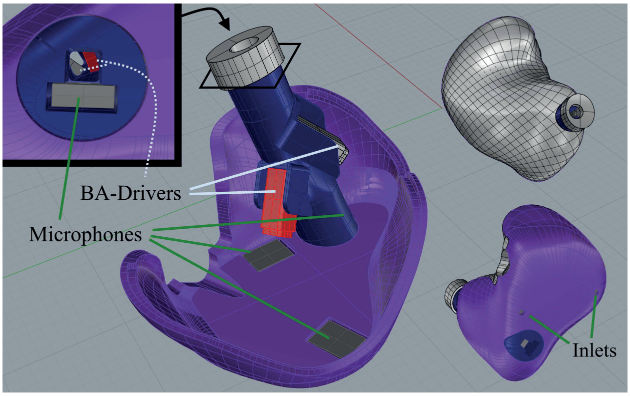


Figure 3.2.1: *CAD model of a left earpiece. Center: Interior view with shell removed, right: Inside and outside view of the complete earpiece, top left: View into the vent from the ear canal side, ring removed (plane and viewing direction indicated by the black square and arrow). See text for further details.*

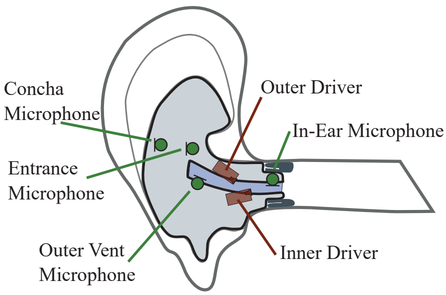


Figure 3.2.2: *Schematic layout and placement of the transducers. The shaded blue area denotes the vent, the light gray area the body of the earpiece. The silicone dome in the ear canal is drawn in dark gray.*

Figure 3.2.3: *Photograph of a right device, outside (left) and inside (right) view. The Entrance and Concha microphones are visible through the transparent acryl, as well as the Outer Vent microphone inside the vent.*

3.2.2.2 Transducer placement and design of the vent

The schematic placement of the transducers is shown in Fig. 3.2.2. As already mentioned, a key feature of the device is the special vent (blue area in Fig. 3.2.2) with several transducers coupling into it. This includes microphones at the inner and outer ends (termed *In-Ear* and *Outer Vent* Microphone, respectively), as well as two separate and different BA drivers that couple into the vent at different positions.

The vent has an overall length of 19.2 mm. The microphone inlets are placed close to both ends at its wall, with a 2-3 mm distance to the end to avoid effects of jumps in cross-section (Stinson and Daigle 2007). The drivers couple into the vent at 8.1 and 11.4 mm from the inner end (excluding the ring and cerumen filter). They are referred to as *Inner* and *Outer* Driver according to their location. Both BA drivers have a flat outlet, and their ends slightly protrude into the vent while minimizing irregularities of the vent surface (see Figs. 3.2.1 and 3.2.2).

The vent is primarily included to increase the wearing comfort and to improve the own-voice perception by reducing the occlusion effect (Kuk et al. 2009). However, a vent unavoidably introduces also some negative acoustic effects (Blau et al. 2008; Winkler et al. 2016): First, it results in a high-pass filtering of sound reproduced by the drivers. Second, it allows low-frequency sound from the outside to leak into the ear canal, which reduces control of the sound generated at the eardrum. This may lead to artifacts due to superposition with the delayed output of the hearing device and may reduce the effectivity of noise reduction algorithms. Third, the acoustic coupling between the drivers and microphones may result in an unstable system and howling sounds. A vent cross-section of ca. 1.5 mm^2 with a roughly quadratic profile that varies slightly over the length was chosen as a compromise between occlusion reduction and the negative effects, particularly sound leakage (Kuk et al. 2009, c.f. Sec. 3.2.3.1).

Two more microphones are mounted on the faceplate of the device in the cavum conchae. One is located in the rear part (termed *Concha* Microphone), the other is located in the frontal part right above the vent, termed *Entrance* Microphone. The distance between both microphones is 11 mm, and their connecting line lies within the horizontal plane for the average ear.

3.2.2.3 One-size-fits-all shell

A photograph of the assembled earpiece is shown in Fig. 3.2.3. Its outer shape is based on the *InEar* ProPhile, a commercially available generic fit earphone (InEar 2019). Its shape resulted from varying a parametric 3D model to fit inside the maximum percentage of several hundred digitized ear impressions. In a second step, average diameters and angles of ear canals were measured and the innermost part of the device was adapted accordingly. With the resulting shape, a fitting rate of about 90% was achieved.

The innermost part of the vent tube sticks out of the shell and is intended to sit in the outer ear canal, between the first and second bend (see Fig. 3.2.3). This part is elliptic (approx. 6x5 mm) and also serves as a mount for a standard silicone dome whose size can be individually selected. The inner side of the vent tube is terminated by a ring that serves as a fixture for the dome (see Fig. 3.2.1). An exchangeable filter (type HF4) is placed in the ring for protection against moisture and cerumen.

For a secure fit, the cable connecting the transducers is lead to the top of and around the ear. This part of the cable is reinforced by a wire and shrink tubing, which enables to bend it to fit the shape of the individual ear (see Fig. 3.2.3).

3.2.2.4 Transducer selection

Due to space constraints and the desired output port locations, BA drivers were selected over dynamic drivers. Specifically, a two-way reproduction system is used, comprising a *Knowles* FK-26768 as a woofer (Outer Driver, see Fig. 3.2.1) as well as a *Knowles* WBFK-30042 as a tweeter (Inner Driver). The tweeter is placed further towards the ear canal to optimize its high-frequency behavior.

For the microphones, a high SNR, stability and a small size is desired. For the in-ear microphone, a hard constraint on the size is imposed since it is placed at the inner surface of the earpiece, i.e., between the first and second bend inside the ear canal. MEMS microphones were selected over electret condenser microphones, which had been utilized in the first prototype (Denk et al. 2018c). While both feature similar SNRs and comparable sizes, the sensitivity of MEMS depends less on temperature and humidity. Also, the variation between devices is usually smaller than with electret condenser microphones (Lewis and Moss 2013). Specifically, the *Knowles* SPH1642HT5H-1 was selected, which is 2.65x3.5x1mm in size, features a convenient top port location and provides an SNR of 65 dB. For ANC applications, it is noted that its low-frequency roll off starts at 55 Hz.

3.2.2.5 In-ear microphone preamplifiers

To increase the robustness against electromagnetic interferences, crosstalk and to provide line output, microphone preamplifiers were included in the earpiece. They are mounted on a folded flexible Printed Circuit Board (flex-PCB). The transducers are directly mounted on the flex-PCB, such that the assembled electronics can be tested prior to installation into the earpiece. The preamplifiers are supplied by the same contact that provides the microphone supply voltage (3-3.6 V). The four preamplifiers are built as one-stage inverting amplifiers with a gain of 10 dB based on two dual-channel OPA1662-Q1 Op-amp chips.

3.2.2.6 Versions

Besides the vented version of the earpiece as described so far, a completely closed version was developed as well. To this end, the vent was completely filled between the coupling point of the outer driver and the outer end of the vent, which implies that the Outer Vent microphone was omitted in the closed version (c.f. Fig. 3.2.2). This version may be useful in applications where the drawbacks of a vent outweigh its benefits (c.f. Sec. 3.2.2.2), e.g., when larger gains in the low-frequency regime or a higher attenuation of external sounds are required.

3.2.3 Acoustic evaluation

Key acoustic parameters were measured for one vented and one closed right-ear device. Assessed parameters include the insertion loss (i.e., the attenuation of external sounds when the device is inserted, see Sec. 3.2.3.1), headphone transfer functions (Sec. 3.2.3.2), harmonic distortion products and maximum sound pressure (Sec. 3.2.3.3), and feedback paths (Sec. 3.2.3.4). It should be noted that for the measurements, a development version without the built-in microphone pre-amplifiers was used. Instead, custom microphone pre-amplifiers were connected to the cable. All measurements were conducted at levels that were well within the dynamic range of the utilized microphones and the BA drivers.

For all experiments the earpieces were inserted in a *G.R.A.S.* KEMAR 45BB-12 with anthropometric pinnae (*G.R.A.S.* KB 5000/5001, Wille and Rasmussen 2016). This assured a realistic and reproducible coupling to the ear. Depending on the acoustic parameter assessed, the KEMAR was equipped with either low-noise (*G.R.A.S.* 43BB) or standard IEC711 (*G.R.A.S.* RA0045) ear simulators, replicating the acoustic properties of an average ear canal and eardrum. The microphones of the ear canal simulator are referred to as (artificial) eardrums in the following.

3.2.3.1 Insertion loss

Methods
The insertion loss, i.e., the attenuation of the transfer function of external sounds to the eardrum by inserting the passive earpiece, was measured on the KEMAR equipped with low-noise ear simulators. The measurements were conducted in an anechoic chamber featuring a 3D array of 94 *Genelec* 8030 loudspeakers, using overlapping exponential sweeps (Majdak et al. 2007) of 4 s length and a frequency range between 30 Hz and 22.05 kHz, i.e., and half the sampling rate of 44.1 kHz.

Head-Related Transfer Functions (HRTF) of the KEMAR were measured for 47 incidence directions equally distributed on a sphere, once with the ear open (nothing inserted) and 5 times with the passive earpiece reinserted to capture variabilities in the fit. The HRTFs were compensated for residual acoustic reflections by frequency dependent truncation (Denk et al. 2018f) and smoothed over 1/6 octave bands. Then, the average power over directions was calculated and the ratio closed/open ear determined, resulting in the insertion loss for an approximated diffuse field (c.f. Denk et al. 2018b).

Results and discussion
Figure 3.2.4 shows the insertion loss for both versions of the earpiece. The measurement variation between reinsertions is low for both versions, which indicates that a tight fit could be achieved reliably.

In the vented version, external sounds are not attenuated below 500 Hz and the attenuation varies between approx. 10 and 30 dB above 1 kHz. The slight amplification around 350 Hz probably results from a Helmholtz resonance of the residual ear canal and the vent. Generally, this attenuation profile matches typical data for a vent with a 2 mm diameter (Kuk et al. 2009).

In the closed version, an insertion loss of around 15 dB is observed for frequencies below 1 kHz, at higher frequencies it varies between roughly 20 and 40 dB. With this attenuation characteristic, it can be assumed that the influence of direct sound leaking into the ear canal can be neglected in comparison to sound electro-acoustically reproduced in a hear-though or hearing aid application. In human

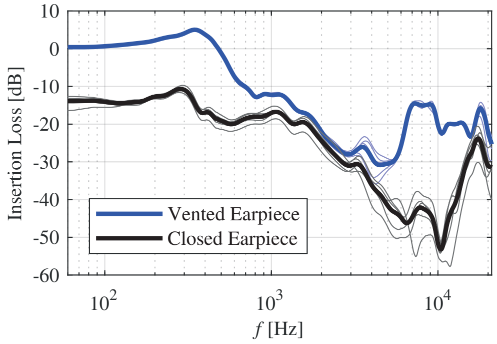


Figure 3.2.4: *Insertion loss for approximated diffuse-field incidence with both versions of the earpiece, obtained in a mannequin with anthropometric ears. Thin lines show results for individual reinsertions of the earpiece, thick lines the average.*

subjects, we expect a larger variation particularly in the low frequencies, since the fit is less controlled (Hiipakka et al. 2010).

3.2.3.2 Headphone transfer functions

Methods

The Headphone Transfer Function (HpTF), i.e. the pressure generated at the artificial eardrum depending on the driving voltage and frequency was measured for both drivers independently in the KEMAR equipped with the standard ear simulators. The measurements were performed using an *Audio Precision* APx525 analyzer and its software (APx500 suite) at a sampling rate of 192 kHz. An exponential sweep covering the range between 30 Hz and 30 kHz with an RMS voltage of 400 mV was used. The BA drivers were fed through a *Lake People* G-103P amplifier adjusted to 0 dB gain.

Results and discussion

Figure 3.2.5 shows the HpTFs for both versions of the earpiece and for both drivers. Differences are evident both between the closed and vented versions as well as between the driver positions/types.

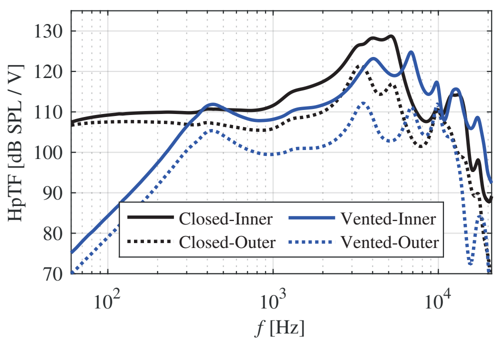


Figure 3.2.5: *Headphone Transfer Functions (HpTFs) for both versions of the earpiece (indicated by the color), and for the inner (solid lines) and outer (dashed lines) driver position, respectively.*

Independent of the driver, the most pronounced difference between the vented and closed version is the low-frequency response. Whereas the low-frequency response is flat in the closed earpiece, for the vented earpiece a bass roll-off with a cut-off frequency of about 300 Hz is observed. This is caused by a reduced acoustic impedance of the unsealed ear canal, which is unavoidable whenever a vent is present. It should be noted, however, that the cut-off frequency for the playback lies around 200 Hz below the cut-off frequency for the attenuation of external sounds (c.f. Fig. 3.2.4). In addition, in a wide frequency range between 800 Hz and ca. 6 kHz, the response is between 5 and 10 dB higher in the closed earpiece as compared to the vented earpiece.

Between both drivers, a general offset in sensitivity of 5-10 dB is evident, where the inner driver has the larger sensitivity. This is mostly caused by selecting different types of drivers featuring different impedances at both positions (c.f. Sec. 3.2.2.4), which has been verified by individual characterization of the driver types. Moreover, their high-frequency responses vary: Whereas the inner driver features a broad peak between 3 and 6 kHz and another peak at about 13 kHz, the response of the outer driver is rather flat below 10 kHz (particularly in the vented earpiece), with only some smaller peaks. Furthermore, the responses of the individual drivers vary between the earpiece versions.

These data verify that the drivers have largely complementary characteristics, as intended by design. High frequencies can better be reproduced by the inner driver, whereas the outer driver mainly supports sound reproduction up to about 10 kHz. The small broadband difference in sensitivity is not expected to be a problem with suitable electronic circuitry.

In both versions of the earpiece, a high reproduction bandwidth far beyond 10 kHz can be achieved, whereas the low-frequency behavior depends on the version.

3.2.3.3 Distortions products and maximum SPL

Methods
Using the same setup and equipment as for the HpTF measurements (see previous section), the Total Harmonic Distortion (THD) generated by the drivers was characterized as a function of the produced sound pressure level (SPL) at the artificial eardrum. To this end, a 1 kHz sine tone was played and the driving voltage was varied between -10 and 10 dBV. The ratio between harmonic distortion components and the playback frequency was determined using the APx525 analyser's software.

Results and discussion
Figure 3.2.6 shows the THD ratios as a function of the produced SPL for both drivers and earpiece versions. As expected, the THD ratios gradually increase with increasing SPL up to the saturation point of the drivers, above which the THD increases rapidly. This point lies at approx. 5% THD and characterizes the maximum SPL that can be generated with a particular driver and earpiece version (dashed horizontal line in Fig. 3.2.6).

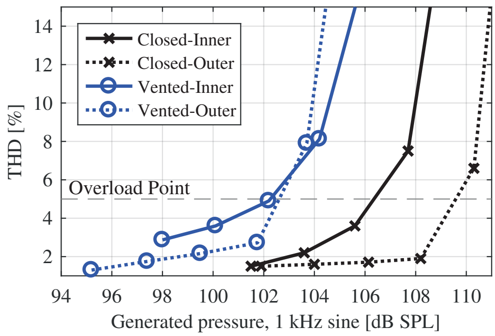


Figure 3.2.6: *Total Harmonic Distortion (THD) ratio as a function of SPL generated at 1 kHz for both versions of the earpiece (indicated by the color), and for the inner (solid lines) and outer (dashed lines) driver, respectively.*

With the vented earpiece, about 103 dB SPL can be generated, independent of the driver. With the closed earpiece, about 109 and 107 dB SPL can be reproduced with the outer and inner driver, respectively. By combining the output of both drivers, a 6 dB increase can, in principle, be achieved. In summary, the achievable SPLs should be sufficient for most applications with normal-hearing or mildly hearing-impaired users. At more moderate SPLs, the THD ratios lie in the typical range for BA drivers.

3.2.3.4 Feedback paths

Methods
The acoustic feedback paths, i.e., the transfer functions between the drivers and the microphones of the earpiece, were measured while the earpiece was inserted into the KEMAR equipped with low-noise ear simulators. Five repetitions with reinserting the device were assessed to capture variabilities in the fit. An exponential sweep with ca. 75 mV RMS, 4 s length and a frequency range between 30 Hz and 22.05 kHz, i.e., half the sampling rate of 44.1 kHz, was used.

Results and discussion
Figure 3.2.7 shows the amplitude of the obtained feedback paths between the inner driver and all microphones for both versions of the earpiece. The feedback paths for the outer driver are very similar. For comparison, the HpTFs to the artificial eardrum (measured together with the feedback paths) are shown. All feedback paths are well reproducible, especially in the vented earpiece. Such reproducibility is typical for in-ear devices as compared to behind-the-ear devices (Sankowsky-Rothe et al. 2015b).

For the vented earpiece, the feedback paths to the microphones in the vent deviate significantly from the feedback paths to the microphones mounted at the faceplate. Especially the broadband 30 dB difference between the feedback paths to the Outer Vent and the Entrance microphone is striking, given the very small distance between both microphones (see Fig 3.2.3). One possible explanation is that the vent emits sound to the outside in a very directional manner. On the other hand, the feedback paths are very similar for the Entrance and Concha microphones.

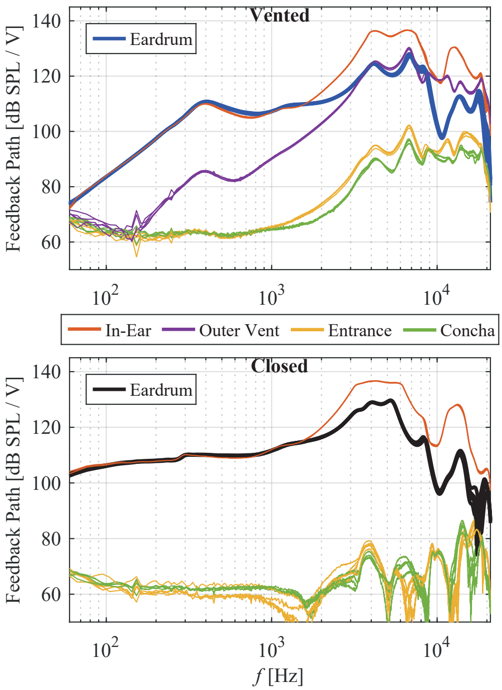


Figure 3.2.7: *Feedback paths between the inner driver and the device's microphones (as indicated by the color). Upper panel: results for the vented earpiece, lower panel: closed earpiece. Transfer functions to the artificial eardrums (HpTF) for comparison (colors consistent with previous figures), individual lines show results for 5 reinsertions.*

For these microphones, the attenuation compared to the HpTF to the eardrum is larger than 20 dB for frequencies < 8 kHz. Given these data, it seems appropriate to utilize only the Entrance and Concha microphones for picking up sound for processing. Future research is required to evaluate whether multi-microphone feedback suppression techniques would also benefit from using the Outer Vent and/or In-Ear microphones, as already shown in the first prototype (Denk et al. 2017d; Schepker et al. 2019a).

For the closed earpiece, the feedback paths are generally smaller in amplitude than for the vented earpiece. Note that in the closed earpiece, there is no Outer Vent microphone. The attenuation compared to the HpTF to the eardrum is larger than 40 dB for frequencies < 8 kHz, and larger than 20 dB below 15 kHz. Thus, we hardly expect problems with feedback for the closed earpiece when using moderate gains.

Independent of the version of the earpiece, the feedback path to the In-Ear microphone is almost identical to the HpTF to the eardrum for frequencies < 1.5 kHz. At higher frequencies, the amplitude is up to approx. 20 dB larger at this microphone than at the eardrum. A similar relative transfer function between both locations was observed also for external sound sources.

3.2.4 Summary and outlook

We presented the design and acoustic evaluation of an earpiece with multiple microphones and drivers that can serve as ear-level hardware for research hearing devices. It contains four microphones and two drivers, which opens up numerous possibilities for signal processing. The earpiece is made up from a one-size-fits-all generic shell that sits shallow and firm in the cavum conchae.

In summary, the acoustic measurements showed that the features intended with the design were largely achieved. That is, the achieved bandwidth, maximum sound pressure level, attenuation characteristics and feedback paths show that the earpiece is suited to construct and study high-fidelity hearing devices for a range of applications. The vented and the closed versions have many comparable, but also complementary properties, particularly regarding leakage of external sounds and the low-frequency behavior.

Future work includes evaluating algorithms for sound pressure equalization, feedback cancellation and active noise/occlusion control on the new earpiece, in particular approaches that worked well on our first prototype (Denk et al. 2017d, 2018c,d; Schepker et al. 2019a; Schepker et al. 2018; Vogl and Blau 2019). There, the two versions of the earpiece allow a direct comparison between open and closed fits without changing any other parameters in a range of research questions.

Availability

Both versions of the earpiece are commercially available from InEar GmbH, Dieburg, Germany. More information can be found at https://www.hoertech.de/en/f-e-products/transparent-earpiece-2.html .

Acknowledgement

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Grant No. 352015383 - SFB 1330 A4 and C1, and the cluster of excellence *Hearing4all*, as well as by the EU/EFRE project *VIBHear*.

3.3 Equalization filter design for achieving acoustic transparency in a semi-open fit hearing device

Outline and context within the thesis

Achieving acoustic transparency in hearing devices requires appropriate equalization of the hearing device output by means of a suitable output filter. In this section, a novel least-squares based approach to compute an output filter for adapting the aided response at the eardrum to an approximated open-ear response is proposed and evaluated. The leakage component and the hearing device delay can be explicitly considered in the design, which reduces spectral ripples in the aided transfer function due to comb-filtering effects significantly. While the aim is identical to the iterative procedure proposed in (Denk et al. 2018c), the proposed algorithm is faster (needs only one set of measurements), more convenient and mathematically motivated. This algorithm was implemented in the real-time demonstrator and is subjectively evaluated on this platform in comparison to the equalization approach from (Denk et al. 2018c) in Sec. 4.3.

Abstract

Acoustically transparent hearing devices should allow hearing equivalent to the open ear while providing the possibility to modify the sound reaching the eardrum in a desired manner. To this end, the output of the device is processed by means of an equalization filter, such that the superposition of the sound played back by the device and an acoustic sound component directly leaking into the ear canal approximates the transfer function to the open eardrum. A particular difficulty in designing the equalization filter is the occurrence of comb-filtering effects due to a superposition of the direct sound and the delayed output of the device. Here, we propose a regularized least-squares design approach with a closed-form solution that takes into account individually measured transfer functions of the device and ear, as well as the processing delay. Experimental results utilizing measured transfer functions from a custom prototype device show good equalization performance, particularly a reduction of comb-filtering effects as a result of an automated frequency-dependent regularization.

This section is a formatted reprint of
F. Denk, H. Schepker, S. Doclo, and B. Kollmeier (2018d). "Equalization filter design for achieving acoustic transparency in a semi-open fit hearing device". *Proc. 13. ITG Conference on Speech Communication.* Oldenburg, Germany, pp. 226–230.

Note: An error in Eq. 3.3.17 has been corrected.

Author contributions: FD developed and implemented the algorithm, HS was involved in helpful discussions. FD performed the simulations, generated the figures and wrote the manuscript; all authors participated in writing the manuscript.

3.3.1 Introduction

Despite a great improvement in hearing technology in the past decades, the acceptance of hearing assistive devices is still limited, also due to a lack in sound quality (Killion 2004b; Sockalingam et al. 2009). This is particularly true in potential first-time users with a mild-to-moderate hearing loss or even (near-to) normal hearing. Although they would benefit from hearing aid features like speech enhancement or amplification in acoustically challenging situations, they are usually not willing to accept a general degradation of the listening quality. To overcome this issue, several contributions aimed at constructing acoustically transparent devices (Denk et al. 2018c; Hoffmann et al. 2013a; Rämö and Välimäki 2012). Such devices allow a listening experience that is the same as with the open ear while having the possibility to modify the sound reaching the eardrum in a desired manner.

To achieve acoustic transparency, the transfer function between external sound sources and the eardrum should be the same for the open ear and the aided case, i.e., with the device inserted. To match the two cases, a so-called equalization filter needs to be computed that spectrally adjusts the output of the device. While in the unaided case, the transfer function includes only the direct sound component, in the aided case the transfer function is a superposition of the device's transfer function and a direct sound component that directly leaks into the partly occluded ear canal. The direct sound is present particularly in *semi-open* fit devices that include a vent to reduce the occlusion effect and improve the wearing comfort (Winkler et al. 2016). Since in digital devices the sound played back is usually delayed compared to the direct sound by some milliseconds, distortions occur due to comb filtering effects (Stone et al. 2008). However, in previous approaches the contribution of the direct sound has been either neglected for the design of the equalization filter (Hoffmann et al. 2013a), or iterative procedures have been utilized to compute the equalization filter (Denk et al. 2018c). Furthermore, a non-iterative approximate all-pass design has been proposed to obtain an equalization filter (Rämö and Välimäki 2014). While this incorporates the direct sound and processing delay, it does not allow to include the electro-acoustic characteristics of the device. Therefore, in this paper we a propose non-iterative design of the equalization filter that takes into account the direct sound, individually measured transfer functions of the device and the processing delay. The proposed equalization filter design method requires a set of transfer functions that can and should be measured in-situ for each individual person (Denk et al. 2018c).

The paper is structured as follows: In Sec. 3.3.2, the acoustic scenario is introduced and problems occurring in designing an appropriate equalization filter are analysed. In Sec. 3.3.3, the filter design methods based on a frequency-domain least squares cost function with various extensions are proposed. Secs. 3.3.4 and 3.3.5 describe verification simulations, as well as results for the introduced filter design methods. Finally, the findings are summarized in Sec. 3.3.6.

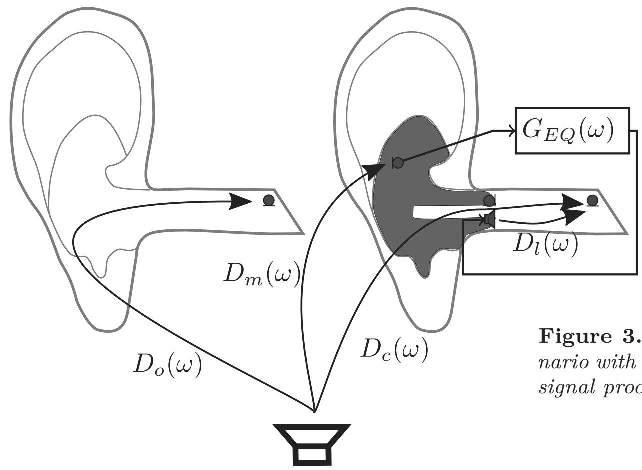


Figure 3.3.1: *Considered acoustic scenario with acoustic transfer functions and signal processing blocks.*

3.3.2 Problem statement

Consider the acoustic scenario shown in Fig. 3.3.1. The loudspeaker at the bottom represents a calibration sound source that is under control for measuring the relevant transfer functions.

In the open ear case shown on the left side of Fig. 3.3.1, the signal at the eardrum is the source signal filtered by the acoustic transfer function to the eardrum of the open ear $D_o(\omega)$ at radial frequency ω. For the aided case as shown on the right side of Fig. 3.3.1, the signal at the eardrum is the superposition of a source signal filtered by the acoustic transfer function of the semi-occluded ear $D_c(\omega)$ and the source signal filtered by the transfer function to the device microphone $D_m(\omega)$, the equalization filter $G_{EQ}(\omega)$ and the transfer function between the device loudspeaker and the eardrum $D_l(\omega)$. For convenience, we assume that all processing delay is included in the transfer function of the device loudspeaker $D_l(\omega)$. The transfer function of the complete system in the aided case (neglecting feedback) is then given by

$$D_{aided}(\omega) = D_m(\omega)G_{EQ}(\omega)D_l(\omega) + D_c(\omega). \tag{3.3.1}$$

While for both cases the sound cannot be measured directly at the eardrum, for the aided case methods exist that employ a microphone at the inner face of the device to obtain an estimate of the sound pressure at the eardrum (Denk et al. 2017d; Vogl et al. 2017). Therefore, in the following we assume availability of the signal at (or the transfer function to) the eardrum in the aided case.

Acoustic transparency means that the transfer function to the eardrum is equivalent in the open and the aided case, i.e.,

$$D_{aided}(\omega) = D_o(\omega). \tag{3.3.2}$$

Note that in practice, the open ear transfer function $D_o(\omega)$ is unknown and needs to be estimated. Such an estimate $\hat{D}_o(\omega)$ can be obtained, e.g., by applying an appropriate transformation function $G_T(\omega)$ to the microphone transfer function (Denk et al. 2018b), i.e.,

$$\hat{D}_o(\omega) = G_T(\omega) D_m(\omega). \tag{3.3.3}$$

In the following, this approximation is referred to as target transfer function. The optimal equalization $G_{EQ}^{(opt)}(\omega)$ is then obtained by requiring

$$\begin{aligned} \hat{D}_o(\omega) &= D_{aided}(\omega) \\ &= D_m(\omega) G_{EQ}^{(opt)}(\omega) D_l(\omega) + D_c(\omega), \end{aligned} \tag{3.3.4}$$

and solving for the equalization filter, yielding

$$\begin{aligned} G_{EQ}^{(opt)}(\omega) =& \frac{\hat{D}_o(\omega) - D_c(\omega)}{D_m(\omega) D_l(\omega)} \\ =& \frac{1}{D_l(\omega)} \left(G_T(\omega) - \frac{D_c(\omega)}{D_m(\omega)} \right). \end{aligned} \tag{3.3.5}$$

Since $D_l(\omega)$ contains a frequency-independent group delay with respect to the other transfer functions, $G_{EQ}^{(opt)}$ is generally acausal, which is not realizable in practice. Also, exact inversion of D_m and D_l might not be possible, since deep notches can occur there. Hence, a filter design method is needed that allows the computation of a realizable (causal) equalization filter while minimizing the differences between the aided transfer function and the open ear transfer function.

3.3.3 Filter design

In this section we present the proposed equalization filter design using a least-squares optimization procedure to obtain a causal filter. While we optimize the time-domain filter coefficients, it is practical to specify the desired transfer functions in the frequency-domain. Therefore, in Sec. 3.3.3.1 we first introduce the frequency-domain representation of the time-domain filter coefficients. In Sec. 3.3.3.2 we formulate the computation of the filter coefficients as a frequency-domain least-squares optimization problem. In Sec. 3.3.3.3 we introduce an acausality management and in Sec. 3.3.3.4 we propose to incorporate a frequency-dependent regularization to reduce comb filtering effects.

3.3.3.1 Frequency-domain optimization of time-domain filter coefficients

Since the target for acoustic transparency is defined as a transfer function, the equalization filter is computed based on a frequency-domain cost function. However, for future implementation on a hearing device, the time domain filter coefficients $\mathbf{g}_{EQ}$ (vector of length N_T) or spectral coefficients decoupled from the spectral resolution of the transfer functions are required. Making the desired length of the time-domain filter N_T independent of the Discrete Fourier Transform (DFT) length $N_F \geq N_T$ used to calculate the transfer functions, we write

$$\mathbf{G}_{EQ} = \underline{\underline{\mathbf{F}}}\mathbf{g}_{EQ} \tag{3.3.6}$$

$$\text{with} \quad \underline{\underline{\mathbf{F}}} = \underline{\underline{\mathfrak{F}}}^{(N_F \times N_F)} \begin{bmatrix} \underline{\underline{\mathbf{I}}}^{(N_T \times N_T)} \\ \underline{\underline{\mathbf{O}}}^{(N_F - N_T \times N_T)} \end{bmatrix}. \tag{3.3.7}$$

$\underline{\underline{\mathfrak{F}}}$ is the DFT matrix, $\underline{\underline{\mathbf{I}}}$ and Identity matrix and $\underline{\underline{\mathbf{O}}}$ a matrix containing only zeros, and $\mathbf{G}_{EQ}$ is the discrete frequency response of the equalization filter.

3.3.3.2 Least-squares cost function

The equalization filter should minimize the difference between the aided and open ear transfer function to the eardrum. Similarly to Eq. (3.3.4), we therefore define a least-squares cost function of the form

$$J_{LS}(\mathbf{g}_{EQ}) = ||(\underline{\underline{\mathbf{D}}}_m \underline{\underline{\mathbf{D}}}_l \mathbf{G}_{EQ} + \mathbf{D}_c) - \hat{\mathbf{D}}_o||_2^2, \tag{3.3.8}$$

where $\underline{\underline{\mathbf{D}}}_m$ and $\underline{\underline{\mathbf{D}}}_l$ are diagonal matrices containing the DFT coefficients of $D_m(\omega), D_l(\omega)$ respectively; $\mathbf{D}_c$ and $\hat{\mathbf{D}}_o$ are according vectors. The optimum with respect to $\mathbf{g}_{EQ}$ is given by

$$\mathbf{g}_{EQ}^{(LS)} = \left(\underline{\underline{\mathbf{A}}}^H \underline{\underline{\mathbf{A}}}\right)^{-1} \underline{\underline{\mathbf{A}}}^H (\hat{\mathbf{D}}_o - \mathbf{D}_c), \tag{3.3.9}$$

where

$$\underline{\underline{\mathbf{A}}} = \underline{\underline{\mathbf{D}}}_m \underline{\underline{\mathbf{D}}}_l \underline{\underline{\mathbf{F}}}, \tag{3.3.10}$$

and $(\cdot)^H$ denotes the hermitian transpose of a matrix.

3.3.3.3 Acausality management

To avoid potential acausality problems, the filter $\mathbf{g}_{EQ}$ is forced to be shifted in time by writing

$$\tilde{\mathbf{G}}_{EQ} = \underline{\underline{\mathbf{z}}}^D \underline{\underline{\mathbf{F}}} \mathbf{g}_{EQ}. \tag{3.3.11}$$

There, $\underline{\underline{\mathbf{z}}}^D$ denotes a diagonal matrix whose elements are the phase coefficients corresponding to a negative shift in time by D samples. D is chosen to be the processing delay, extended by some additional samples that allow for small acausalities in the filter design. The optimization for $\mathbf{g}_{EQ}$ is performed analogous to the previous section, yielding

$$\mathbf{g}_{EQ}^{(LSD)} = \left(\underline{\underline{\mathbf{A}}}_D^H \underline{\underline{\mathbf{A}}}_D\right)^{-1} \underline{\underline{\mathbf{A}}}_D^H (\hat{\mathbf{D}}_o - \mathbf{D}_c), \tag{3.3.12}$$

where

$$\underline{\underline{\mathbf{A}}}_D = \underline{\underline{\mathbf{D}}}_m \underline{\underline{\mathbf{D}}}_l \underline{\underline{\mathbf{z}}}^D \underline{\underline{\mathbf{F}}}. \tag{3.3.13}$$

3.3.3.4 Regularization

Comb filter effects are most pronounced in frequency regions where the direct sound is not attenuated significantly with respect to the target transfer function (Denk et al. 2018c). To include this observation in the design process, a frequency dependent regularization imposing an additional cost for these frequency regions is included in the cost function, similarly to (Kirkeby and Nelson 1999)

$$J_{LSR}(\mathbf{g}_{EQ}) = ||(\underline{\underline{\mathbf{D}}}_m \underline{\underline{\mathbf{D}}}_l \tilde{\mathbf{G}}_{EQ} + \mathbf{D}_c) - \hat{\mathbf{D}}_o||_2^2 + \mu||\underline{\underline{\mathbf{V}}}\tilde{\mathbf{G}}_{EQ}||_2^2, \tag{3.3.14}$$

where $\underline{\underline{\mathbf{V}}}$ is a real-valued diagonal matrix whose diagonal entries are spectral regularization weights for each frequency bin of $\tilde{\mathbf{G}}_{EQ}$, and μ is the regularization parameter. The filter optimizing Eq. (3.3.14) is given by

$$\mathbf{g}_{EQ}^{(LSDR)} = \left(\underline{\underline{\mathbf{A}}}_D^H \underline{\underline{\mathbf{A}}}_D + \mu \underline{\underline{\tilde{\mathbf{V}}}}^H \underline{\underline{\tilde{\mathbf{V}}}}\right)^{-1} \underline{\underline{\mathbf{A}}}_D^H (\hat{\mathbf{D}}_o - \mathbf{D}_c), \tag{3.3.15}$$

with

$$\underline{\underline{\tilde{\mathbf{V}}}} = \underline{\underline{\mathbf{V}}} \underline{\underline{\mathbf{z}}}^D \underline{\underline{\mathbf{F}}}. \tag{3.3.16}$$

We compute the frequency-dependent regularization weight $\mathbf{V}$ based on the level relation of the direct sound $\mathbf{D}_c$ and the target transfer functions $\hat{\mathbf{D}}_o$ in each DFT bin k

$$V[k] = \sum_{k'=0}^{N_F-1} S_k[k'] \min\left(1, \left|\frac{D_c[k']}{\hat{D}_o[k']}\right|\right)^{N_v}. \tag{3.3.17}$$

The constrained relation of target and direct sound transfer function is expanded with an exponent N_v and then smoothed by applying a smoothing vector $\mathbf{S}_k$ (Hatziantoniou and Mourjopoulos 2000). Here, smoothing across 1/6 octave with a rectangular smoothing window was performed and the expansion N_v was set to 5.

3.3.4 Measurements and simulations

The necessary acoustic transfer functions have been measured in a human subject in a free field setup as described in more detail in (Denk et al. 2017b,d). The hearing device as described by Denk et al. (Denk et al. 2018c) was utilized. It is an individual earmould including a vent, 3 microphones and 2 loudspeakers, in an assembly corresponding well to Fig. 3.3.1. Here, as in (Denk et al. 2018c) only one loudspeaker located at the inner face of the device and one microphone (located at the back of the concha) were utilized. Measurements at the eardrum were performed using a probe tube microphone. For the present simulations, the influence of microphone sensitivities were equalized out, i.e., purely acoustic transfer functions are utilized. For all simulations, a single sound source and frontal sound incidence at 1 m distance were considered. Contributions of feedback to the hearing device microphone were neglected.

The utilized transfer functions are shown in Fig. 3.3.2. Below about 1.5 kHz, the device does not attenuate external sounds. Between 500 Hz and 1.3 kHz, the transfer function to the eardrum of the occluded ear canal is larger than the open ear transfer function to the eardrum, which is presumably caused by a Helmholtz resonance of the residual ear canal and the vent. Above 2 kHz, the device attenuates external sounds by about 20 dB on average with respect to the open ear transfer function, however the behaviour is highly frequency-dependent.

The processing delay of the device was set to 6 ms and implemented as a time shift applied to D_l. A sampling rate of 48 kHz was utilized, and spectral analysis was performed with a DFT length of $N_F = 4096$ samples. Using the transfer functions and other parameters, the equalization filters and corresponding aided transfer functions were computed for the approaches described in the previous section.

3.3.4.1 Error metric

As an error metric between two transfer functions $\mathbf{H}_1$ and $\mathbf{H}_2$, a perceptually motivated auditory spectral distance was computed. Therefore, the mean difference between amplitudes in dB is averaged with a weight in a frequency range bounded by the bins k_1 and k_2

$$\Delta H_{\mathrm{Aud}} = \sum_{k=k_1}^{k_2} W[k] \left|10 \log_{10}(|H_1[k]|^2) - 10 \log_{10}(|H_2[k]|^2)\right| . \tag{3.3.18}$$

$W[k]$ is a frequency-dependent weight, which was chosen as the inverse of the ERB-bandwidth depending on frequency to counteract over-representation of high frequencies (Glasberg and Moore 1990). It is normalized such that

$$\sum_{k=k_1}^{k_2} W[k] = 1. \tag{3.3.19}$$

Here, the frequency range in which this error is computed is constrained between 200 Hz and 16 kHz.

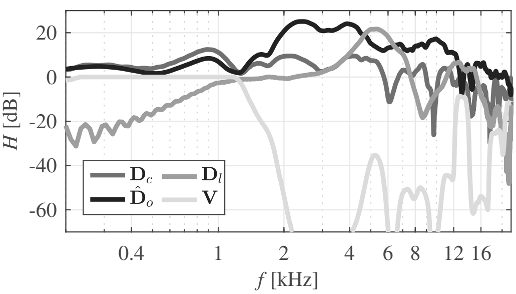


Figure 3.3.2: *Acoustic transfer functions utilized for the simulations. The regularization weight* $\mathbf{V}$ *was calculated according to Eq.* (3.3.17).

3.3.5 Results and discussion

3.3.5.1 Influence of acausality management

Figure 3.3.3 shows the target transfer function $\hat{\mathbf{D}}_0$ together with the aided transfer function and equalization filter impulse responses obtained by Eq. (3.3.9) and Eq. (3.3.12), i.e., not including and including the acausality management. For the acausality management, D was chosen to be the processing delay D_{proc} + 24 samples excess delay, i.e., an additional 0.5 ms. The filter length N_T was 256 samples.

Without the acausality management, the filter coefficients are close to 0, i.e., no sensible filter is computed. In consequence, the aided transfer function is almost equal to the occluded transfer function $\mathbf{D}_c$. Further simulations showed that less excess delay than the utilized 0.5 ms resulted in a poorer performance, while more excess delay did not result in a further improvement. Apparently, besides exploiting knowledge about the processing delay in the optimization process, it is required to allow for some acausality in the filter design to achieve good performance. The acausal taps probably support the partial equalization of the non-minimum phase system.

3.3.5.2 Influence of regularization

Figure 3.3.4 shows the target and aided transfer functions as well as impulse- and frequency responses of $\mathbf{g}_{EQ}$, where $\mathbf{g}_{EQ}$ was calculated according to Eq. (3.3.14), i.e., including the acausality management and variable regularization. The aided transfer functions are shown for 3 different regularization weights $\mu = \{0, 10^{2.5}, 10^8\}$,

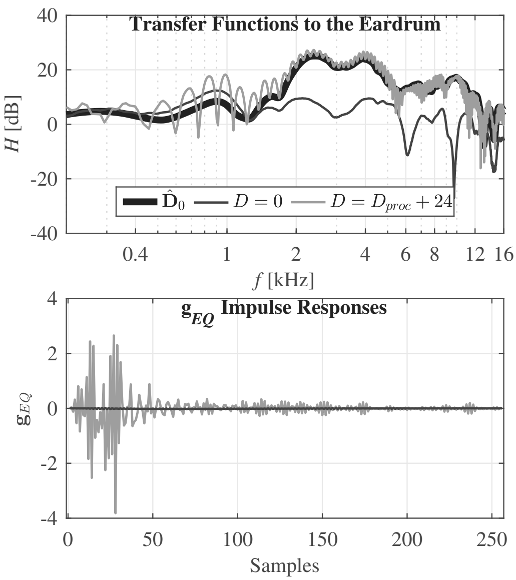


Figure 3.3.3: *Top panel: Target and aided transfer functions, $\mathbf{g}_{EQ}$ computed using Eq.* (3.3.9) *and Eq.* (3.3.12)*, without ($D = 0$) and with ($D = D_{proc} +$ 24, i.e. 0.5 ms excess delay) including acausality management; filter length $N_T = 256$. Bottom panel: corresponding equalization filter coefficients $\mathbf{g}_{EQ}$.*

and as in the previous section using a filter length $N_T = 256$ and D corresponding to the processing delay D_{proc} + 0.5 ms.

The influence of the regularization is generally positive on the performance: Comb-filter effects are significantly reduced as compared to the same setting where no regularization is applied (compare result for $\mu = 0$ against $\mu = 10^{2.5}$). This is because $\mathbf{G}_{EQ}$ reduces the hearing device output in frequency regions where no output is needed, i.e., where the direct sound transfer function $\mathbf{D}_c$ already provides sufficient level (c.f. Fig. 3.3.2). Also, the filter coefficients in time-domain look better behaved when a regularization is applied: Less high-frequency ringing artefacts are noted, and the level is generally lower, additionally resulting in a smaller energy consumption due to a smaller amplification.

However, very large regularization weights result in a poor performance. In Fig. 3.3.4, it becomes clear that for μ as large as 10^8, the frequency response of the equalization filter $\mathbf{G}_{EQ}$ becomes too small in level to provide the desired compensation of attenuation that the device produces by partially occluding the ear.

Figure 3.3.5 shows the auditory spectral distances ΔH_{Aud} according to Eq. (3.3.18) for variable μ and different filter lengths N_T. Independent of the filter length, a

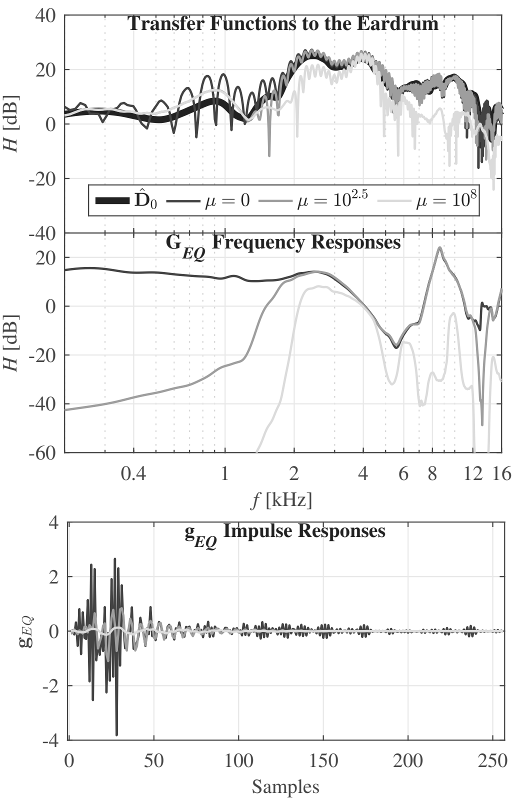


Figure 3.3.4: *Top panel: Target and achieved aided transfer functions, $\mathbf{g}_{EQ}$ computed using Eq. (3.3.15) with different regularization parameters μ, filter length of $N_T = 256$ and $D = D_{proc} + 24$, i.e., and excess delay of 0.5 ms, as in Fig. 3.3.3. Middle Panel: Frequency response of the equalization filters. Bottom Panel: Corresponding time-domain filter coefficients $\mathbf{g}_{EQ}$.*

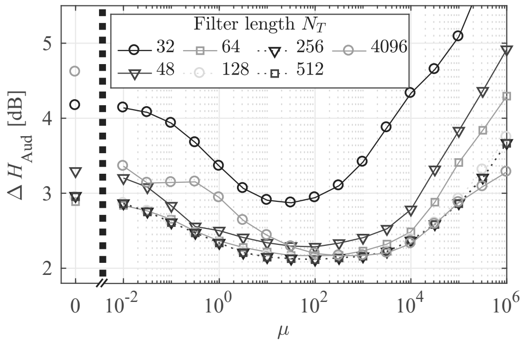


Figure 3.3.5: *Auditory spectral distance between target and aided transfer functions with $\mathbf{g}_{EQ}$ computed using Eq. (3.3.15) depending on the regularization parameter μ, with D set to the processing delay + 24 samples (= 0.5 ms).*

U-shaped error curve over μ is observed. Apparently, there is a trade-off between imposing an additional cost between avoiding signal playback in frequency regions where this is unnecessary, and the regularization aspect of the cost function dominating the optimization: If μ is too small, comb filter effects are not avoided, but if μ is too large, the cost for energy in $\mathbf{G}_{EQ}$ dominates the optimization, resulting in aided transfer functions that are too low. This means that there is an optimal μ, which for the majority of the regarded filter lengths, lies between 10^2 and 10^3.

3.3.5.3 Influence of the filter length

While in the previous sections a fixed filter length was considered, in this section we compare the results for different filter lengths N_T. Again, consider Fig. 3.3.5, which shows the achieved spectral distance ΔH_{Aud} (c.f. Eq. 3.3.18) between the target and the aided transfer function. There, results for different filter lengths N_T and regularization parameters μ are compared.

The filter length has only small influence on ΔH_{Aud}, if a minimal length (about 64 samples) is exceeded. An auditory spectral difference between open and aided transfer function of 2.1 dB seems to be the lower limit for the given filter design and transfer functions. Probably, the observation that the direct sound transfer function $\mathbf{D}_c$ is larger than the target transfer function $\hat{\mathbf{D}}_0$ around 1 kHz contributes to a lower limit that deviates notably from 0. For very short filters with $N_T < 64$ samples, the residual spectral difference is larger, presumably due to a frequency resolution that is too poor. However, a very long filter ($N_T = 4096$, equal to DFT length) does not result in better performance – again, no cancellation of spectral ripples is achieved, although the spectral resolution of the filter would be sufficient to compensate the ripples by means of its amplitude.

It is also worth mentioning that the optimal regularization weight μ interacts with the filter length: The optimum μ increases with increasing filter length. This is understandable in a way that with very short filters, a highly over-determined set of equations is solved by Eq. (3.3.15), which by itself is somewhat equivalent to a regularization.

3.3.6 Conclusions

We presented an approach to design equalization filters for semi-open fit hearing devices with the aim to provide acoustic transparency. The approach is based on a frequency-domain least-squares cost function that takes into account individually measured transfer functions and the processing delay of the device, and has a closed-form solution for the time-domain filter coefficients. Furthermore, using the presented approach it is possible to decouple the spectral analysis length from the desired filter length.

Within this design approach, the results showed hat it is critical to include an acausality management of the filters, where knowledge about the processing delay can be explicitly exploited. Furthermore, frequency-dependent regularization of the energy contained in the equalization filter resulted in a reduction of comb-filtering effects. The regularization parameters were computed automatically from the occurring acoustic transfer functions, which makes the regularization approach easily applicable. In the present work, filter lengths as short as 64 samples or 1.5 ms were sufficient to achieve the best possible performance.

In conclusion, using a least-squares design approach that takes the processing delay and automatic regularization into account, it is possible to compute individualized equalization filters for acoustically transparent hearing devices that produce only little comb filter artefacts.

Acknowledgments

This work was supported by the Research Unit FOR1732 *Individualized Hearing Acoustics* and the Cluster of Excellence EXC1077 Hearing4All, both funded by the German Research Foundation (DFG).

4 Psychophysical verification

4.1 On the limitations of sound localization with hearing devices

Outline and context within the thesis

In this section, results of a subjective localization experiment with virtual stimuli that simulated listening through linear hearing devices are described. In this way, several parameters that can affect localization abilities in hearing devices could be assessed separately and in superposition, namely the microphone position, bandwidth, and processing delay in superposition with a leakage component. The results are thus a contribution to an understanding of the mechanisms that limit the ability to localize sounds while wearing hearing devices. To generate the stimuli, the HRTF database described in Sec. 2.2 is utilized, and the same subjects participated. The study may be seen as an extended subjective replication of the model results from Sec. 2.3.

This section is a formatted reprint of
F. Denk, S. Ewert, and B. Kollmeier (2019b). "On the limitations of sound localization with hearing devices". *The Journal of the Acoustical Society of America* 146(3), pp. 1732–1744.

Author contributions: FD designed, implemented and conducted the experiments, evaluated the results, generated the figures and wrote the manuscript. SE and BK participated in the study design and writing the manuscript.

Abstract

Limited abilities to localize sound sources and other reduced spatial hearing capabilities remain a largely unsolved issue in hearing devices like hearing aids or hear-through headphones. Hence, the impact of the microphone location, signal bandwidth, different equalization approaches as well as processing delays in superposition with direct sound leaking through a vent was addressed in this study. A localization experiment was performed with normal-hearing subjects using individual binaural synthesis to separately assess the above-mentioned potential limiting issues for localization in the horizontal and vertical plane with linear hearing devices. To this end, listening through hearing devices was simulated utilizing transfer functions for six different microphone locations, measured both individually and on a dummy head. Results show that the microphone location is the governing factor for localization abilities with linear hearing devices, and non-optimal microphone locations have a disruptive influence on localization in the vertical domain, and an effect on lateral sound localization. Processing delays cause additional detrimental effects for lateral sound localization; and diffuse-field equalization to the open-ear response leads to better localization performance than free-field equalization. Stimuli derived from dummy head measurements are unsuited for evaluating individual localization abilities with a hearing device.

4.1.1 Introduction

The limited ability to localize sound sources remains an unsolved issue in hearing aids and related devices for normal-hearing users. That is, hearing aids often impose additional sound localization difficulties for hearing impaired subjects, or in the best case provide no benefit in this respect (Akeroyd and Whitmer 2016; Kollmeier and Kiessling 2018). An impairment of spatial hearing due to wearing a hearing device is even more critical for the acceptance of devices targeted to normal-hearing users, like hear-through headphones or electronic hearing protectors (Härmä et al. 2004; Hoffmann et al. 2014; Killion et al. 2011; Marentakis and Liepins 2014; Tikander 2009).

The limitation of sound localization abilities with hearing devices results from an insufficient conservation of directional cues that are usually created by sound transmission through the open ear. These cues are described by the Head-Related Transfer Function (HRTF) and include both interaural time (ITD) and interaural level (ILD) cues as well as monaural spectral cues. Whereas the interaural cues are mostly exploited for lateral (i.e., left-right) localization, monaural spectral cues as well as time-variation of binaural cues originating from head movements are exploited to determine the vertical position of a sound source, which includes the discrimination of the front and rear hemisphere in the horizontal plane (Blauert 1997). A principal limitation on how well directional cues can be conserved is imposed by the directional information captured by the device microphone(s) depending on their location. Other potential sources of inaccuracies are the algorithms operating on the device, the processing delay, and the reproduction bandwidth (Akeroyd and Whitmer 2016; Byrne and Noble 1998; Denk et al. 2018g).

The detrimental effects of hearing devices on sound localization have been assessed mostly by having subjects localize free-field sound sources while wearing hearing devices (Best et al. 2010; Brungart et al. 2003; Brungart et al. 2007; Byrne and Noble 1998; D'Angelo et al. 2001; Hoffmann et al. 2014; Van den Bogaert et al. 2006, 2011). While very realistic results can be obtained in such experiments, it is hard to determine at what stage which cue is distorted and how this affects sound localization. In the present work, we used individual binaural synthesis (Møller 1992) to separate and quantify the influence of the separate stages in a hearing device processing chain on sound localization. To this end, stimuli that resembled listening through simulated hearing devices, where the separate (detrimental) aspects could be turned on and off freely, were presented over headphones. We thereby made use of HRTFs that have been measured at the microphone locations of several device styles in individual subjects (Denk et al. 2018b). Normal-hearing subjects were employed here to assess the principal limitations on sound localization when listening with hearing devices, independent on additional difficulties through a hearing impairment (Noble et al. 1997). Localization performance was assessed for the full horizontal and median planes in separate experiments to assess effects on both lateral and vertical sound localization.

The first stage in the hearing device where directional cues can be distorted is the sound pickup at a non-optimal microphone location. To study the isolated effect of the microphone location, stimuli were convolved with the individual HRTF to the hearing device microphone, which was individually equalized against the HRTF to the eardrum of the open ear (open-ear HRTF). This is the best possible listening condition based on the hearing device microphone signal. The tested microphone locations covered the whole range of devices in use today, from optimum locations in the ear canal over various locations in the cavum conchae to a microphone behind the ear. The impact of the microphone location in isolation has been previously studied in the context of vertical localization by means of computational models (Denk et al. 2018g; Durin et al. 2014). There, a strong detrimental effect of the microphone position was observed, which, however, has not yet been verified in listening tests. In free-field localization experiments with hearing aids, only small effects of the microphone location have been found (Best et al. 2010; Brungart et al. 2003; Jensen et al. 2013; Van den Bogaert et al. 2011). To assess the influence of a restricted bandwidth as in many concurrent devices, the equalized HRTFs were also low-pass filtered at 8 kHz. For vertical localization, previous model studies predicted a generally poorer performance with low-pass stimuli, and smaller differences between microphone locations (Durin et al. 2014). Furthermore, we compared equalization to the open-ear response for frontal against diffuse-field incidence. While our previous work demonstrated that diffuse-field equalization minimizes the spectral differences against the open-ear HRTF (Denk et al. 2018b), the impact on localization is unclear.

The effect of a processing delay was assessed by simulating the superposition of the hearing device output and an acoustic leakage component that directly enters the ear canal through a vent. We chose a delay of 6 ms, which is in the typical range for relevant hearing devices and causes a perceivable disturbance (Groth and Søndergaard 2004; Stone et al. 2008). Since the virtual hearing devices were targeted for normal-hearing subjects, the device output was filtered in a way that the superposition of both sound components approximates the open-ear HRTF (Denk et al. 2018d). Note that this condition does not include compression or synchronization artifacts, i.e., the same constant delay was present at both sides without variations. Also, the processing is restricted to a linear filtering operation. While this is realistic for many devices targeted to normal-hearing users, this condition does not account for the effect of the non-linear processing in most modern hearing aids.

Furthermore, localization performance was assessed where the appropriate HRTFs were measured on a KEMAR to study to what extent such psychophysical studies require the use of individually measured hearing device HRTFs. To the authors' best knowledge, this aspect has not been studied before.

For the isolated effect of the microphone location, we expect an impairment of localization mostly in the vertical domain, since positioning the microphone away from the ear canal results most prominently in a bias of spectral directional cues (Denk et al. 2018g; Durin et al. 2014). Based on previous data discussed above (Best et al. 2010; Van den Bogaert et al. 2011), we expect no influence of the microphone location on lateral localization. Given that the processing delay could potentially disrupt ITD cues, we expect a detrimental effect on lateral localization, but no detrimental effect for vertical sound localization.

4.1.2 Methods

4.1.2.1 Outline

The subjects localized sound sources while sitting inside an anechoic chamber with a visible array of loudspeakers at the positions of the sound sources (see Fig. 4.1.1). The main part of the experiment included stimuli presented over headphones after convolution with appropriate HRTFs, referred to as virtual stimuli. These stimuli simulated listening through idealized linear hearing devices, where the effects of microphone position, bandwidth, a vent and processing delays were assessed. For training purposes, the subjects also localized free-field stimuli presented over the loudspeakers. Individual HRTFs had been measured using the very loudspeaker system utilized for free-field sound presentation (Denk et al. 2018b). When conducting the virtual localization trials, the subjects were not informed that only the headphones were used for playback.

Localization performance in the median and horizontal plane was assessed in separate experiments. In the median plane, elevations between -30° and 90° with a uniform spacing of 15° both in the front and rear hemisphere were used, adding up to 17 incidence directions. The horizontal plane experiment included 24 directions with a uniform spacing of 15°.

4.1.2.2 Apparatus

The study was conducted in the Oldenburg VR-Lab, which is an anechoic chamber featuring a 94 channel three-dimensional array of *Genelec* 8030 loudspeakers with a distance varying between 2.5 and 3m from the center point. Fig. 4.1.1 shows a subject seated in the chamber while conducting the experiment. Thirty-nine loudspeakers were utilized for free-field sound presentation and uniquely labelled; the loudspeakers directly in front and behind the subject were included in both the horizontal and median plane. The loudspeakers were individually equalized such that their pseudo-anechoic frequency response (reflections from the other loudspeakers windowed out, Denk et al. 2018f) was flat between 80 Hz and 18 kHz.

Virtual stimuli were presented from open-coupling *Sennheiser* HD650 headphones, driven by a *Lake People* G-103P amplifier. The headphones were individually equalized to produce a flat frequency response at the eardrum to avoid a double influence of the ear canal, based on individual Headphone Transfer Function (HpTF) measurements that had been conducted with the HRTF measurements. To this end, the HpTF to the eardrum was estimated by transforming the HpTF measured at the blocked ear canal entrance by the transfer function between that location and the eardrum obtained in free-field measurements, assuming that the headphones are free-field equivalent coupling (Møller et al. 1995b).

Figure 4.1.1: *Subject participating in the experiment. The loudspeakers are labelled using a white sticker at the bottom. Note that in the horizontal plane, only every other installed loudspeaker (with white label) is used.*

The subjects indicated the perceived incidence direction by clicking buttons on a Graphical User Interface (GUI) that depicted the labelled loudspeakers, which means the responses were restricted to the actual loudspeaker locations. Different GUIs were utilized for the separate horizontal and median plane localization tasks. The GUI was displayed on a handheld tablet with a 10 in. touchscreen (see in Fig. 4.1.1, median plane GUI displayed). After clicking a response button and a pause that lasted randomly between 1.5 and 2.5 seconds, the next stimulus was presented automatically. By means of a headtracker (*Pohlemus* Patriot) we ensured that before stimulus playback the subject's head was centered in the loudspeaker array and oriented towards the front. If the position exceeded the tolerance of 5° or 5 cm in either degree of freedom, a message was displayed, and the experiment halted until the position was restored.

4.1.2.3 Listeners and individual HRTFs

Eleven normal-hearing subjects (including one of the authors, age 29.8±4.6, five females) participated in the study after giving written informed consent. HRTFs had been measured individually for all subjects at the eardrum and locations in the ear corresponding to microphone locations of a range of hearing devices, including the blocked ear canal entrance. Also, the individual HpTF had been measured at the blocked ear canal entrance. The dataset is publicly available (Denk et al. 2018b).

HRTFs had been measured using the installed loudspeaker system for all directions of interest except those at elevations ±15, +45 and +75° in the median plane. There, HRTFs were computed by linear interpolation in magnitude and phase separately between the next neighbor locations in the median plane (Nishino et al., 1999). For the locations at ±15°, the symmetric distance to the next neighbors was 5° (interpolation between the HRTFs at ±10° and ±20°), for the others 15° (Interpolation between 30° and 60° for 45°, and between 60° and 90° for 75°). In a model evaluation, the directional resolution of the measurements was proven to be sufficient (Denk et al. 2018g).

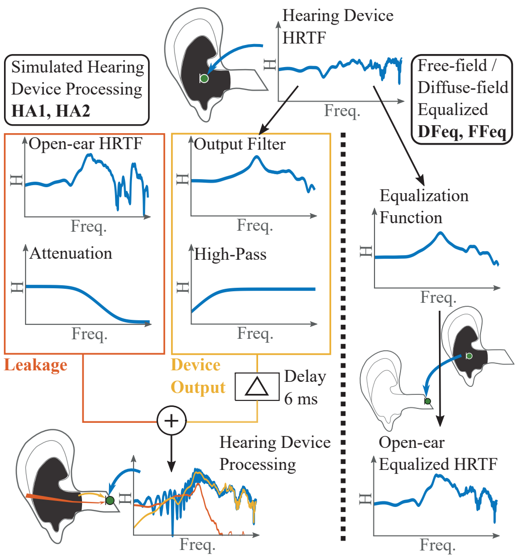


Figure 4.1.2: *Signal flow for equalizing the hearing device HRTFs (right side, processing schemes DFeq and FFeq, curves show DFeq) and simulating linear hearing device processing with delay (left, processing schemes HA1 and HA2, curves show HA2).*

4.1.2.4 Listening conditions and stimuli

Linear hearing devices were simulated aiming for acoustic transparency, i.e., adjusted to approximate the open-ear HRTF as well as possible. A subset of six hearing device microphone locations from the available database was used, for more details and an image of the microphone locations in the ear, the reader is referred to (Denk et al. 2018b). The order below reflects their approximate distance from the eardrum.

- ECEbl: Blocked ear canal entrance. Ideal recording point, equivalent to microphones of in-the-canal devices. This point is also referred to as the reference location in the following.
- InsertHP: microphones placed on small insert headphones (*Sennheiser* CX200).
- ITEind: microphones placed on an earmould filling the concha bottom completely, one near the ear canal entrance (ITEind_Entr), and one in the rear part of the cavum conchae (ITEind_Concha).
- ITEgen: microphone placed on a generic ITE device, near the ear canal entrance (ITEgen_Entr in the database).
- BTE: microphone on a Behind-The-Ear hearing aid dummy. For this study, we chose the middle microphone (BTE_mid) from the database.

Different processing schemes resembling different listening conditions based on the hearing device HRTFs were employed. The basic scheme for the calculations is shown in Fig. 4.1.2, and the individual processing schemes are described in detail below:

— DFeq: Diffuse-field equalization against the eardrum-HRTF. The HRTF at the appropriate microphone location was transformed for each subject by the individual diffuse-field transfer function between the appropriate microphone location and the eardrum (c.f., individual diffuse-field Target Response Correction Function in Denk et al. 2018b).

— DFeq-LP: As DFeq, but with a low-pass filter at 8 kHz (8th order Butterworth) applied to the resulting HRTF.

— FFeq: As DFeq, but utilizing the appropriate free-field correction function, i.e., the transfer function between the appropriate microphone location and the eardrum for frontal incidence.

— HA1: Simulated hearing device processing with a leakage component: The hearing device output is adapted such that the diffuse-field equalized HRTF is approximated, see Fig. 4.1.2. The sound reaching the eardrum includes a leakage component, simulated by the individual open-ear HRTF filtered with a high-shelving filter (2nd order, cut-off frequency 2.5 kHz, high-frequency attenuation 25 dB). An output filter for the hearing device was calculated using a least-squares design that exploits knowledge of the leakage component, as proposed in (Denk et al. 2018d). The design method results in a reduction of hearing device output in frequency regions where the direct sound component is sufficient to produce the desired frequency response at the eardrum, leading to a reduction of comb filter effects. The hearing device output had a delay of 6 ms with respect to the direct sound component, and included a second-order high-pass filter (2nd order Butterworth) at 500 Hz to account for the vent effect.

— HA2: Simulated alternative hearing aid processing with a leakage component: Same acoustic parameters and filter design method as in HA1, but where the leakage component was not considered in the output filter design. In consequence, the leakage component and the device output superimpose in a broader frequency range and generate more comb filter artifacts.

Whereas generally, individual HRTFs were employed to generate the stimuli, the diffuse-field equalized hearing device HRTFs were also computed with HRTFs measured in a KEMAR ear at the appropriate microphone locations. They are implemented to assess the influence of using KEMAR-derived HRTFs for evaluating sound localization with such devices as compared to individual HRTFs. For simplicity, they are treated as further processing schemes:

— DFeq-KEMAR: As DFeq, but taking the HRTF of the KEMAR mannequin.

— DFeq-LP-KEMAR: As DFeq-LP, but taking the HRTF of the KEMAR mannequin.

The conditions included in the localization experiments consist of combinations of microphone locations and processing schemes as listed in Tab. 4.1.1. While localization was tested for all microphone locations in the DFeq processing, the different processing schemes were tested only with a subset of microphone locations,

Table 4.1.1: *Conditions presented in the experiment, marked by "M" if included in the median plane, and "H" if included in the horizontal plane localization experiment (see Sec. 4.1.2.4). Every incidence direction was presented once to each subject, except in the ECEbl-DFeq (reference) condition, where each direction was repeated 3 times.*

		Microphone Location					
		ECEbl Ear canal entrance	**InsertHP** Insert earphone	**ITEind Entr** Individal in-the-ear device	**ITEind Concha** Individal in-the-ear device	**ITEgen** Generic in-the-ear device	**BTE** Behind-the-ear device
Processing scheme	**DFeq** diffuse-field equalized	H M (3x)	H M	H M	H M	H M	H M
	FFeq free-field equalized			H M			H M
	DFeq-LP low-pass	M	M	M	M	M	M
	HA1 Processing w. delay and comb-filter reduction	H M		H M			H M
	HA2 Processing w. delay	H M		H M			H M
	DFeq-KEMAR	H M		H M			H M
	DFeq-LP-KEMAR	M		M			M

namely ECEbl (perfect spatial cues), ITEind_Entr (disturbed cues) and BTE (highly disturbed cues, no pinna effects). The effect of low-pass filtering the stimulus was only tested in the median plane experiment, since no effect on lateral localization is expected. Note that for the ECEbl location, diffuse-field and free-field equalization functions are identical (Denk et al. 2018b), and thus the results of ECEbl-DFeq were copied to ECEbl-FFeq for clarity.

Every incidence direction was presented once to each subject, except in the ECEbl-DFeq (reference) condition. There, each incidence direction was repeated 3 times to provide a larger portion of well-localizable stimuli, as well as being able to assess the performance with the same accuracy as in the training sessions (see Sec. 4.1.2.5). The number of stimuli amount to 456 for the horizontal plane experiment and 476 for the median plane experiment. While the horizontal plane includes more directions, a higher number of conditions was included in the median plane experiment (leading to 28 presentations of each direction) as compared to the horizontal plane (19 presentations of each direction).

The stimulus was a series of two white noise bursts (frequency range 80 Hz to 18 kHz) with individual lengths of 150 ms including 10-ms cosine ramps, separated by 40 ms silence. The short length was chosen to avoid perceivable effects of head motion with the static synthesis, and two bursts were presented to improve the perception of temporal cues (Blauert 1997). The level was 60 dB SPL ±2 dB random variation referenced to the free field.

4.1.2.5 Procedure

Each subject participated in three experimental sessions. First, they underwent a training session with ten runs, five runs for each the median and horizontal localization task. In each run, every incidence direction was presented in random order 3 and 2 times for the median and horizontal plane, respectively, to get a similar number of trials. The first two runs in each plane task were free- field presentations. Subsequently, three runs of virtual presentations were performed where the HRTF at the ear canal entrance (diffuse-field equalized, see Sec. 4.1.2.4) was used, which is the reference condition in virtual presentation. The horizontal and median plane localization runs were performed in subsequent blocks, with the order of these blocks randomized.

The two other sessions included the main experiments for horizontal and median plane localization separately, in randomized session order. Each of these sessions started with one run of free-field presentation, followed by one training run of virtual presentation with the reference condition. Then followed the main experiment, where all test conditions (see Sec. 4.1.2.4) were interleaved and presented in randomized order. After each 80 stimuli, the subjects had the opportunity to take a break, and the experiment was continued upon their notification. Only the data from the main experiment is evaluated in the following. Feedback about the response, i.e., showing the correct incidence direction by button markup, was provided in the training runs but not in the main experiment. The individual sessions lasted between 60 and 90 minutes, depending on the subject.

4.1.2.6 Data analysis and error metrics

For the evaluation, the interaural polar coordinate system was used. It is composed of the lateral angle $\alpha \in$ [-90°,90°], which denotes the lateral displacement from the median plane (negative values indicating the left-hand hemisphere). The polar angle $\beta \in$ [-90°,270°[describes the position inside a sagittal plane - for the median plane, 0° denotes frontal, 90° above and 180° rear incidence.

Two different error metrics were used for the horizontal and median plane experiments independently. Generally, two different metrics were employed to separate small localization errors from gross errors, such as front-back confusions (Carlile et al. 1997). For all metrics, the result for each subject was calculated for each condition independently, i.e. a combination of microphone location and processing scheme. Lateral sound localization was assessed using the lateral error in the horizontal plane experiment. It is defined as the root-mean-square error between presentation and response lateral angles, with possible front/back errors disregarded. Vertical sound localization was assessed using the rate of front-back confusions from the horizontal plane experiment, as well as the local polar error and the quadrant errors based on the median plane experiment. The local polar error is defined as the root-mean-square error angular difference between presentation and response polar angle; where responses with an absolute error $\geq 90°$ were excluded. Contrarily, the quadrant error is defined as the percentage of responses where the deviation to the presented incidence was $\geq 90°$ (Middlebrooks 1999b). Whereas the local polar error measures the performance for fine localization, the quadrant error indicates the ability to perform coarse localization, similar to front-back confusions in the horizontal plane.

The data from subjects were excluded from further analysis if their localization performance with the reference condition in the main experiment was very poor, which would indicate either poor concentration or general spatial hearing deficits. The exclusion boundaries were defined based on the present free-field localization performance as: Lateral error$>15°$, front-back confusions $>25\%$, local polar error $>50°$, and quadrant errors $>8\%$. Local and quadrant polar errors are codependent, therefore data from subjects who exceeded the exclusion boundary in either metric were excluded from both metrics. Contrarily, for lateral errors and front-back confusions the exclusion was made independently.

4.1.3 Results

4.1.3.1 HRTFs and interaural cues

Sample HRTFs of one subject are depicted in Fig. 4.1.3 for three incidence directions in the median plane. In all panels, the top curve (ECEbl, DFeq) is the reference HRTF. Up to about 4 kHz, all HRTFs capture the coarse features of the reference well. Distinct differences are evident for the equalized HRTFs obtained at the ITEind_Entr and BTE microphone locations. The effect of free-field versus diffuse-field equalization is shown for the ITEind_Entr location. As compared to diffuse-field equalization, free-field equalization results in a smaller difference to the ECEbl-HRTF for frontal incidence, but a larger deviation at other incidence directions. Generally, the direction-dependent deviation to the reference HRTF is larger for the BTE than for the ITEind_Entr (Denk et al. 2018g).

The effect of hearing device delay is shown for the ECEbl microphone location in the two curves at the bottom (HA1 versus HA2). While both curves match the reference HRTF in the coarse shape, the delay leads to comb filtering effects visible here as spectral ripple. Above 2.5 kHz, the ripple declines for both conditions as a result of the attenuation of the leakage component. The filter design used in HA1 additionally results in a reduction of the spectral ripple in the low frequencies compared to HA2 (Denk et al. 2018d).

Figure 4.1.4 shows the interaural cues (ILD and ITD) for four sample conditions in subject VP_E1. The spectral power and unwrapped phase of the resulting HRTFs were smoothed independently to auditory resolution of one ERB (Breebart and Kohlrausch 2001) prior to calculating the interaural transfer function, out of which ILD and ITD were calculated for the indicated frequencies (Blauert 1997; Katz and Noisternig 2014). The broadband ILD was calculated by taking the broadband energy ratio between the HRTFs at both ears.

For the diffuse-field equalized ITEind_Entr HRTF, no considerable differences to the reference HRTF (ECEbl) are observed at the assessed frequencies in ITD. For the ILD, small differences are seen for the 8 kHz curve, but not at 1 kHz or in the broadband ILD. At the BTE microphone location with the same processing, differences become apparent in all curves, but more pronounced for the ILD. The differences are larger at high frequencies or in the broadband case and at incidence angles around ±90°.

With the hearing device processing (here HA2), no ILD distortions are apparent. However, large ITD distortions up to several milliseconds appear, most distinct for 1 kHz but also visible at the other frequencies. This corresponds well to the frequency range where the leakage component and the device output have similar amplitudes (c.f., Figs. 4.1.2 and 4.1.3). It should be noted that the oscillating shape of the curves is not an artefact of a wrapping problem, since the increments are not multiples of the appropriate period length. The behaviour probably originates from a variation of the superposition of the two delayed sound sources (leakage vs hearing device output) across incidence directions that differs between the sides.

4.1.3.2 Localization accuracy with virtual stimuli

Most subjects were able to localize the virtual stimuli in the reference condition with comparable accuracy as the free-field stimuli in the training sessions. They also reported a good externalization of sounds in all experimental sessions, which together indicates a satisfactory quality of the binaural reproduction system and

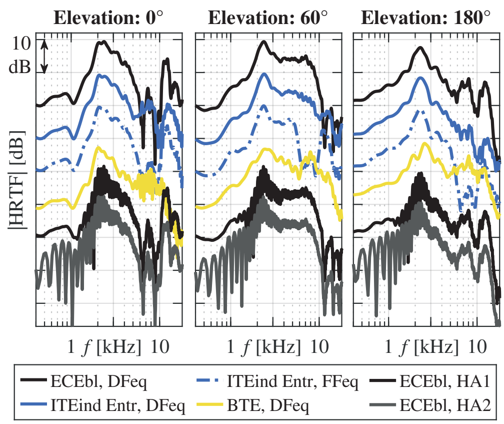


Figure 4.1.3: *Sample HRTFs, left ear of VP_E1 (22 Hz resolution). Three incidence directions from the median plane are shown in the individual panels, colors indicate different combinations of microphone location and processing. Individual curves were shifted by 10 dB for better display. See Tab. 4.1.1 for an explanation of the conditions.*

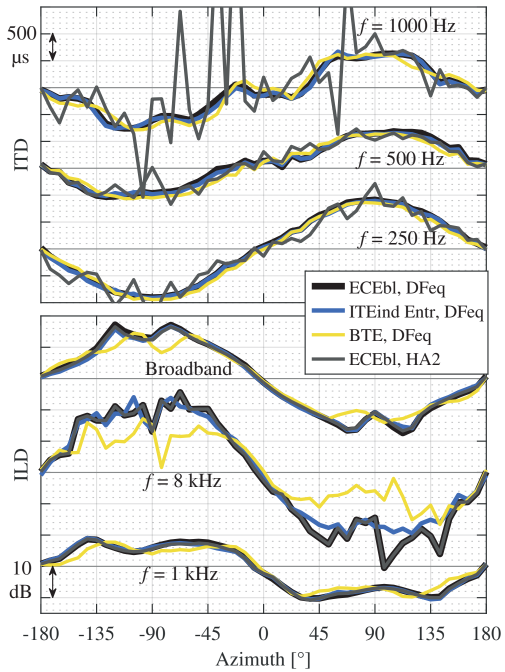


Figure 4.1.4: *ITD (upper panel) and ILD (lower panel) in the horizontal plane for four sample conditions as indicated by the line color, data from subject VP_E1. See Tab. 4.1.1 for an explanation of the conditions.*

the validity of the reference condition. The exclusion criteria led to discarding the data of two subjects for the vertical localization metrics, as well as two from the front-back confusions (in both cases VP_E9 and VP_E13). In consequence, all further evaluation includes data from 11 subjects for the lateral error and 9 subjects for the other metrics.

4.1.3.3 Lateral localization

Figure 4.1.5 shows the lateral errors (see Sec. 4.1.2.6) in the horizontal plane experiment. Within each processing scheme, the best performance is observed with the reference condition (ECEbl-DFeq). An influence of the microphone location is evident, independent of the processing scheme. For each processing scheme, the lateral error generally increases with the pre-sorting according to the distance of the microphone location from the eardrum (left to right).

The variance between subjects in the reference condition is usually smaller as compared to all other conditions (also in the other error metrics, see Figs. 4.1.6 and 4.1.7). This is mostly an artefact of a varying number of samples in the different conditions: Each incidence direction was presented three times in the reference condition but only once in the other conditions, while the error scores for one subject were calculated over all presentations in each condition. We verified that if each set of all possible incidence directions was evaluated independently in the reference condition, the variance between subjects would increase to a similar value as in the other conditions.

A one-way repeated measures analysis of variance (rmANOVA) for the six microphone locations within DFeq revealed a statistically significant main effect [$F(1,5) = 2.93, p = 0.021$], but no significant post-hoc effects between microphone locations (assessed by multiple pair-wise t-tests with Bonferroni correction). Nevertheless, a clear trend was observed with the medians differing up to 3.5° (or 40% with respect to the reference condition) between microphone conditions. The lateral error with the individual HRTF observed at the BTE and ITEgen locations is in the range of the lateral error observed with the HRTF of the KEMAR at the ECEbl location.

Furthermore, a two-way rmANOVA was performed with the factors processing scheme and microphone location (only considering ECEbl, ITEind_Entr and BTE, assessed in all processing schemes). Again, there was a significant effect of the location [$F(2,2) = 12.27, p = 0.0003$]. Although the processing scheme was a significant factor [$F(2,4) = 4.24, p = 0.0059$], the post-hoc test revealed that significant differences only occur when comparing with the KEMAR HRTFs. Specifically, no significant difference was noted between the equalized HRTFs and the conditions including a processing delay (HA1, HA2). However, especially for the ECEbl there is a trend that the lateral error is increased in the HA conditions (as compared to DFeq) by about 2° or 25% with respect to the reference condition.

The results for localization using the KEMAR HRTF deviate considerably from those with individual HRTFs. The errors are generally larger, and the ranking

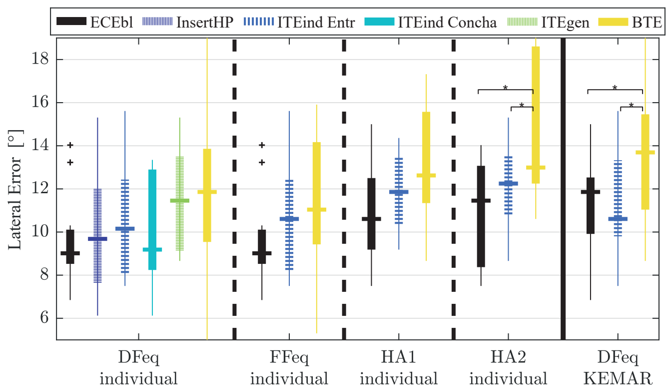


Figure 4.1.5: *Lateral errors from the horizontal plane localization experiment. Boxes indicate the distribution of results across all subjects, the horizontal bar denotes the median, thick vertical lines the 25%-75% quantiles, thin lines the whole data range excluding outliers, which are denoted by crosses. The color indicates the microphone location, whereas the position on the x-axes indicates the processing scheme. See Tab. 4.1.1 for an explanation of the conditions. Stars above horizontal brackets indicate significant post-hoc differences between the connected conditions, *: p <0.05, **: p <0.01, ***: p <0.001. Post-hoc significances are only shown for effects between microphone locations, see text for further details.*

of microphone locations changed — the lateral error is (not significantly) smaller with the ITEind_Entr location than with the ECEbl.

4.1.3.4 Front-back confusions in the horizontal plane

Figure 4.1.6 shows the percentage of front-back confusions (see Sec. 4.1.2.6) in the horizontal plane experiment. For comparison, the confusion rate for chance (i.e., random guessing) is plotted. Again, the best result is observed for the reference condition (ECEbl-DFeq); all other conditions lead to worse performance.

As for the lateral errors, an effect of the microphone location is evident independent of the processing scheme, with a ranking of microphone locations that corresponds well to the pre-sorting with increasing distance from the eardrum (left to right). The medians of the front-back confusion rates obtained with the BTE are very close to chance. A one-way rmANOVA applied to the individual microphone locations for DFeq revealed a significant main effect $[F(1,5) = 25.96, p < 0.0001]$ and significant post-hoc differences as shown in Fig. 4.1.6. This included significant differences between the BTE and all other locations, as well as between the ECEbl and the ITEind_Entr and the ITEgen.

Again, a two-way rmANOVA was performed with the factors processing scheme and microphone location (only considering ECEbl, ITEind_Entr and BTE). Significant main effects of processing $[F(2,4) = 6.66, p = 0.0005]$ and microphone location $[F(2,2) = 94.68, p < 0.0001]$ as well as a significant interaction $[F(2,8) = 3.40, p = 0.0026]$ were found. The post-hoc differences between individual conditions are shown in Fig. 4.1.6, which include mostly differences between microphone locations within the individual processing schemes. Between the processing schemes, the only significant differences included comparisons with the KEMAR HRTF. This means that there was no significant difference between free-and diffuse field equalization and the HA conditions, nor a considerable trend.

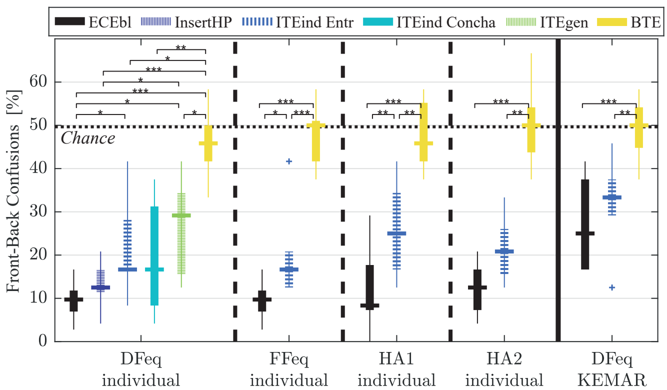


Figure 4.1.6: *Front-Back Confusions from the horizontal plane localization experiment. The dotted horizontal line indicates the result for chance. All other details per Fig. 4.1.5.*

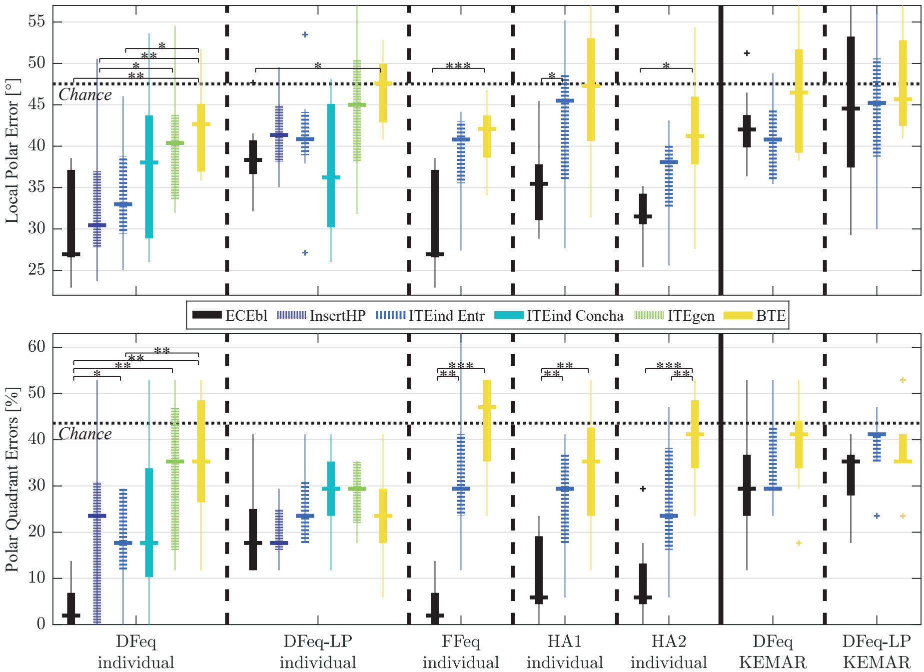


Figure 4.1.7: *Localization errors for the median plane localization experiment: Top: Local polar error, Bottom: Polar quadrant errors. Dotted horizontal lines denote the errors for chance. All other details per Fig. 4.1.5.*

For the ITE_Entr and BTE locations, free-field equalization led to less front-to-back and more back-to-front confusions as compared to diffuse-field equalization.[1] However, the absolute rate of confusions was not affected.

For localization using the HRTFs obtained in the KEMAR, the ranking of microphone locations is consistent with the results of individual HRTFs. However, the differences between microphone locations are much smaller than with the individual HRTFs due to an increase of front-back confusions for the ECEbl and ITEind_Entr while the rate stays constant at chance for the BTE.

4.1.3.5 Localization in the median plane

Both error metrics summarizing the results from the median plane experiment are shown in Fig. 4.1.7. For comparison, the appropriate errors for chance (random guessing) are shown. As in the horizontal plane experiment, the lowest localization error is observed with the reference HRTF (ECEbl-DFeq), all other conditions lead to increased errors.

[1] Percentage of front-to-back reversals within all front-back confusions, mean ± standard deviation across subjects: ITE-DFeq 58±39%, ITE-FFeq 38±24%, BTE-DFeq 56±19%, BTE-FFeq 27±10%. Percentage of back-to-front reversal is 100% minus given values, standard deviations are identical.

For the local polar error, the microphone location is again the governing factor for most processing schemes, and the pre-sorting with increasing distance from the eardrum (left to right) reflects well the localization results. Within the diffuse-field equalized (DFeq) microphone locations, a one-way rmANOVA revealed a significant effect of the microphone location [$F(1,5) = 6.83, p = 0.0001$]. Post-hoc analysis showed that the local polar error in the BTE is significantly different to the ECEbl, InsertHP and ITEind. Also, significant differences are observed between the InsertHP and the ITEgen location (see also Fig. 4.1.7). The median result with the InsertHP is very close to the ECEbl, however at this microphone location a large variance between subjects is observed, with an error corresponding to chance for some subjects. The differences between microphone locations are lower in the low-pass condition (DFeq-LP); but with increased errors in all locations but ITEind Concha.

A two-way rmANOVA with the factors microphone location (only ECEbl, ITEind_Entr and BTE) and processing scheme showed a significant main effect of both factors [location: $F(2,2) = 28.51, p < 0.0001$, processing scheme: $F(2,6) = 7.72, p < 0.0001$], but no significant interaction [$F(2,12) = 1.12, p = 0.35$]. The post hoc-test showed that significant differences between processing schemes are only observed in comparisons involving KEMAR conditions, except for a significant difference between DFeq and DFeq-LP for the ECEbl microphone location. No significant difference between the equalized HRTFs and the HA conditions, nor between the HA conditions, was observed. Within the ECEbl location, the polar error is slightly increased in the HA conditions, more pronounced in HA1. Within some processing schemes, significant differences between microphone locations were noted (as plotted in Fig. 4.1.7).

For the polar quadrant errors, similar trends can be observed. Comparing the diffuse-field equalized (DFeq) HRTFs, the ECEbl has the smallest error, while at all other locations the errors are comparable in size. There, a one-way rmANOVA revealed a significant main effect [$F(1,5) = 13.20, p < 0.0001$] and similar post-hoc differences as for the local polar error as shown in Fig. 4.1.7. In this error metric, smaller differences between microphone locations are observed in the low-pass condition. There, the polar quadrant error increases for the ECEbl, InsertHP and ITEind, but decreases for the other locations. For the FFeq and HA processing schemes, the trend is similar to the DFeq: the results for the BTE microphone location is always very comparable and near chance, but differ considerably from the ECEbl. The results for ITEind_Entr are in between, with an increased error compared to ECEbl. A two-way rmANOVA and post-hoc analysis analogous to the evaluation of local polar errors revealed significant main effects [location: $F(2,2) = 35.72, p < .0.0001$; processing scheme: $F(2,6) = 7.76, p < 0.0001$] as well as a significant interaction [$F(2,12) = 9.04, p < 0.0001$]. Significant post-hoc differences were noted between the ECEbl and BTE for the FFeq and HA processing schemes, and significances between ECEbl and ITEind_Entr for most processing schemes. Between processing schemes, no significant differences were found, except when comparing to KEMAR conditions. Although there is a trend that the quadrant errors with the HA conditions were increased as compared to the equalized HRTFs, this effect did not reach significance.

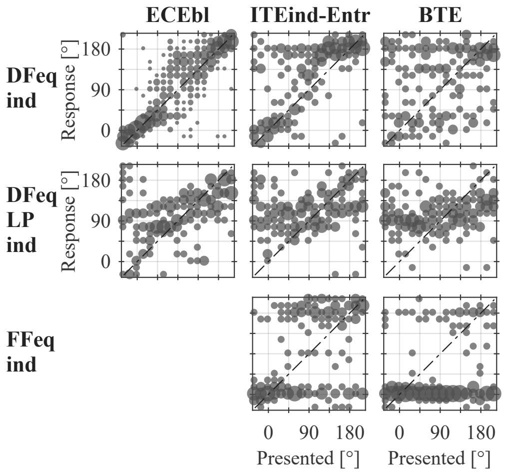


Figure 4.1.8: *Confusion matrices in the median plane experiment, for three microphone conditions (columns) and three processing schemes (rows). The size of the dots is proportional to the numbers of responses. 0° indicates frontal, 90° above and 180° rear sound incidence.*

For localization with the KEMAR HRTFs, the microphone location is irrelevant and both error metrics are at or near chance. This observation does not change in the low-pass condition.

Figure 4.1.8 shows the confusion matrices (i.e., angle-resolved scatter plot for responses over presentations) for all subjects and a subset of conditions. The error metrics from Fig. 4.1.7 are not sufficient to fully describe the localization disruptions. For instance, in the low-pass condition, the localization errors were smaller at the ITEind_Entr and BTE locations as compared to the full bandwidth. As demonstrated by the confusion matrices (second row compared to first row), this does not originate from better localization abilities with the restricted bandwidth. With the low-pass stimuli, subjects localized most stimuli above the head, with only little dependence on the microphone location. Only because the presented directions did not cover the whole circle and were biased to incidence from above, the resulting polar quadrant errors were smaller. Figure 4.1.8 also reveals that the increase of errors with free-field equalized as compared to diffuse-field equalized HRTFs results from systematically different localization patterns. For the FFeq processing, the majority of stimuli was localized at the frontal direction or at the rear direction near the horizontal plane. For the ITEind_Entr location, this means that the ability to perceive a coarsely correct elevation was not observed in about half of the subjects' responses with free-field equalization. For the BTE microphone, free-field equalization of the HRTF resulted in perceiving almost all median plane stimuli in the front.

For the HA1 and HA2 processing schemes, the confusion matrices are very comparable to the DFeq results (1st row) in the appropriate microphone locations. Localization with the KEMAR HRTFs led to an almost uniform distribution of responses, similar to the BTE-DFeq condition.

4.1.4 Discussion

4.1.4.1 Influence of the microphone location and bandwidth: Vertical localization

The location of the microphone alone produced a large disruption of vertical sound localization. The differences between locations were not substantially different when a hearing device delay was present, or when a free-field equalization was applied instead of a diffuse-field equalization. We conclude that the microphone location is the governing factor for vertical localization with linear hearing devices in normal hearing listeners.

Vertical localization depends on spectral directional cues (Blauert 1997). The present psychophysical results are thus well in line with the variation between HRTFs shown in Fig. 4.1.3, and model evaluations in previous studies (Denk et al. 2018g; Durin et al. 2014). This correspondence confirms that computational models are well suited to evaluate the effect of the microphone location on vertical localization (Baumgartner et al. 2013). The present data also revealed much larger differences between microphone locations that are non-optimal, i.e., between ITE and BTE locations, than previous studies (Best et al. 2010; Van den Bogaert et al. 2011). Using the information from a single BTE microphone, virtually all localization metrics were at or near chance. In other words, almost no vertical sound localization seems possible based on one BTE microphone. This is well in line with results from Best et al. (2010). Contrarily, the information from an ITE microphone appears to be sufficient for a reasonable distinction between front and back, and a coarse judgement of elevation. However, the performance with ITE microphone locations was subject to large variations between listeners. The present data also provides evidence that the exact positioning in the ear and the size of the device can directly impact the vertical localization performance, most pronounced for the local polar error, but also seen for front-back confusions in the differences between the InsertHP, ITEind, and ITEgen locations. This explanation fits well with the very large between-subject variance observed in all locations but the reference location, since the exact fit of these device styles is dependent on the individual ear. The present data also shows that there is no relevant difference between two microphones included in the same ITE device with respect to sound localization, which is in line with a previous model evaluation utilizing the same data (Denk et al. 2018g).

When the bandwidth was restricted to 8 kHz, the differences between microphone locations declined, but the trends mostly persisted. Moreover, the confusion matrices also showed that reduced polar quadrant errors for the low-pass condition (e.g., BTE) did not actually result from improved localization abilities, but are an artefact of the evaluation. Generally, the present results show that a restricted bandwidth is detrimental especially for good microphone locations and has no positive impact for others.

4.1.4.2 Influence of the microphone location: Lateral localization

The microphone location alone also had an impact on lateral localization. This aspect of localization is thought to be governed by interaural rather than monaural spectral cues. The impact of the microphone location on the ITD (shown in Fig. 4.1.4 for ITEind_Entr and BTE) was rather small at all assessed frequencies. The impact on the ILD was larger both in single auditory frequency bands as well for the whole audio bandwidth. Roughly, it can be stated that normal ITDs but distorted ILDs were available to the listeners in the DFeq conditions. Since ITD cues are thought to override conflicting ILD (Macpherson and Middlebrooks 2002; Wightman and Kistler 1992), we had expected no influence of the microphone location on the lateral error.

However, the results show that a non-optimal microphone location alone impairs lateral sound localization. While the differences did not reach significance in the present evaluation, clear trends were noted. Specifically, with the BTE and ITEgen microphone locations, the performance in lateral localization was as poor as when listening through a KEMAR. This impairment due to the microphone location results from biased spectral directional cues only and thus probably reflects a reduced spatial fidelity and increased localization blur. Previous studies using hearing-impaired listeners wearing hearing aids (Best et al. 2010; Van den Bogaert et al. 2011) or normal-hearing listeners wearing electronic hearing protectors (Brungart et al. 2003; Brungart et al. 2007) found no or a minor influence of the microphone location on lateral localization. In those studies, these rather subtle errors were probably overwhelmed by more prominent errors related to the devices like bandwidth limitations or nonlinearities, or effects of hearing impairment (Byrne and Noble 1998).

As for vertical localization, microphone positioning in the ear leads to poorer performance than with an optimal location, but to better performance than with a BTE microphone. Small differences are again apparent between the individual ITE microphone locations; mostly showing that the localization errors increase when the microphone sticks out of the ear (as in the ITEgen).

We want to note again that this condition includes no vent, i.e., no acoustic leakage component that might contain unbiased directional information is present. The influence of a vent is discussed in Sec. 4.1.4.4.

4.1.4.3 Free- versus diffuse-field equalization and implications for directional microphones

Diffuse- and free-field equalization produced no considerably different results for lateral localization, but large differences for vertical localization. The application of free-field equalization (for the ITEind and BTE microphone location) made it almost impossible for the subjects to perceive sounds to be displaced from the horizontal plane (Fig. 4.1.8). While for the ITEind_Entr location, this resulted in an impairment of elevation judgements as compared to diffuse-field equalization, for the BTE it resulted in a perception of most stimuli from the front, rather than a very random localization pattern as observed with diffuse-field equalization. The large impact is rather surprising, given that the only difference between the conditions is a constant spectral offset, namely a notch between 4 and 10 kHz that is typical for frontal incidence (c.f. Denk et al. 2018).

The results show that the overall frequency response characteristics of a hearing device have an influence sound localization. Diffuse-field equalization of the hearing device microphone leads to better localization performance than free-field equalization. Whereas diffuse-field equalization seems to better allow utilizing the residual spectral directional cues captured by the device microphone, free-field very effectively "drags" the apparent location of a source towards the front and, more importantly, towards the horizontal plane in general.

If more than one microphone is available, many hearing devices include directional microphones for noise reduction purposes (Doclo et al. 2015). Generally, directional microphones can provide a benefit for localization, mostly front-to-back confusions (Van den Bogaert et al. 2011), even if the generated cues are not necessarily consistent with natural spectral cues. The results for the effect of the microphone location demonstrate that directional microphones should be designed to not additionally disrupt spectral directional cues. Directional microphones could also be designed to restore parts of the pinna cues that are not captured by the device microphones (Schinkel-Bielefeld et al. 2018), which is equivalent to a direction-dependent free-field equalization. The present results demonstrate the potential positive impact of such approaches on sound localization and provide starting points for further refinements.

4.1.4.4 Influence of a vent and delayed hearing device output

The influence of a vent and processing delays generally showed no large effect on the vertical plane localization performance. Apparently, the spectral ripple created by the delay did not significantly bias the relevant spectral directional cues. In particular, the greatest importance of these cues is above 4 kHz (King and Oldfield 1997), where the ripple declined in the present data. Also, the lower spectral resolution of the auditory system with increasing frequency reduces the perceived spectral distortions. Nevertheless, small increments in vertical localization errors were noted between the equalized hearing device HRTF and the delay condition. Given that this increment was smaller for front-back confusions and polar quadrant errors than for local polar errors, an increased apparent source width would be a suitable model to explain the impact of the hearing device delay on vertical sound localization.

For lateral localization, the impact of the vent and hearing device delay was larger. That is, the lateral error increased in HA conditions as compared to the DFeq condition, for each microphone location. This indicates that the disturbing effect of a delay is approximately additive to the effect of the microphone location. Also, there is a trend that the lateral errors were larger when using the filter design from

HA2, i.e., producing a larger frequency region where leakage and hearing device output overlap. In the present results, the leakage component that directly enters the ear canal and carries unbiased directional cues in the low-frequency regime did not improve lateral sound localization, although improving sound localization is one of the many motivations behind using a vent or even open-fit devices (Akeroyd and Whitmer 2016; Noble et al. 1998). The detrimental influence of a vent in the current data can probably be explained by the disturbing effect of the superposition with the delayed hearing device output. The potentially beneficial effects of a vent might better come into effect when spectral overlap between the leakage and the hearing device output is avoided, or the cut-off frequency of the leakage component is increased.

Considering these results and also the evaluation of interaural cues from Fig. 4.1.4 (processing contains delays included in ECEbl-HA2), we assume that the impairment of lateral sound localization due to the delayed hearing device output originates from fluctuations of the ITD. Such uncertainties have been identified as a contributor to increased source widths (Käsbach et al. 2014), and increased source widths have been observed when listening through hearing devices (Cubick et al. 2018). The effects on lateral localization were slightly reduced by minimizing the comb-filtering effects using a suitable output filter design (Denk et al. 2018d), which largely eliminated the low-frequency output of the device in the frequency region of overlap with the direct sound component and thus reduced ITD distortions (see Fig. 4.1.3, bottommost curves). It is hard to assess to what extent the precedence effect plays a role in suppressing the perceptual effects of the delayed hearing device output, given that the spatial location of both signal components is equal, unlike in most studies on the effect of an echo (Brown et al. 2015).

4.1.4.5 Evaluation using dummy head HRTFs

When stimuli based on measurements on a KEMAR were utilized, the obtained localization results changed. More importantly, the variation was different between microphone locations, and also between error metrics. In all cases, the differences between the microphone locations were smaller with the KEMAR data than in the individual evaluation. Also, the performance for each microphone location was worse with the KEMAR, except when it was already at chance with the individual HRTFs (e.g., front-back confusions with the BTE). For the lateral error, the ordering of microphone locations was changed between the individual and KEMAR data. In summary, the data show that a mannequin generally is not suitable as a basis for a psychophysical assessment of localization performance with hearing devices.

Utilizing a different dummy head than the KEMAR is not expected to change or even improve this result. Among commercially available dummy heads, the KEMAR HRTFs produced model localization predictions that were most consistent with individual HRTFs (Denk et al. 2018g). The inconsistency between the present results with dummy head and individual HRTFs is probably caused by a different bias originating from the non-optimal microphone location in different ears, and a superposition with the localization deficits when listening through a non-familiar HRTF.

4.1.5 Conclusions

For normal hearing subjects listening through linear hearing devices without directional microphones, and when no dynamic binaural cues are available, the following conclusions are drawn:

(i) The microphone location is the governing factor for sound localization abilities. A non-optimum microphone location disrupts sound localization in the vertical domain, and reduces the accuracy in lateral localization. Microphones located inside the cavum conchae generally lead to better localization results than BTE microphones or microphones of ITE devices that stick out of the pinna. However, localization performance is decreased as compared to microphones located in the ear canal.

(ii) A bandwidth beyond 8 kHz is a key factor to full localization fidelity, independent of the microphone location.

(iii) A processing delay of 6 ms that is synchronous between the ears, together with a vent that allows low-frequency sound to directly leak into the ear canal, has little to no influence on sound localization in the vertical domain. Such delay artefacts led to a trend towards a negative effect on lateral localization, however it did not reach significance in the present experiment. Reducing the delay artefacts by appropriate filtering of the hearing device output did not affect localization performance as compared to the baseline processing.

(iv) Sound localization is not independent of the frequency response characteristics of the hearing device. To optimally exploit spectral cues captured by the microphone and improve localization performance, the frequency response of hearing devices should be equalized to the diffuse rather than to the free-field response of the open ear.

(v) Dummy head recordings are unsuitable for evaluating the impact of hearing devices on individual localization performance in psychophysical experiments.

Acknowledgement

This research was funded by the German Research Council DFG through the Collaborative Research Center SFB1330 *HAPPAA* – Project number 352015383 A4 and C5, as well as the by the European Regional Development Fund ERDF, Project *VIBHear*. We thank Steven van de Par for helpful discussions regarding the experimental design, Bernhard Eurich for conducting parts of the listening experiments, and our motivated subjects for participation.

4.2 Event-related potentials measured from in and around the ear electrodes integrated in a live hearing device for monitoring sound perception

Outline and context within the thesis

In this section, an experiment where ear-EEG sensors were included in the real-time demonstrator as reviewed in Sec. 3.1 is described. On the one hand, this study was the first to demonstrate that cognitive responses can be obtained from ear-EEG integrated in the hearing device while it is active, which is an important result on the way to cognitive controlled hearing devices. On the other hand, it was shown that a switch in the hearing device processing, here a change of the output filter between presentations of identical stimuli, is clearly perceivable and also measurable using ear-EEG. The EEG even provided an additional information compared to a subjective discrimination task, namely the point in time when the switch was perceived. This demonstrates the potential of ear-EEG as an objective means to study modifications in sound perception with hearing devices in future studies.

This section is a formatted reprint of
F. Denk, M. Grzybowski, S. M. A. Ernst, B. Kollmeier, S. Debener, and M. G. Bleichner (2018e). "Event-Related Potentials Measured From In and Around the Ear Electrodes Integrated in a Live Hearing Device for Monitoring Sound Perception". *Trends in Hearing* 22, p. 233121651878821.

Author contributions: FD, SE, SD and MB designed the experiment and performed pilot studies. FD implemented the experiment, MG conducted the measurements. FD and MB evaluated the data, created the figures and wrote the manuscript, where MB conducted EEG data analysis. All authors participated in writing the manuscript.

Abstract

Future hearing devices could exploit brain signals of the user derived from electroencephalography (EEG) measurements, for example, for fitting the device or steering signal enhancement algorithms. While previous studies have shown that meaningful brain signals can be obtained from ear-centered EEG electrodes, we here present a feasibility study where ear-EEG is integrated with a live hearing device. Seventeen normal-hearing participants were equipped with an individualized in-the-ear hearing device and an ear-EEG system that included 10 electrodes placed around the ear (cEEGrid) and 3 electrodes spread out in the concha. They performed an auditory discrimination experiment, where they had to detect an audible switch in the signal processing settings of the hearing device between repeated presentations of otherwise identical stimuli. We studied two aspects of the ear-EEG data: First, whether the switches in the hearing device settings can be identified in the brain signals, specifically event-related potentials. Second, we evaluated the signal quality for the individual electrode positions. The EEG analysis revealed significant differences between trials with and without a switch in the device settings in the N100 and P300 range of the event-related potential. The comparison of electrode positions showed that the signal quality is better for around-the-ear electrodes than for in-concha electrodes. These results confirm that meaningful brain signals related to the settings of a hearing device can be acquired from ear-EEG during real-time audio processing, particularly if electrodes around the ear are available.

4.2.1 Introduction

Although hearing devices can help to overcome even severe hearing problems, they are often not optimally adapted to the user. To improve the listening outcome in all situations and for optimal individualization, future hearing devices could exploit brain signals of the user derived from electroencephalography (EEG) measurements (Popelka and Moore 2016). Specific applications could be automatic objective fitting of the device based on neural responses (Finke et al. 2017; Lunner and Neher 2013) or real-time steering of signal enhancement algorithms based on the decoded direction of attention (Bleichner et al. 2016; O'Sullivan et al. 2015; O'Sullivan et al. 2017). The integration of EEG into hearing devices needs to be convenient for the user, so that it can be used without stigmatization in everyday situations. In the recent years, several approaches have been presented that allow to record EEG reliably in and around the ears, referred to as ear-EEG (Bleichner and Debener 2017; Debener et al. 2015; Goverdovsky et al. 2016; Looney et al. 2012). It has been shown repeatedly that ear-EEG can record a wide variety of cognitive processes related to auditory perception and auditory attention (Debener et al. 2015; Looney et al. 2012; Mikkelsen et al. 2015; Mirkovic et al. 2016). However, to the authors' best knowledge, ear-EEG has never been evaluated in a situation where it was integrated with a live ear-level hearing device.

We therefore present a first feasibility evaluation of integrating ear-EEG with a live electroacoustic hearing device. We combined a recently presented experimental high-fidelity in-the-ear hearing device (Denk et al. 2018c) with an around-the-ear electrode array consisting of 10 electrodes arranged in a C-shape (cEEGrid; www.ceegrid.com, Bleichner and Debener 2017, Debener et al. 2015) and 3 electrodes distributed in the concha (in-concha electrodes). To evaluate auditory perception through the hearing device, we examined whether perceivable switches in the electroacoustic transmission properties of the hearing device between repeated presentations of otherwise identical stimuli can be identified in the ear-EEG.

For identification of the switches in the EEG data, we specifically focused on the amplitudes in the latency ranges of the N100 and P300 event-related potential (ERP). The N100 is an ERP component that can be detected in response to an auditory stimulus and has a maximal amplitude between 50 to 150 ms after sound onset (Näätänen and Picton 1987). The N100 component shows a clear amplitude reduction for repeated sounds in a sequence, but amplitude increases again when a deviant sound is presented (Barry et al. 1992). The property of an amplitude reduction for repeated identical sounds and the amplitude recovery when a nonidentical (deviant) sound is presented provides an objective means to study whether sequential sounds are perceived as different or identical. The P300 with a maximal amplitude at around 300 to 500 ms after stimulus onset is a second ERP component that is elicited by task-relevant, deviant sounds and reflects the more conscious evaluation and categorization of the stimulus (Polich 2007). For the N100, we expected that the repetition of identical sounds leads to an amplitude reduction, while a deviant sound leads to an amplitude increase. For the P300, we expected a higher amplitude for a repeated deviant sound compared with a repeated identical sound. Importantly for the integration of EEG in hearing devices, these ERPs are comparatively robust responses that can be expected to be also recorded reliably under everyday situations outside of a lab context. ERPs can be exploited, for example, for automated fitting of hearing devices (Finke et al. 2017; Lunner and Neher 2013).

The position of ear-EEG electrodes is a crucial factor for future integration into one combined ear-level hearing device. There appears to be a trade-off between the compact positing of the electrodes to assure ease of use and optimal sensitivity to the brain signal of interest. Therefore, we compared the signal properties of the around-the-ear and in-concha electrodes. Due to the larger interelectrode distance of the around-the-ear electrodes compared with the in-concha electrodes and a finer angular coverage of bipolar channel orientations (Bleichner and Debener 2017), we expected a larger signal amplitude, better channel independence, and a better signal-to-noise ratio (SNR) for around-the-ear than in-concha recordings (Bleichner et al. 2015).

4.2.2 Methods

4.2.2.1 Participants

Seventeen subjects (age 28.4±5.4, 10 male, 7 female) with clinically normal hearing participated in the study. Normal-hearing participants were used here as we were interested in the general feasibility of our approach. The study was conducted in agreement with the declaration of Helsinki and was approved by the local ethical committee of the University of Oldenburg (Drs. 5/2015). Before active participation, each participant gave written informed consent.

4.2.2.2 Acoustic setup

The participants were equipped with a prototype hearing device as presented by Denk et al. (2018c). It consists of an individual soft silicone earmold with an integrated set of electroacoustic transducers as shown in Fig. 4.2.1. External sound is captured with the inbuilt pickup microphone located in the concha, processed, and played back on the inbuilt loudspeaker. Real-time processing was performed on a laptop running the Master Hearing Aid platform (Grimm et al. 2006), which was connected to the transducers through an *Multiface II* soundcard (*RME*, Haimhausen, Germany) with an input–output delay of 7.8 ms.

The hearing device was automatically calibrated in situ for each user to account for individual variations of the external ear (see Denk et al. 2018c for details). The aim was to provide acoustic transparency, that is, the pressure at the eardrum with the device inserted approximates the pressure at the eardrum that can be observed with an open ear (i.e., an unoccluded ear canal). The processing chain (here a finite impulse response filter) was individually adapted in a way that the superposition of electro-acoustically generated sound and a direct sound component leaking through the vented earpiece approximated the open-ear condition. Thereby, an additional

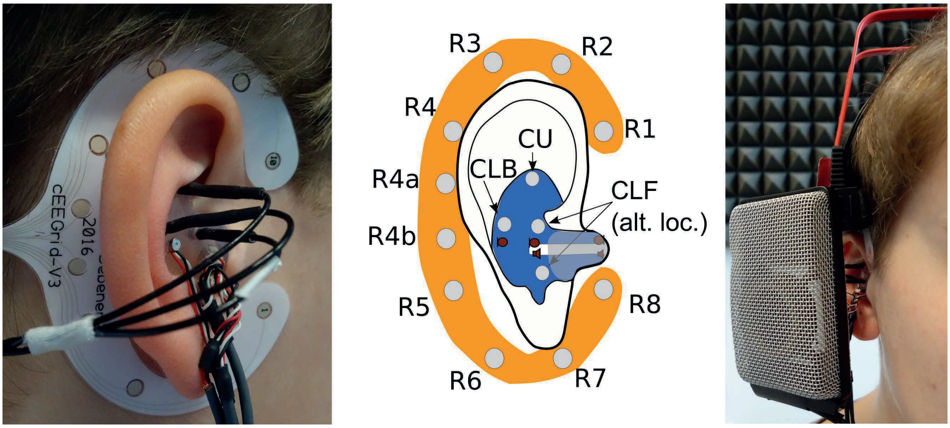


Figure 4.2.1: *Left: Photograph of the setup in the ear of a subject, with the cEEGrid glued around the ear using double-side adhesive tape. In the concha, the earmold containing the transducers of the hearing device (Denk et al. 2018c) and three additional electrodes (black wires) are placed. Center: Schematic view of the layout in the ear. Gray circles indicate electrodes with their according nomenclature; red symbols mark the positions of electroacoustic transducers. The shaded area marks the part of the earmold which is inserted into the ear canal. Electrode CLF has been placed at one of the indicated alternative locations in the individual ears, depending on which was feasible. Right: Subject wearing superaural headphones (AKG K-1000) which provide sufficient free space for the hearing device and electrodes.*

microphone located at the inner surface of the device pointing toward the eardrum was utilized to estimate the pressure at the eardrum.

Stimuli were presented on super-aural headphones (K-1000, *AKG*, Vienna, Austria), which are also shown in Fig. 4.2.1. The special design assured that neither the electrodes nor the hearing device was touched by the headphones. This setup represents a sound-source coupling to the ear similar to the free field (Møller et al. 1995b). The stimuli were presented monaurally on the right ear only, which was equipped with a hearing device, whereas the left ear was fully occluded.

4.2.2.3 EEG setup

EEG was acquired with a cEEGrid including 10 around-the-ear electrodes arranged in a C-shape (Bleichner and Debener 2017; Debener et al. 2015) and 3 electrodes positioned at different positions in the concha (in-concha electrodes) by insertion into bores in the earmold (see Fig. 4.2.1). Only electrodes positioned at the right ear were considered, that is, the side where the hearing device was located, to emulate a fully integrated system. The in-concha electrodes were Ag/AgCl miniaturized (2×4 mm) ton electrodes (i.e., shaped cylindrical, *Easycap* GmbH, Herrsching, Germany) as used in an earlier study (Bleichner et al. 2015). After skin preparation with an abrasive gel and alcohol, a small amount of electrolyte gel (Abralyt HiCl, *Easycap* GmbH, Germany) was applied to the electrodes. The cEEGrids were fixed around the ear with a double-sided adhesive tape and the

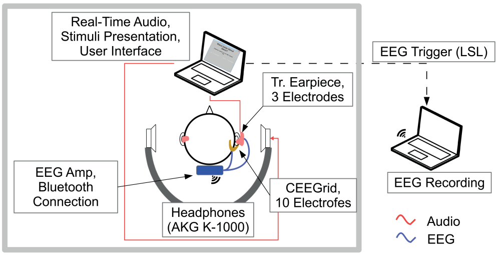


Figure 4.2.2: *Schematic depiction of the general experimental setup. The subjects were seated in a booth (indicated by thick gray line), where they conducted the experiments autonomously on a laptop that controlled the stimuli presentation, real-time audio processing for the hearing device, and EEG triggering through LSL.*

ton electrodes were inserted into the earmold up to skin contact after a drop of electrolyte gel was administered into the bores. All electrodes were connected to a wireless mobile 24-channel DC EEG amplifier (SMARTING, *mBrainTrain*, Belgrade, Serbia; custom modification for cEEGrid acquisition) positioned at the back of the head. EEG data were recorded with 24 bit resolution and 500 Hz sampling rate; electrodes R4a and R4b of the cEEGrid (see Fig. 4.2.1) served as ground and reference, respectively. Signals were wirelessly transmitted to a recording computer through a Bluetooth connection. Although the used system is a laboratory-state prototype, there is no principal reason that electrodes in this layout cannot be included in a real hearing device or a fully mobile prototype.

A schematic drawing of the recording setup is shown in Fig. 4.2.2. Acoustic stimulation and experimental control was implemented in MATLAB (*Mathworks* Inc., Natick, MA) on the same laptop that was also used to send EEG triggers synchronously to audio stimulation via Lab Streaming Layer, a software framework for data acquisition (Swartz Center for Computational Neuroscience and Kothe 2015). The subjects including the laptop were seated in a booth that was acoustically but not electromagnetically shielded. The EEG data as well as the event marker sent by the audio processing laptop were recorded using the recording computer located outside the booth using the LabRecorder Software from Lab Streaming Layer.

4.2.2.4 Paradigm

Participants listened to sequences of four sounds. In each trial, four identical sounds were presented via the headphones. In half of the trials, the filter setting of the hearing device was switched between the third and the fourth sound, resulting in a perceivable deviation. Participants indicated whether they perceived an acoustic difference between the third and fourth sound by pressing buttons (y/n) on a keyboard within 1 s after the end of the fourth sound. Two different filter settings of the hearing device were utilized for each subject; details on the filter design are described in Hearing Device Settings.

Since the EEG evaluation assumes the subjects' attention, data of subjects with a poor task performance were discarded. Only data from subjects whose responses fulfilled the following criteria were included in the analysis: (a) response correct

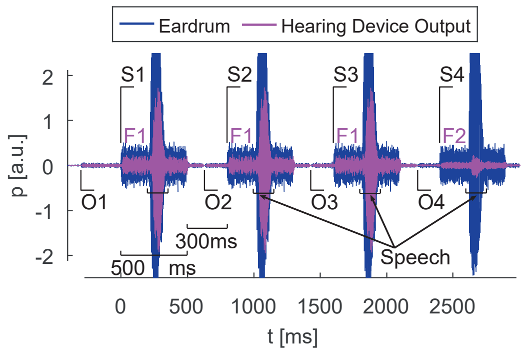


Figure 4.2.3: *Temporal recording of acoustic stimuli, obtained in a measurement on a dummy head. The example shows a trial for the Speech-in-Noise stimulus, including a switch from the hearing device output Filter F1 in the first three sounds (S1 to S3) to F2 in the last sound (S4). The blue line indicates the pressure at the artificial eardrum, which consists out of the direct sound leaking through from headphone playback and the output of the hearing device, indicated by the magenta curve (amplitudes not to scale). The different spectral profiles of F1 and F2 change the apparent power of the hearing device output; however, due to interference with the direct sound, the pressure at the eardrum has a similar level with both filter settings. Stimuli are 500 ms long with 10 ms ramps at the beginning and end, and separated by pauses of 300 ms, the range where the Speech stimulus is active is marked separately. O1 to O4 mark the onsets of the hardware noise before the first stimulus and after briefly turning it off between stimuli; S1 to S4 mark the onsets of the stimuli.*

(identical trials indicated as equal and deviant trials indicated as different) in more than 80% of all identical and deviant trials and (b) response given in more than 90% of all trials.

4.2.2.5 Stimuli

s, Three different stimuli were used: White noise (referred to as Noise), a logatome spoken by a female voice (Sass, from the OLLO corpus referred to as Speech; Meyer et al. 2010), and the combination of both with an SNR of 5 dB (referred to as Speech-in-Noise). White noise was selected since it best supports a detection of the generated deviations, whereas Speech and Speech-in-Noise represent more complex and realistic stimuli. For all stimuli, a bandpass filter between 0.1 and 12 kHz was applied to match the frequency range of the hearing device. The identical waveforms were played in all four repetitions of one trial to avoid random variations of the spectra of the noise stimuli.

Each sound of the Noise condition had a duration of 500 ms. For the Speech-in-Noise condition, the logatome was placed in the middle of the 500 ms noise segment (see Fig. 4.2.3). For the Speech condition, the logatome was placed at the same point in time as in the Speech-in-Noise condition to keep the speech onset time the same for these two conditions. For the Speech condition, this resulted in a longer period of silence before and after the logatome compared with the Speech-in-Noise condition. One trial consisted of four sounds presented sequentially with an interval of 300 ms between the 500 ms stimulus windows (beginning marked by S1 to S4 in Fig. 4.2.3).

The audio stimuli are shown in Fig. 4.2.3, which shows an example of a Speech-in-Noise stimulus measured at the artificial eardrum of a dummy head. The dummy head consisted of a custom adjustable ear canal simulator (Hiipakka et al. 2010) that was attached to KB1065/1066 pinnae (*G.R.A.S.*, Holte, Denmark) and mounted in a modified show-window mannequin. The dummy head was equipped with the same individualized earpieces as the subjects.

Each 16 deviant trials including a filter-switch in both possible orders of the two utilized filters and 16 non-deviant trials in either filter setting were presented. This was repeated for each stimulus (Noise, Speech, and Speech-In-Noise). Thus, in total, 192 trials were presented in randomized order, subdivided in four blocks of the same length. The experiment included additional conditions with a comparable number of trials, the results of which are not considered in this work. One session lasted about 90 min, which included individual adjustments of the hearing device filters prior to presentation of the stimuli (see next section). Prior to each block, the subjects had the opportunity to take a short break while remaining seated.

4.2.2.6 Hearing device settings

The deviation between repeated sounds was created by two different adjustments of the hearing device output filter. In one adjustment, the output filter of the hearing device was adjusted by individual calibration of the hearing device prior to the main experiment (Filter F1, see Sec. 4.2.2.2). In the other adjustment, the output filter resulting from equivalent calibration of the system on a dummy head was used (F2, same dummy head as for the recordings of Fig. 4.2.3). Due to different ear geometries and the tightness of fit, the filter responses varied. Hence, the spectral profile of sound arriving at the eardrum and, in the perceptual domain, the timbre was notably different between the two conditions. Alternative cues that may arise from differences in loudness were compensated through an additional broadband gain applied to the Filter F2, which was individually adjusted by means of an adaptive 1-up-1-down procedure conducted with the Noise stimulus prior to the main experiment.

A life hearing device was used, which made it almost inevitable to have a perceptible noise floor that originates mainly from the pickup microphone and fills the pauses. If the last sound is different, the output filter of the device is switched between playback of stimuli, whereas the operational setting is kept constant in non-deviant trials. To avoid sudden transient modification in the hardware noise coloration in deviant cases, the hearing device output was briefly deactivated 130 ms after presentation of Sounds 1 to 3 (20 ms pause, with 10 ms ramps, referred to as O2 to O4, see also Fig. 4.2.3). The length of the deactivation was adjusted such that the pause was as brief as possible, while it still perceptually separated the noise floor before and after in different sound events. The hearing device was activated 300 ms prior to presentation of the first sound of each trial (referred to as O1) and turned off 130 ms after the last sound.

4.2.2.7 EEG analysis

The analysis was performed offline with EEGLAB (Delorme and Makeig 2004) and MATLAB. For the statistical analysis, we used RStudio (*RStudio*, Boston, MA), for the planned comparisons regarding the stimulus onsets, p-values below 0.05 were considered as indication of statistical significance. For the exploratory analysis regarding the onset of the hardware noise, we corrected for multiple comparisons (i.e., two comparisons). In these cases, p-values below 0.025 were regarded as statistically significant.

The data from each block were filtered prior to segmentation between 0.1 Hz and 20 Hz with a consecutive high-pass filter (filter order 500) and low-pass filter (filter order 100, windowed sinc FIR filters with linear phase). Epochs were extracted for the entire trial (-1000 ms to 4000 ms) relative to the onset of the first sound (S1). Furthermore, epochs were extracted for the individual sounds (-500 ms to 1000 ms) relative to the onset of the respective Sounds S1 to S4. After first inspection of the results (see also Sec. 4.2.4), we also extracted epochs of the same length locked to the onsets of the hearing device O1 to O4, that is, the moments when the hardware noise started. Epochs dominated by artifacts were identified using the probability criteria implemented in EEGLAB (standard deviation: 2) and rejected from further analysis. On average, 27.1% (SD = 6.7%) of the trials were rejected. Averaging across the remaining epochs resulted in ERP waveforms.

The main contrast of interest was between identical and deviant sounds (S4). We expected a difference in amplitude of the N100, with a larger amplitude for the deviant compared with the identical condition. In addition, we expected a larger P300 for the deviant sound compared with the identical condition. Furthermore, we expected a decrease of the N100 amplitude from S1 to S2. The amplitude of the N100 was quantified by computing the mean amplitude for a 40 ms window around the N100 peak latency (±20 ms). The N100 latency was calculated as the peak latency of the N100 amplitude in response to S1. The same temporal window with respect to the stimulus onset was used to compute the amplitude in S2 to S4. The amplitude of the P300 was quantified by computing the mean amplitude for the time window of 230 ms to 430, relative to S4 and O4, respectively. Based on prior experience (Bleichner and Debener 2017), we computed the signal difference between the mean of two electrodes above and the mean of two electrodes below the ear ((R2 + R3)/2) − ((R6 + R7)/2); see Fig. 4.2.1 for the position of the electrodes. The resulting signal is in the following referred to as vertical bipolar cEEGrid channel and utilized to evaluate the ERPs in response to the switch in hearing device processing. All analyses were performed with the electrodes on the right side, that is, on the ear where the stimuli were presented (cf. Sec. 4.2.2.3).

4.2.2.8 Influence of electrode position

A second objective of this study was to compare the signal at different electrode positions in and around the ear. To evaluate this issue, we computed measures of effect size, signal to noise ratio, and between-channel similarities for three different electrode configurations.

Electrode configurations
In addition to the vertical bipolar cEEGrid channel, the signal quality was evaluated for three electrode configurations: (a) all individual electrodes of the cEEGrid referenced to the original reference (Grid), (b) all in-concha electrodes locally rereferenced to the other in-concha electrodes (Concha), and (c) all cEEGrid electrodes rereferenced to the mean of the in-concha electrodes (Grid/Concha).

Effect size
The grand average ERPs (mean over all participants and trials) were computed for identical and deviant sounds relative to O4 for the Speech stimulus. The effect size, that is, the difference between identical and deviant presentation, was measured as Hedges' g (Hedges 1981) for the first negative peak (± 20 ms) and for the P300 window (230 ms to 430 ms). Hedges' g is a variation of Cohen's d (Cohen 1988) but reduces the estimation error for smaller samples by correcting the pooled variance. The interpretation of Hedges' g is analogous to Cohen's d, and effect size above 0.8 (absolute value) is considered as large.

SNR comparison
To assess the quality of the ERPs for the N100 response to the first sound (S1), an SNR was calculated for the Speech stimulus in each participant separately. The SNR was computed for the N100 component by dividing the root mean square of the N100 component at peak latency ± 20 ms by the root mean square of the estimated noise in the same time window scaled to dB ($20 \log_{10}$ (signal/noise)). The noise was estimated using a plus–minus procedure (Schimmel 1967): The time signals of all trials are averaged after the polarity of every other trial is reversed. Assuming a signal that is coherent across repetitions, the resulting average is an estimate of the noise. For each electrode configuration, the individual channel with the maximal SNR was selected for each participant.

Interelectrode correlation
Furthermore, we assessed the influence of the electrode configuration on the between channel similarity. We therefore calculated the between-channel correlation coefficient between the individual channels of the Grid and Concha configurations, and the in-concha electrodes referenced to the recording reference (R4b). As we were interested in the overall signal similarity independent of the task and a specific neural source, all available data of the recording session were used and segmented into epochs of 1 s each. To prevent that the correlation is primarily driven by non-neural artifacts (e.g., eye blinks), epochs that contained artifacts were discarded (EEGLAB joint probability criterion, 2 standard deviations). The correlation coefficient was computed between all channel pairs for each epoch, and then averaged over epochs and participants.

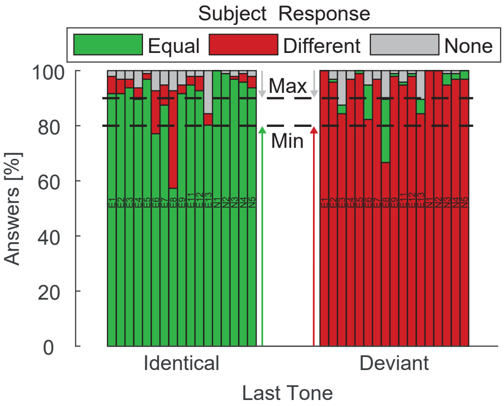


Figure 4.2.4: *Subjective discrimination results, pooled over stimuli. For Identical and Deviant trials, the individual subjects' responses are grouped to equal, different, and none. Data from three subjects (E6, E8, and E13) were excluded from further analysis because at least one of the criteria was not fulfilled: Response correct (i.e., equal for identical and different for deviant trials) >80%, no response (average over both conditions) <10%. The criteria are marked by dashed horizontal lines and colored vertical arrows.*

4.2.3 Results

4.2.3.1 Psychophysical results

Figure 4.2.4 shows the subjective discrimination results as indicated by the individual participants. Generally, identical and deviant trials were recognized with high accuracy. On average, identical sequences were indicated as such in 90.0% of all trials, and sequences where the last sound was deviant were indicated as different with 93.3% accuracy. In 3.6% of all trials, no response was given in the response window. These observations verified the clear audibility of the hearing device processing switch and the practicality of the paradigm. The task performance of three subjects (E6, E8, and E13) did not fulfill the criteria given in the Sec. 4.2.2.6 (also marked in Fig. 4.2.4). Consequently, their data were excluded from further analysis due to a suspected low level of attention.

4.2.3.2 EEG results

Piloting results
Extensive piloting tests had been conducted where the same ear-EEG setup and the same or a similar paradigm, but varying acoustic stimulation were utilized. In particular, sound presentation was also done on distant loudspeakers only, while the hearing device was in place but disconnected from the sound card. Electromagnetic cross-talk from the hearing device or headphone to the EEG electrodes could thus be ruled out by comparing the EEG recordings from loudspeaker stimulation to the hearing device or headphone stimulation. Generally, the piloting results verified that the stimulus-locked signal components observed in the ear-EEG electrodes do originate from brain activity.

N100 amplitude reduction
The N100 amplitude reduction was assessed for the vertical bipolar cEEGrid channel (see Sec. 4.2.2.7). Figure 4.2.5(a, top) shows the grand average ERPs, averaged over all trials for the Noise and Speech-in-Noise stimuli (Mark: Noise and Speech-

in-Noise have the same onset and offset time of the noise). Figure 4.2.5(a, bottom) shows the grand average ERP for the Speech stimulus. A negative deflection around 154 ms after stimulus onset (for S1) is clearly visible. A strong amplitude reduction of the N100 is evident for S2 and S3 relative to S1. Regarding the expected N100 amplitude reduction between S1 and S2, we computed a repeated measures analysis of variance (ANOVA) with the factors Stimulus Type (Speech, Speech-in-Noise,

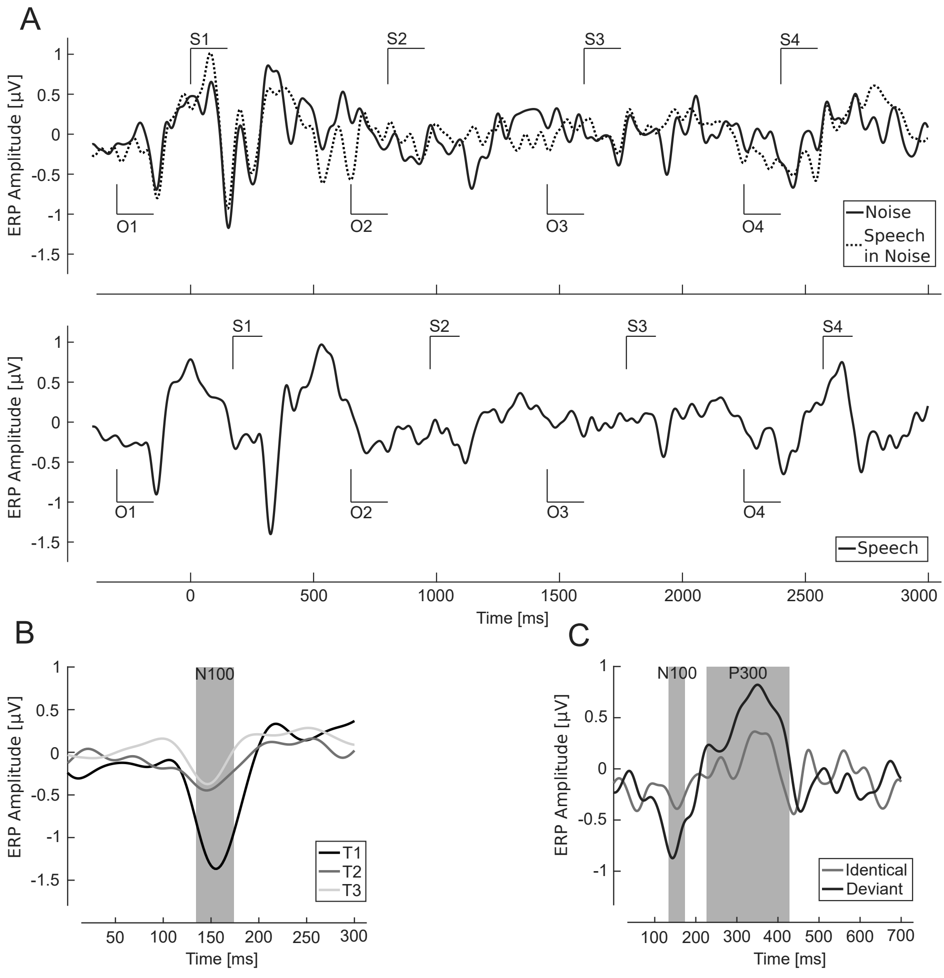


Figure 4.2.5: *A: Top: Grand average ERP for Noise (solid line) and Speech-in-Noise (dashed line) condition, combined for identical and deviant trials. Bottom: Grand average ERP for Speech (solid line). Clearly visible is the N100 response roughly 150 ms after sound onset (S1), as well as after the onset of the hardware noise (O1). B: Grand average ERP for S1 (black), S2 (dark gray), and S3 (light gray) for the Speech stimulus. Indicated in gray is the time window that was used for estimating the N100 amplitude. C: Grand average ERP for identical (gray line) and deviant filters (black line) for the Speech stimulus (O4). Indicated as a gray background is the time window that was used for estimating the N100 amplitude (first) and the P300 amplitude (second).*

and Noise) and Sound (S1 and S2). The results are shown in Fig. 4.2.6a and are also apparent in Fig. 4.2.5b for the Speech condition. We found a significant main effect of Sound on the N100 amplitude $[F(1, 13) = 13.22, p = 0.003]$ with a larger N100 amplitude for S1 compared with S2, and a significant effect of Stimulus Type on the N100 amplitude $[F(2, 26) = 4.31, p = 0.024]$ and no significant interaction $[F(2, 26) = 0.87, p = 0.431]$.

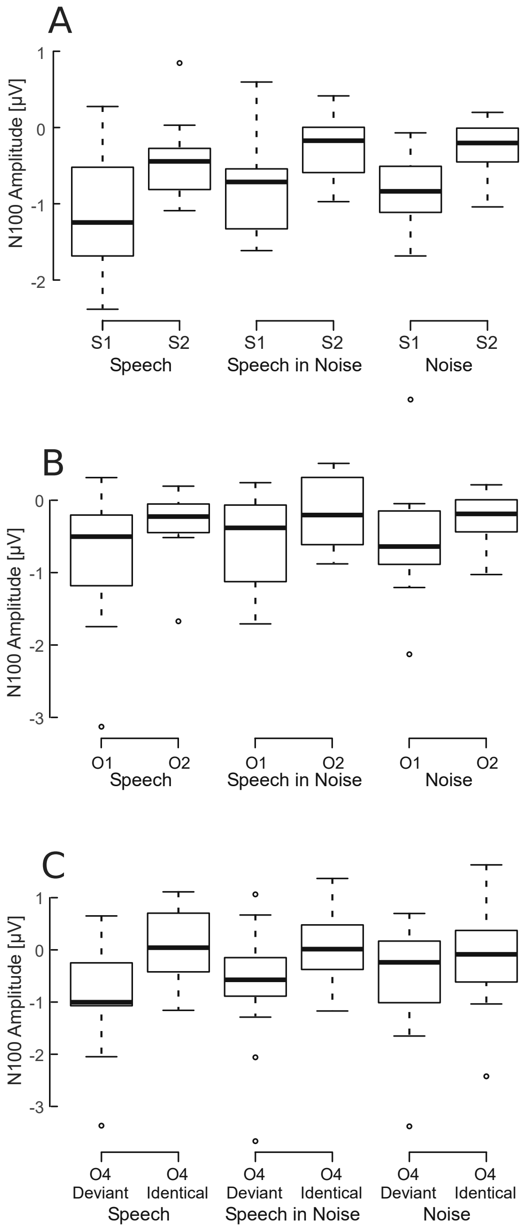


Figure 4.2.6: *A: N100 Amplitude in respect to stimulus onset for S1 and S2 for the individual stimuli. The N100 amplitude is reduced for S2 relative to S1 for all stimuli. B: N100 amplitude in respect to device onset (hardware noise onset) for O1 and O2 for the individual stimuli. C: N100 amplitude for identical and deviant condition in respect to device onset (O4) for the individual stimuli. Boxes and whiskers denote the 25% to 75% quantiles and full data range, respectively; the horizontal line indicates the median, circles denote outliers.*

Also apparent for all stimuli (Fig. 4.2.5a) is a negative deflection prior to stimulus onset with a latency that fits the onset of the hardware noise when the hearing device is switched on (O1). The N100 amplitude was decreased for the subsequent onsets (O1 to O2). We computed a repeated measures ANOVA on the N100 in response to the device onset with the factors Stimulus Type (Speech, Speech-in-Noise, and Noise) and Sound (O1, O2). The results are shown in Fig. 4.2.6b and are also apparent in Fig. 4.2.5b for the Speech condition. The N100 amplitude is larger for O1 compared with O2, analog to what we found for S1 and S2. However, the main effect of Sound on the N100 amplitude does not reach significance for O1 compared with O2 [$F(1, 13) = 4.39, p = 0.056$]. There was no effect of Stimulus Type on the N100 amplitude [$F(2, 26) = 1.32, p = 0.284$] and no significant interaction [$F(2, 26) = 0.02, p = 0.982$].

Filter switch
The identification of the filter switch is first assessed for signals from the vertical bipolar cEEGrid channel. Regarding the expected N100 amplitude reduction of the S4-identical relative S4-deviant condition, we computed a repeated measures ANOVA with the factors Stimulus Type (Speech, Speech-in-Noise, and Noise) and Filter-Switch (S4-identical and S4-deviant). We find a significant main effect of Filter-Switch on the N100 amplitude [$F(1, 13) = 6.16, p = 0.027$] with a larger N100 amplitude for S4-identical as compared with S4-deviant. This is contrary to the expectation of S4-deviant having a larger amplitude compared with S4-identical. There was no main effect of Stimulus Type on the N100 amplitude [$F(2, 26) = 2.32, p = 0.118$] and no significant interaction [$F(2, 26) = 1.69, p = 0.204$].

We also examined the N100 amplitude in response to the device onset (O4) and computed a repeated measures ANOVA with the factors Stimulus Type (Speech, Speech-in-Noise, and Noise) and Filter-Switch (O4-identical, O4-deviant). The results are shown in Fig. 4.2.6c and are apparent for the Speech stimulus in Fig. 4.2.5c. We found a significant main effect of Filter-Switch on the N100 amplitude [$F(1, 13) = 15.36, p = 0.002$] with a larger N100 amplitude for O4-deviant compared with O4-identical. There was no significant effect of stimulus type on the N100 amplitude [$F(2, 26) = 0.32, p = 0.738$] and no significant interaction [$F(2, 26) = 1.21, p = 0.314$]. The independence on the stimulus here was expected, as the hardware noise is independent of the stimulus type.

Finally, we analyzed P300 amplitudes in response to the device onset and computed a repeated measures ANOVA with the factors Stimulus Type (Speech, Speech-in-Noise, and Noise) and Filter-Switch (O4-identical and O4-deviant). The results are shown in Fig. 4.2.7 and are apparent in Fig. 4.2.5c for the Speech stimulus. We found the expected significant main effect of FilterSwitch on the P300 amplitude [$F(1, 13) = 12.25, p = 0.004$], with a larger P300 amplitude for O4-deviant compared with O4-identical. There was a significant effect of Stimulus Type on the P300 amplitude [$F(2, 26) = 6.21, p = 0.006$] and no significant interaction [$F(2, 26) = 0.01, p = 0.984$].

Electrode comparison
The results of the effect size (i.e., difference in ERP between identical and deviant trials on O4) comparison in dependence on the electrode layout are shown in

Fig. 4.2.8a. The largest effect sizes for the N100 and P300 time window were obtained for the Grid configuration (on average of the individual channels). The Grid electrodes behind and above the ear showed a medium effect size (but below the threshold for large effects of 0.8), while the effect size for the below-the-ear channels was relatively small. Some individual Grid channels located behind the ear exceeded the effect size that was observed with the vertical bipolar cEEGrid channel utilized in the previous section. The in-concha electrodes referenced locally to other in-concha and the Grid electrodes referenced to Concha electrodes also showed small effect sizes.

The SNR analysis of electrodes at different positions and referencing schemes is shown in Fig. 4.2.8b. The largest SNR was observed in the cEEGrid electrodes (Grid configuration, $M = 16.18, SD = 6.16$), which was significantly higher than the maximum SNR for inconcha electrodes referenced locally to in-concha (Concha configuration, $M = 7.55, SD = 6.70$); paired t-test: $t(13) = 4.10, p = 0.001$. The maximal SNR for channels in the Grid configuration was not significantly different from the maximal SNR for cEEGrid electrodes referenced against the in-concha electrodes (Grid/ Concha configuration) [$t(13) = -0.98, p = 0.341$].

The analysis of the interelectrode correlation patterns is shown in Fig. 4.2.9a. For the cEEGrid channels, the median correlation was $r = 0.40$ ($\text{min} = -0.16, \text{max} = 0.75$). The smaller the angle between the respective electrodes (relative to the reference electrode), the higher their signals were correlated ($r = -0.94, p < 0.001$; Fig. 4.2.9b). The in-concha channels (referenced to R4b of the cEEGrid) were highly correlated to each other, with a median interchannel correlation of $r = 0.99$ ($\text{min} = 0.99, \text{max} = 0.99$). The correlation between cEEGrid channels and in-concha EEG channels was low, with a median interchannel correlation of $r = 0.12$ ($\text{min} = -0.06, \text{max} = 0.23$). The median interchannel correlation for the

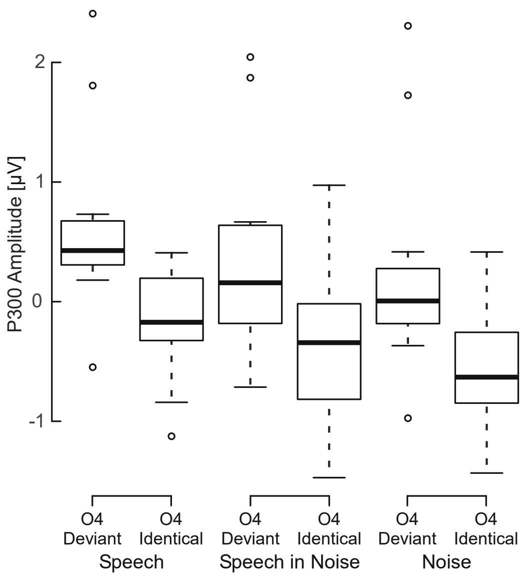


Figure 4.2.7: *P300 in respect to device onset (hardware noise onset) at O4 for the individual stimuli, comparison of identical and deviant trials. Over all stimuli, we observe a higher amplitude in the P300 window for the deviant sounds compared with the identical sounds. Boxes and whiskers denote the 25% to 75% quantiles and full data range, respectively; the horizontal line indicates the median, circles denote outliers.*

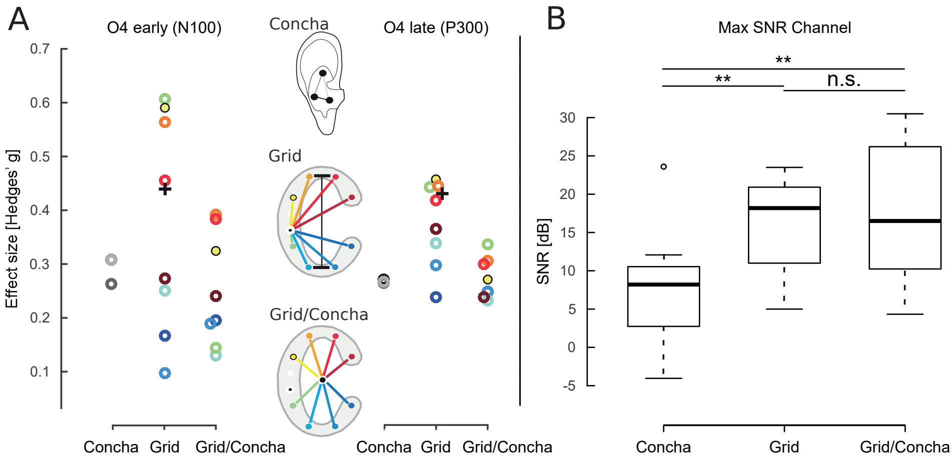


Figure 4.2.8: *A: Mean effect size (Hedges' g, absolute values) for the N100 (left) and P300 (right) between conditions (O4, deviant/identical). The color lines from the insets in the center indicate the electrodes and the respective reference. The cross or black line in the Grid condition marks the vertical bipolar cEEGrid channel that has been utilized for the Filter-Switch discrimination in the previous section. The in-concha channels (Concha) were locally rereferenced to the other in-concha channels. The cEEGrid channels (indicated as colored dots) were either rereferenced to the R4b (white electrode with black dot, Grid condition) or to the mean of the in-concha channels (Grid/ Concha condition). B Maximal SNR (maximum of channels in each condition) for the N100 in respect to the stimulus onset (S1) of the Speech stimulus, individual subjects grouped. The in-concha electrodes were locally rereferenced to other in-concha electrodes (Concha), for cEEGrid channels, references to the below the ear channels or the original reference (Grid), for the Grid/Concha condition the cEEGrid channels were rereferenced to the mean of the in-concha channels.*

three possible in-concha bipolar channels (Concha configuration, i.e., each inconcha channel referenced to the two other channels) was $r = 0.42$, but showed a wide range ($\text{min} = -0.14, \text{max} = 0.99$).

4.2.4 Discussion

4.2.4.1 ERPs and identification of the filter switch

We observed robust ERPs in response to the onset of the stimuli, including N100 and P300-like structures that could be exploited to identify a switch in the filter of the hearing device. For the N100, we found the expected amplitude reduction in response to repeated sounds. For a deviant last sound, the expected increase in the N100 amplitude compared with an identical last sound was not present in our data. However, this result can be fully explained by considering the hardware noise of the hearing device. It is important to note that the manipulation of the filter settings in our experiment lead to a change in the spectral properties, both for the actual stimuli and the hardware noise. That is, a filter-switch between S3 and S4 became already apparent at the moment the device was switched back on (O4). To avoid sudden transient of the audible noise floor on a filter-switch, we

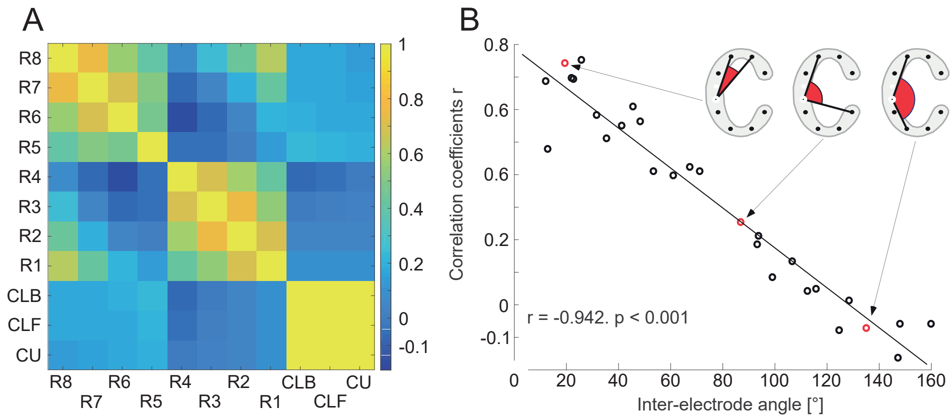


Figure 4.2.9: *A: Between-channel correlation of all around-the-ear and in-concha channels (referenced to electrode R4b). The four belowthe-ear (R5 to R8) and above-the-ear (R1 to R4) electrodes are more highly correlated to each other than to the electrodes of the other group. The in-concha electrodes (CLB, CLF, and CU) are highly correlated to each other and largely uncorrelated to the electrodes around the ear. B: Inter-electrode angle difference (in respect to reference electrode) relative to the inter-electrode correlation. The larger the inter-electrode angle, the lower the inter-electrode correlation ($r = -0.942$). The insets exemplify the relation of inter-electrode correlation and inter-electrode angle.*

had decided to briefly switch the hearing device on and off between presentation of stimuli while the filter was switched (or not). The onset of the hardware noise was very salient, particularly the first onswitch prior to presentation of the first sound which lead to a N100 (Fig. 4.2.5, 100 ms after O1) that was comparable in size to the N100 to stimulus onset (Fig. 4.2.5, 100 ms after S1). The N100 also showed a reduction in amplitude from the first (O1) to the second hardware noise onset (O2). Most importantly, we did find a difference in the N100 amplitude between identical and deviant trials with respect to the moment when the device was switched back on between O3 and O4. The task for the participants was to detect a difference in the last sound of the sequence and to respond with a button press. The expected difference was not observed in the latency range with respect to the onset of the last sound (S4). The P300, which could be associated with the more conscious detection of the difference of the acoustic properties, was consequently also locked to O4. In conclusion, the EEG results suggest that the decision (deviant/identical) could already be made at the time O4, that is, based solely on the hardware noise of the hearing device. Once the change in the hardware noise was detected, no further auditory processing of the actual stimulus was necessary. It should specifically be noted that the time at which the subject was able to make the decision was revealed purely by the EEG analysis and cannot be resolved based on the present psychophysical data.

The conclusion that subjects made the discrimination based on the hardware noise is supported by the comparison of the ERP amplitudes for the different stimuli. The largest amplitude change for both the N100 and P300 between identical/deviant last sounds is present for the Speech condition. The time between device onset and stimulus onset was largest for this condition. For the Speech condition, the hardware noise was audible for a longer duration and could therefore more easily be used to identify the filter condition. Consequently, the respective ERPs could be estimated more reliably as separate ERPs in response to different sound onsets overlapping to a smaller degree compared with the Noise and Speech-in-Noise condition. Especially in the P300 latency range, the ERP in response to O4 for the Noise and Speech-in-Noise stimuli appears to largely overlap with the ERP in response to S4, leading to small P300 median amplitudes (cf. Fig. 4.2.7). This also explains the significant dependence of the P300 amplitude in response to O4 on the stimulus, although the sound the response is locked to is independent of the stimulus.

Admittedly, these results were not intended. We aimed to study the influence of the filter settings on the perception of the stimuli, and the hardware noise was regarded as a disturbance. However, our results provide relevant practical insights for future EEG studies regarding aided hearing. The hardware noise of hearing devices is a common issue that can often be neglected for hearing impaired subjects, but becomes relevant in normal-hearing subjects as in this study. Changes in the acoustic properties of the hardware noise are readily perceivable by the participants, leading to unwanted effects that bias the results. Apparently, avoiding transient cues when switching the processing setting by briefly turning off the hearing device were not sufficient to suppress such disturbances. For future similar studies, we recommend to temporally separate any switches in the device operation from stimulus presentations by several hundred milliseconds, such that the cortical responses to both events are temporally separated. If problems specifically originate from audible hardware noise, another option is to additionally present noise that perceptually masks this cue, as successfully applied in (Denk et al. 2018c).

4.2.4.2 Signal quality in the individual electrode positions

The second objective of our study was to compare the signal quality in the around-the-ear and in-concha electrodes. For around-the-ear recordings, we used a cEEGrid, and for the in-concha recording, we used small ton electrodes. We found the larger effect sizes for the switch identification and the better SNRs for the cEEGrid electrodes. Given the effect sizes for the filter switch discrimination, the electrodes above and behind the ear referenced to a behind-the-ear location seem to be most informative.

Regarding the between-channel correlation analysis, we found that the signals recorded at different cEEGrid electrodes show differences in the correlation scores. The between-channel correlation depends on the angle between two electrodes (relative to the reference electrode), the smaller the angle the higher the signals were correlated. This is also in line with earlier studies where we find that different cEEGrid channels are sensitive to different neural sources (Bleichner and Debener 2017). When referenced against a cEEGrid electrode located behind the ear, the concha channels showed very high between-channel correlations ($r > 0.95$). The correlation to the other cEEGrid electrodes, however, was small. Also, when the in-concha channels were rereferenced to one of the other in-concha channels (i.e., local bipolar reference), the interchannel correlation was much smaller ($r \approx 0.4$). These results underline the importance of the choice of reference for ear-EEG recordings.

The small interelectrode distance in the concha and the small angle between in-concha electrodes relative to the far away cEEGrid reference results in highly correlated signals, by using a local reference; however, distinct signals can be picked up also in the concha. This is in line with earlier studies that have shown that distinct neuronal activity can be picked up at different electrode sites in the ear canal and concha (Mikkelsen et al. 2015). Furthermore, the in-concha electrodes were largely uncorrelated to the around-the-ear-channels. Together with results of earlier studies and the results reported here that show that meaningful information is recorded by the in-concha electrodes (Mikkelsen et al. 2015), we take this as evidence that cEEGrid and in-concha electrodes capture non-redundant information. How much independent information can be obtained at the different recording sites depends on the position of the neural source and interindividual anatomical differences.

The observed SNRs were generally higher in the around-the-ear electrodes than in the in-concha electrodes. However, we also found that the SNR of the in-concha electrodes can be as high as the SNR of around-the-ear electrodes in some participants. We used in-concha electrodes that were not tightly integrated with the earmold but were plugged into a bore, in addition to an electrode technology that was different between around-the-ear and in-concha electrodes. These factors might have contributed to the lower SNR for the in-concha electrodes. Another relevant factor could be amplifier noise (Kidmose et al. 2013). The signal amplitude decreases as the electrode distance decreases, that is, the overall signal amplitude of the concha channels is small. It is hence possible that the amplifier noise (input noise 1 mV, according to amplifier specification) dominated the signal, which could also have led to the observed reduction in SNR. For future work, identical electrode technology should be used to unambiguously compare the SNR between in-concha versus around-the-ear EEG. Furthermore, an amplifier with even lower low input noise should be used to potentially improve the SNR of channels that are spatially close to each other.

4.2.5 Conclusions

We studied the feasibility of using around-the-ear (Bleichner and Debener 2017; Debener et al. 2015) and in-concha EEG electrodes integrated with a live hearing device (Denk et al. 2018c) to objectively monitor sound perception. Although the used system was a laboratory-state prototype, there is no principal reason that electrodes in this layout cannot be included in a real hearing device. Specifically, we switched the processing settings of the hearing device between repeated sounds and studied the effects on the N100 and P300 waves of the ERPs. Besides the identification of such filter switches in the ERPs, we evaluated the signal quality of around-the-ear and in-concha electrodes.

The results show that high-quality ERPs can be obtained from ear-EEG during hearing device activity without crosstalk from the electroacoustic system. The filter switch of the hearing device could be identified in the ERP responses. The EEG results revealed that the subjects detected the switch already in the hardware noise of the hearing device preceding the actual stimulus, although transient changes were avoided—an issue that should be considered in future applications of EEG together with aided hearing. Importantly, the EEG provided additional insight in the perception process that was not apparent in the current psychophysical data, namely the point in time where the deviation was perceived by the subject.

Signals from in-concha and around-the-ear electrodes were largely uncorrelated and appear to capture non-redundant information. Also, individual around-the-ear electrodes were less correlated among each other than the in-concha electrodes. For the correlation between electrode channels, the position of the reference electrode and the angle between electrode-reference vectors seem to be the most dominant factor, rather than the distance. The achievable SNR appears to be higher in the around-the-ear electrodes than in the in-concha electrodes, although the different electrode technologies utilized here may have biased the result. The effect size for the switch discrimination was larger when only around-the-ear electrodes were utilized than for sole utilization of in-concha electrodes. Altogether, according to our results, an array of electrodes distributed around (most importantly above and behind) the ear, potentially extended by one (or more) electrode(s) in the concha or ear canal, is highly promising for applications in hearing devices.

In summary, our results show that using ear-EEG in combination with live hearing devices is possible. EEG signals acquired in close proximity of the ear contain relevant neural information that may be harvested to realize brain-computer interface technology that is integrated into hearing devices of the future.

Acknowledgements

The authors thank Reiner Emkes for technical support, Bojana Mirkovic for practical support in handling the cEEGrids, and Tobias Herzke for support on the Master Hearing Aid.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was funded by the DFG through the Cluster of Excellence EXC 1077 *Hearing4all* (Task Groups 4

"Algorithms for Hearing Instruments" and 7 "BCI for Hearing Aids") and Research Unit FOR1732 *Individualized Hearing Acoustics*.

4.3 Subjective sound quality evaluation of an acoustically transparent hearing device

Outline and context within the thesis

In this section, a subjective evaluation of the revised real-time demonstrator of an acoustically transparent hearing device that includes most results from this thesis, is described. In particular, the demonstrator includes the target definition based on Sec. 2.2, the equalization filter design from Sec. 3.3, all in integration with a feedback cancellation algorithm as described in (Schepker et al. 2019a). Different possible approaches for integrating sound equalization and feedback cancellation were implemented and evaluated. For the subjective evaluation, an approach similar to that of Denk et al. (2016) was chosen, where stimuli were recorded on a dummy head wearing the real-time hearing system in subsequent settings while repeatedly playing the same over loudspeakers stimuli. These recordings were reproduced to subjects via headphones for quality rating in a modified MUSHRA test.

This section is a formatted reprint of
H. Schepker, **F. Denk**, B. Kollmeier, and S. Doclo (2019b). "Subjective sound quality evaluation of an acoustically transparent hearing device". *Proc. 2nd AES Conference on Headphone Technology.* Paper 18. San Francisco, USA, pp. 1–10.

Author contributions: HS and FD contributed equally to this article. FD was in charge of implementing and maintaining the demonstrator; HS and FD designed, implemented and conducted the general study and the acoustic measurements. HS conducted the listening test, HS and FD evaluated the data, HS generated the figures and wrote the manuscript in collaboration with FD. SD and BK participated in the study design and writing the manuscript.

Abstract

In this paper we evaluate the performance of a real-time hearing device prototype that aims at achieving acoustically transparent sound presentation. Acoustic transparency refers to the perceptual equivalence of the sound at the aided ear drum, i.e., with the hearing device inserted and processing on, and the open ear drum, i.e., without the hearing device inserted. The considered hearing device combines a custom earpiece with multiple microphones and signal processing algorithms for robust feedback suppression and sound pressure equalization. We evaluate the perceived overall sound quality of this prototype using dummy head recordings in different acoustic conditions using a multi-stimulus with hidden reference and anchor-like framework with $N = 15$ normal-hearing subjects. Results show that the overall sound quality can be significantly improved for all conditions by using sound pressure equalization, where the processing delay of the device is a crucial limiting factor of the sound quality.

4.3.1 Introduction

In the past decades, major improvements have been made in the area of assistive listening devices like hearing aids and consumer headsets. Nevertheless, the acceptance of these devices remains rather limited, with limited sound quality identified as one of the major reasons (Killion 2004b; Sockalingam et al. 2009), especially for normal-hearing and mild-to-moderately hearing-impaired subjects. While these people would benefit from advanced signal processing in hearing devices (Doclo et al. 2015), e.g., beamforming, dynamic range compression, and dereverberation, they are usually not willing to accept a reduced sound quality (Killion 2004b). Therefore, recently the concept of acoustic transparency has become increasingly popular, which aims at creating the acoustic impression of open ear listening while the device is used (Denk et al. 2017d, 2018c; Hoffmann et al. 2013a; Rämö and Välimäki 2014; Välimäki et al. 2015).

Acoustic transparency is achieved when the sound at the aided ear drum, i.e., with the device inserted and processing on, and the open ear drum, i.e., without the device inserted, is perceptually equivalent. Typically, an equalization filter is used to modify the signal picked up by the hearing device such that in superposition with the sound leaking into the (partially) occluded ear canal the desired acoustic characteristics of the open ear are obtained (Denk et al. 2018c; Hoffmann et al. 2013a). However, since the output of the hearing device is typically delayed, this superposition may cause comb-filtering effects, possibly degrading the perceived sound quality (Stone et al. 2008). For vented hearing devices, it is especially important to take into account the sound leaking into the ear canal (Denk et al. 2018c). A larger vent typically increases the risk of acoustic feedback and may also reduce the effectiveness of noise reduction algorithms (Dalga and Doclo 2011; Gatehouse 1989; Winkler et al. 2016). Hence, for an acoustically transparent hearing device with a larger venting, acoustic feedback suppression is an important component. Since algorithms for equalization and feedback suppression are typically designed independently, combining these algorithms may cause undesired interactions.

The purpose of this paper is twofold. First, we present a prototype real-time acoustically transparent earpiece with integrated single-loudspeaker equalization (Denk et al. 2018d) and multi-microphone acoustic feedback suppression based on an fixed null-steering beamformer (Schepker et al. 2019a). Second, we perform a subjective sound quality evaluation using dummy head recordings and a multi stimulus with hidden reference and anchor (MUSHRA)-like framework (Denk et al. 2016; Völker et al. 2018) with $N = 15$ normal-hearing subjects. We address the following research questions: 1) which equalization algorithm yields the highest perceptual quality compared to the open ear; 2) how is the performance of both the fixed null-steering beamformer and the equalization algorithm affected by different incoming signal directions and reverberation times; 3) how valid is the equalization target used in the equalization filter design; 4) do interactions of the feedback suppression algorithm and the equalization algorithm yield a reduced sound quality.

4.3.2 Methods

In Sec. 4.3.2.1, we first present an overview of the considered acoustic hearing device system. We then briefly review the computation of the null-steering beamformer for feedback suppression in Sec. 4.3.2.2, and the equalization for acoustic transparency in Sec. 4.3.2.3. We describe their real-time implementation in Sec. 4.3.2.4. Finally, we describe the experimental setup for the subjective quality evaluation in Sec. 4.3.2.5.

4.3.2.1 Hearing device system overview

Consider the hearing device system with one loudspeaker and M microphones depicted in Fig. 4.3.1. This block scheme shows the hearing device processing and all acoustic transfer functions between the source and the eardrum. We assume that all acoustic transfer functions are linear and time-invariant and can be modelled as polynomials in the delay operator q (Ljung 1998).

The signal $y_m[k]$ in the mth microphone, $m = 1, \ldots, M$, at discrete time k, consists of the incoming signal component $x_m[k]$ and the feedback component $f_m[k]$, i.e.,

$$y_m[k] = x_m[k] + \underbrace{V_m(q)u[k]}_{f_m[k]}, \tag{4.3.1}$$

where $V_m(q)$ denotes the acoustic feedback path between the loudspeaker and the mth microphone and $u[k]$ denotes the loudspeaker signal. Furthermore, we assume a single directional incoming signal, i.e.,

$$x_m[k] = H_m(q)s[k], \tag{4.3.2}$$

where $H_m(q)$ denotes the acoustic transfer function between the source $s[k]$ and the mth microphone. For convenience we rewrite the microphone signals using

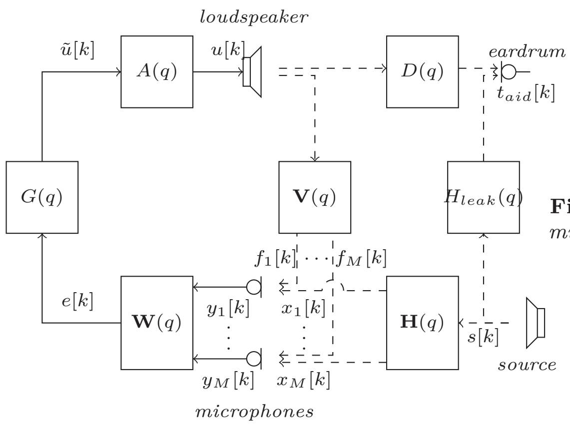


Figure 4.3.1: *Single-loudspeaker multi-microphone hearing device system.*

vector notation, i.e.,

$$\mathbf{y}[k] = \mathbf{H}(q)s[k] + \underbrace{\mathbf{V}(q)u[k]}_{\mathbf{f}[k]}, \tag{4.3.3}$$

where

$$\mathbf{y}[k] = \begin{bmatrix} y_1[k] & \dots & y_M[k] \end{bmatrix}^T, \tag{4.3.4}$$

$$\mathbf{H}(q) = \begin{bmatrix} H_1(q) & \dots & H_M(q) \end{bmatrix}^T, \tag{4.3.5}$$

and $\mathbf{f}[k]$ and $\mathbf{V}(q)$ are defined similarly as $\mathbf{y}[k]$ and $\mathbf{H}(q)$, respectively. The microphone signals are then combined by applying a filter-and-sum beamformer $\mathbf{W}(q)$, i.e.,

$$e[k] = \mathbf{W}^T(q)\mathbf{y}[k]. \tag{4.3.6}$$

Although a beamformer is often designed to reduce ambient noise (Doclo et al. 2015), in this paper we will only use the beamformer $\mathbf{W}(q)$ to reduce the feedback component $\mathbf{f}[k]$ in the beamformer output while preserving the incoming signal (cf. Sec. 4.3.2.2). The transfer function from the source to the output of the beamformer is defined as:

$$H_{dev}(q) = \mathbf{W}^T(q)\mathbf{H}(q). \tag{4.3.7}$$

The beamformer output signal is then processed by the forward path of the hearing device $G(q)$, yielding

$$\tilde{u}[k] = G(q)e[k]. \tag{4.3.8}$$

Aiming to achieve acoustic transparency, an equalization filter $A(q)$ is applied to this signal yielding the loudspeaker signal

$$u[k] = A(q)\tilde{u}[k]. \tag{4.3.9}$$

The signal at the aided eardrum, i.e., with the hearing device inserted and processing the signal, is then defined as

$$t_{aid}[k] = D(q)u[k] + \underbrace{H_{leak}(q)s[k]}_{t_{occ}[k]}, \tag{4.3.10}$$

where $D(q)$ denotes the acoustic transfer function between the hearing device loudspeaker and the eardrum, $H_{leak}(q)$ denotes the acoustic transfer function between the source and the occluded eardrum, e.g., due to leakage through the vent, and $t_{occ}[k]$ denotes the signal at the occluded eardrum. The desired signal at the eardrum is defined as

$$t_{des}[k] = G(q)H_{open}(q)s[k], \tag{4.3.11}$$

where $H_{open}(q)$ denotes the acoustic transfer function between the source and the open eardrum. The goal of an equalization algorithm is then to design the filter $A(q)$ such that $t_{aid}[k]$ is as close as possible to $t_{des}[k]$ (cf. Sec. 4.3.2.3).

4.3.2.2 Acoustic feedback suppression algorithm (Schepker et al. 2019a)

In order to suppress the acoustic feedback component in the microphones, we use a time-invariant beamformer that steers a null towards the location of the loudspeaker and aims at preserving the incoming signal for a specific direction (Schepker et al. 2019a). In particular, this null-steering beamformer (NS-BF) aims to achieve the following two conditions simultaneously

$$\mathbf{W}^T(q)\mathbf{V}(q) = 0, \tag{4.3.12}$$

$$\mathbf{W}^T(q)\mathbf{H}(q) = H_{ref}(q). \tag{4.3.13}$$

While the first condition achieves acoustic feedback suppression, the second condition preserves the incoming signal in a reference microphone in the output of the null-steering beamformer. To compute the null-steering beamformer, we will use the robust least-squares-based design procedure proposed in (Schepker et al. 2019a). This procedure requires multiple sets of measurements of the acoustic feedback paths $\mathbf{V}(q)$ as well as a measurement of the acoustic transfer functions $\mathbf{H}(q)$ for a desired incoming direction.

4.3.2.3 Sound pressure equalization algorithm (Denk et al. 2018d)

Aiming at achieving acoustic transparency, in (Denk et al. 2016, 2018c) an iterative algorithm was used. In this paper we will use the least-squares-based (LS) equalization filter design procedure proposed in (Denk et al. 2018d). In order for the signal at the aided eardrum $t_{aid}[k]$ in (4.3.10) to be equal to the desired signal at the open eardrum $t_{des}[k]$ in Eq. (4.3.11), the equalization filter $A(q)$ needs to satisfy, using Eq. (4.3.7) - (4.3.9),

$$\underbrace{A(q)D(q)G(q)H_{dev}(q) + H_{leak}(q)}_{\text{aided transfer function}} = G(q)H_{open}(q). \tag{4.3.14}$$

To compute the equalization filter, measurements or estimates of the transfer functions $H_{open}(q)$ from the source to the open ear, $H_{leak}(q)$ from the source to the occluded ear, $H_{dev}(q)$ from the source to the microphone(s) and $D(q)$ from the loudspeaker to the eardrum are required.

We will exploit different possibilities for the transfer function $H_{dev}(q)$. To achieve perfect equalization for the considered setup in Fig. 4.3.1, it should be chosen as $H_{dev}(q) = \mathbf{W}^T(q)\mathbf{H}(q)$. However, when the beamformer is not known a-priori, $H_{dev}(q)$ could be chosen to be the acoustic transfer function between the sound source and a reference microphone of the hearing device, i.e., $H_{dev}(q) = H_{ref}(q)$.

4.3.2.4 Real-time implementation

Both the fixed null-steering beamformer for feedback suppression as well as the equalization filter for acoustic transparency were implemented on the Master

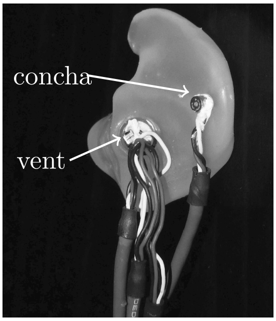


Figure 4.3.2: *Custom earpiece used in the hearing device prototype (Denk et al. 2018c). The microphone in the concha and in the outer side of the vent are indicate. Both loudspeakers are inside the vent, as well as an additional microphone at the inner side of the vent.*

Hearing Aid (MHA, Grimm et al. 2006), which is a software platform for real-time signal processing. The MHA was run on an Intel NUC personal computer using an *RME* Fireface UCX soundcard with a sampling rate of 32 kHz. As earpieces, two custom vented prototypes as described in (Denk et al. 2018c) were used that were inserted in the ear of a dummy head (see Fig. 4.3.2). The custom earpieces consist of two loudspeakers located in the vent (diameter 4.5 mm, effective diameter due to the transducers ≈1.5 mm) and three microphones, one located at the inner side of the vent in the ear canal, one at the outer side of the vent and one in the concha. Although the earpieces has two loudspeakers, in this study we only use the loudspeaker located at the inner side of the vent. The processing delay of this setup was approximately 6.5 ms, which is in the range of tolerable delays for open fittings (Stone et al. 2008). Furthermore, as forward path of the hearing device, a broadband gain $G(q) = 1$ was applied.

In the following we will describe the measurements used for the computation of both the null-steering beamformer $\mathbf{W}(q)$ and the equalization filter $A(q)$. Note that all required acoustic transfer functions could be measured a-priori, e.g., on a dummy head in an anechoic chamber. However, when fitted to human subjects, some of these measurements should be individualized, while others are expected to be less sensitive to individual variations or difficult to measure. Therefore, for some of the acoustic transfer functions we will use estimates, whose influence on the performance will be investigated.

In order to compute the robust null-steering beamformer (cf. Sec. 4.3.2.2), in this study we used two sets of acoustic feedback paths $\mathbf{V}(q)$ per ear of the dummy head, which were measured using sine sweeps (Müller and Massarani 2001). The first set was measured without any objects in the close vicinity of the dummy head and the second set was measured with hands covering the ears. Furthermore, we used a set of acoustic transfer functions $\mathbf{H}(q)$ measured a-priori in an anechoic chamber for a source in front of the dummy head (0°).

Figure 4.3.3 shows the directional responses of the left ear for the null-steering beamformer and for the reference microphone at the outer side of the vent, which the null-steering beamformer aims to preserve. In general, the frontal direction is preserved for all frequencies. For different directions the directional responses are very similar for frequencies up to 2000 Hz, while small differences can be observed for higher frequencies. Nevertheless, these results show that the null-steering beamformer does not alter the directional responses of the reference microphone to a large extent.

In order to design the equalization filter (cf. Sec. 4.3.2.3), in this study we performed measurements of the required acoustic transfer functions $D(q)$, $\mathbf{H}_{leak}(q)$, and $H_{dev}(q)$ using sine sweeps (Müller and Massarani 2001). While for $D(q)$ the sine sweeps were played back from the device, in order to measure $\mathbf{H}_{leak}(q)$ and $H_{dev}(q)$ we used *Sennheiser* HD650 headphones that were put onto the dummy head with the hearing device inserted. Note that a measurement of $D(q)$ in human subjects is difficult and estimation procedures could be used, e.g., based on an electro-acoustic models (Vogl and Blau 2019). However, using electro-acoustic models is beyond the scope of this paper. Furthermore, since the open ear transfer function $H_{open}(q)$ is not available with additional measurement equipment and effort in individual subjects, we used the average diffuse-field equalization function obtained from several subjects from (Denk et al. 2018d) to estimate the open-ear transfer function $H_{open}(q)$ from the acoustic transfer function between the headphones and the concha microphone. Note that the effect of using this estimate will be investigated in the experimental evaluation. For a discussion on the effect of using headphones to measure the required transfer functions, the reader is referred to (Denk et al. 2018c).

Figure 4.3.4 shows the aided transfer function of the left hearing device system with feedback suppression and sound pressure equalization when using headphones. As can be observed, the aided transfer function and the estimated open-ear transfer function match well across the whole frequency range indicating a successful computation of the equalization filter. Nevertheless, comb-filtering effects due to the processing delay are clearly visible that may affect the perceived quality. Note that the comb-filtering effects mainly occur in the frequency range where the leakage component and the output of the hearing device have a similar level.

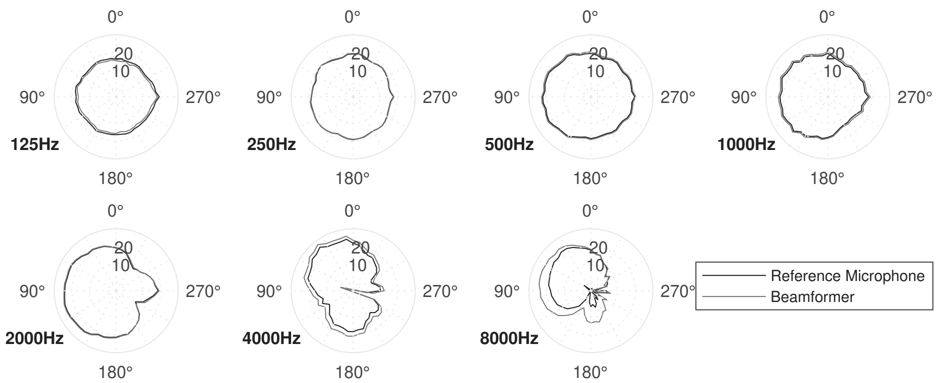


Figure 4.3.3: *Directional responses of the left ear for the null-steering beamformer and for the reference microphone for several frequencies.*

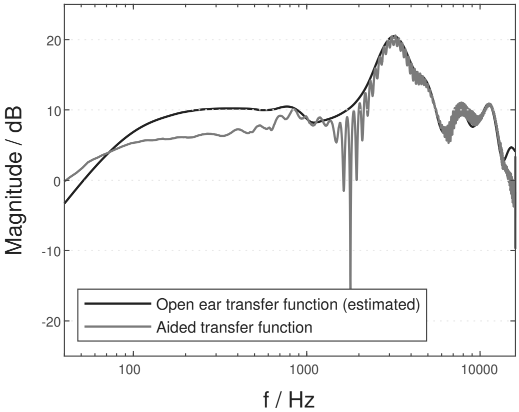


Figure 4.3.4: *Open ear transfer function and aided transfer function of the combined system using a null-steering beamformer and the equalization filter in the real-time prototype using headphones.*

4.3.2.5 Subjective quality evaluation

To evaluate the presented acoustically transparent hearing device system, we conducted a formal listening test with $N = 15$ self-reported normal-hearing subjects (none of the authors participated). The task of the subjects was to rate the overall quality of the processed stimuli compared to the open ear reference in a MUSHRA-like framework using a drag-and-drop interface (Völker et al. 2018). Note that in contrast to (Denk et al. 2016), in the present study the reference was explicitly presented to the subjects. Two subjects had to be excluded from the data analysis since they were not able to reliably identify the hidden reference, resulting in a total of 13 subjects (age 28.2±4.0 years). Stimuli were pre-recorded at a sampling rate of 48 kHz using a *G.R.A.S.* KEMAR 45BB-12 Head & Torso with low-noise ear simulators with the hearing device prototype inserted. The dummy head was placed in a lab with variable acoustics (cf. Fig. 4.3.5), where the reverberation time can be varied using absorber panels mounted at the walls and the ceiling. The recordings were played back to the subjects using MATLAB through an *RME* ADI-2 Pro FS headphone amplifier and *Sennheiser* HD650 headphones, which were equalized for a flat magnitude response at the average eardrum.

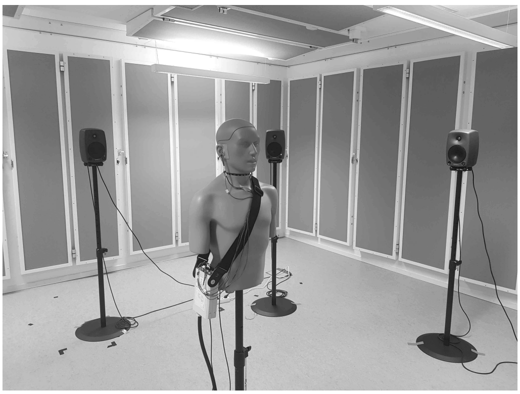


Figure 4.3.5: *Dummy head with inserted hearing device prototypes in a lab with variable acoustics. The green absorbing panels can be flipped to make them highly reflective.*

Table 4.3.1: *Processing conditions used in the experimental evaluation.*

Cond.	Feedback Suppression	Equalization Algorithm	$H_{dev}(q)$	Sound Recording	Processing Delay
A	open ear		n/a	$t_{des}[k]$	none
B	none	DF eq. (Denk et al. 2018b)	$H_{ref}(q)$	simulated	none
C	NS-BF (Schepker et al. 2019a)	DF eq. (Denk et al. 2018b)	$H_{ref}(q)$	$e[k]$	none
D	NS-BF (Schepker et al. 2019a)	Iterative (Denk et al. 2018c)	$\mathbf{W}^T(q)\mathbf{H}(q)$	$t_{aid}[k]$	6.5 ms
E	NS-BF (Schepker et al. 2019a)	LS (Denk et al. 2018d)	$\mathbf{W}^T(q)\mathbf{H}(q)$	$t_{aid}[k]$	6.5 ms
F	NS-BF (Schepker et al. 2019a)	LS (Denk et al. 2018d)	$H_{ref}(q)$	$t_{aid}[k]$	6.5 ms
G	NS-BF (Schepker et al. 2019a)	none	n/a	$t_{aid}[k]$	6.5 ms
H	occluded ear		n/a	$t_{occ}[k]$	none

The goal of the listening test was to assess the impact of different equalization filters and the estimate of the open ear transfer function, the impact of the processing delay of the real-time implementation, as well as potential interactions of the beamformer and the equalization filter. Therefore, the following 8 processing conditions were judged by the subjects (cf. also Tab. 4.3.1), where processing conditions B and C were simulated (without processing delay) and conditions D–G used real-time processing (including a processing delay):

A The open-ear reference condition, i.e., without the hearing device inserted to the ear.

B A fully simulated system that uses the measured acoustic transfer functions from the source to the vent microphone $H_{ref}(q)$ and artificially maps it to the open ear using a diffuse field equalization function presented in (Denk et al. 2018b).

C A partly simulated system that uses the output $e[k]$ of the null-steering beamformer for feedback suppression and artificially maps it to the open ear using a diffuse field equalization function presented in (Denk et al. 2018b).

D Using the null-steering beamformer for feedback suppression algorithm and the iterative equalization filter design presented in (Denk et al. 2018c). Note that the iterative equalization filter design implicitly exploits knowledge about the null-steering beamformer.

E Using the null-steering beamformer for feedback suppression and the LS equalization filter design presented in Sec. 4.3.2.3 computed using $H_{dev}(q) = \mathbf{W}^T(q)\mathbf{H}(q)$.

F Using the null-steering beamformer for feedback suppression presented in Sec. 4.3.2.2 and the LS equalization filter design presented in Sec. 4.3.2.3 computed using $H_{dev}(q) = H_{ref}(q)$.

G Using only the null-steering beamformer for feedback suppression and no equalization filter, i.e., $A(q) = 1$, where the gain of the hearing device was changed compared to conditions D–F to achieve the same broadband level for the aided ear as for the open ear.

H The occluded ear, i.e., with the hearing device prototypes inserted but without processing, providing a low-quality anchor signal.

Using these conditions allows to assess the effect of the diffuse field equalization (Denk et al. 2018b) and possible directional distortions (Denk et al. 2018g, comparing A and B), the effect of sensor noise (comparing B and C), the effect of the processing delay (comparing C and D, E, and F), and the effect of using equalization (comparing D, E, and F and G). Furthermore, potential interactions of the null-steering beamformer and the equalization filter can be assess (comparing E and F).

The stimuli and acoustic conditions used in the subjective evaluation are shown in Tab. 4.3.2. As stimuli we used two speech signals (male and female) taken from (Vincent et al. 2012) and two music signals (an excerpt from a jazz song[1] and an excerpt from a classical piano recording[2]). The stimuli were played back from three different directions (0°, 90°, 225°) using *Genelec* 8030 loudspeaker for three different reverberation times: low ($T_{60} \approx 0.35\,\mathrm{s}$), mid ($T_{60} \approx 0.45\,\mathrm{s}$), and high ($T_{60} \approx 1.4\,\mathrm{s}$). The loudspeaker were placed at distance of $\approx 2\,\mathrm{m}$ from the dummy head and adjusted in height to be at ear level with the dummy head (approximately 1.6 m). Note that the use of our lab with variable acoustics allowed us to change

[1] J. Redman: Timeless tales for changing times, 1. Summertime
[2] K. Jarret: Bach, Wohltemperiertes Klavier, Book 1, prelude no. 3

Table 4.3.2: *Overview on acoustic conditions and signals.*

Reverberation	Signal Direction	Signals
low $T_{60} \approx 0.35\,\mathrm{s}$	0°	piano, jazz female+male speech
	90°	piano, jazz female+male speech
	225°	piano, jazz female+male speech
mid $T_{60} \approx 0.45\,\mathrm{s}$	0°	jazz female speech
	225°	jazz female speech
high $T_{60} \approx 1.4\,\mathrm{s}$	0°	jazz female speech
	225°	jazz female speech

the reverberation time without changing the physical setup of the loudspeakers and the dummy head.

4.3.3 Results

Figure 4.3.6 shows the quality ratings (QRs) of the listening test. Individual panels show the results for the different acoustic conditions of Tab. 4.3.2. First, consider the condition with a frontal source (0°) and low reverberation in the top left panel. As can be observed, most subjects were able to identify the hidden reference (processing condition A) and rated the occluded ear with the lowest scores (processing condition H). Furthermore, all signals (piano, jazz, female speech, male speech) were rated similarly for all processing conditions, where generally the piano signal was rated slightly better. Comparing the different processed signals, the fully simulated processing condition B yields the highest QRs that are similar to the open ear, while artificially mapping the output of the null-steering beamformer to the eardrum (processing condition C) yields only a slightly lower QR. Comparing the three real-time processing conditions that include an equalization filter (D–F), it can be observed that all three processing conditions yield similar quality compared to the open ear, with ratings in the range of medium and good QRs. When using no equalization filter (processing condition G), the quality is rated slightly higher compared to the occluded ear (processing condition H). Statistical analyses of the results were conducted using a three-factor analysis of variance (ANOVA) for each of the signal directions with processing condition, reverberation time, and signal as factors using Huynh-Feldt correction for sphericity violation. For the frontal incoming direction the ANOVA showed a significant effect of all three main factors [Proc. condition: $F(2.8, 34.2) = 151.9$, $p < 0.001$; Reverb: $F(0.8, 9.8) = 286.1$, $p < 0.001$; Signal: $F(1.2, 14.7) = 335.6$, $p < 0.001$] as well as their interactions [Proc. condition×Reverb: $F(5.7, 68.4) = 28.1$, $p < 0.001$; Proc. condition×Signal $F(8.5, 102.6) = 46.2$, $p < 0.001$; Reverb×Signal $F(2.4, 29.3) = 219.5$, $p < 0.001$; Proc. condition×Reverb×Signal: $F(17.1, 205.1) = 114.7$, $p < 0.001$]. For the 90° direction the ANOVA showed a significant effect of both main factors [Proc. condition: $F(4.1, 49.5) = 142.2$, $p < 0.001$; Signal: $F(1.8, 21.2) = 4.9$, $p < 0.01$] as well as their interaction [Proc. condition×Signal $F(12.4, 148.6) = 2.3$, $p < 0.01$]. For the 225° incoming direction the ANOVA showed a significant effect of all three main factors [Proc. condition: $F(2.7, 32.4) = 156.7$, $p < 0.001$; Reverb: $F(0.8, 9.2) = 461.4$, $p < 0.001$; Signal: $F(1.2, 13.9) = 283.5$, $p < 0.001$] as well as their interactions [Proc. condition×Reverb: $F(5.4, 64.7) = 29.9$, $p < 0.001$; Proc. condition×Signal $F(8.1, 97.1) = 51.0$, $p < 0.001$; Reverb×Signal $F(2.3, 27.7) = 178.1$, $p < 0.001$; Proc. condition×Reverb×Signal: $F(16.2, 194.1) = 132.3$, $p < 0.001$]. Post-hoc analysis of the three main factors using Bonferroni correction showed for all three signal directions that QRs in low reverberation were significantly higher than in both mid and high reverberation. Furthermore, QRs for the piano signal were significantly higher than for all other signals for the 0° and 225° directions, while for the 90° direction the quality of the jazz signal was rated significantly lower compared to the female speech signal. For all signal directions, QRs were significantly different for all processing conditions, except when comparing

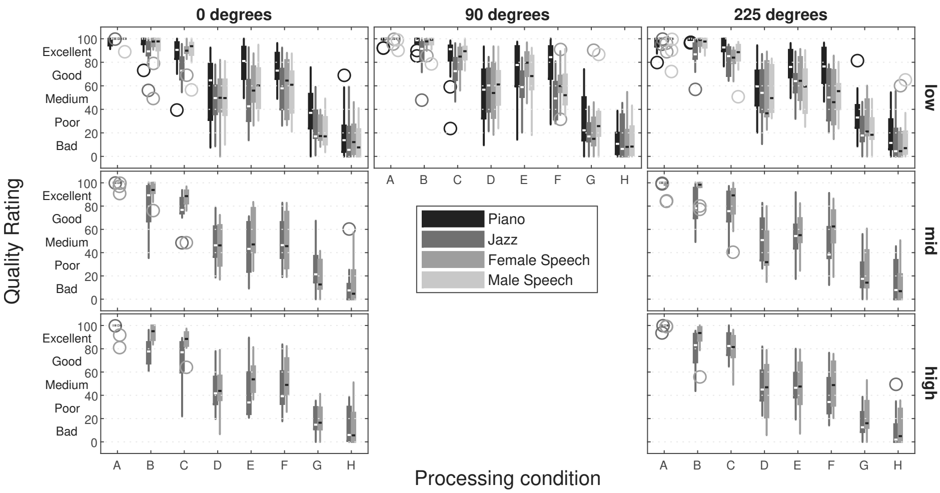


Figure 4.3.6: *Results of the formal listening test for different directions (columns) and reverberation times (rows) for the different processing conditions (cf. Tab. 4.3.1) and signals. Lines show the median, boxes show the interquartile ranges, whiskers indicate the last point included within 1.5 times the interquartile range, and circles show outliers.*

processing conditions D, E, and F as well as comparing the processing conditions G and H. Additionally, for the signal directions of 0° and 225°, QRs of processing conditions B and C were not significantly different.

In order to easier visualize the differences between the processing conditions, Fig. 4.3.7 shows the distribution of the median ratings per subject across all acoustic conditions and signals. Similar trends as in Fig. 4.3.6 are observed, where a small improvement of the equalization algorithm proposed in (Denk et al. 2018d, processing condition E) compared to the equalization algorithm proposed in (Denk et al. 2018c, processing condition D) is observed.

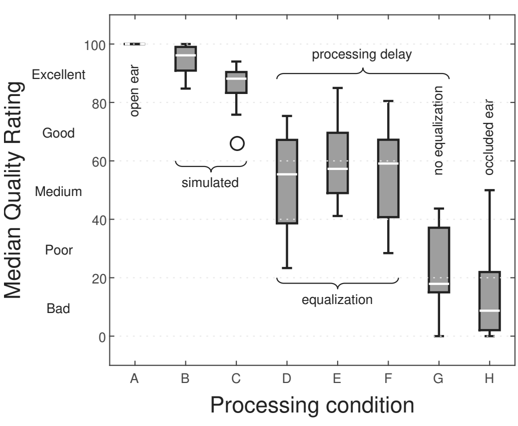


Figure 4.3.7: *Median results across all acoustic conditions and signals.*

4.3.4 Discussion

From the median QRs in Fig. 4.3.7 it can be observed that, on the one hand, all subjects were able to reliably identify the open ear reference (processing condition A). On the other hand, the occluded ear (processing condition H) was rated worst showing the necessity for sound processing. In the following we will first discuss the results for the real-time processing conditions (D–G). We will then present arguments for the observed significant differences between these processing conditions and the open ear reference condition based on the simulated processing conditions (B, C).

As revealed by the statistically similar ratings for processing conditions G and H, only suppressing the feedback using the null-steering beamformer does not yield a significant improvement compared to the occluded ear. This supports the need for additional processing, i.e., equalization, of the played back signal to achieve a sound quality that is comparable to the open ear, i.e. to achieve acoustically transparent sound presentation. When using an additional equalization filter, a significant improvement in sound quality can be achieved for all considered equalization filters (processing conditions D–F). While there is no significant difference between the different equalization filters, the equalization filter design procedure presented in (Denk et al. 2018d, processing conditions E and F) is much faster to compute than the iterative procedure presented in (Denk et al. 2018c, processing condition D). Furthermore, incorporating a-priori knowledge about the null-steering beamformer did not yield an improvement in sound-quality (processing condition E vs F). However, it should be noted that the null-steering beamformer aims at preserving the signal of the reference microphone in its output and thus no large differences were expected.

Even though the proposed hearing device system (processing conditions E–F) achieves a significant improvement in sound quality compared to the occluded ear (processing condition H), the sound quality was still rated lower compared to the open ear (processing condition A). Potential reasons are comb-filtering effects as well as sensor noise. As can be observed by comparing processing conditions A and B, using a precomputed estimate of the open ear transfer function based on the diffuse field equalization (Denk et al. 2018b) is able to achieve almost the same (excellent) quality compared to the open ear. While sensor noise and potentially the feedback cancellation algorithm degrade the quality slightly (processing conditions B vs C), the quality is still perceived as excellent. Comparing processing conditions D–F and G reveals that the processing delay of the real-time system is by far the most important factor that degrades the sound quality. This indicates that when acoustic transparency is desired, the processing delay should be as small as possible to counteract undesired comb-filtering effects. This is especially important in scenarios were the levels of the leakage component and the played back sound are similar, as was the case in the present study where the gain of the hearing device was $G(q) = 1$ (cf. Sec. 4.3.2.4). If the gain of the hearing device is larger, comb-filtering effects are expected to be smaller.

Comparing the different acoustic conditions (reverberation times and signal direction), no large differences can be observed. This indicates that the directional response of the null-steering beamformer (cf. Fig. 4.3.3) does not largely impact the results. Even though a significant effect of reverberation time was found, the small interaction effect between processing condition and reverberation time indicates that evaluating a limited number of (or even a single) reverberation times is presumably sufficient to investigate difference between hearing devices for acoustically transparent sound presentation.

A similar evaluation was conducted in (Denk et al. 2016), where equalization using the iterative procedure described in (Denk et al. 2018c) was combined with additional signal enhancement algorithms. In contrast to the present study, however, in (Denk et al. 2016) the subjects were not provided with an explicit open ear reference and signals were bandlimited to a frequency range of 8 kHz. Results in (Denk et al. 2016) showed that the quality of transparent sound presentation and the open ear canal were considered similar. Thus it is expected that if subjects were not provided with an explicit reference in the present study, smaller differences between the ratings of the open ear and the presented acoustically transparent hearing device could have been achieved.

4.3.5 Summary

In this paper we presented an evaluation of a real-time prototype for acoustically transparent sound presentation. The prototype combines a custom earpiece with multiple integrated microphones with a null-steering beamformer for acoustic feedback suppression and an equalization algorithm taking into account the sound leaking into the ear canal. The results of a formal listening test show that the median QRs of the proposed approach for acoustic transparency are significantly better than the occluded ear (i.e., no processing) and not using sound pressure equalization. Nevertheless, the processing delay of 6.5 ms causing comb-filtering effects, is the main limiting factor for sound quality.

In future work, we aim to investigate the requirements on the processing delay in assistive listening devices with acoustic transparency features as well as using a model-based approach to estimate the sound pressure at the eardrum (Vogl and Blau 2019).

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project ID 352015383 – SFB 1330 A4 and C1, and the Cluster of Excellence 1077 "Hearing4all" - Project ID 390895286 - EXC 2177/1.

5 General discussion and conclusions

In this chapter, the main results of the individual sections are summarized and discussed, and general conclusions are drawn. Furthermore, suggestions for future research on questions that arise from the results of this thesis are given.

Limitations on acoustic transparency: The target definition problem

In Chapter 2, the limitations on achievable acoustic transparency imposed by the hearing device microphone location were assessed. Specifically, Sec. 2.2 dealt with how well a so-called target that approximates the signal for the open-ear case can be estimated from the hearing device microphone signal. Section 2.3 assessed how well directional cues are captured, depending on the hearing device style and the microphone location. Correction functions termed Target Response Correction Function (TRCF) that transform the hearing device microphone signal to the eardrum of the open ear were defined and evaluated based on the direction-dependent Relative Transfer Function (RTF) between the hearing device microphone and the eardrum of the open ear. In Sec. 4.3, the validity of the approximated open-ear signal were psycho-acoustically evaluated.

Generally, the results revealed that the TRCF depends significantly on the device style. This can be largely attributed to a distorted capturing of the direction-independent aspects of the ear acoustics, mostly the destruction of the cavum conchae resonance by filling. Furthermore, the TRCF is subject to large inter-individual deviations of up to 20 dB and more. Although exact determination of signal at the open eardrum generally requires a direction-dependent transformation of the hearing device microphone signal, an accurate and perceptually convincing approximation can be computed using only a single direction-independent TRCF, at least for In The Ear (ITE) devices.

Regarding the reproduction of the open-ear transmission properties, the results from Sec. 2.2 showed a benefit of employing individual correction functions. This result is generally in line with previous results that showed differences in the non-directional properties between individual ears. Nevertheless, the differences between individual and non-individual TRCFs were smaller than those between free- and diffuse-field equalization, and also decreased with increasing distance of the microphone from the eardrum. It should be noted that there is no practically applicable approach in sight to determine the individual open-ear response at the eardrum with sufficient accuracy, other than by highly controlled measurements using probe tubes or other specialized equipment (Hiipakka et al. 2012). An intermediate solution for individualization could be the derivation of a set of typical TRCFs for a given device style, and (self-) selecting one TRCF from this set for the individual user. Future studies should evaluate the perceptual benefit of individualized response targets. The author hypothesizes that the deviation in defining the target over a non-individualized TRCF is the smallest of all issues, and satisfactory transparency can be achieved using average correction functions tailored to the device style.

Regarding the practical implementation of acoustic transparency, the result showed that the TRCF should be defined for the diffuse field to minimize spectral differences to the open-ear response. Specifically, equalization to the diffuse field should generally be preferred over the frontal direction. For the evaluation of spectral cues in Sec. 2.3, the differences between microphone locations were therefore assessed after compensation of diffuse-field differences. It was implicitly assumed that the applied equalization and spectral directional cues are independent, i.e., a frequency response difference would not impact directional cues and sound localization.
This assumption was tested in Sec. 4.1 by comparing the influence of diffuse-field and free-field equalization of the hearing device microphone signal on sound localization, i.e., when listening to the defined target signal. As it turned out, the impact of the correction function on sound localization is surprisingly high, and localization patterns deviate significantly between the free- and diffuse-field equalized Head-related Transfer Function (HRTF) (c.f. Fig. 4.1.8). Due to free-field equalization, directional cues that are characteristic for frontal incidence are imprinted on the general frequency response of the device. This resulted in sound sources being "drawn" towards the front and horizontal plane. The size of this effect increases with increasing distance of the hearing device microphone from the ear canal entrance, i.e., with increasing destruction of pinna cues in the hearing device signal. Also, the general performance in median-plane localization decreased when free-field equalization instead of diffuse-field equalization was employed.
In summary, for either accuracy in timbre conservation and spatial hearing, equalizing the hearing device to the diffuse-field response of the ear was found to be the best option in all experiments conducted in this thesis. However, the considerations were restricted to exploiting only one microphone. When several microphones are available, the directional features of the open-ear HRTF that are missing at the hearing device microphone location could be replicated by the directionality of a beamformer (Kuk et al. 2013). This can be understood as a direction-dependent equalization.

The results also demonstrate possible improvements for hearing aid fitting using probe tube measurements (Mueller 2001). First, the individuality of all occurring transfer functions seen in the present data show that verification of real-ear performance using probe tube measurements is generally in place. This should motivate audiologists to include these measurements in their regular fitting process (Aazh and Moore 2007; Mueller and Picou 2010). Second, the variability of the open-ear transfer function in general shows that hearing aids should be fitted to a prescribed real-ear insertion gain rather than a predefined real-ear aided response (Dillon 2012). Third, the results of this thesis show a benefit of conducting real-ear measurements in a diffuse sound field, or at least not for frontal sound incidence. Diffuse-field measurements, even if crudely approximated, would certainly also result in an improved reliability of the real-ear measurements, since the impact of movements would be reduced (c.f. Killion and Revit 1987).

Limitations on acoustic transparency: Equalization and processing delay

Even when neglecting the principal issues discussed before, and it is assumed that a perfect target signal is known, equalizing the hearing device to produce a target response is another limiting issue. Within this thesis, a novel method to compute an equalization filter was proposed in Sec. 3.3. The performance of this approach was evaluated in Sec. 4.3 and compared to a previous approach from (Denk et al. 2018c).

As it turned out, if there is a processing delay even of only several milliseconds, the equalization problem is confounded with comb-filtering artifacts. If these are not an issue, e.g., when the leakage component can be reduced or the hearing device delay is in the order of tens of microseconds, the equalization problem can be regarded as solved with the present approaches (Denk et al. 2018d; Fabry et al. 2019; Hoffmann et al. 2013a; Schepker et al. 2018). Specifically, if the leakage component is negligible, the response of the device's driver at the eardrum should equal the appropriate TRCF. In practice, a delay might still be imposed by additional sound processing algorithms that are required for a certain application, or simply by limitations of the signal processing platform. Therefore, possible future approaches to reduce comb-filter artifacts in spite of a delay are briefly discussed.
The most salient consequence of delays in the range below the echo threshold (i.e., up to ca. 15 ms) are spectral ripples due to alternating constructive and destructive interferences across frequency. Arguably, it is very difficult and not robust to adapt the phase of the delayed hearing device output to the phase of the leakage. The approach proposed in this thesis is to avoid (by appropriate equalization) hearing device output in frequency regions where leakage and hearing device output would be similar in level to achieve the response target (Denk et al. 2018d). This markedly reduced the spectral ripple, however, this came at the cost of having no control over sound reaching the eardrum in these typically low-frequency regions. The approach will therefore have negative effects on noise reduction algorithms, which operate mostly in this frequency region. It will be worthwhile to explore further possibilities to reduce comb-filtering effects without reducing the control of the sound at the eardrum in non-occluding fit devices. One independent possibility is to reduce the leakage component using active noise cancellation without reducing the wearing comfort by completely occluding the ear canal. Further, some effects of delay may not always be noticed or perceived as disturbing by the user. Basic psychoacoustic research is required to better understand the perceptual principles in perceiving these distortions, to guide the way to algorithms that equalize for comb-filtering effects on the perceptual level.

For the individualized equalization filter design as used throughout this thesis, it was assumed that all relevant transfer functions to the eardrum are known, which is not the case in practice. Nevertheless, different methods to estimate these transfer functions can be employed. For the open-ear HRTF to the eardrum, the TRCF concept (c.f. Sec 2.2) delivers reasonable results. To obtain the transfer functions to the eardrum in the aided case (both for external sound sources and the hearing device drivers), an in-ear microphone together with electro-acoustic models can be employed, as outlined in Sec. 3.1 and in (Sankowsky-Rothe et al. 2015a; Vogl and Blau 2019). It has been demonstrated that these transfer functions are subject to large inter-individual differences. Therefore, individualized model predictions are expected to improve the performance significantly, as compared to utilizing average data. A requirement for these approaches to work is that the RTF between an in-ear microphone and the eardrum is not dependent on the sound source. How well this is fulfilled in the newly developed earpiece (c.f. Sec. 3.2), and general optimizations regarding this issue, are under current investigation (Roden et al. 2019a,b). Finally, future studies should evaluate the perceptual benefit of individualization of the equalization filter design, compared to a predefined generic filter. This includes the evaluation of the accuracy and robustness of electro-acoustic models that have to be employed in practice.

Disturbances on spatial hearing

The results of Secs. 2.3 and 4.1 showed that hearing device microphones often capture largely distorted spectral directional cues, which translates into a decreased ability to localize sounds. It should be noted here again that this distortion is largely a destruction and no alteration of the directional information (c.f. 2.3). Therefore, adaptation to the modified HRTFs, as demonstrated with altered ear shapes, is unlikely to completely compensate for this effect (Hofman et al. 1998; Mendonça 2014).
However, some of the perceptual impacts of the distorted HRTF could be compensated by the availability of head movement cues, which were not available in the localization experiment from Sec. 4.1. It is possible that the localization performance with hearing devices is not as heavily affected for longer stimuli as for short stimuli, which were included in the experiment (Blauert 1997).

For an ITE device, as utilized in the developed real-time demonstrator, a disturbance of spatial cues due to the microphone location was clearly demonstrated by the localization experiments in Sec. 4.1. Contrarily, the microphone location of the real-time demonstrator itself did not result in a considerable loss of sound quality in the subjective listening quality evaluation (c.f. Fig. 4.3.7, condition A vs. B). However, the subjective listening quality assessment was based on recordings made in a KEMAR mannequin. The results in Sec. 4.1 showed that the localization performance between an optimal and an ITE microphone location were not different for KEMAR-based data, contrary to individual data. Therefore, these two results are in no contradiction, and the spatial distortions may still be noticeable in experiments with real-time devices or analogous stimuli generation using individualized transfer functions.

For understanding the impact of the microphone location and other spatial distortions in hearing devices, especially on overall sound quality, further experiments should be conducted. The interplay between spatial perception (as assessed explicitly in Sec. 4.1) and perceived sound quality has to be understood more deeply. Future experiments could be based on real-time hearing devices with different geometries, as it has been done previously in (Cubick et al. 2018; Van den Bogaert et al. 2011). An alternative possibility is to simulate listening through hearing devices using individual binaural synthesis, as in Sec. 4.1 and (Lundbeck et al. 2018; Mueller et al. 2012), with the extension to include (or not) the effects of head movement by dynamic synthesis (Begault et al. 2001; Brungart et al. 2004; Grimm et al. 2019; Wallach 1940). This approach would make the presented stimuli more controllable and manipulable without the restrictions of real-time devices.

So, what's the optimum hearing device style?

The results of this thesis verified once more that the full spatial information is only captured in a hearing device when the ear is not obstructed and the microphone is placed in the ear canal (Durin et al. 2014; Hammershøi and Møller 1996; Hoffmann et al. 2013b). Given that spatial cues are distorted in all other device styles, it seems straightforward to conclude that the best device style for realizing acoustically transparent hearing devices is generally an In The Canal (ITC) device. Given the size constraints for ITC devices and the restrictions regarding batteries and acoustic coupling, 'pseudo-ITC' devices could be made up of a Behind The Ear (BTE) unit and a wired external microphone and driver positioned in the ear canal.

For sole hear-through applications and electronic hearing protectors, this statement holds, and such devices should be designed to obstruct the ear as little as possible. However, a device in the ear canal is usually restricted to one microphone. Directional microphones for noise reduction are a standard feature whenever any kind of hearing support is the aim, which usually require at least two microphones per side. The results of this thesis demonstrated a consistent benefit of ITE over BTE devices with respect to conservation of external ear cues. Compared to ITC devices, ITE microphones are subject to a small principal impairment of achievable transparency, which on the other hand is significantly smaller than in BTE devices and mostly an issue of directional cues rather than achievable timbre accuracy. The results even showed one practical benefit of ITE over ITC devices, namely that the appropriate TRCFs are subject to smaller inter-individual variations.

If the application requires more than one microphone, an ITE device therefore seems to be the best choice — if it sits shallow in the cavum conchae and fills if uniformly, like the ITEind device from the present data (c.f. Fig. 2.2.2). If it sticks considerably out of the ear, the acoustic restrictions are similar to a BTE device. It should be noted again that no differences between microphone locations in one ITE device were evident. A different approach would be to extend a BTE device by an in-ear microphone, which may be integrated into the earmould or a receiver-in-canal unit (Gomez et al. 2016; Jensen et al. 2013). In such a setting, the combined conservation of directional cues of the in-ear microphone with beamforming including BTE microphones are interesting and challenging signal processing problems.

Practical outcomes of thesis

Part of the work from Chapter 2 established and extended state-of-the-art methods on ear-related acoustic measurement technology. That is, measurements of the HRTF with and without relation to hearing devices are now routinely performed in the Virtual Reality Laboratory of the University of Oldenburg. Extensions of state-of-the-art HRTF measurement techniques (Enzner et al. 2013) in the greater scope of this thesis was to improve low-frequency accuracy by the Frequency Dependent Truncation (FDT) technique (c.f. Sec. 2.1, Denk et al. 2018f), the development of an earplug for measurements at the ear canal entrance (in collaboration with AK group, TU Berlin, Denk et al. 2019c), as well as the development of a visual feedback system to stabilize the head position (Denk et al. 2017a).

Some of the methods and data of this thesis have been made available to the public, with the aim to improve reproducibility in research as demanded by the Open Science movement. This includes the database of hearing-device related HRTFs[1] (including the TRCFs), a reference implementation of the FDT method,[2] and the design files of the the earplug used for measurements at the ear canal entrance.[3] Furthermore, the multi-microphone earpiece that was developed for the own real-time demonstrator has been made publicly available as a commercial product.[4] It should be stated here that the motivation behind the commercialization was by no means profits, but accelerating research by providing otherwise unavailable but necessary experimental hardware. Compatibility of the earpiece with a recent mobile hearing device signal processing platform (Pavlovic et al. 2018) has been established in cooperation with its developers. The availability of these components in combination constitutes a highly versatile tool, which will enable many researchers to accelerate their progress using a near-to-ideal hearing device platform for both the lab and real life.

Final remarks

In the near future, hearing devices that are not officially hearing aids, will likely provide a variety of hearing support features that are traditionally reserved for medical products (Sabin 2018). Thus, a clear separation of devices for hearing-impaired and normal-hearing users will no longer exist. Such a device could be used as a multimedia device (e.g., to present augmented reality sound objects), but scaled up to a full hearing aid whenever necessary, simply by adjusting the signal processing setting (Kollmeier et al. 2014; Kollmeier and Kiessling 2018). It might even be possible to achieve 'superhuman' hearing by transferring techniques like directional microphones or amplification of otherwise inaudible sounds to the needs of normal-hearing people. Acoustically transparent reproduction of the acoustic environment is a key requirement for this class of devices to provide convincing functionality and to gain user acceptance. Arguably, the advent of hearing devices

[1] https://medi.uni-oldenburg.de/hearingdevicehrtfs, see Sec. 2.2
[2] https://github.com/floriandenk/FDT, see Sec. 2.1
[3] https://zenodo.org/record/2574395, see (Denk et al. 2019c)
[4] https://www.hoertech.de/de/f-e-produkte/transparent-earpiece.html, see Sec. 3.2

targeted at normal hearing people might also contribute to reduce the persisting stigma about hearing aids as a medical product for elderly people with a severe sensory impairment (David and Werner 2016). For hearing aid users, it would be optimal if they were perceived like glasses, which are even worn as a sole fashion accessory. If ear-worn devices became more common in (near-to) normal-hearing users and socially accepted, it would not be obvious to others whether a person is wearing a hearing aid or pair of headphones - especially if such devices are often capable of both functionalities (Dillon 2012, Sec. 3.11). Only if such hearing devices get accepted by the user without hard restrictions imposed by cosmetics that limit the acoustic performance unavoidably, better hearing support for everyone can be achieved in the long run. Contributing to this ambitious goal was the main motivation behind the work in this thesis, and the author hopes that the presented results are useful for other researchers and developers in the field.

Bibliography

Aazh, H. and B. C. Moore (2007). “The Value of Routine Real Ear Measurement of the Gain of Digital Hearing Aids”. *Journal of the American Academy of Audiology* 18(8), pp. 653–664.

Akeroyd, M. A. and W. M. Whitmer (2016). “Spatial Hearing and Hearing Aids”. *Hearing Aids*. Ed. by G. R. Popelka, B. C. J. Moore, R. R. Fay, and A. N. Popper. Cham: Springer International Publishing, pp. 181–215.

Albrecht, R., J. Jaatinen, and T. Lokki (2017). “Electronic hearing protection for musicians”. *Proc. 14th Sound and Music Computing Conference*. Espoo, Finland, pp. 306–313.

Algazi, V. R., C. Avendano, and D. Thompson (1999). “Dependence of subject and measurement position in binaural signal acquisition”. *Journal of the Audio Engineering Society* 47(11), pp. 937–947.

Algazi, V., R. Duda, D. Thompson, and C. Avendano (2001). “The CIPIC HRTF database”. *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*. New Paltz, USA, pp. 99–102.

Allen, J. (1977). “Short-term spectral analysis, and modification by discrete Fourier transform”. *IEEE Transactions on Acoustics Speech and Signal Processing* 25(3), pp. 235–238.

Anovum (2018). *EuroTrak Germany 2018*. Tech. rep. EHIMA - European Hearing Instrument Manufacturers Association.

Barry, R. J., K. I. Cocker, J. W. Anderson, E. Gordon, and C. Rennie (1992). “Does the N100 evoked potential really habituate? Evidence from a paradigm appropriate to a clinical setting”. *International Journal of Psychophysiology* 13(1), pp. 9–16.

Baumgartner, R., P. Majdak, and B. Laback (2013). “Assessment of Sagittal-Plane Sound Localization Performance in Spatial-Audio Applications”. *The Technology of Binaural Listening*. Ed. by J. Blauert. Modern Acoustics and Signal Processing. Springer Berlin Heidelberg, pp. 93–119.

Baumgartner, R., P. Majdak, and B. Laback (2014). “Modeling sound-source localization in sagittal planes for human listeners”. *The Journal of the Acoustical Society of America* 136(2), pp. 791–802.

Baumgartner, R., D. K. Reed, B. Tóth, V. Best, P. Majdak, H. S. Colburn, and B. Shinn-Cunningham (2017). “Asymmetries in behavioral and neural responses to spectral cues demonstrate the generality of auditory looming bias”. *Proceedings of the National Academy of Sciences* 114(36), pp. 9743–9748.

Begault, D., E. M. Wenzel, and M. R. Anderson (2001). “Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source”. *Journal of the Audio Engineering Society* 49(10), pp. 904–916.

Benjamin, E. (2004). “Extending quasi-anechoic measurements to low frequencies”. *Proc. Audio Engineering Society Convention 117*. Convention Paper 6218. San Francisco, USA, pp. 1–16.

Bentler, R. A. and C. V. Pavlovic (1989). “Transfer Functions and Correction Factors Used in Hearing Aid Evaluation and Research:” *Ear and Hearing* 10(1), pp. 58–63.

— (1992). “Addendum to ’Transfer Functions and Correction Factors Used in Hearing Aid Evaluation and Research’”. *Ear and Hearing* 13(4), pp. 284–286.

Best, V., S. Kalluri, S. McLachlan, S. Valentine, B. Edwards, and S. Carlile (2010). “A comparison of CIC and BTE hearing aids for three-dimensional localization of speech”. *International Journal of Audiology* 49(10), pp. 723–732.

Bisgaard, N. and S. Ruf (2018). “Findings From EuroTrak Surveys From 2009 to 2015: Hearing Loss Prevalence, Hearing Aid Adoption, and Benefits of Hearing Aid Use”. *American Journal of Audiology* 26(3S), pp. 451–461.

Blau, M., T. Sankowsky, A. Stirnemann, H. Oberdanner, and N. Schmitt (2008). “Acoustics of open fittings”. *The Journal of the Acoustical Society of America* 123(5), pp. 3011–3011.

Blau, M., T. Sankowsky, P. Roeske, H. Mojallal, M. Teschner, and C. Thiele (2010). “Prediction of the Sound Pressure at the Ear Drum in Occluded Human Cadaver Ears”. *Acta Acustica united with Acustica* 96(3), pp. 554–566.

Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge: MIT Press.

Bleichner, M. G., M. Lundbeck, M. Selisky, F. Minow, M. Jäger, R. Emkes, S. Debener, and M. De Vos (2015). “Exploring miniaturized EEG electrodes for brain-computer interfaces. An EEG you do not see?” *Physiological Reports* 3(4), e12362.

Bleichner, M. G. and S. Debener (2017). “Concealed, Unobtrusive Ear-Centered EEG Acquisition: cEEGrids for Transparent EEG”. *Frontiers in Human Neuroscience* 11(163), pp. 1–14.

Bleichner, M. G., B. Mirkovic, and S. Debener (2016). “Identifying auditory attention with ear-EEG: cEEGrid versus high-density cap-EEG comparison”. *Journal of Neural Engineering* 13(6), p. 066004.

Brandenburg, K., E. Cano, F. Klein, T. Köllmer, H. Lukashevich, A. Neidhardt, J. Nowak, U. Sloma, and S. Werner (2018). “Personalized Auditory Reality”. *Fortschirtte der Akustik - DAGA*. München, pp. 1430–1432.

Breebart, J. and A. Kohlrausch (2001). “Perceptual (ir)relevance of HRTF magnitude and phase spectra”. *Proc. Audio Engineering Society Convention 110*. Convention Paper 5406. Amsterdam, Netherlands, pp. 1–9.

Brinkmann, F. and S. Weinzierl (2017). “AKtools—An Open Software Toolbox for Signal Acquisition, Processing, and Inspection in Acoustics”. *Proc. Audio Engineering Society Convention 142*. Convention e-Brief 309. Berlin, Germany, pp. 1–6.

Brown, A. D., G. C. Stecker, and D. J. Tollin (2015). “The Precedence Effect in Sound Localization”. *JARO: Journal of the Association for Research in Otolaryngology* 16(1), pp. 1–28.

Brungart, D., A. Kordik, C. Eades, and B. Simpson (2003). “The effect of microphone placement on localization accuracy with electronic pass-through earplugs”. *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on.* New Paltz, USA, pp. 149–152.

Brungart, D. S. and W. M. Rabinowitz (1999). “Auditory localization of nearby sources. Head-related transfer functions”. *The Journal of the Acoustical Society of America* 106(3), pp. 1465–1479.

Brungart, D. S., B. W. Hobbs, and J. T. Hamil (2007). “A comparsion of acoustic and psychoacoustic measurements of pass-through hearing protection devices”. *Applications of Signal Processing to Audio and Acoustics, 2007 IEEE Workshop on.* New Paltz, USA, pp. 70–73.

Brungart, D. S., B. D. Simpson, R. L. McKinley, A. J. Kordik, R. C. Dallman, and D. A. Ovenshire (2004). “The Interaction between Head-tracker Latency, Source Duration, and Response Time in the Localization of Virtual Sound Sources”. *Proc. 10th International Conference on Auditory Display (ICAD).* Sydney, Australia, pp. 1–7.

Buhl, M., **F. Denk**, M. Bodenstein, N. Wiedemann, S. Jacobsen, M. R. Schädler, and B. Kollmeier (2019). “Calibration of a mobile hearing aid prototype and its validation: Towards transparent listening with commodity hardware”. *Fortschirtte der Akustik - DAGA.* Rostock, Germany, pp. 927–930.

Bujacz, M. and P. Strumiłło (2016). “Sonification: Review of Auditory Display Solutions in Electronic Travel Aids for the Blind”. *Archives of Acoustics* 41(3), pp. 401–414.

Butler, R. A. and K. Belendiuk (1977). “Spectral cues utilized in the localization of sound in the median sagittal plane”. *The Journal of the Acoustical Society of America* 61(5), pp. 1264–1269.

Byrne, D. and W. Noble (1998). “Optimizing Sound Localization with Hearing Aids”. *Trends in Amplification* 3(2), pp. 51–73.

Carlile, S. (2014). “The plastic ear and perceptual relearning in auditory spatial perception”. *Frontiers in Neuroscience* 8(237), pp. 1–13.

Carlile, S., P. Leong, and S. Hyams (1997). “The nature and distribution of errors in sound localization by human listeners”. *Hearing Research* 114(1–2), pp. 179–196.

Christensen, F., C. B. Jensen, and H. Møller (2000). “The Design of VALDEMAR - An Artificial Head for Binaural Recording Purposes”. *Proc. Audio Engineering Society Convention 109.* Los Angeles, USA, pp. 1–8.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* 2nd ed. Hillsdale, NJ: L. Erlbaum Associates.

Cubick, J. and T. Dau (2016). “Validation of a Virtual Sound Environment System for Testing Hearing Aids”. *Acta Acustica united with Acustica* 102(3), pp. 547–557.

Cubick, J., J. M. Buchholz, V. Best, M. Lavandier, and T. Dau (2018). "Listening through hearing aids affects spatial perception and speech intelligibility in normal-hearing listeners". *The Journal of the Acoustical Society of America* 144(5), pp. 2896–2905.

D'Angelo, W. R., R. S. Bolia, P. J. Mishler, and L. J. Morris (2001). "Effects of CIC Hearing Aids on Auditory Localization by Listeners With Normal Hearing". *Journal of Speech, Language, and Hearing Research* 44(6), pp. 1209–1214.

Dalga, D. and S. Doclo (2011). "Combined feedforward-feedback noise reduction schemes for open-fitting hearing aids". *Proc. Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop on.* IEEE, pp. 185–188.

David, D. and P. Werner (2016). "Stigma regarding hearing loss and hearing aids: A scoping review." *Stigma and Health* 1(2), pp. 59–71.

Debener, S., R. Emkes, M. De Vos, and M. Bleichner (2015). "Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear". *Scientific Reports* 5, p. 16743.

Delorme, A. and S. Makeig (2004). "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis". *Journal of Neuroscience Methods* 134(1), pp. 9–21.

Deng, Y., I. Choi, B. G. Shinn-Cunningham, and R. Baumgartner (2019). "Impoverished auditory cues fail to engage brain networks controlling spatial selective attention". *bioRxiv* 533117.

Denk, **F.** (2018a). *The Oldenburg Hearing Device (OlHeaD) HRTF Database.* URL: medi.uni-oldenburg.de/ hearingdevicehrtfs.

Denk, **F.**, B. Kollmeier, and S. Ernst (2016). "High-Fidelity Hearing Instruments: Evaluating Listening Quality of a New Prototype Using a Method for Evaluating Modified Listening (MEML)". *Proc. AES Conference on Headphone Technology.* Aalborg, Denmark, pp. 206–213.

Denk, **F.**, J. Heeren, S. D. Ewert, B. Kollmeier, and S. M. A. Ernst (2017a). "Controlling the Head Position during individual HRTF Measurements and its Effect on Accuracy". *Fortschritte der Akustik - DAGA*. Kiel, Germany, pp. 1085–1088.

Denk, **F.**, L. Haverkamp, B. Kollmeier, and S. M. Ernst (2017b). "Evaluation of a High-Fidelity Hearing Device Prototype using Real-Ear Measurement". *Presented at DGA Deutsche Gesellschaft für Audiologie Annual Conference.* Aalen, Germany.

Denk, **F.**, M. Grzybowski, S. M. A. Ernst, B. Kollmeier, S. Debener, and M. G. Bleichner (2017c). "Measuring hearing instrument sound modification using integrated ear-EEG". *Proceedings of the International Symposium on Auditory and Audiological Research ISAAR, vol 6.* Nyborg, Denmark, pp. 351–358.

Denk, **F.**, S. Vogl, H. Schepker, B. Kollmeier, M. Blau, and S. Doclo (2017d). "The Acoustically Transparent Hearing Device: Towards Integration of Individualized Sound Equalization, Electro-Acoustic Modeling and Feedback Cancellation".

Proc. 1st International Workshop on Challenges in Hearing Assistive Technology (CHAT-2017). Stockholm, Sweden, pp. 89–94.

Denk, **F.**, S. M. A. Ernst, S. D. Ewert, and B. Kollmeier (2018b). "Adapting Hearing Devices to the Individual Ear Acoustics: Database and Target Response Correction Functions for Various Device Styles". *Trends in Hearing* 22, p. 2331216518779313.

Denk, **F.**, M. Hiipakka, B. Kollmeier, and S. M. A. Ernst (2018c). "An individualised acoustically transparent earpiece for hearing devices". *International Journal of Audiology* 57(S3), pp. 62–70.

Denk, **F.**, H. Schepker, S. Doclo, and B. Kollmeier (2018d). "Equalization filter design for achieving acoustic transparency in a semi-open fit hearing device". *Proc. 13. ITG Conference on Speech Communication*. Oldenburg, Germany, pp. 226–230.

Denk, **F.**, M. Grzybowski, S. M. A. Ernst, B. Kollmeier, S. Debener, and M. G. Bleichner (2018e). "Event-Related Potentials Measured From In and Around the Ear Electrodes Integrated in a Live Hearing Device for Monitoring Sound Perception". *Trends in Hearing* 22, p. 233121651878821.

Denk, **F.**, B. Kollmeier, and S. Ewert (2018f). "Removing reflections in semianechoic impulse responses by frequency-dependent truncation". *Journal of the Audio Engineering Society* 66(3), pp. 146–153.

Denk, **F.**, S. D. Ewert, and B. Kollmeier (2018g). "Spectral directional cues captured by hearing device microphones in individual human ears". *The Journal of the Acoustical Society of America* 144(4), pp. 2072–2087.

Denk, **F.**, M. Lettau, H. Schepker, S. Doclo, R. Roden, M. Blau, J. Bach, J. Wellmann, and B. Kollmeier (2019a). "A one-size-fits-all earpiece with multiple microphones and drivers for hearing device research". *Proc. AES Conference on Headphone Technology*. Paper 13. San Francisco, USA, pp. 1–9.

Denk, **F.**, S. Ewert, and B. Kollmeier (2019b). "On the limitations of sound localization with hearing devices". *The Journal of the Acoustical Society of America* 146(3), pp. 1732–1744.

Denk, **F.**, F. Brinkmann, A. Stirnemann, and B. Kollmeier (2019c). "The PIRATE: an anthropometric earPlug with exchangeable microphones for Individual Reliable Acquisition of Transfer functions at the Ear canal entrance". *Fortschirtte der Akustik - DAGA*. Rostock, Germany, pp. 635–638.

Dillon, H. (2012). *Hearing Aids*. 2nd ed. Turramurra: Boomerang Press.

Doclo, S., W. Kellermann, S. Makino, and S. E. Nordholm (2015). "Multichannel Signal Enhancement Algorithms for Assisted Listening Devices: Exploiting spatial diversity using multiple microphones". *IEEE Signal Processing Magazine* 32(2), pp. 18–30.

Durin, V., S. Carlile, P. Guillon, V. Best, and S. Kalluri (2014). "Acoustic analysis of the directional information captured by five different hearing aid styles". *The Journal of the Acoustical Society of America* 136(2), pp. 818–828.

Enzner, G., C. Antweiler, and S. Spors (2013). “Trends in Acquisition of Individual Head-Related Transfer Functions”. *The Technology of Binaural Listening*. Ed. by J. Blauert. Modern Acoustics and Signal Processing. Springer Berlin Heidelberg, pp. 57–92.

Ernst, S. M., B. Kollmeier, M. Hiipakka, and **F. Denk** (2015). “Method for operating an electroacoustic system and electroacoustic system”. US20180084328A1-1. Patent application.

Fabry, J., F. König, S. Liebich, and P. Jax (2019). “Acoustic Equalization for Headphones Using a Fixed Feed-forward Filter”. *Proc. 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Brighton, UK, pp. 980–984.

Fastl, H. and E. Zwicker (2007). *Psychoacoustics: facts and models*. 3rd. ed. Springer series in information sciences 22. Berlin ; New York: Springer.

Fincham, L. R. (1985). “Refinements in the impulse testing of loudspeakers”. *Journal of the Audio Engineering Society* 33(3), pp. 133–140.

Finke, M., M. Billinger, and A. Büchner (2017). “Toward Automated Cochlear Implant Fitting Procedures Based on Event-Related Potentials”. *Ear and hearing* 38(2), e118–e127.

Gardner, M. B. and R. S. Gardner (1973). “Problem of localization in the median plane: effect of pinnae cavity occlusion”. *The Journal of the Acoustical Society of America* 53(2), pp. 400–408.

Gatehouse, S. (1989). “Limitations on insertion gains with vented earmoulds imposed by oscillatory feedback”. *British Journal of Audiology* 23(2), pp. 133–136.

Genuit, K. (1984). “Ein Modell zur Beschreibung von Aussenohrübertragungseigenschaften”. PhD Thesis. RWTH Aachen.

Glasberg, B. R. and B. C. Moore (1990). “Derivation of auditory filter shapes from notched-noise data”. *Hearing Research* 47(1-2), pp. 103–138.

Gomez, G., V. M. Hoening, and B. U. Seeber (2016). “Spatial Sound Perception with Hearing Aids – Examining Localization, Distance, Width, Elevation and Internalization”. *Presented at the Internatinal Hearing Aid Research Conference (IHCON)*. Tahoe City, CA, USA.

Goverdovsky, V., D. Looney, P. Kidmose, and D. P. Mandic (2016). “In-Ear EEG From Viscoelastic Generic Earpieces: Robust and Unobtrusive 24/7 Monitoring”. *IEEE Sensors Journal* 16(1), pp. 271–277.

Grimm, G., T. Herzke, D. Berg, and V. Hohmann (2006). “The Master Hearing Aid: A PC-Based Platform for Algorithm Development and Evaluation”. *Acta Acustica united with Acustica* 92(4), pp. 618–628.

Grimm, G., B. Kollmeier, and V. Hohmann (2016). “Spatial Acoustic Scenarios in Multichannel Loudspeaker Systems for Hearing Aid Evaluation”. *Journal of the American Academy of Audiology* 27(7), pp. 557–566.

Grimm, G., J. Luberadzka, and V. Hohmann (2019). “A Toolbox for Rendering Virtual Acoustic Environments in the Context of Audiology”. *Acta Acustica united with Acustica* 105(3), pp. 566–578.

Groth, J. and M. B. Søndergaard (2004). “Disturbance caused by varying propagation delay in non-occluding hearing aid fittings”. *International Journal of Audiology* 43(10), pp. 594–599.

Hammershøi, D. and H. Møller (1996). “Sound transmission to and within the human ear canal”. *The Journal of the Acoustical Society of America* 100(1), pp. 408–427.

Harder, S., R. R. Paulsen, M. Larsen, S. Laugesen, M. Mihocic, and P. Majdak (2015). “Reliability in Measuring Head Related Transfer Functions of Hearing Aids”. *Acta acustica united with acustica* 101(5), pp. 1064–1066.

Härmä, A., J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, J. Hiipakka, and G. Lorho (2004). “Augmented reality audio for mobile and wearable appliances”. *Journal of the Audio Engineering Society* 52(6), pp. 618–639.

Hassager, H. G., A. Wiinberg, and T. Dau (2017). “Effects of hearing-aid dynamic range compression on spatial perception in a reverberant environment”. *The Journal of the Acoustical Society of America* 141(4), pp. 2556–2568.

Hatziantoniou, P. D. and J. N. Mourjopoulos (2000). “Generalized Fractional-Octave Smoothing of Audio and Acoustic Responses”. *Journal of the Audio Engineering Society* 48(4), pp. 259–280.

Hedges, L. V. (1981). “Distribution Theory for Glass’s Estimator of Effect Size and Related Estimators”. *Journal of Educational Statistics* 6(2), pp. 107–128.

Hellstrom, P.-A. and A. Axelsson (1993). “Miniature microphone probe tube measurements in the external auditory canal”. *The Journal of the Acoustical Society of America* 93(2), pp. 907–919.

Hiipakka, M., M. Tikander, and M. Karjalainen (2010). “Modeling the external ear acoustics for insert headphone usage”. *Journal of the Audio Engineering Society* 58(4), pp. 269–281.

Hiipakka, M., T. Kinnari, and V. Pulkki (2012). “Estimating head-related transfer functions of human subjects from pressure–velocity measurements”. *The Journal of the Acoustical Society of America* 131(5), pp. 4051–4061.

Hirahara, T., H. Sagara, I. Toshima, and M. Otani (2010). “Head movement during head-related transfer function measurements”. *Acoustical Science and Technology* 31(2), pp. 165–171.

Hoffmann, P., F. Christensen, and D. Hammershøi (2013a). “Insert Earphone Calibration for Hear-Through Options”. *Proc. Audio Engineering Society Conference 51: Loudspeakers and Headphones*. Helsinki, Finland, pp. 1–8.

— (2013b). “Quantitative assessment of spatial sound distortion by the semi-ideal recording point of a hear-through device”. *Proceedings of Meetings on Acoustics*. Vol. 19. Acoustical Society of America, p. 050018.

Hoffmann, P., A. K. Møller, F. Christensen, and D. Hammershøi (2014). “Sound localization and speech identification in the frontal median plane with a hear-through headset”. *Proc. Forum Acusticum*. Krakow, Poland, pp. 1–6.

Hofman, P. M., J. G. Van Riswick, and A. J. Van Opstal (1998). “Relearning sound localization with new ears”. *Nature neuroscience* 1(5), pp. 417–421.

Hudde, H., A. Engel, and A. Lodwig (1999). “Methods for estimating the sound pressure at the eardrum”. *The Journal of the Acoustical Society of America* 106(4), pp. 1977–1992.

Hudde, H. (1983). “Measurement of the eardrum impedance of human ears”. *The Journal of the Acoustical Society of America* 73(1), pp. 242–247.

InEar (2019). “InEar ProPhile Series”. https://www.inear-monitoring.eu/en/produkte/inear-universelles-monitoring/monitoring-prophile-8.htm.

Jensen, N. S., T. Neher, S. Laugesen, R. B. Johannesson, and L. Kragelund (2013). “Laboratory and Field Study of the Potential Benefits of Pinna Cue-Preserving Hearing Aids”. *Trends in Amplification* 17(3), pp. 171–188.

Jessen, A. H., L. Baekgaard, and H. P. Ander (2014). “What is a Good Hearing Aid Sound Quality, and Does it Really Matter?” *AudiologyOnline*. Article 12340.

Karjalainen, M. and T. Paatero (2001). “Frequency-dependent signal windowing”. *Proc. Applications of Signal Processing to Audio and Acoustics (WASPAA), 2001 IEEE Workshop on*. New Paltz, USA, pp. 35–38.

Käsbach, J., T. May, G. Oskarsdottir, C.-H. Jeong, and J. Chang (2014). “The effect of interaural-time-difference fluctuations on apparent source width”. *Proc. Forum Acusticum*. Krakow, Poland, pp. 1–5.

Katz, B. F. G. and M. Noisternig (2014). “A comparative study of interaural time delay estimation methods”. *The Journal of the Acoustical Society of America* 135(6), pp. 3530–3540.

Kayser, H., S. Ewert, J. Anemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier (2009). “Database of Multichannel In-Ear and Behind-the-Ear Head-Related and Binaural Room Impulse Responses”. *EURASIP Journal on Advances in Signal Processing* 2009(1), p. 298605.

Keefe, D. H. (1984). “Acoustical wave propagation in cylindrical ducts: Transmission line parameter approximations for isothermal and nonisothermal boundary conditions”. *The Journal of the Acoustical Society of America* 75(1), pp. 58–62.

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”. *International Journal of Audiology* 46(10), pp. 554–568.

Kidmose, P., D. Looney, M. Ungstrup, M. L. Rank, and D. P. Mandic (2013). “A Study of Evoked Potentials From Ear-EEG”. *IEEE Transactions on Biomedical Engineering* 60(10), pp. 2824–2830.

Kießling, J., B. Kollmeier, and U. Baumann (2018). *Versorgung mit Hörgeräten und Hörimplantaten*. 3rd ed. Stuttgart: Thieme.

Killion, M. C. (1979). “Design and Evaluation of High-Fidelity Hearing Aids”. PhD Thesis. Evanston, Illinois: Northwestern University.

— (2004a). “Myths About Hearing Aid Benefit and Satisfaction”. *Hearing Review* (August), pp. 14–66.

— (2004b). “Myths that Discourage Improvements in Hearing Aid Design”. *Hearing Review* 11(4), pp. 32–70.

Killion, M. C. and E. L. Monsor (1980). “CORFIG: Coupler Response For Flat Insertion Gain”. *Acoustical Factors Affecting Hearing Aid Performance*. 1st ed., pp. 149–168.

Killion, M. C., E. H. Berger, and R. A. Nuss (1987). “Diffuse field response of the ear”. *The Journal of the Acoustical Society of America* 81(S1), S75–S75.

Killion, M. C. and L. J. Revit (1987). “Insertion Gain Repeatability versus Loudspeaker Location: You Want Me to Put My Loudspeaker WHERE”. *Ear and Hearing* 8(5), pp. 68–73.

Killion, M. C., T. Monroe, and V. Drambarean (2011). “Better protection from blasts without sacrificing situational awareness”. *International Journal of Audiology* 50(S1), pp. 38–45.

King, R. B. and S. R. Oldfield (1997). “The impact of signal bandwidth on auditory localization: Implications for the design of three-dimensional audio displays”. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 39(2), pp. 287–295.

Kinkel, M. (2016). “Hearing aid (HA) demographics: What did change during the last two decades?” Tahoe City, CA, USA.

Kirkeby, O. and P. A. Nelson (1999). “Digital Filter Design for Inversion Problems in Sound Reproduction”. *Journal of the Audio Engineering Society* 47(7/8), pp. 583–595.

Kollmeier, B., J. Peissig, and V. Hohmann (1993). “Real-time multiband dynamic compression and noise reduction for binaural hearing aids”. *Journal of Rehabilitation Research and Development* 30(1), pp. 82–94.

Kollmeier, B., M. Hiipakka, G. Grimm, T. Neher, T. d. Taillez, J. Schröder, J. Anemüller, and V. Hohmann (2014). “Toward a scalable, binaural hearing device: Acoustics, interaural processing, and scene decomposition”. *The Journal of the Acoustical Society of America* 135(4), pp. 2427–2427.

Kollmeier, B. and J. Kiessling (2018). “Functionality of hearing aids: state-of-the-art and future model-based solutions”. *International Journal of Audiology* 57(S3), pp. 3–28.

Kuhn, G. F. (1977). “Diffuse-field sound pressure transformation to a manikin’s ear, head surface, and torso surface”. *The Journal of the Acoustical Society of America* 61(S1), S96–S96.

— (1979a). “Acoustic diffraction and resonance effects for the aided versus the unaided listener”. *The Journal of the Acoustical Society of America* 66(S1), S16–S16.

Kuhn, G. F. (1979b). "The pressure transformation from a diffuse sound field to the external ear and to the body and head surface". *The Journal of the Acoustical Society of America* 65(4), pp. 991–1000.

Kuhn, G. F. and E. D. Burnett (1977). "Acoustic pressure field alongside a manikin's head with a view towards in situ hearing-aid tests". *The Journal of the Acoustical Society of America* 62(2), pp. 416–423.

Kuk, F., D. Keenan, and C. Lau (2009). "Comparison of Vent Effects between a Solid Earmold and a Hollow Earmold". *Journal of the American Academy of Audiology* 20, pp. 480–91.

Kuk, F., P. Korhonen, C. Lau, D. Keenan, and M. Norgaard (2013). "Evaluation of a Pinna Compensation Algorithm for Sound Localization and Speech Perception in Noise". *American Journal of Audiology* 22(1), pp. 84–93.

Langendijk, E. H. A. and A. W. Bronkhorst (2002). "Contribution of spectral cues to human sound localization". *The Journal of the Acoustical Society of America* 112(4), pp. 1583–1596.

Lewis, J. and B. Moss (2013). "MEMS Microphones, the Future for Hearing Aids". *Analog Dialogue* 47(11).

Liebich, S., C. Anemuller, P. Vary, P. Jax, D. Ruschen, and S. Leonhardt (2016). "Active noise cancellation in headphones by digital robust feedback control". *Proc. 24th European Signal Processing Conference (EUSIPCO)*. Budapest, Hungary, pp. 1843–1847.

Lindau, A. and S. Weinzierl (2006). "FABIAN - An instrument for software-based measurement of binaural room impulse responses in multiple degrees of freedom". *24. Tonmeistertagung - VFT International Conference*. Leipzig, Germany, pp. 621–625.

Lindau, A. and F. Brinkmann (2012). "Perceptual Evaluation of Headphone Compensation in Binaural Synthesis Based on Non-Individual Recordings". *Journal of the Audio Engineering Society* 60(1/2), pp. 54–62.

Ljung, L. (1998). *System identification: theory for the user*. 2nd ed. Prentice Hall.

Lokki, T., H. Nironen, S. Vesa, L. Savioja, A. Härmä, and M. Karjalainen (2004). "Application Scenarios of Wearable and Mobile Augmented Reality Audio". *Proc. Audio Engineering Society Convention 116*. Convention Paper 6026. Berlin, Germany, pp. 1–9.

Looney, D., P. Kidmose, C. Park, M. Ungstrup, M. Rank, K. Rosenkranz, and D. Mandic (2012). "The In-the-Ear Recording Concept: User-Centered and Wearable Brain Monitoring". *IEEE Pulse* 3(6), pp. 32–42.

Lundbeck, M., L. Hartog, G. Grimm, V. Hohmann, L. Bramsløw, and T. Neher (2018). "Influence of Multi-microphone Signal Enhancement Algorithms on the Acoustics and Detectability of Angular and Radial Source Movements". *Trends in Hearing* 22, p. 233121651877971.

Lunner, T. and T. Neher (2013). "Automatic real-time hearing aid fitting based on auditory evoked potentials". US201161547079P. Patent.

Macpherson, E. A. and J. C. Middlebrooks (2002). “Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited”. *The Journal of the Acoustical Society of America* 111(5), pp. 2219–2236.

Madaffari, P. L. (1974). “Pressure variation about the ear”. *The Journal of the Acoustical Society of America* 56(S1), S3–S3.

Majdak, P., P. Balazs, and B. Laback (2007). “Multiple exponential sweep method for fast measurement of head-related transfer functions”. *Journal of the Audio Engineering Society* 55(7/8), pp. 623–637.

Majdak, P., T. Walder, and B. Laback (2013). “Effect of long-term training on sound localization performance with spectrally warped and band-limited head-related transfer functions”. *The Journal of the Acoustical Society of America* 134(3), pp. 2148–2159.

Majdak, P., R. Baumgartner, and B. Laback (2014). “Acoustic and non-acoustic factors in modeling listener-specific performance of sagittal-plane sound localization”. *Frontiers in Psychology* 5(00319), pp. 1–10.

Majdak, P., R. Baumgartner, and C. Jenny (2019). “Formation of three-dimensional auditory space”. *arXiv:1901.03990 [q-bio]*. arXiv: 1901.03990.

Marentakis, G. and R. Liepins (2014). “Evaluation of hear-through sound localization”. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Toronto, Canada, pp. 267–270.

Martin, A., C. Jin, and A. v. Schaik (2009). “Psychoacoustic evaluation of systems for delivering spatialized augmented-reality audio”. *Journal of the Audio Engineering Society* 57(12), pp. 1016–1027.

Martin, R. and R. V. Cox (1999). “New speech enhancement techniques for low bit rate speech coding”. *1999 IEEE Workshop on Speech Coding Proceedings. Model, Coders, and Error Criteria (Cat. No.99EX351)*, pp. 165–167.

Mazur, K. and J. Voix (2013). “Implementing 24-hour in-ear dosimetry with recovery”. *Proceedings of Meetings on Acoustics*. Vol. 19. Acoustical Society of America, p. 040016.

Mehrgardt, S. and V. Mellert (1977). “Transformation characteristics of the external human ear”. *The Journal of the Acoustical Society of America* 61(6), pp. 1567–1576.

Mendonça, C. (2014). “A review on auditory space adaptations to altered head-related cues”. *Frontiers in Neuroscience* 8(00219), pp. 1–14.

Meyer, B. T., T. Jürgens, T. Wesker, T. Brand, and B. Kollmeier (2010). “Human phoneme recognition depending on speech-intrinsic variability”. *The Journal of the Acoustical Society of America* 128(5), pp. 3126–3141.

Middlebrooks, J. C. (1999a). “Individual differences in external-ear transfer functions reduced by scaling in frequency”. *The Journal of the Acoustical Society of America* 106(3), pp. 1480–1492.

Middlebrooks, J. C. (1999b). “Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency”. *The Journal of the Acoustical Society of America* 106(3), pp. 1493–1510.

Mikkelsen, K. B., S. L. Kappel, D. P. Mandic, and P. Kidmose (2015). “EEG Recorded from the Ear: Characterizing the Ear-EEG Method”. *Frontiers in Neuroscience* 9(438), pp. 1–8.

Minnaar, P., S. K. Olesen, F. Christensen, and H. Møller (2001). “Localization with Binaural Recordings from Artificial and Human Heads”. *Journal of the Audio Engineering Society* 49(5), pp. 323–336.

Mirkovic, B., M. G. Bleichner, M. De Vos, and S. Debener (2016). “Target Speaker Detection with Concealed EEG Around the Ear”. *Frontiers in Neuroscience* 10(349), pp. 1–11.

Møller, H. (1992). “Fundamentals of Binaural Technology”. *Applied Acoustics* 36(3/4), pp. 171–218.

Møller, H., M. F. Sørensen, D. Hammershøi, and C. B. Jensen (1995a). “Head-related transfer functions of human subjects”. *Journal of the Audio Engineering Society* 43(5), pp. 300–321.

Møller, H., D. Hammershøi, C. B. Jensen, and M. F. Sørensen (1995b). “Transfer Characteristics of Headphones Measured on Human Ears”. *Journal of the Audio Engineering Society* 43(4), pp. 203–217.

Møller, H., M. F. Sørensen, C. B. Jensen, and D. Hammershøi (1996). “Binaural Technique: Do We Need Individual Recordings?” *Journal of the Audio Engineering Society* 44(6), pp. 451–469.

Moore, B. C. J. (2012). *An Introduction to the Psychology of Hearing*. Bingley, UK: Emerald Publishing Group.

Moore, B. C. J. and C.-T. Tan (2003). “Perceived naturalness of spectrally distorted speech and music”. *The Journal of the Acoustical Society of America* 114(1), pp. 408–419.

— (2004). “Development and validation of a method for predicting the perceived naturalness of sounds subjected to spectral distortion”. *Journal of the Audio Engineering Society* 52(9), pp. 900–914.

Morimoto, M. (2001). “The contribution of two ears to the perception of vertical angle in sagittal planes”. *The Journal of the Acoustical Society of America* 109(4), pp. 1596–1603.

Mueller, H. G. (2001). “Probe Microphone Measurements: 20 Years of Progress”. *Trends in Amplification* 5(2), pp. 35–68.

— (2006). “Opens is in”. *The Hearing Journal* 59(11), pp. 11–12.

Mueller, H. G. and E. M. Picou (2010). “Survey examines popularity of real-ear probe-microphone measures”. *The Hearing Journal* 63(5), p. 27.

Mueller, M. F., A. Kegel, S. M. Schimmel, N. Dillier, and M. Hofbauer (2012). “Localization of virtual sound sources with bilateral hearing aids in realistic acoustical scenes". *The Journal of the Acoustical Society of America* 131(6), pp. 4732–4742.

Müller, S. and P. Massarani (2001). "Transfer-function measurement with sweeps". *Journal of the Audio Engineering Society* 49(6), pp. 443–471.

Näätänen, R. and T. Picton (1987). "The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an Analysis of the Component Structure". *Psychophysiology* 24(4), pp. 375–425.

Nelder, J. A. and R. Mead (1965). "A Simplex Method for Function Minimization". *The Computer Journal* 7(4), pp. 308–313.

Neubauer, W. G. and L. R. Dragonette (1967). "Monostatic Reflection from a Rigid Rectangular Plane". *The Journal of the Acoustical Society of America* 41(3), pp. 656–661.

Niedrist, G. (1993). "Echo suppression for loudspeaker-microphone system measurements". *Journal of the Audio Engineering Society* 41(3), pp. 143–153.

Noble, W., D. Byrne, and K. Ter-Horst (1997). "Auditory localization, detection of spatial separateness, and speech hearing in noise by hearing impaired listeners". *The Journal of the Acoustical Society of America* 102(4), pp. 2343–2352.

Noble, W., S. Sinclair, and D. Byrne (1998). "Improvement in aided sound localization with open earmolds: observations in people with high-frequency hearing loss". *Journal of the American Academy of Audiology* 9(1), pp. 25–34.

O'Sullivan, J. A., A. J. Power, N. Mesgarani, S. Rajaram, J. J. Foxe, B. G. Shinn-Cunningham, M. Slaney, S. A. Shamma, and E. C. Lalor (2015). "Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG". *Cerebral Cortex* 25(7), pp. 1697–1706.

O'Sullivan, J., Z. Chen, J. Herrero, G. M. McKhann, S. A. Sheth, A. D. Mehta, and N. Mesgarani (2017). "Neural decoding of attentional selection in multi-speaker environments without access to clean sources". *Journal of Neural Engineering* 14(5), p. 056001.

Orfanidis, S. J. (1997). "Digital parametric equalizer design with prescribed Nyquist-frequency gain". *Journal of the Audio Engineering Society* 45(6), pp. 444–455.

Pavlovic, C., V. Hohmann, H. Kayser, L. Wong, T. Herzke, S. R. Prakash, z. Hou, and P. Maanen (2018). "Open portable platform for hearing aid research". *The Journal of the Acoustical Society of America* 143(3), pp. 1738–1738.

Polich, J. (2007). "Updating P300: An integrative theory of P3a and P3b". *Clinical Neurophysiology* 118(10), pp. 2128–2148.

Popelka, G. R. and B. C. J. Moore (2016). "Future Directions for Hearing Aid Development". *Hearing Aids*. Springer Handbook of Auditory Research. Springer, Cham, pp. 323–333.

Rämö, J. and V. Välimäki (2014). "An allpass hear-through headset". *Proc. 22nd European Signal Processing Conference (EUSIPCO)*. Lisbon, Portugal, pp. 1123–1127.

Rämö, J. and V. Välimäki (2012). “Digital Augmented Reality Audio Headset”. *Journal of Electrical and Computer Engineering* 2012, p. 457374.

Rayleigh, J. W. S. B. (1896). *The Theory of Sound*. 2nd ed. London: Macmillan.

— (1907). “XII. On our perception of sound direction”. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 13(74), pp. 214–232.

Riederer, K. A. (1998). “Repeatability analysis of head-related transfer function measurements”. *Audio Engineering Society Convention 105*. San Francisco, USA, pp. 1–62.

— (2004). “Part IVa: Effect of cavum conchae blockage on human head-related transfer functions”. *Proceedings of 18th International Congress on Acoustics ICA*. Kyoto, Japan, pp. 1–4.

Rife, D. D. and J. Vanderkooy (1989). “Transfer-function measurement with maximum-length sequences”. *Journal of the Audio Engineering Society* 37(6), pp. 419–444.

Roden, R., N. Wulbusch, **F. Denk**, H. Schepker, A. Chernov, S. Doclo, and M. Blau (2019a). “Optimization of vented earpiece designs using the finite element method”. *Presented at International Congress on Acoustics ICA*. Aachen, Germany.

Roden, R., N. Wulbusch, A. Chernov, **F. Denk**, and M. Blau (2019b). “Zur optimalen Positionierung eines Vent-Mikrophons in akustisch offenen Hörsystemen - eine numerische Studie”. *Fortschirtte der Akustik - DAGA*. Rostock, Germany, pp. 1284–1287.

Sabin, A. T. (2018). “Trends influencing hearing devices and real-world efficacy of a self-fit method”. *Presented at International Hearing Aid Research Conference (IHCON)*. Tahoe City, CA, USA.

Sankowsky-Rothe, T., M. Blau, S. Köhler, and A. Stirnemann (2015a). “Individual Equalization of Hearing Aids with Integrated Ear Canal Microphones”. *Acta Acustica united with Acustica* 101(3), pp. 552–566.

Sankowsky-Rothe, T., M. Blau, H. Schepker, and S. Doclo (2015b). “Reciprocal measurement of acoustic feedback paths in hearing aids”. *The Journal of the Acoustical Society of America* 138(4), EL399–EL404.

Schepker, H., L. T. T. Tran, S. Nordholm, and S. Doclo (2017a). “Combining null-steering and adaptive filtering for acoustic feedback cancellation in a multi-microphone earpiece”. *Proc. 25th European Signal Processing Conference (EUSIPCO)*. Kos, Greece, pp. 231–235.

Schepker, H., S. E. Nordholm, L. T. T. Tran, and S. Doclo (2019a). “Null-Steering Beamformer-Based Feedback Cancellation for Multi-Microphone Hearing Aids With Incoming Signal Preservation”. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 27(4), pp. 679–691.

Schepker, H., L. T. Tran, S. Nordholm, and S. Doclo (2016a). “A Robust Null-Steering Beamformer for Acoustic Feedback Cancellation for a Multi-Microphone Earpiece”. *Proc. 12. ITG Conference on Speech Communication*. Paderborn, Germany, pp. 1–5.

— (2016b). “Acoustic feedback cancellation for a multi-microphone earpiece based on a null-steering beamformer”. *Proc. IEEE International Workshop on Acoustic Signal Enhancement (IWAENC)*. Xi’an, China: IEEE, pp. 1–5.

— (2017b). “Null-steering beamformer for acoustic feedback cancellation in a multi-microphone earpiece optimizing the maximum stable gain”. *Proc. Conference on Acoustics, Speech and Signal Processing*. New Orleans, USA, pp. 341–345.

Schepker, H., **F. Denk**, B. Kollmeier, and S. Doclo (2018). “Multi-loudspeaker equalization for acoustic transparency in a custom hearing device”. *Proc. 13. ITG Conference on Speech Communication*. Oldenburg, Germany, pp. 36–40.

— (2019b). “Subjective sound quality evaluation of an acoustically transparent hearing device”. *Proc. 2nd AES Conference on Headphone Technology*. Paper 18. San Francisco, USA, pp. 1–10.

Schimmel, H. (1967). “The (+) reference: accuracy of estimated mean components in average response studies”. *Science (New York, N.Y.)* 157(3784), pp. 92–94.

Schinkel-Bielefeld, N., C. Oreinos, and H. P. Kamkar (2018). “Improving localization in binaural beamforming for hearing aid wearers”. *Presented at Speech In Noise Workshop*. Glasgow, UK.

Shaw, E. A. G. (1974). “Transformation of sound pressure level from the free field to the eardrum in the horizontal plane”. *The Journal of the Acoustical Society of America* 56(6), pp. 1848–1861.

— (1980). “The Acoustics of the External Ear”. *Acoustical Factors Affecting Hearing Aid Performance*. Baltimore: University Park Press.

Shaw, E. A. G. and R. Teranishi (1968). “Sound Pressure Generated in an External Ear Replica and Real Human Ears by a Nearby Point Source”. *The Journal of the Acoustical Society of America* 44(1), pp. 240–249.

Shaw, E. A. G. and M. M. Vaillancourt (1985). “Transformation of sound-pressure level from the free field to the eardrum presented in numerical form”. *The Journal of the Acoustical Society of America* 78(3), pp. 1120–1123.

Shinn-Cunningham, B. G., N. Kopco, and T. J. Martin (2005). “Localizing nearby sound sources in a classroom: Binaural room impulse responses”. *The Journal of the Acoustical Society of America* 117(5), pp. 3100–3115.

Siqueira, M. G. and A. Alwan (2000). “Steady-state analysis of continuous adaptation in acoustic feedback reduction systems for hearing-aids”. *IEEE Transactions on Speech and Audio Processing* 8(4), pp. 443–453.

Sockalingam, R., J. Beilin, and D. L. Beck (2009). “Sound Quality Considerations of Hearing Instruments”. *Hearing Review* 16(9), pp. 22–28.

Spriet, A., I. Proudler, M. Moonen, and J. Wouters (2005). “Adaptive feedback cancellation in hearing aids with linear prediction of the desired signal”. *IEEE Transactions on Signal Processing* 53(10), pp. 3749–3763.

Spriet, A., S. Doclo, M. Moonen, and J. Wouters (2008). “Feedback Control in Hearing Aids”. *Springer Handbook of Speech Processing*. Ed. by J. Benesty, M. M.

Sondhi, and Y. A. Huang. Springer Handbooks. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 979–1000.

Stinson, M. R. and G. A. Daigle (2007). “Transverse pressure distributions in a simple model ear canal occluded by a hearing aid test fixture”. *The Journal of the Acoustical Society of America* 121(6), p. 3689.

Stone, M. A. and B. C. Moore (1999). “Tolerable hearing aid delays. I. Estimation of limits imposed by the auditory path alone using simulated hearing losses”. *Ear and Hearing* 20(3), pp. 182–192.

Stone, M. A., B. C. J. Moore, K. Meisenbacher, and R. P. Derleth (2008). “Tolerable Hearing Aid Delays. V. Estimation of Limits for Open Canal Fittings:” *Ear and Hearing* 29(4), pp. 601–617.

Struck, C. J. and S. F. Temme (1994). “Simulated free field measurements”. *Journal of the Audio Engineering Society* 42(6), pp. 467–482.

Swartz Center for Computational Neuroscience, C. and C. Kothe (2015). *Lab Streaming Layer (LSL).*

Tikander, M. (2009). “Usability Issues in Listening to Natural Sounds with an Augmented Reality Audio Headset”. *Journal of the Audio Engineering Society* 57(6), pp. 430–441.

Union, I. T. (2003). *ITU-R BS.1284-1 - General methods for the subjective assessment of sound quality*. Tech. rep. ITU Recommendation.

Välimäki, V., A. Franck, J. Rämö, H. Gamper, and L. Savioja (2015). “Assisted Listening Using a Headset: Enhancing audio perception in real, augmented, and virtual environments”. *IEEE Signal Processing Magazine* 32(2), pp. 92–99.

Van den Bogaert, T., T. J. Klasen, M. Moonen, L. Van Deun, and J. Wouters (2006). “Horizontal localization with bilateral hearing aids: Without is better than with”. *The Journal of the Acoustical Society of America* 119(1), pp. 515–526.

Van den Bogaert, T., E. Carette, and J. Wouters (2011). “Sound source localization using hearing aids with microphones placed behind-the-ear, in-the-canal, and in-the-pinna”. *International Journal of Audiology* 50(3), pp. 164–176.

Vincent, E. et al. (2012). “The signal separation evaluation campaign (2007–2010): Achievements and remaining challenges”. *Signal Processing* 92(8), pp. 1928–1936.

Vogl, S., T. Sankowsky-Rothe, and M. Blau (2016). “Elektroakustische Modellierung eines Ohrpassstücks mit integrierten Mikrofonen und Lautsprechern [Electroacoustic modeling of an earpiece with integrated microphones and loudspeakers]”. *Fortschritte der Akustik - DAGA*. Aachen, Germany, pp. 1–4.

Vogl, S., M. Blau, and T. Sankowsky-Rothe (2017). “Model based prediction of sound pressure at the ear drum for an open earpiece equipped with two receivers and three microphones”. *The Journal of the Acoustical Society of America* 141(5), pp. 3957–3957.

Vogl, S. and M. Blau (2019). “Individualized prediction of the sound pressure at the eardrum for an earpiece with integrated receivers and microphones”. *The Journal of the Acoustical Society of America* 145(2), pp. 917–930.

Völk, F., C. Landsiedel, and H. Fastl (2011). “Auditory adapted exponential transfer function smoothing (AAS)”. *Proc. Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop on.* New Paltz, USA, pp. 49–52.

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”. *International Journal of Audiology* 57(sup3), S92–S104.

Vorländer, M. (2004). “Past, present and future of dummy heads”. *Acústica.* Guimarães, Portugal, pp. 1–6.

Wahba, G. (1981). “Spline interpolation and smoothing on a sphere”. *SIAM Journal of Scientific and Statistical Computing* 2(1), pp. 5–16.

— (1982). “Erratum: Spline interpolation and smoothing on a sphere”. *SIAM Journal of Scientific and Statistical Computing* 3(3), pp. 385–386.

Wallach, H. (1940). “The role of head movements and vestibular and visual cues in sound localization”. *Journal of Experimental Psychology* 27(4), pp. 339–368.

Waterschoot, T. v. and M. Moonen (2011). “Fifty Years of Acoustic Feedback Control: State of the Art and Future Challenges”. *Proceedings of the IEEE* 99(2), pp. 288–327.

Wenzel, E. M., M. Arruda, D. J. Kistler, and F. L. Wightman (1993). “Localization using nonindividualized head-related transfer functions”. *The Journal of the Acoustical Society of America* 94(1), pp. 111–123.

Wiener, F. M. and D. A. Ross (1946). “The Pressure Distribution in the Auditory Canal in a Progressive Sound Field”. *The Journal of the Acoustical Society of America* 18(2), pp. 401–408.

Wightman, F. L. and D. J. Kistler (1989). “Headphone simulation of free field listening. I: Stimulus synthesis”. *The Journal of the Acoustical Society of America* 85(2), pp. 858–867.

— (1992). “The dominant role of low-frequency interaural time differences in sound localization”. *The Journal of the Acoustical Society of America* 91(3), pp. 1648–1661.

Wille, M. and P. Rasmussen (2016). “IEC 60318-4 Ear Simulator for Low Noise Measurements & Anthropometric Rubber Pinna”. *Proc. AES Conference on Headphone Technology.* Aalborg, Denmark, pp. 96–102.

Winkler, A., M. Latzel, and I. Holube (2016). “Open Versus Closed Hearing-Aid Fittings: A Literature Review of Both Fitting Approaches”. *Trends in Hearing* 20, p. 2331216516631741.

Danksagung

Birger Kollmeier danke ich für Möglichkeit, diese Promotion durchzuführen, sowie für die langjährige Betreuung mit wichtigen Anstößen, Anleitung wo sie nötig war, umfassender Unterstützung einer Vielzahl von Ideen mit den nötigen Ressourcen, aber auch mit großen Spielräumen zur freien Entfaltung. Bedanken will ich mich auch für das Schaffen einer generell positiven und offenen, lockeren aber auch ausgesprochen produktiven Arbeitsatmosphäre. Die zügige Fertigstellung dieser Arbeit wurde durch die jährlichen Schreibworkshops auf Brač maßgeblich vorangetrieben.

Dorte Hammershøi danke ich für das Interesse an dieser Arbeit und die Übernahme des Koreferats.

Stephan Ewert danke ich für die Mitbetreuung vieler Teile dieser Arbeit mit einem regelmäßigen Austausch, vielen Diskussionen, und hilfreicher Anleitung.

Simon Doclo danke ich im selben Maße für eine mehr als gute Zusammenarbeit, regen Austausch und lehrreiche Diskussionen.

Henning Schepker danke ich für eine fruchtvolle Zusammenarbeit und viele spannende Diskussionen, besonders bei der gemeinsamen Weiterentwicklung der Equalization-Algorithmen und des Echtzeit-Demonstrators und deren Evaluation.

Steffen Vogl, Reinhild Roden und Matthias Blau danke ich für die Zusammenarbeit, vor allem zu den Hardware-Aspekten des transparenten Earpiece, danken. In gemeinsamen Diskussionen ist so manche Schwachstelle inklusive deren Lösung gefunden worden.

Martin Bleichner und Stefan Debener danke ich für die spannende und erfolgreiche interdisziplinäre Zusammenarbeit bei der Realisierung des Ear-EEG-Earpiece Projekts.

Jan Heeren danke ich für die Unterstützung bei Messungen im VR-Lab, sowie gemeinsame Konstruktions- und Umbauaktionen dort. Auch wenn ich mittlerweile selbst fast jedes der zahlreichen Kabel beim Namen kenne, er hat sie mir vorgestellt.

Hendrik Kayser und Giso Grimm danke ich für schnelle Hilfestellungen bei praktischen Problemen mit Echtzeitverarbeitung, Linux Audio oder dem Master Hearing Aid.

Stephan Ernst danke ich für die Betreuung der vorangegangenen Bachelor- und Masterarbeit, die mich nachhaltig für die hier behandelte Thematik begeistern konnte.

Ein besonderes Dankeschön geht auch an die TAs der Medi, ohne die vieles sicher nicht so rund laufen würde wie es tut: Anita Gorges, Frank Grunau, Ralf Hauptmann, Jennifer Köllner und Katja Warnken.

Das Department für Medizinische Physik und Akustik der Universität Oldenburg, das Exzellenzcluster *Hearing4all* und insbesondere die AG Medi waren für mich in den letzten Jahren ein inspirierendes Umfeld, von dem ich in vielerlei Hinsicht stark profitieren konnte. Ich empfinde es als ein großes Glück, hier als Forscher tätig zu sein. Besonders meinen Büro-Mitbewohnern Tobias de Taillez, Lena Haverkamp und Julia Schütze aus NeSSy-104, sowie dem Rest der Medi danke ich für eine stets gute Stimmung, erfrischende Pausen, wildes Brainstorming, Rat und Tat bei den kleinen und größeren Problemen des Alltags und abwechslungsreiche Diskussionen über Gott und die Welt.

Als letztes möchte ich meiner Familie und Freunden danken. Meinen Eltern für den Rückhalt und die Unterstützung auch während des Studiums, und natürlich meiner Partnerin Stefanie. Danke für die Hilfe, das Verständnis und dafür, dass ich alle Höhen und Tiefen der gemeinsamen Doktorandenzeit auf sämtlichen Ebenen mit teilen durfte.

Hearing devices are electro-acoustic instruments worn in the ear to modify the heard environment, like hearing aids, electronic hearing protectors or hearables. An ideal hearing device would allow for natural sound perception just as with the open ear, except for the desired modification, like an amplification of sound in a hearing aid. This property is referred to as acoustic transparency. Current hearing devices usually do not fulfil this requirement but introduce additional "side-effects" that lead to a reduced ability to localize, separate and understand different sound sources, and a reduced sound quality. In this thesis, extensive acoustic and psychophysical measurements revealed how the presence of a hearing device alters the acoustic transmission properties of the external ear, leads to increased difficulty in localizing sounds, and reduces the sound quality. Based upon these findings, methods for providing optimum acoustic transparency and thus reducing such side-effects to a minimum were proposed and evaluated using a custom prototype hearing device. The results of this work show how the mechanical design of a hearing device imposes principal limitations on the achievable grade of acoustic transparency, and that for best possible performance the hearing device should be adapted to the individual ear.

ISBN 978-3-8142-2387-2