Fakultät für Elektrotechnik und Informationstechnik
Technische Universität München

Consolidating natural spatial perception and improved SNR in hearing aids: Jackrabbit, a new method

Gabriel Gomez

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender:
Prof. Dr.-Ing. Eckehard Steinbach

Prüfende der Dissertation:
1. Prof. Dr.-Ing. Bernhard U. Seeber
2. Prof. Dr.-Ing. Werner Hemmert

Die Dissertation wurde am 22.11.2018 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 23.06. 2019 angenommen.

Acknowledgements

To my wife and daughters, who gave me the strength for this journey

I would like to thank Dr. Marko Takanen and Dr. Claudia Freigang, who gave me helpful feedback on this document. I also thank my students Valerie Hoening, Norbert Kolotzek, Beike Lu, Philipp von Unold and Simon Conrady, who helped me with the implementation of the experiments. Special thanks to all my colleagues and students who had to endure countless hours of listening tests, which laid the foundations for this work. I thank Prof. Dr. Bernhard Seeber for teaching me good scientific practice. Finally, I am very grateful to Dr. Peter Derleth, Dr. Markus Hofbauer and Siddhartha Jha from the Phonak AG, who made this PhD project possible with funding, hardware, software and great support.

Abstract

This PhD thesis deals with the natural spatial perception of sounds with hearing aids. These devices help hearing impaired persons by making sounds audible again, reducing noise and improving speech understanding. Yet, these aids may also distort cues that are important for perceiving sounds naturally and identifying their correct spatial location. This work investigates situations in which hearing aids can significantly deteriorate the spatial perception of sounds. First, a psychoacoustic experiment on spatial perception shows a deterioration of sound when behind-the-ear (BTE) or beamformer (BF) hearing aid conditions are used, compared to the in-the-ear (ITE) condition which exploits cues of the outer ear (pinna). Based on these findings, two novel methods are presented that attempt to combine pinna cue preservation with noise reduction by beamforming. The first method imposes pinna cues upon a beamforming signal, while the Jackrabbit method reduces disturber energy from the ITE signals by directional filtering. Four psychoacoustic experiments with normal hearing participants are presented which validate the novel algorithms, compared to the ITE and BTE conditions and to a static beamformer condition. The experiments covered (a) localization accuracy in the front and back, (b) speech understanding in two different noise conditions, (c) spatial sound quality with concurrent target and disturber in a reverberated cafeteria scenario, and (d) the externalization perception of sounds presented from the front. The results show a benefit in spatial hearing by preserving pinna cues using the ITE microphone position in many of the tested spatial dimensions. An additional benefit can be achieved when the ITE microphone position signals are combined with noise reduction using directional filtering, such as in the Jackrabbit method which performed best in all four experiments.

Zusammenfassung

Diese Doktorarbeit befasst sich mit der natürlichen räumlichen Wahrnehmung von Schallen mit Hörgeräten. Diese helfen, Geräusche wieder hörbar zu machen, verringern Störgeräusche und verbessern das Sprachverstehen. Sie können jedoch bestimmte Informationen verzerren, die wichtig für die natürliche und korrekte räumliche Wahrnehmung von Schallen sind. Diese Arbeit untersucht Situationen, in denen Hörgeräte signifikant die räumliche Wahrnehmung verzerren. Zuerst wird ein Experiment zur räumlichen Wahrnehmung von Schallen präsentiert, welches für die Hinter-dem-Ohr und Richtmikrofon Bedingungen merkliche Verschlechterungen gegenüber der Im-Gehörgang Bedingung, welche die Merkmale des Außenohrs (Pinna) einbezieht, zeigt. Basierend auf diesen Erkenntnissen werden zwei neue Methoden vorgestellt, welche Pinna-Merkmale bewahren und zusätzlich Störgeräuschreduktion auf Basis von Richtmikrofonie anwenden. Die erste Methode prägt Pinna-Merkmale auf ein Richtmikrofonsignal auf, während die Jackrabbit-Methode ausgehend vom Signal im Gehörgang eine Filterung der Störenergie durch räumliche Trennung durchführt. Vier Experimente zur Validierung der neuen Methoden werden vorgestellt, in denen diese mit den Signalen Im-Gehörgang, Hinter-dem-Ohr sowie mit statischen Richtmikrofonen verglichen werden. Die vier Experimente befassen sich mit (a) der Lokalisierung von Schallen von vorne und hinten, (b) dem Sprachverstehen für zwei verschiedene Störgeräuschszenarien, (c) der räumlichen Qualitätswahrnehmung von Schallen für konkurrierende Ziel- und Störsignale in einer verhallten Cafeteria-Umgebung, sowie (d) dem Externalisierungsempfinden von Schallen von vorne. Ergebnisse zeigen für die Mikrofonposition im Gehörgang ein besseres räumliches Hören für viele der Dimensionen die getestet wurden. Zusätzliche Vorteile ergeben sich durch die Kombination der Mikrofonposition im Gehörgang mit Störgeräuschreduktion basierend auf Richtmikrofonie wie in der Jackrabbit-Implementation, die in allen Experimenten die besten Ergebnisse lieferte.

Contents

Acronyms

1N	One-Noise.
2N	Two-Noise.
AD	Analog-to-Digital.
ANOVA	Analysis of Variance.
ANSI	American National Standards Institute.
ASW	Apparent Source Width.
at	Attack Time.
BF	Beamformer.
BRIR	Binaural Room-Impulse Response.
BTE	Behind-The-Ear.
dB	Decibel.
DRR	Direct-To-Reverberant Ratio.
ER	Externalization Rating.
ERB	Equivalent Rectangular Bandwidth.
HA	Hearing Aid.
HI	Hearing-Impaired.
HRIR	Head-Related Impulse Response.
HRTF	Head-Related Transfer Function.
IC	Interaural Coherence.
IIR	Infinite Impulse Response.
ILD	Interaural Level Difference.
IPD	Interaural Phase Difference.
IQR	Interquartile Range.
ITD	Interaural Time Difference.
ITE	In-The-Ear.
ITF	Interaural Transfer Function.
IVS	Interaural Vector Strength.
JND	Just Noticeable Difference.

MLE	Microphone Location Effect.
MLS	Maximum-Length Sequence.
MUSHRA	Multi-Stimulus Test with Hidden Reference and Anchor.
NH	Normal-Hearing.
OLSA	Oldenburger Satztest.
ProDePo	Proprioception Decoupled Pointer.
REF	Reference.
RMS	Root-Mean-Square.
SAQI	Spatial Audio Quality Inventory.
SMR	Signal-to-Masker Ratio.
SOFE	Simmulated Open Field Environment.
SSN	Speech Shaped Noise.
STA	Short Time Average.
STFT	Short-Time Fourier Transform.

1 Introduction

The auditory sense is used as the main receptor channel for communication between people, allowing complex social interaction. Unfortunately, hearing is a sense that for many humans deteriorates with time, having multiple possible causes. When a person's auditory sense degenerates, it becomes harder for that person to hear and understand speech. Hearing is important not only for communication but also for environmental awareness in all directions. Having two ears allows us to localize sound sources in space far better than we could with only one ear. It is especially beneficial to have two ears when there are multiple people participating in a conversation, since this allows to follow the conversation(s) and to localize individual speakers accurately in space. Hearing-impaired people not only have a hard time understanding speech in such situations, but also their spatial localization abilities deteriorate, because hearing and localizing sources are related processes. With the evolution of technology, hearing aid devices have been developed to partially compensate for the hearing loss - thereby improving speech comprehension of hearing-impaired people when they are wearing such devices. The success of hearing aids is based on two processes: first, hearing aids make sounds audible again by enhancing those parts of the sound that would otherwise be inaudible due to the hearing impairment. Secondly, hearing aids can reduce noise, e.g. irrelevant speech from a different direction than the speech the hearing aid user is trying to understand. This clarifies the desired signal, making it easier for the user to hear and understand the target speaker. However, there is an important drawback when using hearing aids. Most hearing aids on the market are so-called behind-the-ear (BTE) hearing aids, which have microphones on the upper side of the casing behind the outer ear. On the one hand, BTEs use two or more sufficiently separated microphones, facilitating a powerful noise reduction method called beamforming. On the other hand, beneficial alterations of sounds by the outer ears are not represented in the signals picked up by the hearing aid microphones located behind the outer ear. Thus, important spatial information gets lost when using behind the ear hearing aids. Specifically, a sound reaching the ears of a listener is altered in very specific ways by the outer ears before entering the ear canals. A unique modification of the incoming sound occurs for each direction in space. In people with healthy audition, the brain learns to distinguish these small modifications individually from each ear, which enables accurate localization and spatial perception of sounds. By using microphones placed behind the ear, much of the information conveyed by the outer ear is lost. The use of beam-forming for directional noise reduction worsens the situation even further, by not only failing to capture much of the spatial information, but even distorting it. Thus, normal BTE hearing aids have the advantage of making sounds, especially speech, more audible and easier to understand, but do so at the expense of deteriorating the spatial perception of sounds significantly.

The aforementioned problem is addressed in the present thesis and an attempt is made to overcome the loss of spatial information while still retaining the advantages of behind-the-ear hearing aids, especially of directional noise reduction by beamforming. To that end, two novel methods were developed to preserve the natural ear cues while reducing noise using beamforming. The developed methods were evaluated in a series of listening experiments that tackled the following questions:

- What is the influence of the position of the hearing aid microphone on the spatial perception of sounds?
- How does directional noise reduction affect the spatial perception of sounds?
- Which dimensions of spatial perception are influenced by different hearing aid algorithms or microphone positions (conditions)?
- How do different hearing aid conditions perform in speech understanding tests in the context of noise?
- How is spatial sound quality of sounds affected by different hearing aid conditions?
- Can the new methods perform similar to or better than other hearing aid conditions by combining the advantages of directional noise reduction by beamforming while maintaining natural outer ear (pinna) information?

This thesis is structured as follows: Chapter 2 presents a literature review on spatial hearing, speech understanding and hearing aid devices, illustrating the differences between normal-hearing and hearing-impaired listeners. Results of a psychoacoustic experiment on the spatial perception of acoustic stimuli in distance, azimuth, externalization, elevation and apparent source width in a reverberant environment are presented in Chapter 3. Signals from the microphone position behind the ear (of BTEs) were compared with those from in-the-ear (ITE) devices and of static beamforming using the BTE microphones. Chapter 4 presents two new methods that try to combine the advantages of beamforming and the naturalness of the ITE microphone position. The first method, called the STA BF method, attempts to apply natural pinna cues to a beam-formed signal, while the second method, called the Jackrabbit method, goes the opposite way by preserving the naturalness of the ITE microphone position and simultaneously increasing the Signal-to-Noise Ratio (SNR) by reducing disturber energy using directional filtering. Chapter 5 presents the experimental results of four psychoacoustic studies evaluating the different hearing aid conditions in the dimensions of localization, speech understanding, spatial sound quality and externalization of sound. This chapter also analyzes binaural cues that may be the cause of a reduced externalization percept with static beamforming, using the auditory model of Dietz et al. (2011) [Dietz et al., 2011]. Finally, Chapter 6 summarizes the findings of this work.

2 Literature Overview

One of the goals of this dissertation is to investigate the influence of hearing aids on how humans perceive sounds in the spatial dimension. This chapter gives a brief overview of the auditory system, hearing loss and hearing aids, introduces several aspects of the auditory spatial perception for normal-hearing and hearing-impaired listeners, and gives a review on speech understanding.

2.1 The Auditory System in a Nutshell

The auditory system is a chain of consecutive sound processing steps, comprising the outer ear, middle ear, inner ear and higher-level processing in the brain [Pickles, 1988]. The outer ear consists of the pinna and ear canal, where incoming sound is filtered depending on its incoming direction and frequency. The end of the ear canal is closed by a flexible membrane, called the tympanic membrane, that vibrates in tune with the incoming sounds. Fixed to the tympanic membrane on the other side of the ear canal is a little chain of three bones (ossicles), called the malleus, incus and stapes. The purpose of this chain of bones is to transform the vibrations of the tympanic membrane to vibrations on a smaller membrane called the oval window, which is the sound entrance to the inner ear, the liquid-filled cochlea. Since liquid is incompressible, a second membrane at the outside of the cochlea, called the round window, deflects in tune to the oval window deflection, allowing for the pressure wave to travel through the cochlea [Pickles, 1988]. Thus, sound vibrations in air are transformed into sound vibrations in liquid by the middle ear. The middle ear has the same air pressure as the environment allowing for optimal deflection of the membranes, which is achieved by a little tube connected to the oral cavity (eustachian tube). The cochlea is a spiral shaped bone, filled with liquid and separated into different chambers by membranes along the spiral. The most important membrane to mention here is the basilar membrane, which vibrates along its length depending on the frequency of the sound. Thus, high frequency sounds make the basilar membrane vibrate at its start, close to the oval window, while low frequency sounds travel further inside the cochlea and make the membrane vibrate close to its end, at the tip (apex) of the spirally shaped cochlea. This transformation of a vibrating oval window at all frequencies, into a vibration at different locations across the basilar membrane is called tonotopy and allows for a separation of sound waves into its frequency components, giving each frequency in the range of roughly 20 Hz to 20 kHz a specific location on the basilar membrane to vibrate (at that specific frequency!) [Pickles, 1988, Yost, 2001]. Fixed to the basilar membrane are cells that have elements that bend by the deflection of the membrane, and thereby incite firing of neural impulses along nerves that bundle into the auditory nerve and travel higher into the brain. These cells on the basilar membrane

are called hair cells. There are two types of hair cells, one row of inner hair cells and three rows of outer hair cells. It is the inner hair cells that incite neural spiking for sounds to be perceived by the brain, while the role of most of the outer hair cells is that of actively enhancing the deflection of the membrane by stretching and contracting in tune to the naturally occurring deflection, increasing the sensitivity [Yost, 2001]. Neural activity of the left and right auditory nerve is processed in the brain at higher levels in different regions, where differences in time, level and frequency between the left and right signals are compared, such that we can hear and understand our acoustic environment. Sound processing in the cochlea is done separately in logarithmically spaced and extended frequency bands, called critical bands ([Zwicker, 1961, Zwicker and Fastl, 2013]).

2.2 Monaural and Binaural Cues

The alterations of sounds by the physiognomy of the outer ear, and by reflections at head and torso before reaching the ear canal are called monaural cues. Monaural cues are different for each incoming direction of a sound, such that a sound wave arriving at a certain direction is filtered differently than from any other direction. These characteristic differences between different directions are learned by the brain, enabling a discrimination of whether a sound comes from the front, back, above or below. Binaural cues, on the other hand, are the characteristic differences between the right and left ear signals. A sound arriving from the left of a person will reach the left ear at a slightly earlier time than it reaches the right ear. Furthermore, the sound wave will be attenuated by the head, such that the sound wave will have a lower amplitude than at the left ear. These interaural time differences (ITDs) and interaural level differences (ILDs) are frequency dependent, since the size of the head as an obstacle between both ears will affect high frequencies (short wavelengths) more than low frequencies. Binaural cues are important for the localization of sounds on the horizontal plane, while the combination of monaural and binaural cues is crucial for a natural percept of space in the 3-dimensional space.

2.3 Hearing Loss

There are several different kinds of hearing loss. Hearing loss can have its origin at the outer ears (e.g. obstruction of the ear canals), at the middle ear due to a stiffening of the ossicles, disruption of the chain of bones, infections, loss of pressure compensation with the environment (e.g. obstruction of the eustachean tube) or due to liquid in the cavity. Also, problems at the auditory nerve, e.g. due to tumors or degenerative neural diseases can cause hearing loss at the brain level. But the most common type of hearing loss is sensorineural due to problems of the inner ear in the cochlea or auditory nerve fibres. The inner or outer hair cells can be damaged with time or due to very loud noise exposure. Since the hair cells are unable to regenerate, their loss means a reduction of signal generation to the brain. In the case of the inner hair cells which incite the neural impulse responses, not only the amount of information is reduced, leading to lower sensitivity, but also the place on the basilar membrane where the loss of inner

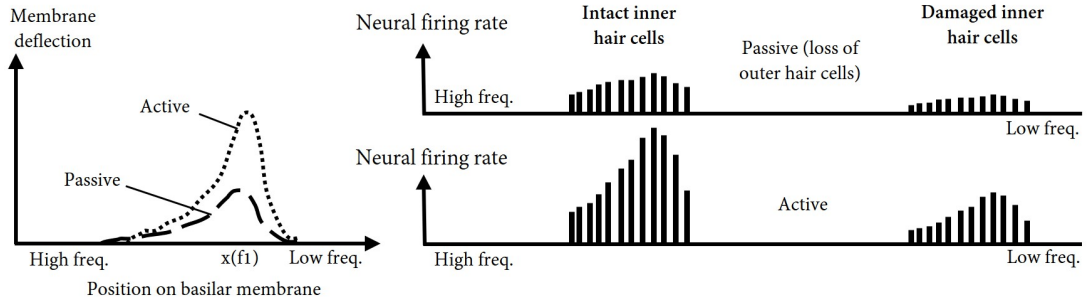

Figure 2.1
*Simplified schematic diagram of (**left**) the basilar membrane deflection of an exciting sound with frequency f1 when there is no active enhancement by outer hair cells (dashed line) and with active enhancement by outer hair cells (dotted line). The **middle** part shows exemplary neural firing rates for intact inner hair cells, the **right** part shows exemplary neural firing rates for damaged inner hair cells.*

hair cells appears varies between subjects, such that some frequency regions are more affected than others. With loss of outer hair cells, the active enhancement of the basilar membrane vibration is reduced, which leads to a reduced vibration amplitude (less firing by the inner hair cells). Also, reduced active enhancement at the corresponding location on the basilar membrane leads to a spatial extension of the basilar membrane vibration relative to its amplitude, which causes a broadening of the auditory filters, since inner hair cells at neighboring frequencies are excited as well, and the firing rate at the exciting frequency compared to the firing rate at neighboring frequencies of the sound decreases, thus also reducing the sensitivity (Fig. 2.1). While hearing loss cannot be treated, some of the drawbacks can be compensated for by hearing aids, or in extreme cases by cochlear implants.

2.4 Hearing Aids

Hearing aids are devices that help people with hearing impairment to make sounds more audible, enhancing those parts of a sound that became too soft to be heard without hearing aids. Specifically, hearing aids take an acoustic signal that arrives at a microphone, convert it to a digital signal by sampling and discretizing the sound wave, applying digital signal processing on the signal, and then converting that processed digital signal back to an acoustic signal, adapted for each hearing aid wearer's individual hearing loss [Dillon, 2001] (Fig. 2.2).

The basic steps of the digital signal processing comprise firstly a transformation of the signal into the frequency domain, splitting the signal into frequency bands either linearly spaced with equal bandwidth, or into logarithmically spaced frequency bands resembling the natural bands of the auditory system, of 1 bark width ([Zwicker, 1961]). For each band, further processing consists of an environmental analysis for an automatic adjustment of parameters, such as the recognition of the inside of a car, a quiet or noisy

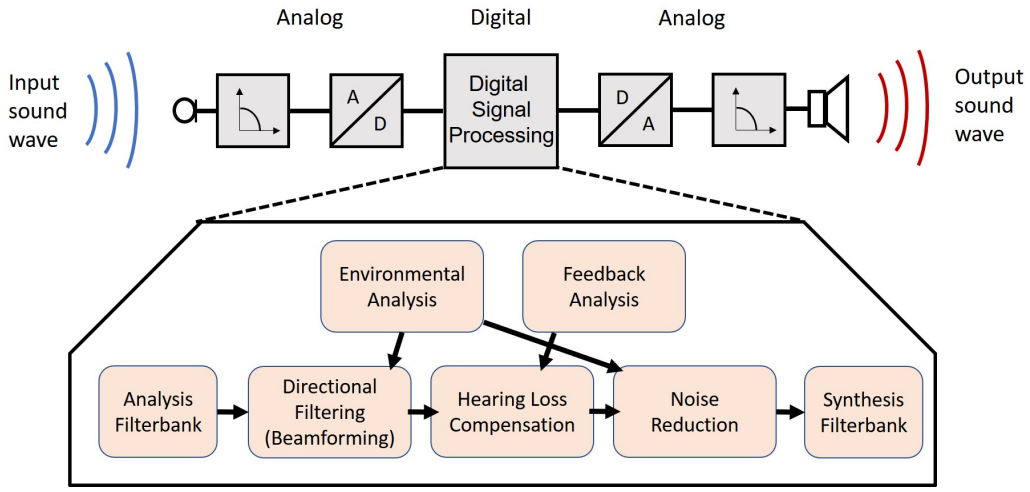

Figure 2.2
Schematic diagram of exemplary processing steps in a digital hearing aid. Firstly, an input sound wave is picked up by one or more microphones, the signal is then low-pass filtered and digitized. A variety of digital signal processing steps are conducted before the signal gets transformed into an analog signal and played back by the loudspeaker (receiver) of the hearing aid.

environment. Then, attenuation of noise sources from different directions than the front or the location of a dominant speaker is conducted with a method called beamforming, that subtracts two or more microphone signals for directional attenuation [Dillon, 2001]. There are two types of hearing aids that are mostly used. BTE hearing aids and ITE hearing aids, as shown in Fig. 2.3.

While hearing aids have become smaller and better through time, more and more focus has been placed on using digital signal processing algorithms for noise reduction, while maintaining the naturalness and accurate spatial perception of sounds has been given lower priority. Hearing aids quickly reach their limits in busy restaurants, bars, social gatherings, or other difficult acoustic scenarios consisting of multiple sound sources. In such situations, there is no stationary background noise that could be filtered out, no unique voice pattern towards which one could steer an automatic beam-former, nor do listeners even typically face a speaker directly - just to name but a few problems hearing aids and their users face in such situations. Normal-hearing listeners also face a hard time understanding what is said in such situations, but they succeed mainly because of the remarkable ability to focus their attention on a sound source and ignore competing ones as noise. This ability to focus auditory attention becomes worse for hearing-impaired listeners (without hearing aids) already due to the restrictions imposed by the hearing impairment itself. Specifically, with hearing loss, the amount of hair cells is reduced, the auditory filters in the inner ear therefore become broader and the sensitivity at individual frequency regions worsens. Also, the dynamic range decreases (recruitment), such that soft sounds are not audible, then there is a compressed region where hearing impaired subjects hear sounds well and the loudness increases rapidly with increasing sound level.

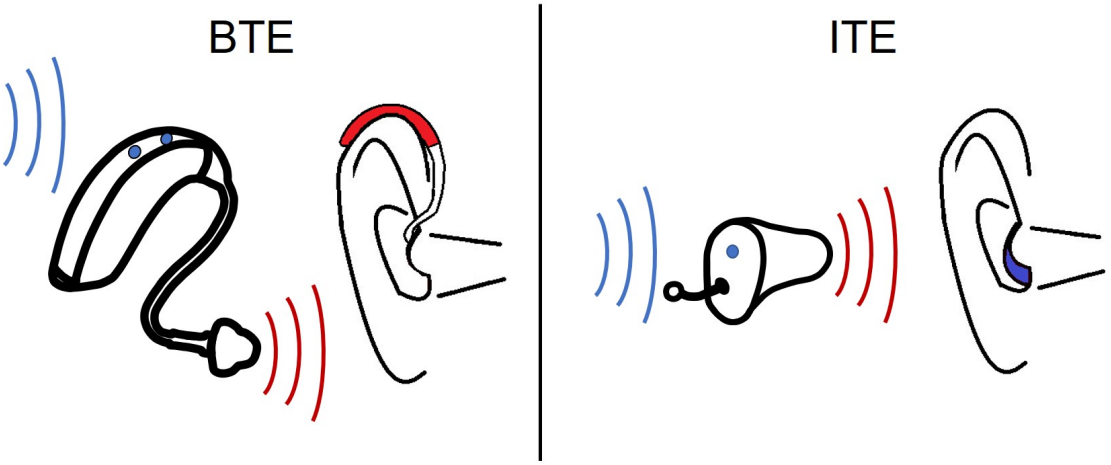

Figure 2.3
Schematic diagram of the two main types of hearing aids, behind-the-ear (BTE) and in-the-ear (ITE) hearing aids.

At higher sound pressure levels sounds become uncomfortably loud. With hearing aids, some of these detrimental effects of the hearing impairment can be compensated for, e.g. sounds can be made audible again, on the expense of other aspects, such as of the spatial perception and naturalness of sounds. Most hearing aids worn by the users are BTE devices. BTEs usually have two microphones per device, which allows for combining the microphone signals to perform directivity recordings (beamforming). However, a microphone pickup at a location behind the ear causes a loss of spectral information of the outer ear, information which is crucial for many aspects of spatial hearing. While hearing aids work well in quiet environments, it is often this sub-optimal microphone position that hinders the above-mentioned ability of the brain to focus on auditory objects. This is because sounds and reflections from all directions arrive more-or-less unfiltered to the microphones placed behind the ear. It is therefore naturally difficult for the brain to separate auditory objects into different streams ([Shinn-Cunningham, 2008]) from such a signal mixdown, leading to a significant degradation of the ability to focus auditory attention.

2.5 Speech Understanding

2.5.1 Factors influencing speech understanding

The main problem for people with hearing loss is usually that speech becomes harder to understand, especially in difficult acoustic conditions such as in restaurants or reverberant environments. This often leads to HI listeners avoiding such problematic situations, which are often social events. Thus, they often become more isolated and avoid social gatherings. NH people, on the other hand, understand talkers even in very difficult acoustic conditions. Part of the reason is that they can

selectively focus the attention on a specific talker. It is believed that perceiving sounds as different objects by their spectro-temporal congruency helps them separate the desired object from objects not considered interesting ([Bregman and Pinker, 1978], [Shinn-Cunningham, 2008], [Dau et al., 2009]). Second, these objects can be perceived at different locations in space, such that focusing on a desired sound is a combination of following a spectro-temporally congruent sound, and its unique spatial location ([Bregman, 1994]). When two or more sound sources are emitting sounds from different locations in space, rather than from a single position, masking effects between them are reduced ([Bronkhorst and Plomp, 1988]). This so called spatial release from masking is caused on the one hand by better-ear listening, resulting in an improvement of SNR at one ear for the target source due to head-shadow effects. On the other hand, binaural unmasking is involved, a noise suppression mechanism based on different interaural phase and level differences of the sources, which aids the brain to group or segregate these sources creating an attention-based SNR benefit ([Durlach, 1972], [Colburn et al., 2006]). The release from masking when sound sources are located at different angles ([Freyman et al., 2001], [Litovsky, 2012]) is bigger than for sources from differing distances (at the same angle) from the listener ([Westermann and Buchholz, 2013], [Chabot-Leclerc and Dau, 2014]), since in the latter case the ITD and ILD differences occur mainly due to different reflections of the sources. It has also been proven beneficial for speech intelligibility when listeners have previously been in the room before the listening task, possibly due to an adaptation or learning of the reflection patterns and acoustics of the room ([Brandewie and Zahorik, 2010]). These astounding properties of selective attention, sound object formation and higher-level processes like adaptation get deteriorated with hearing loss, with the broadening of auditory filters, inaudibility of high-frequency components of consonants which are important for speech understanding, and often correlated, age-related cognition problems. In addition, when using BTE hearing aids, important spectral cues by the filtering of the pinna get lost or disturbed, such that it becomes even harder to separate objects in space. These drawbacks can only partially be compensated by hearing aids. The hearing aid industry has tried to tackle these problems, finding ways to restore audibility, reduce background noise and improve the SNR with beamforming methods. Wiggins and Seeber (2013) showed that fast-acting compression in hearing aids can affect the speech intelligiblity in steady state noise conditions. Linked compression across left and right device improved the long-term apparent speech-to-noise ratio (SNR) at the ear with better SNR compared to unlinked compression [Wiggins and Seeber, 2013]. State of the art hearing aids can stream entire audio signals between a pair of hearing aids, improving beamforming even more than when using independent beamformers on each ear separately. While hearing aids are of great help in quiet environments, they are still far from restoring normal hearing's good speech understanding in reverberant, complex acoustic scenes or multitalker environments.

2.5.2 Assessment of speech understanding

There exist several different speech understanding tests to rate how well a HA user can understand speech with a given HA algorithm or without wearing HAs (unaided baseline). In Germany, the most commonly used speech tests are the *Marburger Satzverständnis Test*, the *Freiburger Einsilber Test*, the *Göttinger Satztest (GÖSA)* and the *Oldenburger Satztest (OLSA)*. The Marburg speech test contains ten lists, each of which consists of ten sentences. The words are chosen to mimic the average phoneme distribution of the German language ([Meier, 1964]). Since some of the sentences are incorrect in their syntax, this test is not ideal for testing speech understanding. The Freiburger test uses monosyllabic words that are presented in noise, where the participants must repeat the words they heard. The answer is supervised by the examiner and marked as either correct or incorrect. The words used for the test are not restricted, i.e. it is an open set of words with several lists. While this test is commonly used when fitting hearing aids and for the assessment of the degree of hearing loss, being a short duration test, it is criticized since some of the lists are much easier than others and the test re-test results show large variance, such that a comparison of results or studies can be difficult. Also, manipulation of results is easy, e.g. for showing the benefit of a specific algorithm over a different one, the exact selection of a list can have a big impact on results. By increasing the duration of the test when using multiple lists, a more accurate result can be achieved, yet often there is not enough time at the hearing aid dispensers or audiologists to perform longer tests. The GÖSA is a closed sentence test with 20 lists of 10 sentences each. Due to the correct semantics and meaningful sentences, they are easily remembered and a list should not be used twice in a subject. The OLSA (which is used in this work) uses a closed set matrix of words that are combined to form 5-word sentences. These are semantically and syntactically correct while often meaningless, and thus context does not bias the results. The structure of each sentence is name – verb – number – adjective – object. For each of the names, verbs, numbers, adjectives or objects there are 10 specific possibilities from which a sentence can be built, as seen in Table 2.1.

This test was originally proposed by Hagerman in 1982 [Hagerman, 1982] in Sweden, and was adapted to German by Wagener ([Wagener et al., 1999]). It is nowadays available in many different languages with the speaker either male or female. The noise used as a masker sound is speech shaped noise (SSN), composed of a superposition of several randomly concatenated OLSA sentences, thus having the same long-time spectrum as the speech. Usually, the masker is presented from 90° with respect to the target speech at 0°. The level of the target and masker start at 65 dB SPL. The level of the target is adapted relative to the amount of correct or incorrect words recognized from the sentence, with a decrease in target level when the sentence was correctly repeated, and an increase in level when mistakes were made. Each of the 20 lists consists of 30 sentences. The 50 percent word recognition threshold is taken as the average signal-to-masker level ratio of the last 20 sentences of a list. Although this test is well suited for speech intelligibility assessment of hearing impaired subjects, it requires much more time than other tests, as a preliminary training session is required since there is an increase in

Table 2.1
Base list for the German OLSA matrix speech test.

Name	Verb	Number	Adjective	Object
Peter	bekommt	drei	große	Blumen
Kerstin	sieht	neun	kleine	Tassen
Tanja	kauft	sieben	alte	Autos
Ulrich	gibt	acht	nasse	Bilder
Britta	schenkt	vier	schwere	Dosen
Wolfgang	verleiht	fünf	grüne	Sessel
Stefan	hat	zwei	teure	Messer
Thomas	gewann	achtzehn	schöne	Schuhe
Doris	nahm	zwölf	rote	Steine
Nina	malt	elf	weiße	Ringe

performance during the first two test lists while the test subject gets familiarized with the test and words used within [Wagener et al., 1999]. Additionally, people in real-life situations usually spend most of the time in positive SNRs and only rarely in extreme negative SNRs less than -10 dB, at which the Speech Reception Thresholds (SRTs) are usually obtained in the OLSA or similar speech tests. Thus, one should be careful to analyze the behaviour of noise reduction algorithms using the OLSA since the operating point is not necessarily one encountered in everyday situations, and most noise reduction algorithms depend on a given ratio of noise floor to signal peaks. On the other hand, directional noise reduction techniques like beamforming are well suited to be tested in this test, since a directional benefit can be well examined and comparison between different directional conditions made.

2.6 Spatial Hearing

The human brain learns from early childhood to discriminate and locate sound sources in space and to associate the perceived auditory images with inputs from other senses, such as vision. Having two ears gives rise to two different acoustic signals of slightly delayed pathways and different pressure levels that arrive at our ears. Arriving sounds are also altered by the outer ears and torso depending on the incoming direction. These differences in the sound signals between both ears help us to have an accurate spatial localization ability regardless of a sound's incoming direction, while the visual field is restricted to the front. Spatial hearing comprises different perception dimensions, including localization, elevation, distance perception, externalization, apparent source width and diffuseness. The remainder of this chapter explains the sound playback system used in this work and gives a literature overview on these aforementioned dimensions of

spatial hearing. It also summarizes how spatial hearing is affected by hearing aids, and introduces the term *spatial sound quality*.

2.7 Spatial Sound Presentation

Scientific research on the perception of sounds needs a controlled environment with reproducible stimulus playback. Research about how we perceive realistic sound scenes in a laboratory usually involves a playback system that can accurately play back reflections to deliver spatial sounds. There are different ways for playing back spatial sounds, such as wave field synthesis or ambisonics for loudspeaker reproduction, or binaural synthesis for headphone reproduction. While these methods deliver realistically sounding acoustic scenes, they have some drawbacks concerning the accuracy of the sound field. Wave field synthesis and ambisonics are prone to spatial aliasing at high frequencies or coloration from combfilters by coherent wave reproduction. For headphone reproduction non-individual HRTFs, unnatural head motion relative to the simulated sound field and non-congruent auditory and visual cues are the main problems. When studying reflections and reflection patterns, it is of great importance to accurately reproduce individual reflections in amplitude and phase over a wide frequency range. The present work makes use of the Simulated Open Field Enviroment (SOFE v3, [Seeber et al., 2010]) due to its ability to accurately simulate and reproduce a sound field in amplitude and phase using an arbitrary calibrated loudspeaker configuration surrounding the listener in an ideally anechoic room. The SOFE is a method and apparatus to calculate, simulate and auralize spatial sound fields based on the image source model, which enables to place source and listener positions in a virtual space of arbitrary geometry ([Borish, 1984]). Reflections are simulated by mirroring the virtual room on each wall manifold times (fig. 2.4) and tracing the virtual sources of each mirrored room to the listener position in the original virtual room, taking into consideration absorption coefficients of surface materials and the air, and phase changes at wall encounters. The SOFE calculates the room impulse response and uses a backtracking algorithm ([Vorländer and Summers, 2008]) to remove hidden or invalid reflections from mirrored rooms. Each reflection arriving at the listener is played back by a single loudspeaker. A room impulse response for each loudspeaker in the given configuration is calculated and used for convolution of the stimulus for playback.

2.8 Localization

Acoustic localization of sounds is usually an easy and effortless task for normal-hearing (NH) listeners. They can accurately determine the horizontal azimuth angle of a source, being able to tell whether the sound comes from the front or the back. The main localization cues are the differences in level (ILDs) and in time (ITDs) between the sounds reaching the two ears ([Thompson, 1882], [Rayleigh, 1907], [Blauert, 1997]). Humans are even able to distinguish between sound sources that are separated by only 1° when sounds are presented in the front of them. For sound sources on the sides,

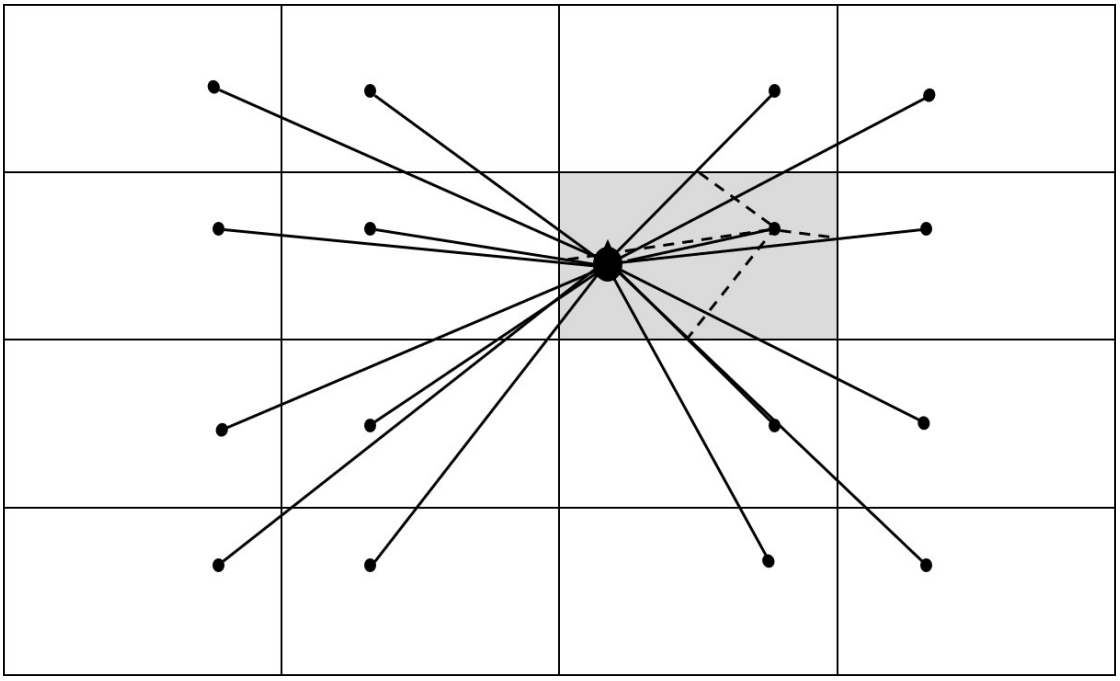

Figure 2.4
Schematic diagram of the image source model principle for a virtual room and some adjacent mirrored rooms. Black dots represent mirrored positions of the original source.

these so-called just noticeable differences (JNDs) increase up to 8° ([Akeroyd, 2014]). Sound localization is crucial for spatial orientation ([Noble et al., 1998]) and thus for evaluating one's surroundings. Obstruction of sound by one's head and the spatial separation between the ears leads to interaural level (ILD) and time differences (ITD). These binaural cues are responsible for sound localization in the horizontal plane and, roughly speaking, ITDs are responsible for localization in the lower frequencies up to 1.5 kHz, while ILDs are more important at higher frequencies above 1.5 kHz with an intermediate region where both cues contribute ([Thompson, 1882], [Rayleigh, 1907]). While ITDs usually dominate sound localization in quiet environments, Lorenzi et al., (1999) [Lorenzi et al., 1999] found that this dominance of ITDs does not apply in noisy situations where subjects were found to rely more on ILDs for accurate localization. On the other hand, monaural cues are determined by the shape of the pinna and reflections at the head and torso that induce spectral notches to the incoming sound depending on the direction of sound's incidence. These monaural cues are needed for vertical localization of the sound source (elevation) as well as the distinction between front and back ([Howard and Angus, 2017]). Dynamic cues from head movements are additionally beneficial for localizing sounds [Wallach, 1940]. Numerous studies on localization in the horizontal and vertical plane have shown the impact of binaural and monaural cues in anechoic and reverberant environments ([Good and Gilkey, 1996], [Hartmann, 1983], [Middlebrooks et al., 1989], [Makous and Middlebrooks, 1990],

[Musicant and Butler, 1984], [Zhang and Hartmann, 2010]) However, there are additional influences on sound localization, such as previous knowledge on the environment and stimulus, the ventriloquist effect or psychoacoustic effects such as loudness or suppression of concurrent disturbers ([Blauert, 1997]). Localization is an especially difficult task for hearing-impaired (HI) people ([Noble et al., 1998]). Both high-frequency ILDs and low-frequency ITDs are important for localization, and both of those cues are affected by hearing aids. Yet, acclimatization plays a role as people tend to localize better with the devices they are accustomed to than with other devices, and subjects can adapt with time to distorted localization cues ([Hofman et al., 1998]). The effect of hearing loss on localization differs for vertical and horizontal localization depending on the degree and type of hearing loss, and on which frequencies are affected. Unilateral hearing impairement deteriorates localization cues [Humes et al., 1980] and unilateral amplification is worse than bilateral amplification in terms of localization accuracy [Köbler and Rosenhall, 2002]. Localization is more severely affected in subjects with conductive- or mixed hearing loss than in subjects with sensorineural hearing loss because of the effect of bone conducted sound (if perceivable) altering the time differences in the case of conductive hearing loss [Noble et al., 1994]. Both the individual hearing loss and the used hearing aid affect the spatial sound perception. The main acoustic difference between BTEs and ITEs for omni-directional signals stems from the different positioning of the microphones, resulting in spectral notches originating from the pinna reflections mostly missing in BTE signals. Since these spectral notches play a crucial part in sound localization, some experiments have shown poorer performance for BTE fitted participants. [Keidser et al., 2006], [Van den Bogaert et al., 2006] and [Noble and Byrne, 1990] conducted localization studies with HA, showing a detrimental effect of HA on spatial perception.

For noise stimuli, the human binaural auditory system works as a level meter that considers the average level at each ear to assess ILDs ([Hartmann and Constan, 2002]). Instantaneously looking into the temporal fine structure would require much faster integration constants in our auditory system. A meter model based on loudness rather than level accounted for bandwidth dependencies of ILD thresholds in Hartmann's and Constan's experiments [Hartmann and Constan, 2002] that further showed that binaural coherence has an almost negligible effect on the use of ILDs. Coherent signals at both ears are rare in real life situations and are mostly limited to brief moments of time when the sound reaches the ears of the listener for the first time via the direct path. In reverberant environments, directional sensitivity is better during the (modulation) onsets of a stimulus, since auditory localization is mostly encoded by an integration of rate responses of subcortical auditory neurons over time that deliver robust estimates of source positions due to onset dominance firing ([Devore et al., 2009]). This coherence-based localization, at onsets of a stimulus, is also the basis of some binaural auditory models (e.g. [Faller and Merimaa, 2004]). As reverberation increases, the coherence between the ears decreases and also the modulation depth of the envelope and the slope of the modulation flancs is reduced. As such, the sensitivity of the auditory system to ITD cues in reverberant environments is reduced with significantly increased ITD discrimination thresholds both for low frequency fine structure ITDs and

high frequency envelope ITDs [Monaghan et al., 2013], making localization less accurate and more dependant on ILD cues. Previous studies have also found that in addition to the binaural cues, monaural spectral cues can influence horizontal localization of sounds ([Musicant and Butler, 1984], [Slattery and Middlebrooks, 1994]). The auditory perception of a sound source's origin can be described by spatial attributes such as the angle in azimuth and elevation relative to the head's coordinate system, and a specific distance from the listener. While this perception is often unambiguous in everyday life for NH listeners, the sound's location can become less distinct for HI listeners, especially when they are wearing hearing aids. Moreover, localization abilities of HA users can be affected in various ways due to the signal processing in the devices that is designed to restore audibility and reduce noise. Especially when the hearing aids are not linked between the ears the spatial perception of sounds is influenced due to binaural cue alterations [Akeroyd, 2014], [Gomez and Seeber, 2015b]. When fast-acting compression is used, spatial perception, including localization, can suffer severely ([Wiggins and Seeber, 2011], [Wiggins and Seeber, 2012]. Wiggins and Seeber (2011) reported a strong influence of dynamic range compression on the lateralization of stimuli. While fast-acting compression shifted the sound image of sounds with abrupt onsets and offsets towards the center, sounds containing gradual onsets and offsets (such as speech) were perceived more lateralized, moving, broader or even as a split image when high frequency envelope ITDs and ILDs drifted appart due to the compression. Recently, Akeroyd and Whitmer (2016) presented an overview on localization with hearing aids, summarizing studies on localization with bilateral hearing aids [Akeroyd and Whitmer, 2016]. There they stated that, in general, previous studies have not shown large within-subject differences in azimuthal localization in the front between unaided and aided test conditions - RMS (root-mean-square) localization error being only about 1° worse in the aided condition. Only very few studies have investigated localization with directional microphones ([Keidser et al., 2009], [Van den Bogaert et al., 2006], [Picou et al., 2014]), and among those, only Keidser et al. (2009) [Keidser et al., 2009] tested both horizontal and front-back localization with directional hearing aids. They found that HI listeners took more advantage of ITD cues than of ILD cues for horizontal localization and could utilize spectral cues below 2 kHz for front-back localization. When directional microphones were applied in the hearing aids, localization became significantly poorer in the left-right dimension, but front-back discrimination improved slightly. For real-life situations, where head movements are usual, head motion is (usually) beneficial for localization ([Perrett and Noble, 1997]). Yet, directional microphones can potentially lead to large errors when localizing sounds that are not in the visible field of vision because head movements can result in confusing orientation-dependent changes in the signal-to-noise ratio, contradictory to natural changes in levels of the ear canal signals due to head turns ([Brimijoin et al., 2010]).

2.9 Confusions

When using hearing aids to compensate for hearing loss, front-back confusions often occur when the head is held still. This is caused by equality of binaural cues within a cone of confusion, resulting in hearing aid wearers confusing sounds coming from the front as coming from the back and vice versa. This phenomenon is observed especially when wearing BTE devices that have their microphones located behind the ear and do not therefore convey the natural spectral information from the pinna, information which normally helps to resolve these types of confusions ([Wallach, 1940, Wenzel et al., 1993, Perrett and Noble, 1997, Wightman and Kistler, 1997, Brimijoin and Akeroyd, 2012]). One possible solution to overcome often occurring front-back confusions is the use of directional microphones ([Noble et al., 1998]). Directionality can be achieved by multi-channel recording with microphone arrays ([Bader, 2014]). This method called beam-forming (BF) is implemented in HAs ([Mueller et al., 2010, Korhonen et al., 2015]) and is used to improve the signal-to-noise ratio (SNR) ([Zhang and Hartmann, 2010]) or modify spectral cues in a direction dependant manner, and thereby enable better distinction between front and back. The number of front-back confusions increases significantly for HAs with closed fitting, especially for microphone positions that do not enable capturing of spectral pinna cues. Such sub-optimal configurations include BTE devices employing beam-formers. In general, directional microphones show about 3°-larger RMS left-right localization errors compared to the unaided setting, and even 10°-larger RMS errors compared to the omnidirectional mode for high-frequency stimuli ([Akeroyd and Whitmer, 2016]). A few studies have tested localization both in the front and back ([Keidser et al., 2009, Van den Bogaert et al., 2011, Byrne and Noble, 1998, Noble et al., 1998]) comparing the unaided and aided case. These studies differed substantially, mainly due to differences in the hearing aid types, open or closed fitting, and whether they were linked or unlinked. In general, front-back localization errors were about 2.5 times higher than for left right azimuthal localization ([Akeroyd and Whitmer, 2016]).

2.10 Distance Perception

Auditory distance perception is important in everyday spatial navigation tasks. While auditory localization is crucial for determining the direction of an auditory source, distance estimation is also important for optimal communication and the judgement of importance or danger of a sound emitting object. Sounds of warning like a car or motorcycle honking will draw more or less attention, depending on the perceived distance of the sound. People might react faster to sounds related to danger when they judge sounds being closer. Another advantage of being able to discriminate different distances takes place when a listener is placed in a complex sound field with multiple talkers. While there is a benefit in speech understanding from spatial release from masking when the sound sources are spatially separated from each other in azimuthal direction ([Freyman et al., 2001, Litovsky, 2012]), there is also a benefit

from differing them in distance from the listener ([Westermann and Buchholz, 2013, Chabot-Leclerc and Dau, 2014]). Egocentric auditory distance perception has been studied extensively in the past decades. Loomis et al. (1998) [Loomis et al., 1998] measured distance responses in open field with participants walking to the perceived location. Mershon and Hutson (1991) [Mershon and Hutson, 1991, Mershon, 1997] used geometric measures to indirectly determine the perceived sound distance by measuring different parameters like the participant's horizontal displacement and the angles at which they pointed. While distance perception has been measured in real rooms ([Mershon and Bowers, 1979, Mershon et al., 1989, Calcagno et al., 2012]), virtual acoustics has been used as well using either head-related transfer functions (HRTFs) or with individually-measured binaural room-impulse responses (BRIRs), presenting the stimuli over headphones either in a sound booth ([Bronkhorst and Houtgast, 1999, Zahorik, 2002, Cubick et al., 2015]), or in the same room the BRIRs were measured in ([Cubick et al., 2014]). Virtual acoustics has also been used in relation with distance perception experiments, using loudspeakers to auralize a simulated room ([Akeroyd et al., 2007, Gomez and Seeber, 2015a]). The influence of light sources on auditory distance perception was studied by Min and Mershon (2005) [Min and Mershon, 2005] and Calcagno et al. (2012) [Calcagno et al., 2012], while distance perception with blind participants was tested by Kolarik et al. (2013) [Kolarik et al., 2013]. Further information about the findings on distance perception can be found, e.g. in ([Zahorik et al., 2005]). Previous studies on auditory distance perception differ greatly in terms of methodology, response measures, room size, range of tested distances, presentation angles and stimulus choice. Common results suggest an overestimation of perceived distances for closer sound sources, and an underestimation for distances farther away. Inaccuracies for both close and far distances can be approximated by a compressive power function of the form $\hat{\delta} = k \cdot \delta^{\alpha}$ where $\hat{\delta}$ denotes the perceived distance, δ is the presented distance, α is a power exponent and k is a constant ([Zahorik et al., 2005]). Perceptual distance experiments may show either boundary, regression or range effects ([Gomez and Seeber, 2015a]), which are also present in visual psychophysical magnitude estimation studies ([Petzschner and Glasauer, 2011, Petzschner et al., 2015]). In addition, auditory distance perception in the back has been found to differ from the one in the front of the listener ([Gomez and Seeber, 2015a]).

2.11 Externalization

Externalization is the perception of a sound as originating from the outside world, contrary to internalized sound perceptions that are located inside the listener's head. The latter happens typically when listening to music over headphones. Localization is also defined differently for internalized sounds as azimuthal angles are mapped onto a lateral axis between the ears and distance is often non-existent as it cannot be perceived/judged. According to Hartmann and Wittenberg (1996) [Hartmann and Wittenberg, 1996], sound externalization is a perceptual continuum between fully internalized and fully externalized, the latter meaning that the sound source is perceived to be at the actual

position of the source. Externalization is thus closely related to distance perception, since distance can be judged for any externalized sound. There are several reasons why a sound can be perceived less externalized or even fully internalized. When listening to sounds over headphones, the main influence behind internalization percepts is believed to be the lack of information about the sound path from an external source to the ears, which is characterized by the head-related transfer function. HRTFs can be used to externalize sounds presented over headphones, and include spectro-temporal information about reflections at the head, torso and outer ears. This spectro-temporal information is direction dependent. Externalization is improved further using HRTF filtering (binaural synthesis) when the acoustic reflections from the walls and objects in a room are also simulated using a room simulator and corresponding HRTFs used for the individual reflections. When such HRTF-filtered information is missing or when the applied HRTFs deviate from the given individual's own HRTFs, sounds can be perceived less externalized or fully internalized when using headphones. It is still very difficult to achieve the same degree of externalization perception with binaural synthesis than perceived in natural real-life listening situations. The main reasons are auditory incongruencies between the recording and playback room, i.e. the acoustic signal played back does not match the room's acoustical cues where the listener is physically located ([Gil-Carvajal et al., 2016]), and non-natural movements of the acoustic signals relative to a listener's own head movements, i.e. the expected change in the sound does not match with the perceived one when turning the head ([Brimijoin and Akeroyd, 2012, Brimijoin and Akeroyd, 2014, Brimijoin and Akeroyd, 2016, Brimijoin et al., 2013]). There is increasing research on externalization, and how hearing impairment and hearing aids affect externalization. Ohl et al. (2010) [Ohl et al., 2010] conducted a study where the sensitivity to externalization cues was investigated by representing sounds over headphones to both NH and HI listeners. Individual BRIRs were measured in the test room using loudspeakers, and test signals were convolved with these BRIRs. Such a binaural synthesis resulted in a perception that the signals were actually coming from the loudspeakers. Then, two psychoacoustic experiments were conducted with a 2-alternative forced-choice method, presenting two speech signals of which one had to be rated as being more likely to originate from the loudspeaker than the other. The first experiment assessed how many externalization cues were needed for a change from a complete internalization (unprocessed speech signal presented over headphone, being the reference for internalization). The gradual changes were implemented by mixing the BRIR-processed signal with the un-processed signal in discrete mixing ratios. The second experiment tested the opposite, by mixing the un-processed signal with the completely externalized BRIR-processed signal, and used the latter as the reference signal to be compared with. The results showed a difference of about 10% between NH and HI subjects in the externalized/internalized ratio that they needed to reach a point at which they could not perceive any further change in either internalization or externalization. NH subjects were not able to perceive changes in internalization or externalization when the mixing ratio between processed and unprocessed signals was below 19% or above 79%, respectively. HI subjects were considerably less sensitive as their corresponding limits were 31% and 69%. Catic et al. ([Catic et al., 2013, Catic et al., 2015]) studied to what extend interaural cues from

room reverberation affected externalization. The spectro-temporal behavior of room reverberation on the ILD cues was investigated by analyzing dummy-head recordings of speech from different distances. Their analysis showed a reduction of ILD fluctuations with decreasing distance to the sound source. Afterwards, the effect of ILD fluctuations on externalization was investigated in a psychoacoustical experiment with NH subjects wearing headphones. A sound was simulated to come from a distant source in the room using binaural synthesis employing individual binaural impulse responses. In the experiment, ILDs were modified by restricting the naturally occurring ILD variation. The ILD fluctuations were restricted separately for different frequency bands using a Gammatone filterbank for the band-pass filtering and reconstructing the stimulus following the ILD-variation restrictions. The results showed that such restriction leads to reduced externalization. The authors therefore concluded that an alteration of naturally occurring ILDs by hearing-aid algorithms, such as by compression or noise reduction, could reduce the externalization of sounds for hearing aid users. Boyd et al. (2012) [Boyd et al., 2012] investigated the ability to externalize impulsive signals, using one distracting talker, played over a loudspeaker from the back of the test subject at a distance of two meters. Ten HI and three NH participants were asked to describe their hearing impression using a discrete externalization scale. The impulsive signals to be recognized were played back from four different positions (0°,30°, 60°, 90°) at 3 meters distance at random time intervals from 2-5 seconds. The results of their study show that externalization decreases drastically for NH subjects from the nominal unaided condition when the subjects wear hearing aids, as impulsive sounds were always perceived to locate either somewhere in the room or at the loudspeakers in the unaided condition. For HI subjects, externalization did not change when listening with or without hearing-aids, with the perceived distance corresponding always to some location in the room, but never at the loudspeakers. The results also show a consistent decrease of externalization with the angle of origin, 0° being the worst for externalization (perceived inside the head) and 60° and 90° being the best for externalization. This effect, which was also noticed by Ohl et al. (2010) [Ohl et al., 2010], shows that larger ILDs and ITDs, which are maximal for directions at the sides, have a positive influence on the degree of externalization. While externalization is a naturally occurring sensation for normal-hearing listeners, the above- mentioned studies have shown that externalization of sounds decreases with hearing impairment. That is, sound sources are no longer perceived at their actual location but closer or even internalized within the head due to the hearing impairment and use of hearing aids ([Boyd et al., 2012, Ohl et al., 2010, Catic et al., 2013]). The reason is believed to be the combined effect of reduced bandwidth, broadening of auditory filters, lack of pinna cues and a decrease of naturally occurring interaural level difference (ILD) fluctuations in reverberant environments. However, it should be noted that most of these studies have used headphone reproduction, which itself, as stated above, can lead to deteriorated spatial perception due to lack of head-movement compensation, and thus to an increased internalization of sounds ([Catic et al., 2013]). Ideally, externalization should be studied using listener's own ears and real-time test methods in realistic environments.

2.12 Apparent Source Width and Diffuseness

One additional important aspect on the spatial perception of sounds is the extension in space which a sound takes on. The apparent source width (ASW) of a sound can vary significantly between a very point-like to extremely wide impression, even in natural every-day listening situations for normal hearing listeners. The degree of ASW varies depending on multiple reasons. On one hand, reflections from walls and objects in a room that reach the ears are delayed and arrive in a certain time range with varying energy. With varying direct-to-reverberant ratio (DRR) of sounds, the ASW can also be affected. For large DRRs, a source can be mostly perceived as focused and accurately located in the auditory space. For small DRRs on the other hand, where reflection energy dominates over the direct sound, the sound image will more likely be perceived as broader since energetically dominant reflections arrive from different directions at the two ears. When a sound image is no longer accurately locatable in space and its extension uncertain, one often speaks of diffuseness, since sounds in a diffuse sound field, i.e. emanating from all directions, are also undefinable in their extension and exact location. The definition of a diffuse sound perception is difficult to describe and agree on. Diffuseness is a sense of uncertainty of the location and extension of a sound source, and for the present work, it will be related to ASW and localization in that for increasing ASW, the diffuseness also increases and localization is reduced, rather than being a binary percept of either completely diffuse or localizable. ASW and diffuseness can occur when ILDs and ITDs are affected either by the listening environment or by alteration of the sound signals at the ears, e.g. by a blocked ear or by wearing hearing aids. Whitmer et al. (2014) [Whitmer et al., 2012] report a diffuse and wide perception of sounds in older hearing impaired. Wiggins and Seeber (2012) [Wiggins and Seeber, 2012] reported influences of unlinked dynamic-range compression on spatial aspects for NH, leading to increased diffuseness, movement perception, image splits and reduced externalization. Gomez et al. (2016) [Gomez et al., 2016] also reported influences of the hearing aid devices on spatial perception in NH. They found that hearing aids increase the perceived diffuseness and apparent source width, reduce source separability and lead to a significant reduction in externalization perception for static heads.

2.13 Spatial Perception with Hearing Aids

While most studies on spatial perception with hearing aids were conducted with hearing impaired participants and commercial devices, there is little known about the effect that hearing aids have on spatial perception when disregarding hearing loss compensation. The reduction of spectral information due to a microphone position behind the ear is called the microphone location effect (MLE), an undesired drawback of BTE hearing aids ([Jensen et al., 2013, Gomez and Seeber, 2015b]). The assessment of the spatial perception of sounds with hearing-aids is very important for understanding the problems that millions of hearing aid users face every day when using their hearing aid devices. Usually, the performance of hearing aids is measured by how much the devices and their

algorithms help to better understand speech in noise, which is perhaps the most important contribution of hearing aids for helping the hearing impaired (HI) to communicate in their day to day lives. But there is more to hearing than understanding speech. The perception of where sound sources are located in space, how wide or narrow the sounds are perceived, and whether the visual and auditory information is congruent are very important aspects of hearing.

The NH have learned over the years to exploit monaural and binaural cues from their individual anthropomorphic characteristics. Thus, any information reduction of the received sounds due to hearing-aids has a detrimental effect on their natural perception of the environment. Among these detrimental influences are the different microphone position picking up sound behind the ear rather than at the ear-drum, a reduced bandwidth, internal noise, non-linear frequency responses of the components, occlusion effects to name but a few. How dynamic-range compression affects the lateral perception of sounds was investigated by Wiggins and Seeber (2011) [Wiggins and Seeber, 2011]. While HI may not notice some of the influences that NH do, such as the internal noise or reduced bandwidth, they experience many additional drawbacks from the combination of their individual hearing-loss and the hearing aid device limits on spatial perception. For NH, one must keep in mind the normal unaided case against which they rate spatial quality measures. On the other hand, HI using hearing-aids get accustomed to their devices, such that a spatial quality rating will always be compared to their own devices' sound. Since their overall spatial perception is already lower than that of NH, much clearer differences between hearing-aid conditions and influences of the devices on spatial perception can be expected by collecting NH data for spatial quality ratings. This allows to separate the effects of the hearing aid devices from the effects of hearing impairment on the quality of spatial perception.

2.14 Spatial Sound Quality

Experiments on localization in the horizontal and vertical plane, distance perception and apparent source width have been conducted for NH and HI for the past years ([Akeroyd, 2014, Keidser et al., 2009, Wiggins and Seeber, 2012]). There is yet a different way of testing how sounds are perceived in space, by assessing subjective measures to rate the spatial sound quality of sounds. When examining spatial sound quality, firstly considering sound quality per se seems appropriate. According to Fastl (2002) [Fastl, 2002], sound quality is a magnitude to describe relationships between characteristics of a sound stimulus in the physical domain and subjective impressions of that stimulus. One goal of sound quality research is thus to determine the perceptual impressions of the sound of a product and optimize that sound to be preferred by most customers. A prominent example hereof is the sound design of a car engine. Spatial sound quality, on the other hand, uses the methods of sound quality testing to investigate perceptual spatial impressions of sounds. A method widely used in psychology studies to acquire proper spatial sound quality ratings is the semantic differential. This method uses pairs of adjectives, related but opposite to each other, where participants rate whether the

measured aspect is closer to one or the other adjective from the pair. This method uses a seven-value rating scale, which is, depending on the application, either labelled between [-3; 3], from [0 – 7] or as in the present work, using only markers without any value labels. With the help of carefully selected descriptor adjective pairs, one can measure the highly individual judgements of the quality of spatial dimensions such as externalization, diffuseness, source separation, width and locatability. The rating can be discrete, at the given markers between the adjectives of a pair, or in a continuous manner as used in the present work. The Spatial Audio Quality Inventory (SAQI, [Lindau et al., 2014]) gives a good overview of perceptual quality descriptors that cover diverse dimensions of the acoustics and spatial perception of acoustical environments, such as timbre, tonalness, spatial geometry, room aspects, the dynamics and time behaviour of stimuli, artifacts and other general descriptors thereof. Colsman et al. (2016) [Colsman et al., 2016] published a questionnaire to assess the spatial perception of 3D audio reproduction systems. Important work in the investigation of spatial sound quality with hearing aids was done by Wiggins and Seeber [Wiggins and Seeber, 2011, Wiggins and Seeber, 2012]. They investigated how different settings of compression in hearing aids affect the spatial perception of stimuli for normal hearing subjects. They found that fast-acting compression, when acting independently in each ear, deteriorates spatial attributes, increasing diffuseness, movement, image split and internalization, especially for sounds with gradual on- and offsets like speech. Some authors have since published their investigations on how bilateral hearing aids affect the spatial perception of sounds, especially when compression alters binaural cues (e.g. [Ernst et al., 2013, Schwartz and Shinn-Cunningham, 2013, Hassager et al., 2017a, Hassager et al., 2017b]).

3 Spatial Perception with Hearing Aids in Reverberation

3.1 Summary

This chapter deals with the investigation of how sounds are perceived spatially in a reverberated environment when using hearing aids. It summarizes the methods, results and outcomes of a follow up study of Gomez and Seeber (2015) [Gomez and Seeber, 2015a] on spatial sound perception with hearing-aids and a static head position, where differences between three aided conditions were examined and compared to an unaided baseline. The three aided conditions were BTE, ITE and static beamformer (BF) signals, computed on a real-time Simulink model that processed the signals from custom made, individualized hearing-aid prototypes. While most previous studies on spatial perception with hearing aids were conducted with hearing impaired participants and commercial devices, there is little known about the effect that hearing aids have on spatial perception when disregarding hearing loss compensation. Consequently, spatial sound perception with different kinds of HA conditions was tested for normal hearing participants using HA prototypes with linear gain, to examine the device's influence on spatial perception separately from hearing loss or hearing loss compensation algorithms [Hoening, 2016]. We presented reverberated speech material at distances ranging from 0.75 m to 9 m, in the front and back, to eight normal hearing listeners using virtual acoustics in the Simulated Open Field Environment (SOFE v3, [Seeber et al., 2010]). We asked our participants to rate the perceived distance, azimuth, apparent source width (ASW), elevation and internalization using an intuitive GUI on a touchscreen. Results show a strong influence of microphone directivity, i.e. relative level differences between the front and back, on distance and elevation perception. In all aided conditions, a strong azimuthal lateral shift was observed for sounds from the front, but not from the back. In the BF condition, ASW and internalization were worse than in the ITE and BTE conditions. For front-back confusions, BF and BTE conditions performed equally bad, with most sounds from the front perceived as coming from the back. Overall, we observed a deterioration of spatial sound perception with hearing-aids compared to the unaided baseline, with the BF condition having the biggest negative effect on spatial quality.

3.2 Methods

3.2.1 Virtual Room

The room simulated for this experiment was a large rectangular shaped room of dimensions 15.5 m x 18.5 m x 10 m, chosen in ratio as suggested by Cox et al. (2004) [Cox et al., 2004], with an average reverberation time of 1.4 seconds. The virtual listener was positioned at [5 m, 10 m, 1.3 m] in a non-symmetrical position, for which we verified that it was not located at a room mode maximum in a range up to 1 kHz using room mode simulations. The critical distance of the room after Sabine is 2.6 m, for which the direct-to-reverberant ratio (DRR) is 1. We chose positions in the room at 0.75 m, 1.5 m, 3 m, 6 m and 9 m distance from the virtual listener at 30° in the front and 150° in the back for which we calculated the room impulse responses using an image source model up to the 20th order, with time jittering of individual reflections of up to 5% from the 5th order onwards, which provided a very realistic sound.

3.2.2 Auralization

We presented the stimuli to normal hearing listeners over 48 loudspeakers of the loudspeaker ring of the SOFE [Seeber et al., 2010, Völk, 2010]). Listeners sat in the center of the ring with their heads resting on a custom-made headrest to minimize head movements. We equalized the loudspeakers (BOSE Freespace 3) in amplitude and phase for a bandwidth of 200 Hz to 10 kHz.

3.2.3 Stimuli

We used ten different sentences spoken by ten different male speakers as stimuli. We limited their bandwidth from 200 Hz to 8 kHz, and to a duration of 2-3 seconds. We further normalized the mono signals to 58 dB SPL RMS and slightly adjusted for equal loudness, after which we convolved them with the precomputed room impulse responses. Additionally, we used three different sentences for a short familiarization session, as described below. Even though we ruled out effects of the position of the virtual listener in the room on distance perception in a previous study ([Gomez and Seeber, 2015a]), for sound presentation from the back we mirrored the room at the time of playback. Therefore, presented stimuli were always equal in distance and in the reflection pattern for the front and back, with the advantage of avoiding symmetric positioning of the listener in the room.

3.2.4 Hearing Aid Sound Presentation

The hearing aids used in the experiment were custom-made ITE and BTE prototypes by Phonak, comprising only microphones (two per BTE shell and one per ITE shell) and a receiver in the ITE shells. They were connected over cables to a PC on which we ran a real-time Simulink model with a total delay of 7.8 ms. We applied a frequency response equalization of the microphones and receivers in the frequency domain. In addition,

we applied a 5 dB gain to the signals after loudness compensation to mask direct sound leaked through the hearing aid shells. The hearing aid conditions used in this experiment were the processed ITE and BTE signals, and a static delay-and-subtract beamforming signal (BF) with maximum attenuation at 180° in the back. These were compared to a reference unaided condition.

3.2.5 Participants

Eight normal hearing participants (7 male, 1 female, average age 22 – 34 years) took part in the experiment, all of which had normal hearing thresholds as verified with a calibrated Békésy tracking audiometer in our sound booth [Seeber et al., 2003]. All participants had previously taken part in hearing experiments and had used their hearing aid prototypes previously. The TUM ethics committee approved this study.

3.2.6 Response Measure

In a previous study ([Gomez and Seeber, 2015b]) on distance perception we had presented the same ten stimuli from nine different distances in the front (30°) and back (150°) simulated in the same virtual room as in the present study. We had restricted the responses to be a forced choice between front and back, and participants had to rate the perceived distance on logarithmically scaled axes from a top view in a graphical user interface using a touch screen. We became aware of participants hearing sounds as coming from the right at 90°, very diffuse or even internalized. We therefore designed the current study to be comparable concerning presentation angles and distances in the same virtual room, but allowing for a much more detailed sound perception response comprising azimuth, distance, elevation angle, source width and whether heard as internalized. Participants used a touch screen to input the responses on a GUI. A schematic diagram of the GUI is shown in Fig. 3.1. The left view input mask for distance and azimuth showed logarithmically spaced rings at distances 1m, 2m, 4m, 6m, 8m, 10m, 12m and 14m.

To get familiarized with the GUI before the experiment, we conducted a short session where we played back sounds from 1 m, 4 m and 8 m, giving feedback on the distance but not the direction by highlighting a ring at the presented distance after user input. To allow participants to experience an internalized sound perception that was not point-like as in a diotic case, we routed the right ITE signal to the left ear, and the right BTE front microphone signal to the right ear as one of the presentation conditions, besides ITE, BTE and BF in this familiarization session. This gave a broad internalized sound perception that still contained distance information due to the reverberation and intensity of the signals, but no localization information on the sound source.

Participants were instructed to respond on perceived distance and azimuth, while source width, elevation and internalization were only set when applicable. We also encouraged them to set the azimuth to 90° and maximize the source width when no judgement of the direction of the source was possible. Since we presented the stimuli only

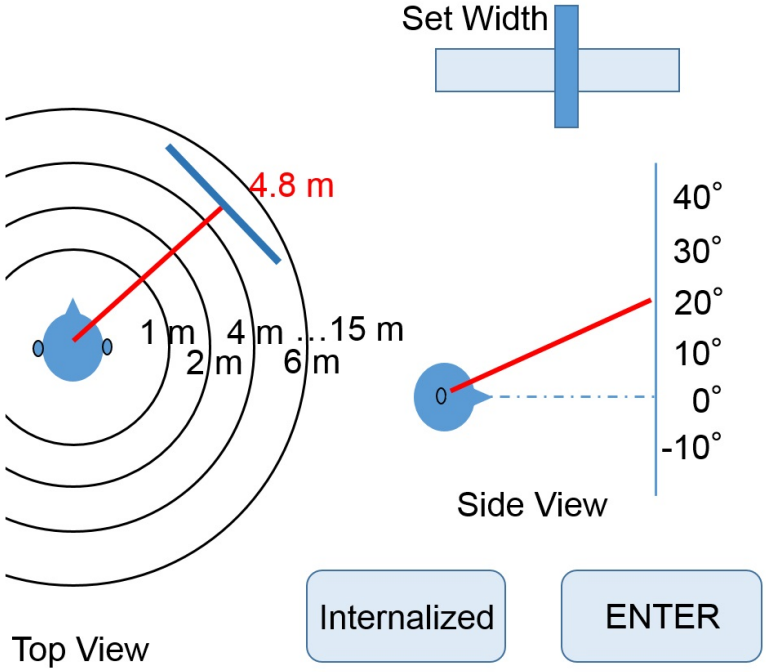

Figure 3.1
Schematic diagram of the GUI for input of perceived spatial parameters.

from 30° and 150°, sounds were always perceived as coming from the right hemisphere, never from the left, due to the ITDs and ILDs of the direct path and early reflections.

3.3 Results

3.3.1 Distance perception

Figure 3.2 shows median curves of reported distance as a function of the presented distance. Absolute distances are shown depending on the direction of sound presentation, but not depending on the perceived direction of the sounds, i.e. disregarding front-back confusions. For all conditions (BTE, ITE, BF and REF) distance perception for frontal presentation is similar, showing a compression of distance perception for distances further away. For sound presentation from the back, we observe two main differences to the responses in the frontal part. The first difference is that for the BTE condition sounds were perceived closer in the back than from the front. The second difference is that distances were perceived much further away in the BF condition than from the front, as seen in Fig. 3.3.

We performed the statistical analysis of perceived distance (after logarithmic transform of the data) using a multifactorial ANOVA with the random factor *subjects (Subj)*, and main factors *presented distance (Dists)*, *direction (Dirs)* and *condition (Conds)*. It showed significant differences for *Dists* ($F(4,229) = 218.45$, $p < 0.0001$) and *Conds*

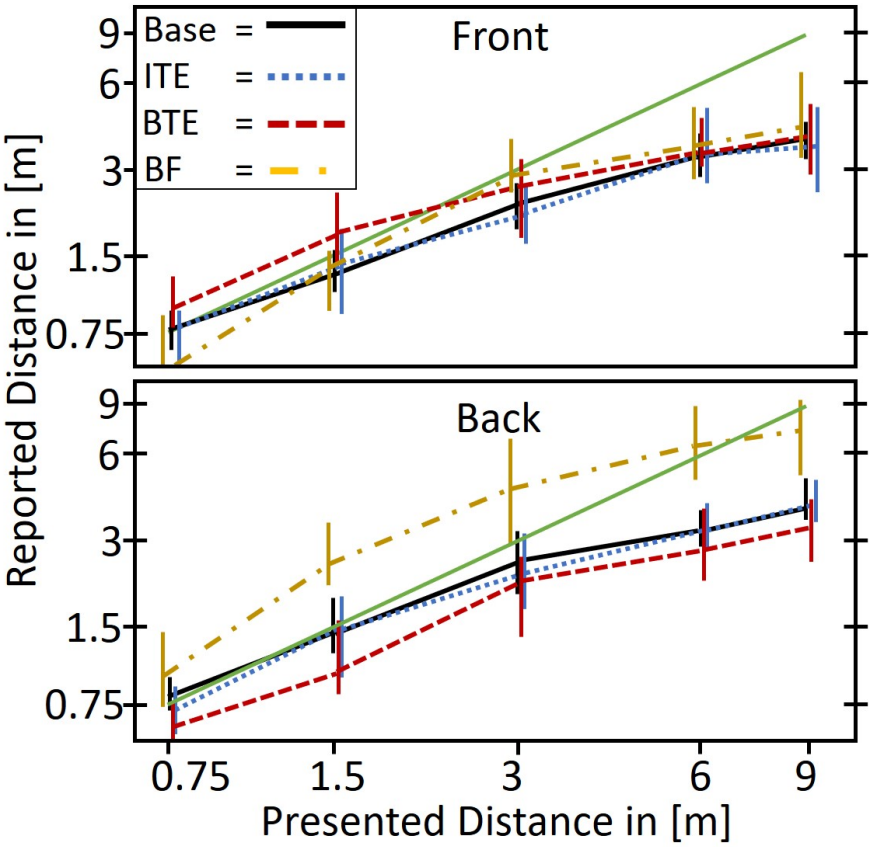

Figure 3.2
Median distance results for eight normal hearing subjects for BTE (dashed red), ITE (dotted blue), BF (dash-dotted yellow) and the unaided reference (black) for sound presentation for the front (top) and back (bottom). The green lines show the presented distance. Error bars show the 25% and 75% quartiles.

($F(3,229) = 15.39$, $p < 0.0001$) and in the interaction terms *Dirs*Conds ($F(3,229) = 73.95$, $p < 0.0001$), Dists*Conds ($F(12,229) = 3.56$ $p < 0.001$)* and the interaction of the random factor *Subj* with all fixed main effects *($p < 0.0001$)*. A post-hoc analysis after Tukey showed significant differences for the factor *Conds*, in that the BF condition in general differed from the other three conditions. For the factor *Dists*, results for all presented distances differed from each other. In the interaction term *Dirs*Conds*, the reported distance in the BF condition in the back differed from all the others in that the stimuli were significantly perceived further away. Also, in the interaction term *Dists*Conds*, significant differences occurred in the BF condition for 3 m, 6 m and 9 m compared to the other conditions.

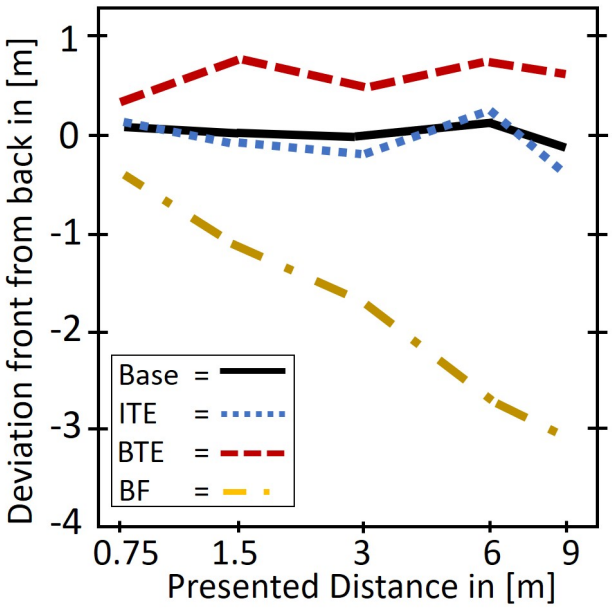

Figure 3.3
Difference between front and back for distance results as shown in Fig. 3.2, for BTE (dashed red), ITE (dotted blue), BF (dash-dotted yellow) and the unaided reference (black)

3.3.2 Azimuth

Fig. 3.4 shows the distribution of responses in azimuth and distance on a polar diagram for each condition separately. Responses for the back (blue) were converted to the range 0°-90° for comparison. The reference unaided condition gives the narrowest distribution around 30°. In the ITE and BTE conditions, the distributions start to get broader and it is noticeable how sounds presented in the front are lateralized more (closer to 60°), while sounds presented from the back remain closer to 30°. In the BF condition, azimuthal localization is worst, covering the entire right hemisphere. Also, for the BTE and especially for the BF condition, a high number of results lie close to 90° showing that many sounds were perceived somewhere on the right, when participants were not able to discriminate between front and back.

We performed a multifactorial ANOVA on perceived azimuth, with the same main and random factors as for distance perception. Here, we converted azimuthal responses to a range between 0° and 90° to compare distributions for the front and back. We found significant differences for *Dirs* ($F(1,229) = 42.97$, $p < 0.001$), *Dists* ($F(4,229) = 3.73$, $p < 0.05$) and *Conds* ($F(3,229) = 8.73$, $p < 0.001$) and in the interaction terms *Dirs*Conds* ($F(3,229) = 17.57$, $p < 0.0001$). Also, the interaction of the random factor *Subj* with all fixed main effects ($p < 0.05$) was statistically significant. Post-hoc analysis after Tukey revealed that significant differences exist between front and back, where sounds presented from the front were perceived more lateral than in the back. Also, with increasing distance, sounds were perceived less lateralized. When comparing

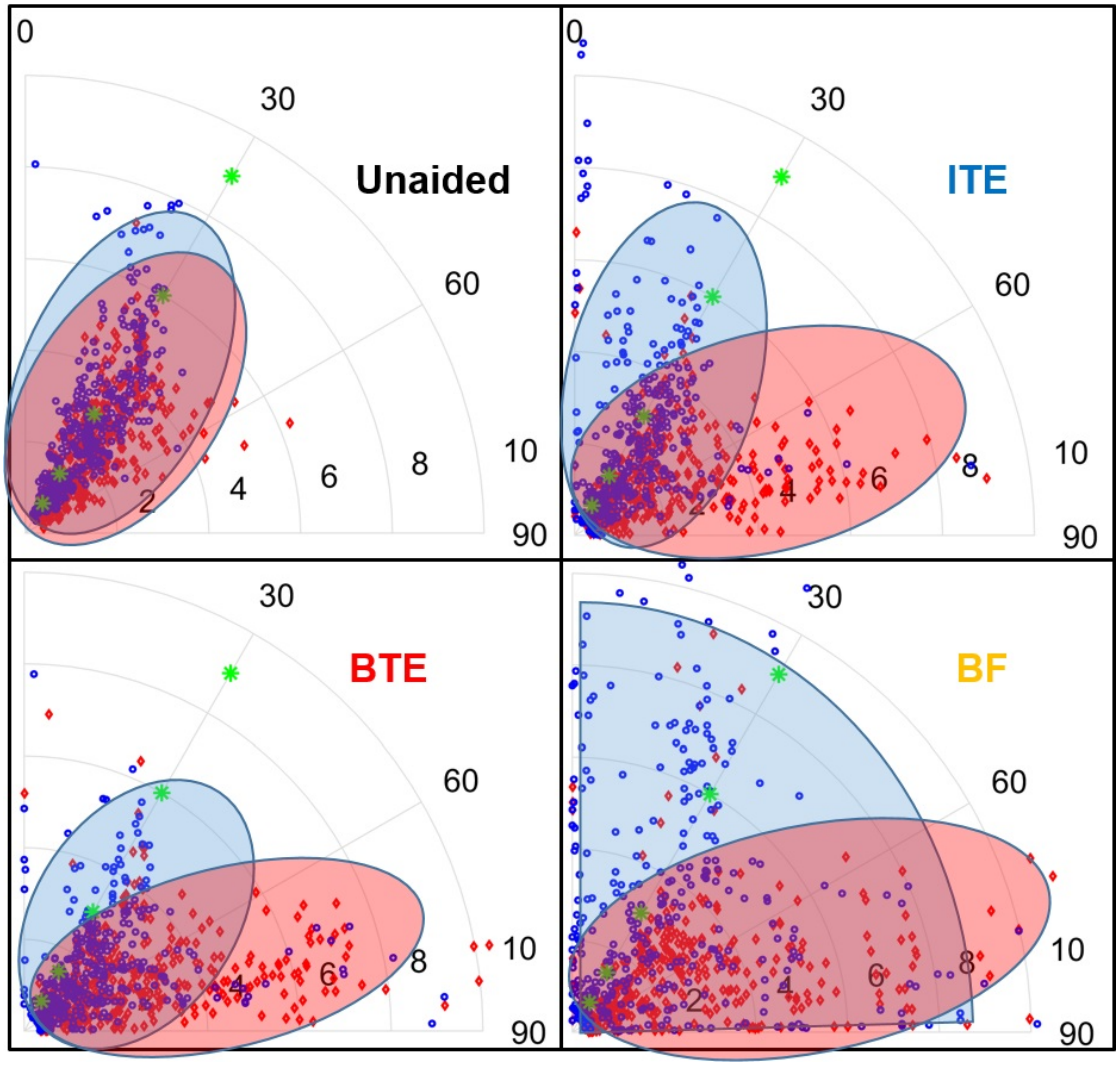

Figure 3.4
Azimuth responses over ten trials per distance and direction for eight normal hearing subjects. The four tested conditions are shown separately. Sounds presented in the front are marked as red dots, sounds from the back as blue dots after range conversion to the front. The green dots show the actual presented azimuth and distance.

azimuth between conditions, only ITE and BF did not differ from each other. Also in the interaction term *Dirs*Conds* only ITE and BF were not different from each other.

3.3.3 Front-Back Confusions

In contrast to our previous study [Gomez and Seeber, 2015b] where the direction was a forced choice between front and back, in this study the perceived azimuth was set separately. When participants were not able to hear whether the sound was coming from the front or back, they set the azimuth to 90° and maximized the width. If applicable, they marked the sound as internalized. For front-back confusion analysis we therefore only regarded responses where the azimuth was not set in the range of 85° - 95° or internalized. Fig. 3.5 shows the percentage of front-back and back-front confusions for the three tested conditions (BTE, ITE and BF) for the front (left figure) and the back (right figure). In the unaided case, no confusions were made. Clearly, most confusions were made for sounds presented from the front, especially for the BTE and BF conditions, where most sounds were heard in the back. These results show the actual number of confusions disregarding guesses when the direction was not possible to be determined, since in that case the azimuth was set to 90°.

We analyzed the number of confusions in two ways. Firstly, we performed a multifactorial ANOVA taking the percent of confusions for each subject across all ten sentences. We assume that the number of confusions would be normally distributed for a whole population, from which we had a subset of eight participants. As expected, we found significant differences between *Dirs* ($F(1,229) = 8.87$, $p < 0.05$) and *Conds* ($F(3,229) = 24.6$, $p < 0.0001$) and in the interaction terms *Dirs*Conds* ($F(3,229) = 12.84$, $p < 0.0001$), *Dirs*Subj* ($F(7,229) = 10.24$, $p < 0.0001$) and *Conds*Subj* ($F(21,229) = 2.83$, $p < 0.0001$). A post-hoc analysis confirmed that the difference in confusions between conditions lies between the ITE and BTE, and the ITE and BF conditions, but not between BTE and BF. Also, more confusions were made for frontal sound presentation than in the back. In the interaction term *Dirs*Conds* confusion results for the BTE and BF did not differ in the back nor in the front, while the other conditions differed from each other.

Since the main help to discriminate sounds from the front or back are monaural pinna cues, we additionally checked the BTE and BF conditions (both lacking pinna cues) separately using McNemar's test [McNemar, 1947] for dependent trials. As mainly front back confusions occurred, we considered the effect of the beamformer attenuation in the back as not relevant for sounds coming from the front regarding confusions. We wanted to know whether we see significant changes in the number of confusions on a trial by trial basis between BTE and BF conditions for otherwise identical conditions. Therefore, we compared the BTE and BF responses for each spatial position and sentence. The McNemar test analyses whether the number of confusions made in the BTE and not in the BF condition, or vice versa, is statistically significantly different. We analyzed separately responses for the front and for the rear sound presentation. Contrary to the ANOVA results, McNemar analysis showed significant differences in the front and also in the back, between BTE and BF conditions. This is due to confusions occurring in

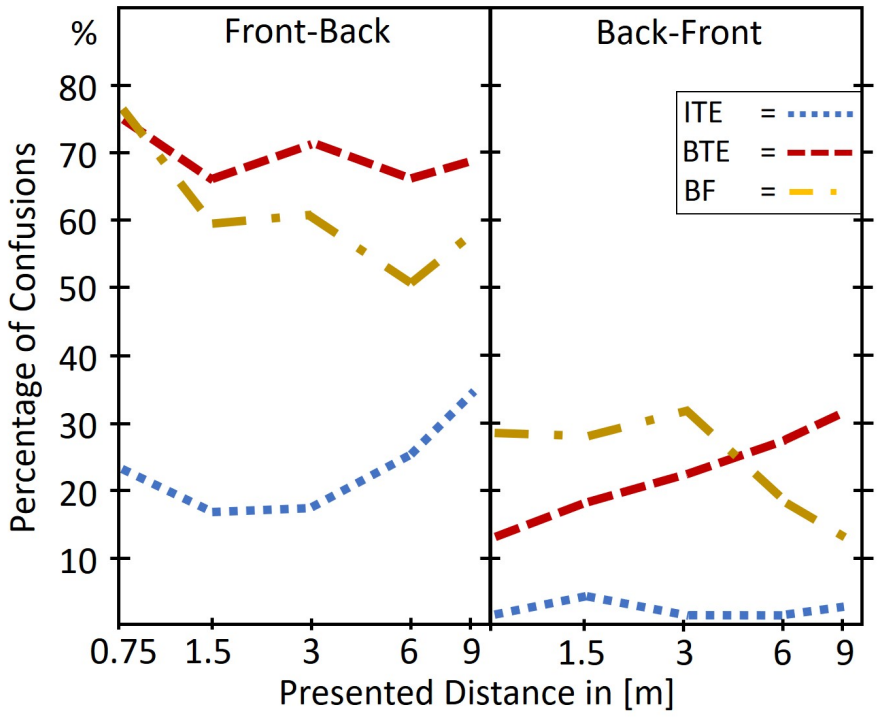

Figure 3.5
Percent of front-back confusions (left) and back-front confusions (right) averaged over eight normal hearing participants as a function of distance. Results for BTE (dashed red), ITE (dotted blue) and BF (dash-dotted yellow) are presented separately. No confusions were experienced in the unaided case.

different trials (stimulus sentences). Therefore, the BF does not perform better than the BTE in number of confusions for the front or back, but the trials on which confusions were made differ.

3.3.4 Internalization

Fig. 3.6 shows a diagram of the percentage of internalized sounds for each condition as a function of distance. Internalization occurs mainly for the BF condition and decreases with distance.

Similar to the front-back confusions, the internalization reported by this study's participants is a binary value, such that we took the percent of internalized sounds across all ten sentences presented at each spatial location for each condition. ANOVA results confirm significant differences in internalization between *Dists* ($F(4,229) = 7.73$, $p < 0.001$) and *Conds* ($F(3,229) = 5.98$, $p < 0.01$) , in the interaction terms *Dirs*Conds* ($F(3,229) = 3.79$, $p < 0.05$) and *Dists*Conds* ($F(12,229) = 4.92$, $p < 0.0001$), and all interaction terms of the main factors with *Subj* ($p < 0.001$). Post-hoc analysis after Tukey showed that internalization results for the closest distance (0.75 m) significantly

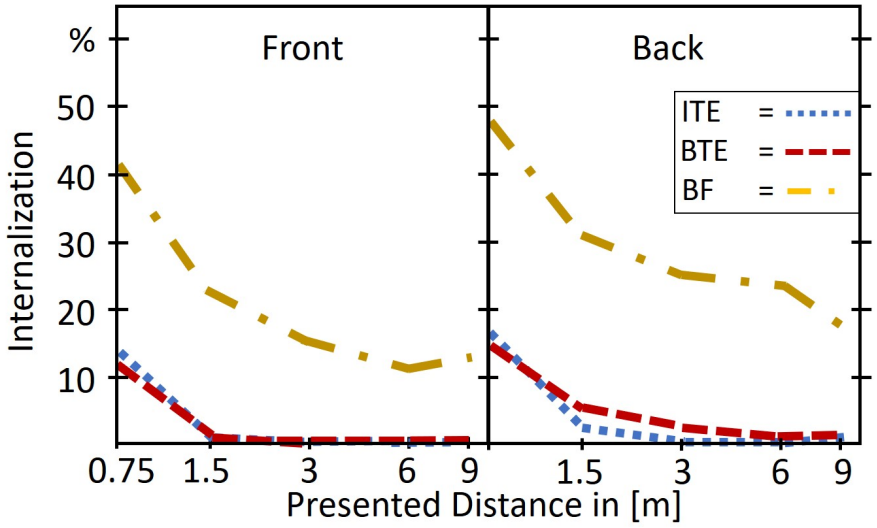

Figure 3.6
Percentage of internalization results for eight normal hearing participants as a function of distance. Colours represent the different conditions tested, with BTE in dashed red, ITE in dotted blue and BF in dash-dotted yellow. The left part shows results for sound presentation from the front, the right part for sound presentation from the back. No internalization was experienced in the unaided case.

differed from the other distances, and the BF condition differed from the rest. For the interaction term *Dirs*Conds* the BF condition differs from the rest, and within the BF condition front differs from the back. For the interaction term *Dists*Conds* the BF condition differs from the rest, and within the BF condition the closest distance differs from the other distances. McNemar analysis also confirmed differences between front and back for the BF condition but not for ITE and BTE.

3.3.5 Elevation

Fig. 3.7 shows mean results over 10 trials of perceived elevation angles for all conditions for the front (left side) and back (right side) sound presentation. Elevation results are non-uniformly distributed, with most sounds as not perceived elevated, some few sounds perceived as coming from below and otherwise perceived elevated across a range of elevation angles up to 90°. From Fig.3.7 we see the dependence of elevation angle on presented distance, with more sounds perceived elevated and at greater elevation angles at close distances than at distances further away.

Due to the lack of normality of the distribution of elevation results, a randomization test with within-subject bootstrapping was used to statistically analyse the data. We wanted to compare the obtained results with a pseudorandom distribution, where bootstrapped samples, taken from the combined population of samples of one tested parameter, were randomly reassigned. To give an example, for each subject, for sounds

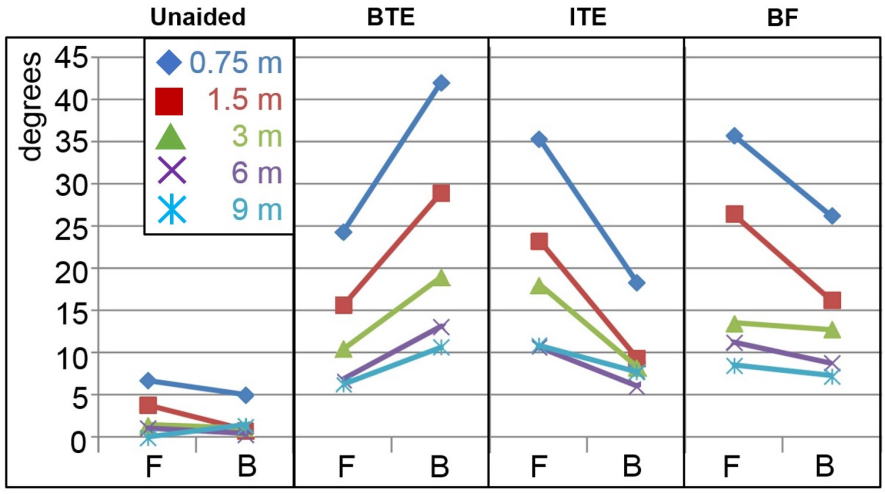

Figure 3.7
Mean perceived elevation angle as a function of distance for all four conditions. Each line connects the mean result of frontal sound presentation (F) and sounds presented from the back (B)

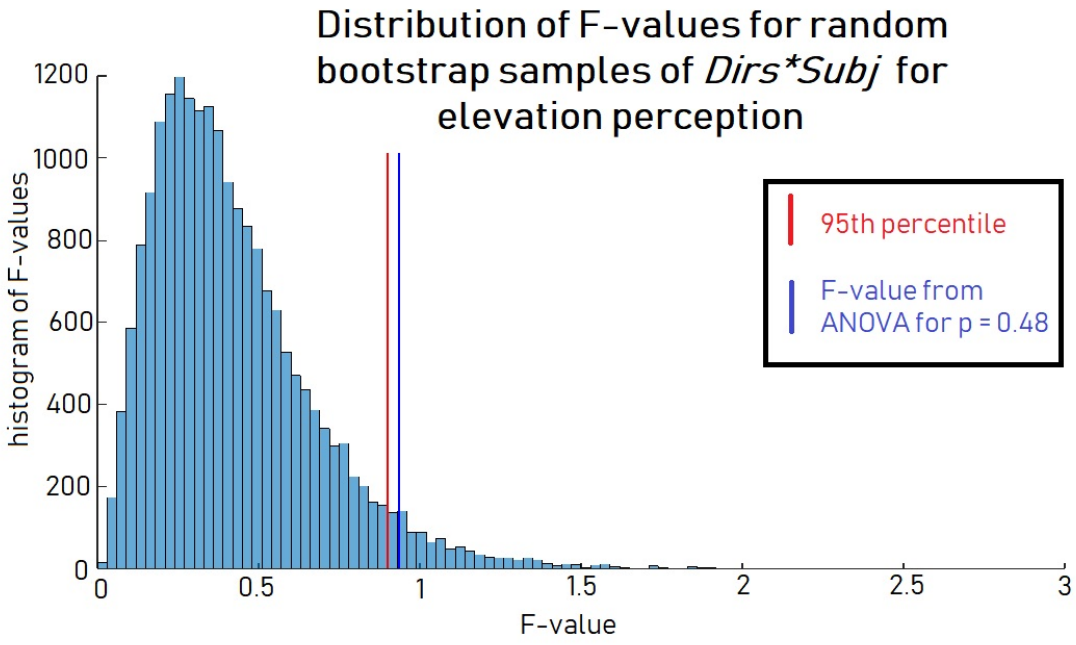

Figure 3.8
*Exemplary F distribution of 20.000 bootstrap sets for the interaction term Dirs*Subj, with random distribution of bootstrap samples into either front or back (Dirs) for perceived elevation. The blue vertical line shows the F value for the original data, the red vertical line the 95% percentile of the F distribution. If the blue line is on the right side of the red line, the results differ significantly from each other.*

from the front and for the BTE condition, we drew ten bootstrap samples out of the entire corresponding population of distance results (50 values: from 5 distances at ten trials each) and randomly assigned these ten bootstrap trials to one distance. We repeated this process 20 000 times for each possible combination of parameters. We calculated the F-statistic of each of the 20.000 bootstrap sets, getting a distribution of F values. By comparing the F values of the original data with the 95% percentile of the bootstrap distribution of F values, we gained insight whether a given factor was statistically significantly different, as shown in Fig. 3.8 for an exemplary interaction term *Dirs*Subj*. Here, this bootstrap analysis differs from significance results obtained with an ANOVA of the mean elevation results, since the data clearly violates ANOVA assumptions. As seen in Fig. 3.8, the ANOVA did not mark this interaction term as significant, with $p = 0.48$. By comparing the F value to our F distribution, we can see that actually $p < 0.05$. Thus, using the bootstrap randomization test, we find significant differences in *Dists* ($p < 0.0001$), *Conds* ($p < 0.01$), and in the interaction terms *Dirs*Conds* ($p < 0.0001$), *Dirs*Subj* ($p < 0.05$), *Dists*Conds* ($p < 0.001$) and *Dists*Subj* ($p < 0.001$). Post-hoc analysis after Tukey revealed that elevation differs between close distances (0.75 m and 1.5 m) and distances further away, and between both closest distances as well. In addition, elevation differs significantly between aided and unaided conditions when analysing the factors *Conds* and the interaction term *Dists*Conds*. For the interaction term *Dirs*Conds* the ITE and BF conditions do not differ, while they differ from the BTE and unaided condition for both the front and back.

3.3.6 Apparent Source Width

Over 50% of all presented sounds were perceived wider than in the reference unaided condition. The actual width in angles of the perceived sounds varied strongly between conditions. Especially for the BF condition sounds were perceived wider than for the ITE and BTE in the back, while for the front the width for BF and BTE was similar and wider than for the ITE signals, as shown in Fig. 3.9 where ASW is presented as an opening angle, thus independent of distance. For all conditions, mean ASW remains unchanged with distance for the front and back, except for the BF condition in the back. For many sounds, and especially for the BF condition, the AWS was set to angles greater than 45° and up to the maximum possible angle of 90°. We also found a dependence of the ASW on azimuth, showing a maximum at 90° azimuth, which is when sounds were perceived as coming from the right (due to the present ITDs and ILDs), but no judgment on front or back was possible, and thus perceived as very diffuse.

Proceeding in the same manner as with elevation results, we observed significant differences in the ASW for *Dirs* ($p < 0.05$), *Dists* ($p < 0.01$), *Conds* ($p < 0.001$) and for the interaction term *Dirs*Conds* ($p < 0.0001$). Post-hoc analysis showed that sounds in the back were perceived broader than in the front, conditions differed between each other except for ITE and BTE, and sounds from 9 m distance were perceived broader than sounds at the three closest distances (0.75 m, 1.5 m and 3 m). For the interaction term *Dirs*Conds* the analysis confirmed that the unaided condition differed in ASW

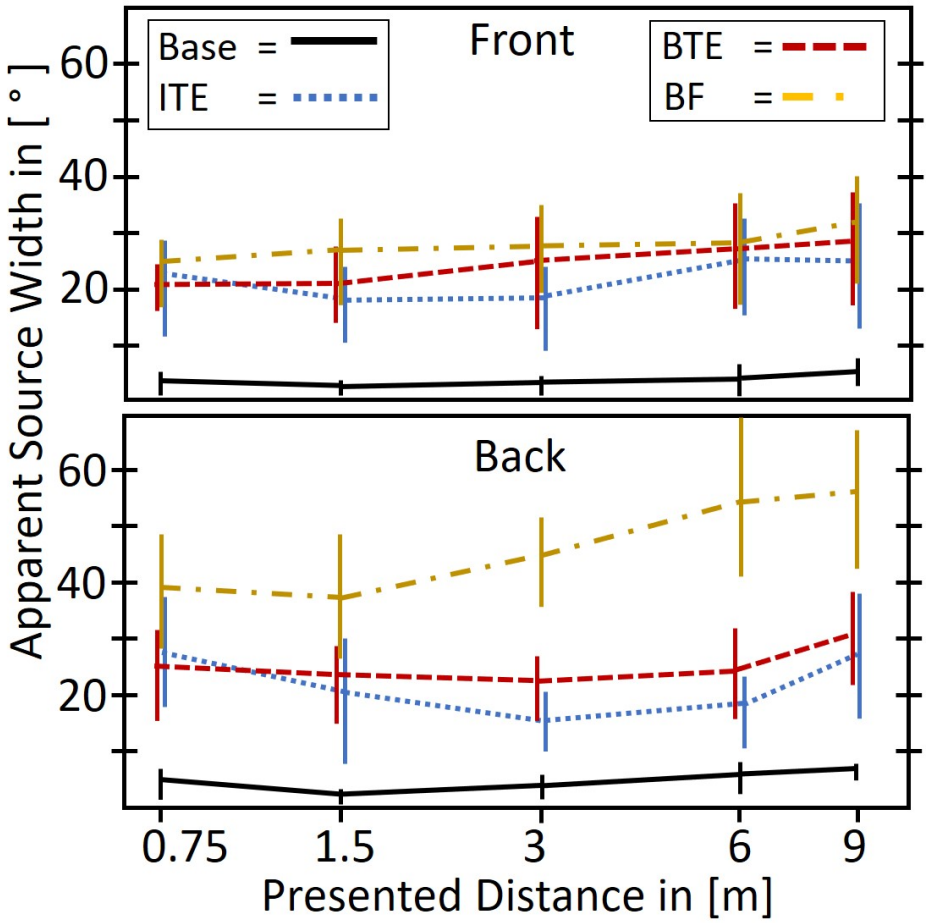

Figure 3.9
Mean perceived ASW opening angle as a function of distance for BTE (dashed red), ITE (dotted blue), BF (dash-dotted yellow) and unaided reference in black. Error bars show the 25% and 75% quartiles.

from the aided conditions, which otherwise do not differ between each other except for the BF condition in the back, which significantly differed from the rest.

3.4 Discussion

To our knowledge, this is the first study to extensively test for spatial sound perception both in the front and back simultaneously. This study shows very interesting effects of the microphone position and static beamforming on spatial perception. Firstly, we can show that all spatial aspects we tested in this study were affected using hearing aids for a static head position. Thus, we could exploit much more data than in the previous study on distance perception. Even though we presented speech sentences from different

distances in the front and back, they came always from 30° or 150° azimuth and at almost eye level, dependent on the subject's torso height while seated in the center of the loudspeaker ring. Yet the use of hearing aids massively affected the perceived azimuth, elevation and apparent source width of the sounds when compared to the reference unaided condition. The influence of level, and more specifically of the directivity of the BTE and BF conditions on distance and elevation is apparent when comparing results for the front and back. The relative level differences due to the microphone directivity lead to a shift in distance perception towards the higher level (louder) hemisphere, i.e. sounds were perceived closer with higher level, and to more elevated sounds at higher levels shown as elevation differences between front and back. The azimuthal tilt towards 60° for sounds presented from the front lies in different binaural cues than in the unaided case. While azimuthal localization in the unaided case is good, large deviations from the presented directions appear in the aided case. As will be discussed in Chapter 5 in detail, this tilt can be attributed to higher ILDs and ITDs from the microphone positions of the ITE and especially the BTE shells. The larger binaural cues caused by the microphone pick-up lead to perceived localization away from the midline for sounds presented from the front, but to a localization shift towards the midline for sounds presented in the back. Monaural spectral cues seem not to affect the localization shift for sounds from the front. In that case, we would see more accurate localization in the ITE case, comprising full spectral pinna information compared to the BTE and BF conditions. We can also rule out an influence on localization by the virtual room acoustics or asymmetric reflection patterns, since we mirrored the room for each sound emanating from the back, such that the identical reflection pattern is present in all sound presentations regardless of its direction. Additionally, it would equally affect the unaided results. The azimuthal shift seen in our results may have two possible, related reasons. The high number of front-back confusions shows that sounds from the front are more difficult to localize correctly in space with hearing aids. Therefore, the more lateralized sounds might be related to participants being less confident of their responses. We think that when listeners are unsure whether a sound came from the front or back, they will tend to answer more to the side, closer to the "average" of their perceived sound image. Additionally, but related to the mentioned uncertainty, we think that naturalness of sounds can affect localization. Humans are very sensitive to how sounds from the front should sound like, and can discriminate very accurately the location of sounds. The spectral change and level change due to the hearing-aids alters the sounds from what we are used to or expect to hear. Small changes in the ITDs and ILDs by the hearing aids lead to a loss of naturalness and exagerated binaural cues that cause large localization errors. It is known that spectral alteration can cause differences in azimuth ([Musicant and Butler, 1984]). This change of sound naturalness and binaural cues due to the hearing-aids is of course a special case for our normal hearing participants. Aided hearing impaired will be less affected since the get used to their own devices' particular sound. Also, very rarely do people encounter acoustic situations in real life where they must accurately localize a sound with their head static in place. Even very small orienting movements help resolve ambiguities and localization inaccuracy intrinsic to static head situations. When asked about the internalization perception, all participants re-

ported having experienced internalization, yet different to the internalization perception inherent to stereo playback over headphones. Here, some participants reported to be able to discriminate distance, the reverberation of the room and even lateralization, yet at times what they heard was so diffuse, that they were unable to think of any place outside their heads to place the source, which led to an internalized perception. The definition of internalization, for sounds as being perceived inside the head, should possibly be expanded by the inability to assign any spatial position in the outer space to the perceived sound. One very important outcome we can see in our data compared to our previous study ([Gomez and Seeber, 2015a, Gomez and Seeber, 2015b]) is what we believe to be the influence of the response method on distance perception. For comparability, in this study [Gomez et al., 2016] we used the exact same presentation angles, the same virtual room and the same range of distances in the front and back. Also, the hardware used was the same and the same number of participants took part in both experiments, of which seven out of eight participated in both experiments. Yet the results for perceived distance from this study differ greatly from the previous results. While before we found that in general distances were overestimated up to about 5-6 meters, and underestimated at distances further away (for our specific test scenario), our present results show a consistent underestimation of distance, except for the beamformer condition for sounds presented in the back. This was expected since sounds were attenuated greatly in the back by up to 15 dB. On the other hand, the observed compression of distance in both experiments is the same, where the factor between perceived and presented distance decreases with distance. Also, as in our previous study, we observe that sounds in the BTE condition were perceived further away in the front than in the back, while the curves for the ITE and reference condition are almost identical between each other. The observed shift in the BTE condition seems to be due to a gain of about 3 - 8 dB (frequency dependent) to the back caused by the design and the microphone position of the BTE devices, thus making sounds from behind being perceived louder than their counterparts in the front. Interestingly, we did not observe the opposite shift of distance perception for the ITE and reference condition as in our previous study on distance perception. We believe it to be the effect of training, since the overestimation of distances from the back compared to the front became less during the course of our previous studies and we could rule out effects of the position of the virtual listener in the room ([Gomez and Seeber, 2015a]). We cannot say what led to the different results in distance perception between our comparison studies, except that there must be an influence of the response method that is much greater than we would have expected for otherwise identical conditions, possibly to the higher complexity and higher dimensionality of our new interface. These differences show how difficult it is to compare studies among each other, and why results between different studies differ so greatly. The experimental method, which determines how the perceived sound object in a participant's mind is transformed into a numerical result used by the experimenter for data analysis has, in our experience, as much of an influence on the results of a study as has the selection of the rooms, stimuli or other parameters in the experimental design.

3.5 Conclusions

The presented study analyzed the effect of hearing aid devices on spatial sound perception. Having normal hearing participants taking our test, we could separate the sole effect of the hearing aid devices from any effects related to hearing loss or individual hearing loss compensation methods. We compared results from the aided conditions BTE, ITE and BF to the unaided baseline condition. Additionally, we tested for spatial perception both in the front and back, which to our knowledge has not been tested before to this extend. We used reverberated stimuli in virtual acoustics for a range of distances at 30° in the front and 150° in the back, and tested for the perception of distance, azimuth, apparent source width, elevation, internalization and implicitly for front-back confusions. In general, we found a large alteration of all spatial dimensions between the aided conditions and the unaided baseline. The ITE condition showed the most natural results of the aided conditions, including most of the spectral pinna cues due to its microphone position in the ear canal, followed by the BTE and the BF conditions. While beamforming is beneficial for improving SNR and speech understanding, it becomes clear from the present results that regarding spatial aspects it performs the worst. Thus, if hearing aid manufacturers want to significantly improve the perception of spatial aspects in their hearing aids while maintaining good speech understanding, they should ideally find a way to combine the naturalness of the microphone position in the ear canal and keep the SNR advantages of beamforming possible with BTE devices.

4 Application of Pinna Cues to Beamforming Signals

4.1 Short-Time Averaged Pinna Cue Filtering for Beamforming Signals (STA BF)

The previous chapter gave an introduction of the spatial perception of sounds with hearing aids. While beam-forming based noise reduction is a powerful aid to improve SNR of a target speaker in noisy environments, HA user's spatial perception severely suffers because such noise-reduction methods make use of BTE microphones behind the ear – being therefore unable to convey important pinna cues. This chapter presents a method that strives to combine both advantages. That is, to provide the SNR gain due to BTE beamforming, while imposing dynamic spectro-temporal pinna cues into the beam-formed signals that are delivered to the eardrums. Imposing of dynamic pinna cues into the beam-formed signals is achieved as follows. Firstly, the monaural ITE signal is temporally smoothed by filtering it in the frequency-domain with an infinite impulse response (IIR) averaging filter, the "short-time average" (STA) depicted in Fig. 4.1. The STA filter is a first-order low-pass filter with a time constant ß, which allows adjustment of the temporal memory of the decaying filter. Small values of ß (i.e., close to zero) lead to weak filtering, since the averaging is then performed only across the recent parts of the signal, while large values of ß (i.e., close to one) result in averaging across all previous values of the signal, with the weight of the values decaying exponentially as a function of time into the past. Therefore, depending on the value of ß, the STA filter will smooth out rapid spectral changes but allow slow spectral changes to pass through.

Thus, we can present the difference equation of the low-pass filter output as:

$$y[n] = x[n] + ß(y[n-1]–x[n]) \tag{4.1}$$

and to express the transfer function of the filter in the z-domain as:

$$H(z) = \frac{1-\beta}{1-\beta \cdot z^{-1}} \tag{4.2}$$

Alternatively, we can use the impulse response

$$h(k) = \beta^{k-1}(1-\beta) \tag{4.3}$$

to describe the filter.

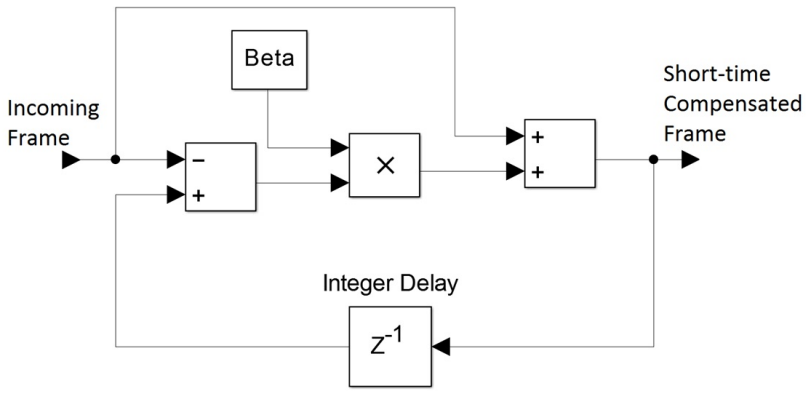

Figure 4.1
Diagram of a first-order low-pass filter. The input signal takes a vector of length 65 samples representing positive frequency bins. The low-pass filter acts upon each individual bin independently.

It is noteworthy to mention that filtering a signal with such a low-pass filter will not affect the DC component (i.e. $f = 0\ Hz$), as can be proven using the Eq. (4.4)

$$H(z = e^{j\omega}|_{\omega=2\pi f=0}) = \frac{1-\beta}{1-\beta \cdot 1^{-1}} = 1 \tag{4.4}$$

In the present work, the STA filter is implemented as a Simulink model and the input consists of individual amplitudes of STFT (short-time Fourier transform) bins that are obtained upon converting the microphone signal (sampled with a sampling rate of 22050 Hz) into the frequency domain. Specifically, the STFT is computed over a time-frame of 128 samples with $\frac{3}{4}$ frame overlap. Considering the frame size and the sampling rate, the attack- and release times of the filter can be adjusted depending on the parameter β, as listed in Table 4.1 ([ANSI, 2003]).

The STA processing is applied on the beam-formed signals following Eq. (4.5), separately for the two ears. Here, BF_{comp} denotes the compensated output of a static delay-and-subtract beam-former that is designed to attenuate signals from the back (at 180°) using both microphones of the BTE hearing aid shell. This signal is divided by a short-time averaged signal from the frontal microphone of the BTE (STA(BTE_{front})). In theory, the BTE_{front} and BF signal are highly similar for frontal sound sources and therefore, the division will result in values close to 1. On the other hand, sounds from the back are attenuated by the beam-former, resulting in the division of BF by STA(BTE), yielding values between 0 and 1. To impose dynamic pinna cues into the beam-formed signal, the division is multiplied with the short-time averaged signal from an ITE HA shell (STA(ITE)).

$$BF_{comp} = \frac{STA(ITE)}{STA(BTE_{front})} BF \tag{4.5}$$

In practice, the denominator in Eq. (4.5) must be limited to remain within a meaningful range. Otherwise, sharp peaks could be imposed on the spectrum, depending on

Table 4.1
Attack- and release times to 3 dB and 4 dB from the end level respectively, for a level difference step of 35 dB (according to ANSI S3.22 [ANSI, 2003] from 55 dB SPL to 90 dB SPL or back). The time in ms is calculated using the number of frames taken times the frame update rate (32 samples) divided by the model sampling frequency of 22050 Hz.

β	Attack time (signal increase by 35 dB, 3dB from maximum value)		Release time (signal decrease by 35 dB, 4dB from minimum value)	
0.1	1 frame	(1.5 ms)	2 frames	(2.9 ms)
0.2	1 frame	(1.5 ms)	3 frames	(4.4 ms)
0.3	2 frames	(2.9 ms)	4 frames	(5.8 ms)
0.4	2 frames	(2.9 ms)	5 frames	(7.3 ms)
0.5	2 frames	(2.9 ms)	7 frames	(10.2 ms)
0.6	3 frames	(4.4 ms)	9 frames	(13.1 ms)
0.7	4 frames	(5.8 ms)	13 frames	(18.9 ms)
0.8	6 frames	(8.7 ms)	21 frames	(30.5 ms)
0.9	12 frames	(17.4 ms)	44 frames	(63.9 ms)
0.95	24 frames	(34.8 ms)	89 frames	(129.2 ms)

the difference between the short-time averaged ITE and BTE signals, which would then result in artefacts in the output audio signal.

In order to find optimal values of β for the algorithm, a formal listening experiment was performed, assessing the perceived spatial sound quality of sounds that were processed using five different values of β. The values, listed in Table 4.1, were chosen as they result always in a doubling of the attack time. In the experiment, a compensation was however performed on the BTE signals (Eq. (4.6)) rather than the beam-formed signals (Eq. (4.5)). This means that STA pinna cues were applied onto the BTE signal. This modification was done to fully analyse the effect of the compensation in a fixed acoustic scenario, with a target and disturber in different hemispheres (front and back) and rate the perceived spatial sound quality of both target and disturber independently. Had we used the beam-forming compensation, our results for the target and disturber would have been influenced by the impact of the beam-former on them and we could not have properly investigated certain aspects of spatial perception deterioration due to the time constants alone.

$$BTE_{comp} = \frac{STA(ITE)}{STA(BTE_{front})} BTE_{front} \tag{4.6}$$

A detailed explanation of the spatial sound quality experiment to determine proper time constant values is given in Chapter 5, together with a follow up experiment.

4.2 The Jackrabbit method

Above, a new method was presented to combine the advantages of having the good SNR due to beam-forming and of considering the dynamic pinna cues. It will be shown in Chapter 5 that the least amount of spatial deterioration is achieved when the time constant β has a small value, keeping the attack time under 3 ms. As discussed in Chapter 3, the use of a static delay-and-subtract beam-forming in hearing aids can however result in an internalized sound percept. Moreover, Chapter 5 will show that the externalization is only partially improved with the STA-BF method. Another drawback of such a static beam-forming is the attenuation of low frequencies, which results from the subtraction of two signals that are captured with two rather closely-spaced microphones. Especially at low frequencies, the distance between the microphones is small in comparison to the wavelength and thus the signals picked by the microphones are very similar. Consequently, the subtraction attenuates such frequencies drastically This roll-off can be compensated for by applying a low-frequency amplification, which can be implemented, for instance with a low-pass filter that is calibrated for the specific device. While such processing compensates for the low frequency roll-off, it also enhances the low-frequency microphone noise to become clearly audible.

To circumvent the drawbacks of the static beam-former based approach, a novel method was designed. This method still brings the benefits of the STA BF but overcomes the internalization and low-frequency issues of it. While the above-presented STA-BF method applies dynamic pinna cues to the beam-formed signal, the new Jackrabbit method (stands for the large ears of the jackrabbit animals that allow for directional hearing using the outer ears) takes an opposite approach. The approach is motivated by the assumption that the pure ITE signal is the best possible signal regarding spatial quality and naturalness as it is the most similar one to the signal arriving at the eardrum in normal conditions. This is the signal that the brain has learned to use and to interpret, allowing for correct front-back localization, elevation perception, and perception of a focused and externalized sound image. In other words, the ITE signal includes most of the HRTF cues that are essential for correct localization. Thus, the idea behind the new method is to use the ITE signal and attenuate the disturbing parts of it, the parts which are caused by interferers located at different spatial locations.

The present method assumes that the target and the interferer(s) differ from each other at any time instant in terms of spectral information. That is, the amplitude- and/or phase spectra of any two or more sound sources are always different. Thus, the sound pressure at the eardrum becomes a sum of the sound pressures caused by individual sound sources and their reflections, which arrive from different directions in space and are, therefore, filtered with the corresponding HRTFs. The linear superposition of sound pressure fields implies that the spectral information is also a sum of individual components of the sources. Consequently, selective subtraction of spectral components of a given disturber leads to an effective attenuation of selected components from the entire signal reaching the eardrum, assuming that the spectrum is known.

The seemingly stringent requirement of a prior knowledge about the energetic contributions of different disturbers can be elegantly met with beamforming, which, by

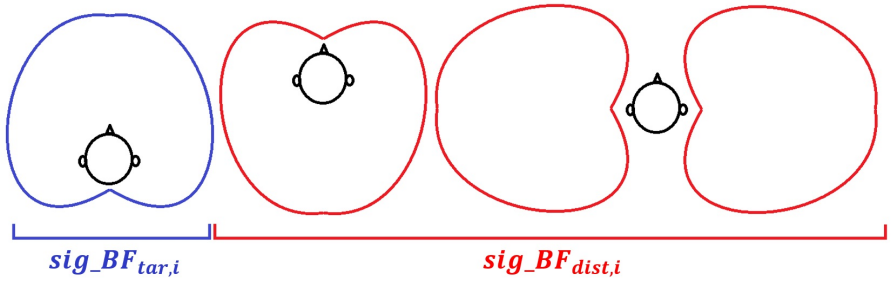

Figure 4.2
Exemplary diagram of possible beamforming patterns to be used for separating spectral information of the target and disturbers.

combining two or more microphone signals, allows to enhance or attenuate sounds from specific directions. For hearing aid users, the target is often assumed to be in the front. Thus, a beam-former directed towards the target attenuates sounds from directions other than the one of the target. Contrary, a beam-former with (maximum) attenuation aimed towards the target will exclude most of the target's sound energy (except for reflections from the target arriving from different directions). When combining the spectra of the two beam-former signals of which one emphasized the target and the other attenuated it, one can obtain relative contributions of the entire signal and can then separate the entire signal at the location of the beam-former microphones into a target and (one or multiple) interferer parts (see Eq. (4.7) and Fig. 4.2).

Since any real-life sound field varies dynamically in time, the spectral analysis of the beamformer signals must be performed continuously using short time frames. Therefore, the entire sound scene can be approximately separated into individual signals of a target and one or multiple disturbers using beamforming signals

$$sigBF_{sum}(t,f) \approx sigBF_{tar}(t,f) + sigBF_{dist}(t,f) \tag{4.7}$$

where $sigBF_{dist}(t,f)$ includes all the spectral information (which frequencies, their amplitude and phase) about the disturbers at the location of the beam-former microphones at a time instant t, and $sigBF_{tar}(t,f)$ contains the same information about the target. In addition, a function $\mathbf{F}$ needs to be defined, with which one can filter out the disturber energy from the summed signal at the eardrum

$$sigEAR_{tar}(t,f) = \mathbf{F} \cdot sigEAR_{sum}(t,f) \tag{4.8}$$

which includes the HRTF filtering of the individual sound sources and reflections. Here, $sigEAR_{tar}(t,f)$ corresponds to the signal that would be present at the eardrum, if only the target sound source was present with its corresponding HRTF filtering. In reality, it will be very difficult to completely eliminate disturbing sound sources. Yet by using the filter function $\mathbf{F}$, disturber sound sources should be attenuated enough to significantly increase the signal-to-noise ratio (SNR). Most importantly, energetic attenuation of disturbers should not alter the relevant HRTFs and monaural cues of the target sound.

The filter function **F** can be, for example a weighting function, that compares and weights the energetic components of individual frequency bands i of the target and disturber signals. By taking the ratio of weighted energies of individual frequency bands, an effective attenuation of the disturber signals can be achieved.

$$\mathbf{F} = \alpha \cdot [\omega_1, \omega_2, \ldots, \omega_i, \ldots \omega_m] \tag{4.9}$$

The weights ω can be defined either as,

$$\omega_i(t = \tau_j) = \frac{\mathbf{Energy} sigBF_{tar_i}(t = \tau_j)}{\mathbf{Energy} sigBF_{tar_i}(t = \tau_j) + \mathbf{Energy} sigBF_{dist_i}(t = \tau_j)} \cdot H_{ear \to BF,i} \tag{4.10}$$

or as

$$\omega_i(t = \tau_j) = \frac{\mathbf{Energy} sigBF_{tar_i}(t = \tau_j)}{\mathbf{Energy} sigBF_{dist_i}(t = \tau_j)} \cdot H_{ear \to BF,i}. \tag{4.11}$$

Here, $t = \tau_j$, denotes a given time instant or frame, and

$$H_{ear \to BF,i} = \frac{\beta \cdot \mathbf{Energy} sigEAR_i(t = \tau_j)}{\gamma \cdot \mathbf{Energy} sigBFn_i(t = \tau_j)} \tag{4.12}$$

corresponds to a correction term that is used to compensate for spectral differences between the signal at the eardrum and the signal at the BF microphone n. Values for the constants α, β, and γ can additionally be set individually to optimize the filter function **F**.

In addition, it makes sense to limit the weighting factors ω and correction term H to a specific range of values, e.g. [0 – 1], or to apply a compressing function $\int$, in order to ensure that the signal at the eardrum $sigEAR_{sum}(t, f)$ is not deteriorated in a perceptual sense. In the worst case, the summed signal at the eardrum should remain almost untouched when processed with the filter function **F**. For example, if there are no disturbers present, the weighting factors ω (Eq. (4.11)) would become greater than one if no limiting is applied, which would corrupt the otherwise clean target signal. By limiting the range of ω to a maximum value of one, the eardrum signal will not be altered in a given frequency bin if the target energy is greater or equal to the energy from disturber directions. An additional weighting restriction can be set to handle situations where the target and disturber energies are similar. In other words, a 3-dB attenuation could be set to compensate for the summation of incoherent sound source energies of similar value from target and disturber signals.

Also, the weighting factors ω and correction term H should be temporally smoothed, for example using a weighted moving-average filter with a forgetting factor. This smoothing can be done at an arbitrary rate, regardless of the signal processing rate of the time signals at the eardrum or that of the beam-formers (which can be placed at a different location than at the eardrums, e.g. in a separate device). Such a smoothing is beneficial

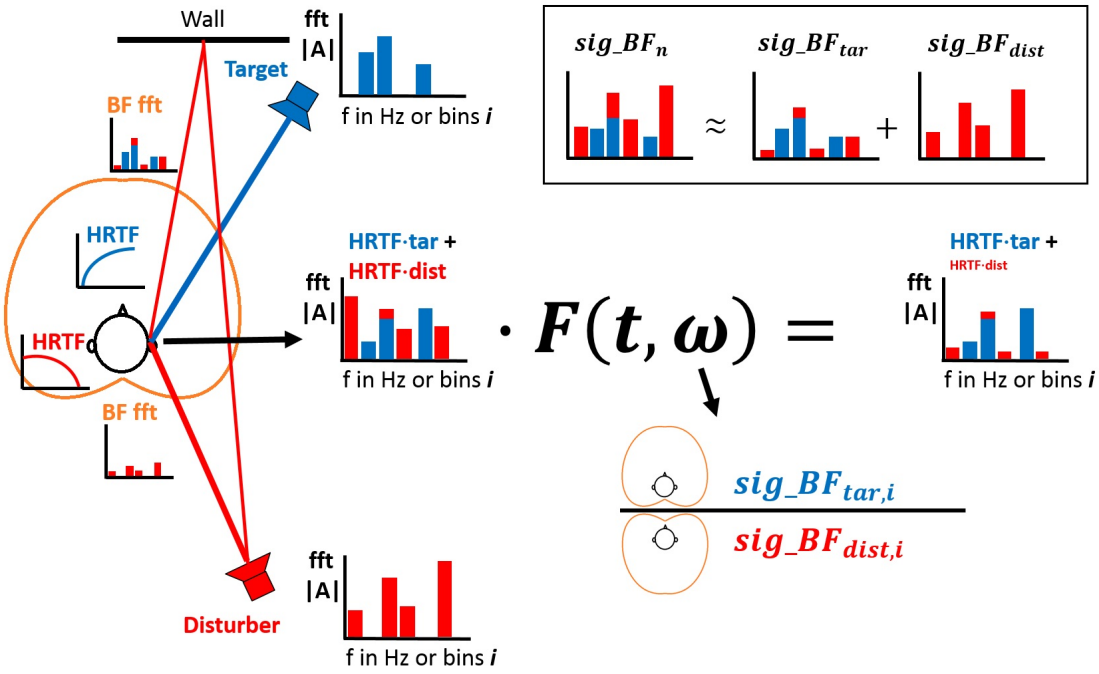

Figure 4.3
Schematic diagram of an exemplary acoustic situation and usage of the Jackrabbit method for improvement of spatial hearing with ITE hearing aids. The overall SNR gain originates from processing the ITE signals at the ear canals (HRTF-filtered sounds) with a filter function F that computes energy ratios of target and disturber signals and uses beam-formers to dynamically attenuate such spectral bins of the ITE signal that are affected by the disturbers.

not only for suppressing any processing artifacts or discontinuities, but also for preserving the target's spectral information, at least for a short period of time. In theory, the smoothing allows early reflections of the target signal to go through without being affected, even when those reflections do not originate from the direction of the target, while reflections of interferer signals are attenuated, also when they would indeed originate from the direction of the target source (see Fig. 4.3). Consequently, optimization of the smoothing function is likely to play an important role in the sound percepts evoked by the processed signal.

When applying the aforementioned steps and restrictions, the spectral components of the summed signal will remain unaffected when the energy of the disturber signal is smaller than that of the target signal in the given frequency band. On the other hand, if the energy of the disturber is greater than that of the target, the energy within such frequency bands will be attenuated by applying the weighting factors $\omega_i < 1$.

One important advantage of the Jackrabbit method is that it does not impose any restrictions on the beam-former microphones. That is, the number and placement of microphones and the beam-forming method can be chosen freely, depending on the

Figure 4.4
Close-up picture of an ear with the custom-made BTE and ITE prototype hearing-aid shells. The shells are connected via cables to a PC that runs a real-time Simulink model.

optimization of $\mathbf{F}$, as long as it is possible to selectively attenuate sounds (within specific frequency bands) from specific directions for calculation of the weighting functions ω_i. Thus, the beam-former microphones can be positioned on a separate device that can then be either placed somewhere in the room or worn by the user (as long as real-time signal transmission is possible) or integrated into the hearing aids and/or glasses to monitor the user's head movements. It is possible to implement the presented method also by relying only on ITEs with two microphones per device, as is common in many commercial devices. This is also the most probable application scenario for a real product. BTE devices would require an additional microphone to be positioned at the ear canal to obtain the relevant monaural cues that are so important for spatial hearing. In that case, proper feedback-suppression algorithms or closed ear molds should be used to inhibit feedback.

The remaining parts of this chapter explain how the Jackrabbit method was implemented in this thesis. Custom-made prototype ITE shells and BTE shells by Phonak were used, as shown in Fig. 4.4. Both microphones of the left- or right-ear BTE device were used for a delay-and-subtract beam-forming that was performed in the frequency domain. The right-ear BTE beam-former signals were used for the right-ear ITE signal processing and vice versa for the left ear.

Here, only one interferer-oriented beam-forming signal was used per ear to get maximum attenuation in the front. This was accomplished by inverting the beam-former

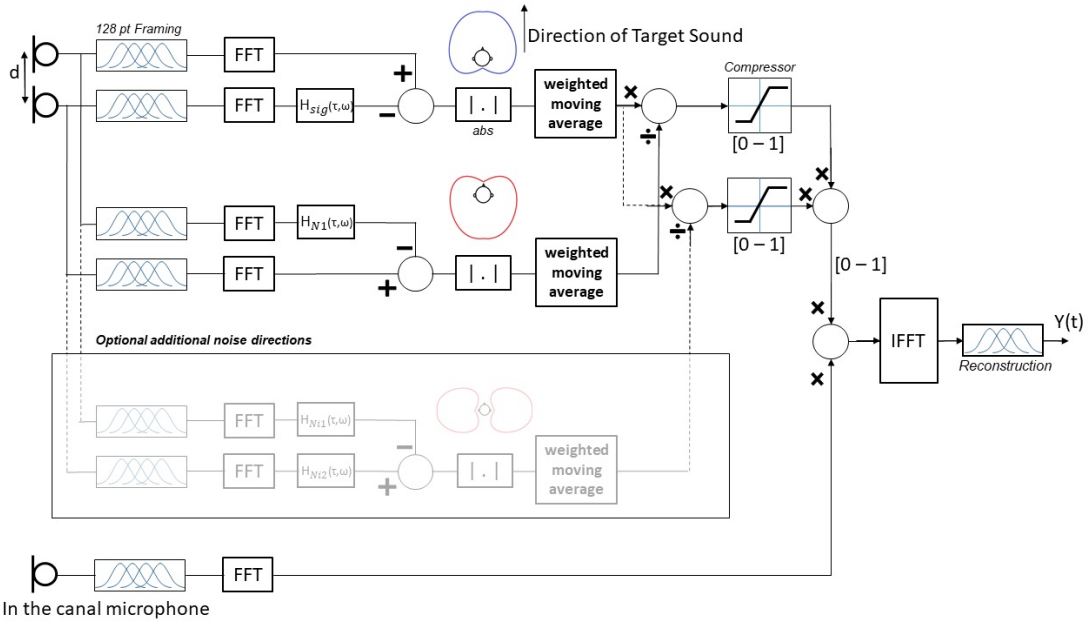

Figure 4.5
Schematic diagram of the applied implementation of the Jackrabbit method in a real-time Simulink model.

directivity. Due to this simple approach and the weighting factors ω_i being computed simply as the ratio between the energies of the target and the interferer, no roll-off compensation nor any equalization filtering is needed. The signals picked up by the BTE and ITE microphones are sampled at a sampling rate of 22050 Hz and buffered into 128-samples-long frames, with $\frac{3}{4}$ overlap between frames. An FFT is performed on each time frame to obtain 64 frequency bins with amplitude and phase. An additional BF-compensation filter is then applied to the bins of one of the microphone channels to compensate for any level and/or time differences between the two BTE microphones before beam-forming. Then, the frequency-domain signals are subtracted from each other, giving rise to a directivity pattern that is oriented either to the front or to the back, depending on what type of BF-compensation filter is applied. The magnitude spectra of both BF signals are first smoothed with a moving-average filter, after which ratios of target-to-disturber energies are calculated within each frequency bin. Specifically, an STA-compensation filter (see Chapter 3) with $\beta = 0.9$, corresponding to an attack time of 18 ms, is used for the smoothing. This value was deemed to be optimal after extensive listening and qualitative assessment of Jackrabbit signals. The final weights for different frequency bands are obtained by limiting the obtained ratios of target-to-disturber energies to lie within the range [0.1 – 1]. Finally, the weights are then applied on the STFT amplitude spectrum of the ITE signal, and the weighted spectrum is converted back to a time-domain signal with inverse Fourier transformation and overlap-and-add method. A schematic diagram of the implementation is found in Fig. 4.5.

4.3 Summary

This chapter presented two methods that aim to improve how sounds are perceived with hearing aids. Both these methods, the $STA(BF)$ and the Jackrabbit method attempt to preserve pinna cues while applying noise reduction by beam-forming, increasing the SNR. The $STA(BF)$ method imposes pinna cues onto beam-forming signals, while the Jackrabbit method takes a natural signal at the ear drum and attenuates energy originating from disturbers at different directions than the target sound. The following chapter will present the validation of the developed methods in comparison to traditional beamforming and the pure ITE- and BTE signals. There, a thorough quantitative experimental validation of localization, externalization and speech understanding is presented to highlight the advantages of the new methods. Additionally, an experimental validation on the subjective rating of perceived spatial quality with these hearing-aid conditions will be given.

5 Experimental Validation

This chapter presents four validation experiments in which different hearing-aid algorithms were evaluated. The results demonstrate the usefulness and importance of preserving natural spatial cues in the processed signals that are delivered to the listeners ears. Moreover, the results show that the newly developed Jackrabbit algorithm does not only preserve the spatial cues but is also a simple yet powerful selective noise-reduction method that can be applied to audio signals in such a manner that disturber noises are attenuated to aid the hearing-aid wearers in their listening tasks.

5.1 Localization accuracy with hearing aid algorithms preserving spatial cues

5.1.1 Summary

An experiment on localization accuracy was conducted with eight normal-hearing participants who wore custom-made bilateral ITE and BTE hearing aids. Five different hearing-aid processing schemes were tested in random order of presentation, i.e. each new stimulus was randomly presented with a different HA condition. The conditions were the ITE and BTE omnidirectional microphone signals, a static delay-and-subtract beam-former directed to the front, a beam-forming signal with applied dynamic pinna cues (STA BF), and the novel Jackrabbit algorithm that preserves pinna cues and attenuates energy of disturbers. The localization accuracy was investigated with broadband noise bursts using loudspeakers that were spaced 15° apart within the range of ±60° in the front and between ±30° in the back. An unaided baseline test was conducted for comparison. The study revealed that the BTE microphone position with and without static beam-formers leads to large localization errors both in the front and back [Kolotzek, 2017]. Analysis of HRTF cues revealed this to be caused by the BTE signals conveying too large binaural cue values for frontal sources and too small ones for sources at the back. While the beam-former with applied pinna cues (STA BF) showed also large localization errors and a high rate of front-back confusions, the Jackrabbit method performed as well as the ITE condition, with small localization errors and the lowest confusion rate of all aided conditions.

5.1.2 Methods

5.1.2.1 Participants

Eight NH participants (aged 22-35 years, male) took part in the experiment. All had normal hearing as verified with a calibrated Békésy tracker procedure

([Von Békésy and Wever, 1960, Seeber et al., 2003] in a sound-isolated listening booth ([Frank, 2000]). All participated voluntarily and were not paid for taking part in the experiment. The ethics committee of the Technical University of Munich approved of this study.

5.1.2.2 Stimuli

Broadband white Gaussian noise bursts (200 Hz – 8 kHz) of 500 ms total duration were used as stimuli. The pulse duration was set at 30 ms and the inter-pulse interval at 70 ms. Gaussian-shaped ramp with 10-ms-long on-and offsets were applied to the pulses. The noise signal was first generated at 60 decibel sound pressure level (dB SPL) and the pulse-train envelope applied to.

5.1.2.3 Hearing Aids and Algorithms

The subjects wore custom-made ITE shells and BTE shells from Phonak AG that were designed individually for each participant. The shells consisted of similar casings, microphones and receivers as in commercially available products, while all signal processing was performed on a laptop that ran a real-time Simulink model of 7.8 ms delay. Five different hearing aid conditions were tested, namely the ITE and BTE omnidirectional signals, a static delay-and-subtract beam-former with attenuation towards the back, a short-time averaged (STA) pinna cues filter applied to the beam-former, as in Eq. (4.5), and the novel Jackrabbit method that preserves pinna cues and attenuates unwanted sound sources. The signals were band-pass filtered in the frequency domain to contain energy only within the range from 200 Hz up to 8 kHz. In addition, frequency bins outside this range were set to zero in the hearing aid model. The frequency responses of the microphones and receivers of the ITEs and BTEs were compensated for ([Gomez and Seeber, 2015b]). All output signals were presented to the listener using the receiver of the ITE devices.

5.1.2.4 Experimental Setting

A ring of 96 loudspeakers of the Simulated Open Field Environment (SOFE v3, [Seeber et al., 2010]) was used for the experiment. Fourteen evenly-spaced loudspeakers were used to emit sounds, of which nine were in the front within the range $\pm 60°$, and five at the back within $\pm 30°$ in the back with $15°$ spacing between neighbouring presentation angles, as shown in Fig. 5.1. The loudspeakers were calibrated to have a flat frequency response and linear phase response within the range of 180 Hz – 10 kHz. Participants were seated in darkness in the middle of the loudspeaker ring. They wore hearing aid shells and a magnetic head tracker (Polhemus Fastrack, [Mine, 1993]) to ensure that no head movements were made during stimulus presentation.

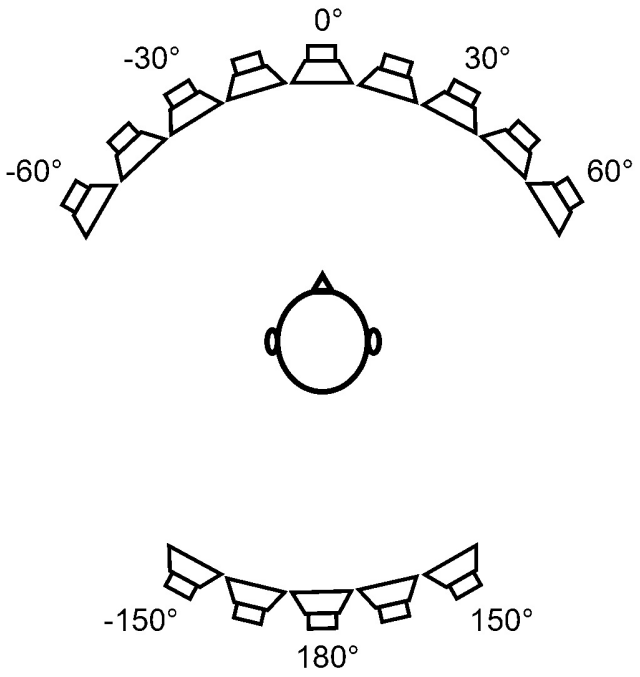

Figure 5.1
Loudspeaker setup used for the localization experiment.

5.1.2.5 Response Method

The participants indicated the perceived azimuthal direction of the presented stimuli using the Proprioception Decoupled Pointer (ProDePo, [Seeber, 2002]), with which they could move a laser pointer to the desired direction by moving the ball, positioned on the upper part of a mouse, and click the left mouse button to confirm. To indicate positions in the back without turning their heads, this being difficult due to the dummy hearing aid's cabling, they were instructed to indicate the corresponding angle in the front (i.e., the mirrored angle) and to click the right button of the trackball-mouse to tell the software (MATLAB) that the direction was to be mirrored.

5.1.2.6 HRTF Measurement

Since localization strongly depends on the ITD and ILD cues ([Rayleigh, 1907]), the binaural cues conveyed in the signals delivered to the listener's ears were investigated for all (baseline and HA) conditions and for all stimulus angles. To that end, HRTFs of one subject were measured using the microphones of the hearing-aid shells, as well as, for comparison, with miniature microphones (Sennheiser KE4-211) positioned in the ear canals. The HRTFs were measured in the same acoustically dry room where the experiments were conducted. As the room was not anechoic, the impulse responses that

were measured using MLS (Maximum-Length Sequence) signals, were faded out before the arrival of the first reflections.

5.1.2.7 Baseline

Five participants took part in a separate unaided baseline experiment that aimed to test the accuracy of the ProDePo response system and to pilot-test the conditions for the main experiment. Here, localization was tested in the whole azimuthal range from 0° to 360° using mainly 15° spacing between the active loudspeakers (i.e., the ones used for stimulus presentation). Motivated by the better localization accuracy of humans in the frontal region ([Akeroyd, 2014]), a denser spacing of 7.5° was chosen for directions within ±30° in the front. Otherwise, the same stimuli and experimental setup as in the aided conditions were used [Kolotzek et al., 2018]. Results of the localization errors of the baseline experiment are shown in the upper left corners of Figs. 5.2 and 5.3 (for the same range as the HA conditions).

5.1.3 Results

Figures 5.2 and 5.3 show median localization errors across all participants for the unaided baseline condition (five NH subjects) and all aided conditions (eight NH subjects). The dashed horizontal line in each graph represents the ideal response pattern. Here, the responses have been corrected for front-back confusions before plotting, by mirroring responses of the incorrect hemisphere to lie also in the hemisphere the sound was presented from. As can be seen, localization results for the front in the unaided condition are very good with less than three degrees localization error. For the aided conditions, the ITE and Jackrabbit conditions show low localization errors of less than ten degrees, while the BTE, STA BF and BF conditions all show localization errors of up to twenty degrees. All aided conditions show a positive error slope. For sound presentation from the back (smaller angle range), localization is worst for the STA BF and BF conditions, while all other aided conditions (ITE, BTE and Jackrabbit) show small localization errors of less than eight degrees. The unaided condition shows significantly higher localization errors than for the front. All aided conditions show a negative error slope for sounds from the back.

For the statistical analysis, individual localization errors were computed for the participants as the median across the four trials for each sound direction. A multifactorial ANOVA analysis was performed to investigate the effect of the hearing-aid algorithms on the absolute localization errors. The aided conditions and direction from which the stimulus was presented, were modelled as fixed factors while the subject was modelled as a random factor. Separate analyses were performed on the frontal and rear stimulus presentation scenarios. The localization errors depended significantly on the type of HA processing ($F(2419, 4) = 7.51, p < 0.001$), as shown in Figures 5.2, 5.3 and Table 5.1. Tukey's HSD post-hoc analysis revealed further that there was no statistically significant difference ($alpha = 0.05$) between the ITE and Jackrabbit conditions for frontal sound sources, while both distinguish themselves from the BTE, STA BF and BF conditions.

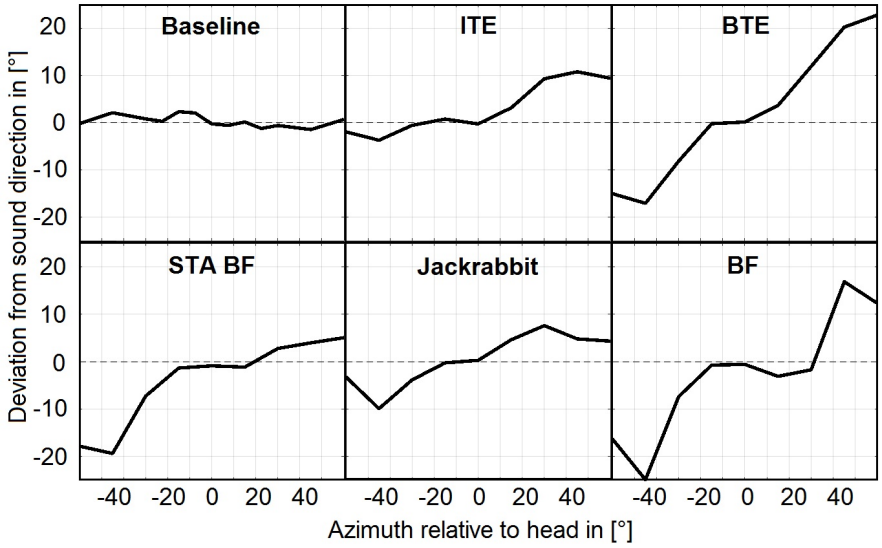

Figure 5.2
Median localization errors for the baseline and the main localization experiment for sounds presented within ±60° from the front. Five and eight normal-hearing listeners participated in the baseline and main experiments, respectively. The values in the graphs were calculated correcting for front-back confusions.

No significant differences were found between the BTE, STA BF and BF conditions either. The stimulus direction had also a significant effect on the localization error ($F(8022, 8) = 9.44, p < 0.001$). The localization error depended also on the interaction between stimulus direction and HA condition ($F(3221, 32) = 2.18, p < 0.001$). Tukey's HSD post-hoc test revealed that the significance of the interaction stems from the differences between HA conditions being significant only at certain stimulus angles: No differences exist between the conditions for angles within ±30°, nor when the stimuli were presented from −60°. At +45°, the BTE and BF conditions significantly differed from the ITE, STA BF and Jackrabbit conditions, and at +60° the BTE condition significantly differed from the ITE, STA BF and Jackrabbit conditions. An ANOVA analysis on the localization errors for sound source directions in the back did not reveal any significant effects for the HA condition, the stimulus direction nor for the interaction between the two terms.

Table 5.1 shows the mean localization errors for all 6 conditions for corrected data disregarding front-back confusions. For the frontal ±30° region all conditions show relatively small errors, with only the BTE condition exhibiting some greater errors. When looking at the broader range of azimuthal directions between ±60°, we start seeing greater errors, with the ITE and Jackrabbit conditions showing better accuracy than the other aided conditions. In the back, the two beamforming algorithms STA BF and BF show large errors. The baseline errors are far greater than in the front showing a similar accuracy than in the aided conditions without beamforming. While in the front we observe shifts of localization away from the center, expanding the auditory

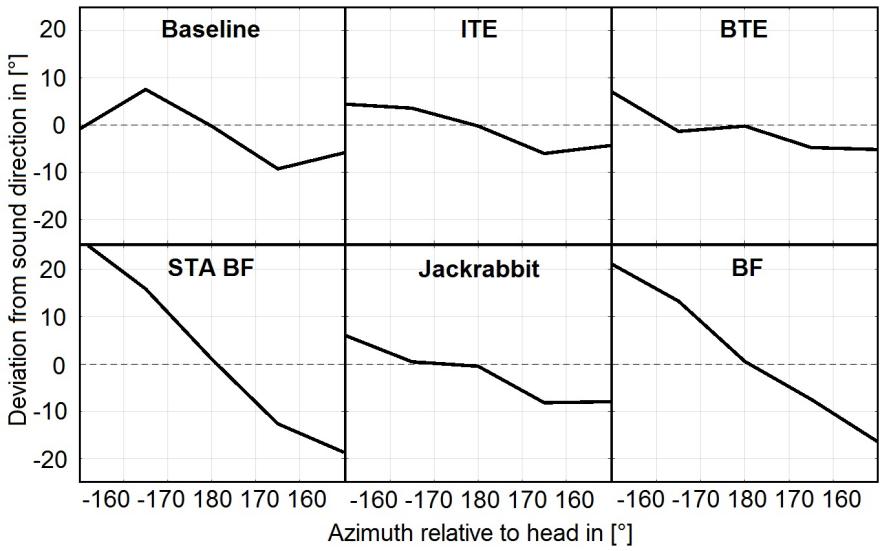

Figure 5.3
Median localization errors of the baseline and the aided experiment for sounds from the back between −150° and 150° for eight normal hearing participants, disregarding front-back confusions.

space, shown by the positive slope of the lines going from negative values to positive values as the presentation angle increases, the opposite seems to be the case for sounds presented in the back. There, the slope is negative which means that lateral sounds are mapped towards the center rear (180°), thus exhibiting a compression of the auditory space. To further investigate the reason for the shifts seen in the localization results, HRTF measurements were performed to inspect the differences in binaural cues (ITDs and ILDs) between the aided conditions and the baseline condition. ITDs were extracted from low-pass filtered head-related impulse responses (HRIRs) that were measured for one of the participants. The ILDs were extracted from the hearing aid HRTFs and the median ILD for a given direction within different frequency bands was computed. Specifically, four frequency bands were used for analysis: 0.2 – 1.5 kHz, 1.5 – 3.5 kHz, 3.6 – 5.6 kHz and 5.7 – 8 kHz. Fig. 5.4 shows the obtained ITD estimates in microseconds for BTE (red), ITE (blue) and the binaural microphones (i.e. the baseline condition; black) extracted as a linear regression of measured ITDs (since the ITDs in that region resemble a sine function which can be approximated by a linear function). On the left side of the dashed vertical line, ITDs are shown for sounds coming from the front. Here, larger than normal ITDs are found for the ITEs and even larger for the BTEs. On the right side of the dashed vertical line, the ITDs are shown for sounds from the back. Here, in contrast to what was found for frontal sources, the hearing-aid ITDs are smaller than the ones extracted from HRTFs measured with binaural microphones inside the ear canals. Moreover, the ITE and BTE ITDs are almost identical for sources behind the listener.

Table 5.1
Mean deviation from target for all hearing-aid conditions across data that was corrected for front-back confusions. Values are computed across errors for sound directions between ±30° and ±60° in the front, and ±30° in the back.

	Baseline	BTE	ITE	STA BF	Jackrabbit	BF
Mean Error front ±30°	0.9° sd: 0.8°	4.8° sd: 5.1°	2.8° sd: 3.8°	2.7° sd: 2.7°	3.3° sd: 3.1°	2.7° sd: 2.8°
Mean Error front ±60°	1.0° sd: 0.8°	11.0° sd: 8.5°	4.4° sd: 4.2°	6.6° sd: 7.1°	4.3° sd: 3.1°	9.3° sd: 8.7°
Mean Error back ±30°	4.8° sd: 4.0°	3.7° sd: 2.8°	3.7° sd: 2.2°	14.9° sd: 9.3°	4.6° sd: 3.9°	11.7° sd: 8.0°

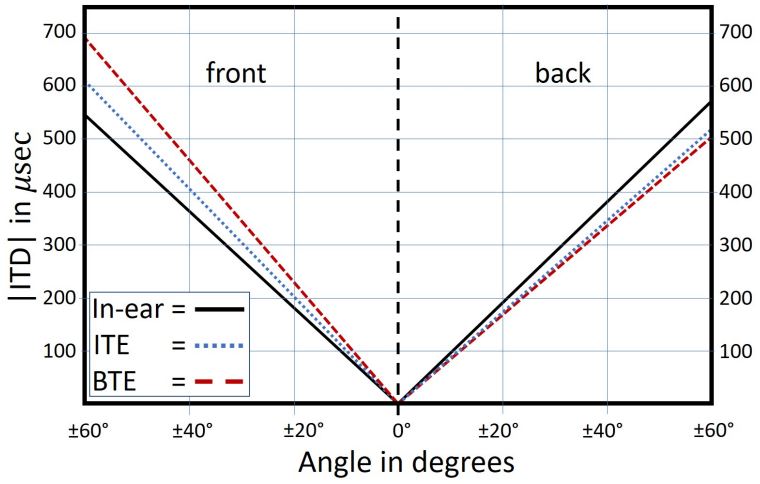

Figure 5.4
Absolute value of the ITD linear regression in µs for the binaural microphones (black), ITE (dotted blue) and BTE (dashed red) microphone positions for sounds from the front (left side) and back (right side) between ±60°.

Fig. 5.5 shows the obtained ILD estimates as a function of the azimuthal angle for the binaural recordings (black), the ITEs (blue) and BTEs (red). Separate values were extracted for four different frequency regions because the of frequency-dependency of ILDs for a given azimuthal direction ([Kuhn, 1987, Middlebrooks et al., 1989, Musicant and Butler, 1984]). The lowest frequency region from 200 Hz to 1.5 kHz is the region that is usually dominated by ITDs in terms of localization ([Hartmann et al., 2016]). For higher frequencies, ITDs become ambiguous and the ILDs are the dominating binaural cue for localization ([Macpherson and Middlebrooks, 2002]). Fig. 5.5 shows that the ILDs of the 2nd and 3rd frequency bands of 1.5 – 3.5 kHz and 3.6 – 5.6 kHz closely resemble the ITD pattern seen before in Fig. 5.4. That is, larger ILDs

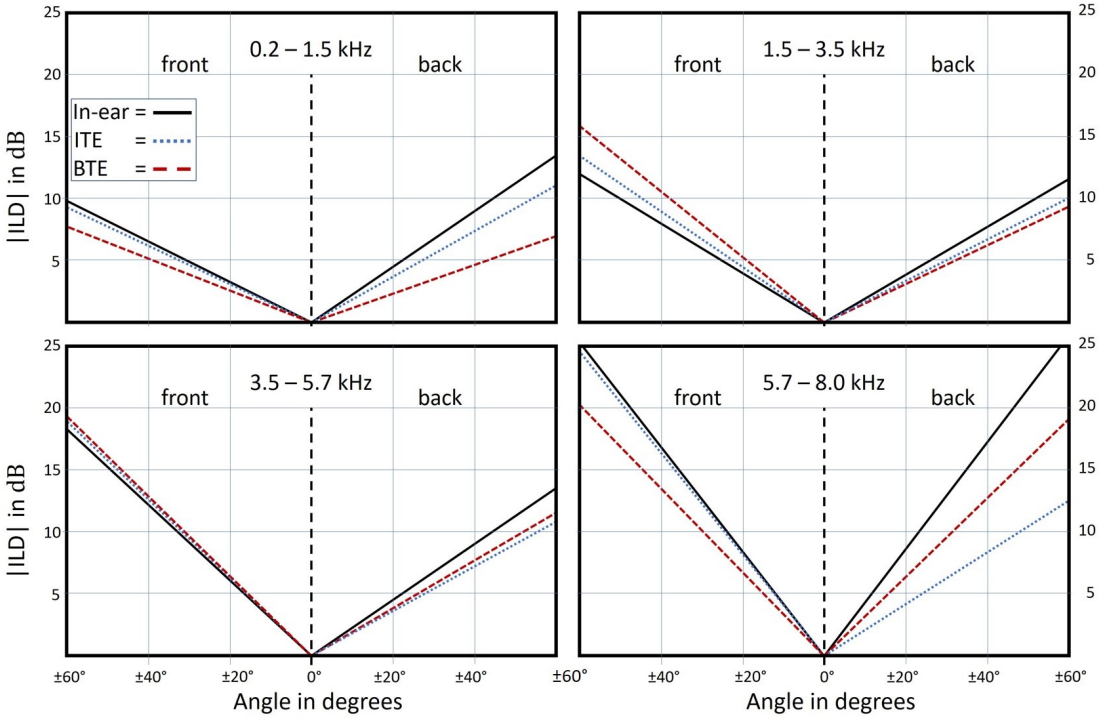

Figure 5.5
Absolute values of the ILD estimates[dB] obtained via linear regression of the extracted values from HRTFs measured with binaural microphones (black), and with microphones of the ITE (dotted blue) and BTE (dashed red) hearing aids. Values were computed for sounds using four different frequency-analysis regions between 200 Hz and 8 kHz.

are observed for HRTFs measured with the hearing-aid microphones for frontal sounds than for HRTFs measured with binaural microphones, while the opposite is observed for sound sources behind the listener.

Considering frontal sounds, the performance of the BTE and BF conditions is the worst, with 40% of sounds being localized to the back (Fig. 5.6). The STA BF has a reduced amount of confusions, but it still falls short of the performance of the ITE and Jackrabbit conditions, having the smallest amount (about 12%) of front-back confusions. All but the BTE condition perform well with sounds presented from behind the listener, with less than 12% confusions. In the BTE condition, 55% of the sounds were perceived as coming from the front.

5.1.4 Discussion

Results from this experiment show three important aspects of sound localization with hearing aids. First, as has been repeatedly reported in the literature ([Van den Bogaert et al., 2006, Akeroyd, 2014]), localization accuracy drops significantly from the unaided condition when hearing aids are worn. Here, the average error

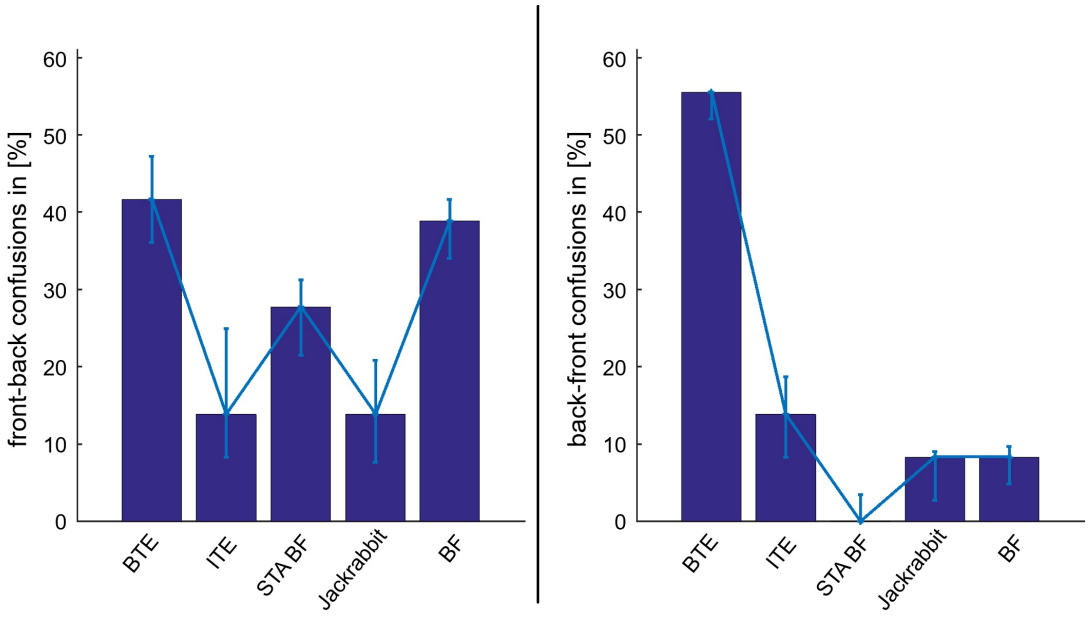

Figure 5.6
Median percentage of front-back and back-front confusions for the different hearing aid conditions. The error bars represent the 25^{th} and 75^{th} percentiles of the confusions over the presentation angles ±60° in the front and ±30° in the back.

for the baseline condition was only about 0.9° - 1°, which was far smaller than the average errors for the aided conditions, being approximately 2.7° to 4.8° for sounds presented within±30°, and about 4.3° - 11° for sounds presented within ±60°. Yet, this was found to hold only for the frontal region. For sounds presented from behind the listener, the errors in the baseline condition are equal to those of the ITE and BTE conditions. Second, an expansion of the auditory space was found for sources presented in the front when listening through hearing aids, while the opposite, a compression of the auditory space was observed for rear sources, at least for the angles that were tested here (between−150° and 150°). In general, out of all the aided conditions, the ITE and Jackrabbit performed the best, with the smallest localization errors both in the front and back. Third, the amount of front-back confusions and of back-front confusions is very high for the BTE condition. With up to 42% - 55% of front-back and back-front confusions, respectively, the front-back discrimination with BTE shells is close to random guessing for a static head position. The beam-forming algorithms STA BF and BF also had a high number of front-back confusions (between 30% - 40%) but only in a small number of back-front confusions. This discrepancy could be due to the large attenuation in the STA BF and BF algorithms that helps to discriminate back from front due to the large level differences, since sounds from the back were significantly attenuated by the beamformer's notch at 180 degrees compared to the sounds from the front. Spectral differences might have also played a role in the discrimination task. The lower number of confusions in the

STA BF condition, compared to the BF condition, both in the front and back is most likely due to the additional filtering that introduces pinna cues into the beam-formed signal, thus increasing the spectral frequency-dependent level differences between front and back and thus helping to distinguish whether a sound comes from the front or back. The 10% rate of confusions for the ITE and Jackrabbit conditions are in accordance with previous experiments, (e.g. [Gomez and Seeber, 2015b, Gomez, 2016]), and occur more often in participants with little experience on wearing hearing aids compared to more experienced subjects.

The localization errors seen for the aided conditions can mainly be explained by the ILD and ITD values that were extracted from the measured HRTFs. The localization results of the aided conditions correlate well with increased ILD and ITD differences between the hearing-aid microphone positions and the binaural recordings in the ear canal. The larger ILD and ITD values found for hearing-aid recordings explain the perceived expansion of auditory space for frontal sound sources and the smaller values explain the contraction in the back. The about 5-dB difference in ILD between the BTE and the binaural recordings in the frequency band of 1.5 - 3.5 kHz for sounds at 60° and the over 150μs ITD difference explain the localization errors rather well. We know from literature (e.g. [Akeroyd, 2014]) and from here performed binaural recordings that there is a linear about 10μs per degree correspondence in localization in the range from $-60°$ to 60°. For ILDs, in that azimuthal range, there is roughly a 0.25-dB per degree correspondence ([Middlebrooks et al., 1989, Wightman and Kistler, 1997]). The overly large binaural cues observed for the BTE microphone position suggest a localization bias of 15° – 20° towards more lateral angles when a sound is presented from 60°, which is very well reflected in the localization results (see Fig. 5.2). On the other hand, for sounds from the back, the binaural cues conveyed by hearing aids are smaller than the ones captured with microphones in the ear canals. For sounds at 120° (i.e., at 60° in the back), ITD values for BTEs are about 70μs smaller than the nominal ones and ILD values are also about 3-dB smaller in the frequency bands of 1.5 – 3.5 kHz and 3.6 – 5.6 kHz. This implies about 6° - 7° compression in spatial perception of sounds from the back. In other words, with hearing aids, sounds are expected to be perceived closer to the centre (180°) than in the unaided case. These findings from the HRTF analysis are also very well reflected in the localization results in Fig. 5.3.

Fig. 5.5 shows that the ILD values are very different in the highest frequency region (between 5.7 – 8 kHz) between the binaural, ITE and BTE microphone positions. Such large differences could result in much larger localization errors than what was observed in the listening test results. Therefore, this frequency region does not seem to play a dominant role for localization, at least not for broadband stimuli such as the ones used here. The differences in ILDs within the frequency region of 1.5 – 3.5 kHz and possibly also the 3.6– 5.6 kHz region seem to best reflect the results of our localization experiment. The ITD values were extracted from low-pass filtered HRIRs, containing information up to 1.5 kHz, and the differences in those values are also well reflected in the listening test results. The large ILD and ITD distortions are well in accordance with the measurements by Udesen [Udesen et al., 2013].

Recently, Kolotzek et al. (2018) performed an extended experiment, investigating the effect of head-orientation on localization. They found that an eccentric head orientation significantly helps to solve front-back confusions due to the asymmetric reflection pattern at the torso [Kolotzek et al., 2018]. Pertaining to the findings of the present study, they also found that the expansion of the auditory space with increasing azimuth is even more prominent for eccentric head positions, shifting the perceived azimuth of sounds originating in the frontal hemisphere away from the centre. The experiment in the present study was conducted with eight normal-hearing participants that wore hearing aid prototypes and held their heads still during the sound stimulus presentation. Apart from having a short familiarization session, the participants did not go through any training for the task. While very high localization errors and a high number of front-back confusions were observed, these errors are likely to be smaller for people accustomed to wearing hearing aids. HA users get used to the altered binaural cues and learn to map the new binaural cues to the correct location of sounds, especially with the help of visual feedback of sound source locations in space. As for hearing-impaired listeners, wearing different types of hearing aids with open or closed fittings, their localization accuracy can vary significantly between subjects. Open fittings may allow for natural ITDs, but problems can still arise due to comb filtering that results from the signal from the hearing aid reaching the eardrums later than the sound wave traversing naturally through the open ear canal entrance. Furthermore, as shown here, the delayed signal from the hearing aid conveys contradictory ITDs and ILDs. In addition, alteration of ILDs due to unlinked compression and other signal processing algorithms can also play a significant role in the localization performance of hearing aid users [Wiggins and Seeber, 2011, Musa-Shufani et al., 2006].

5.1.5 Conclusions

The present study investigated azimuthal localization of normal-hearing participants wearing hearing aids. Different hearing-aid conditions, based on either the ITE, BTE microphone position or combination of those, were tested and compared against an unaided baseline condition. Left-right localization (without considering front-back confusions) of sounds from the front showed large errors (up to 20° in the ±60° range) for the BTE and beam-forming conditions. The ITE condition and novel Jackrabbit method performed the best, achieving similar results and having only small localization errors of 5°-10° in the ±60° range. The baseline results showed almost perfect localization with errors under 3°. In the aided conditions, an expansion of the perceived auditory space was observed as positive slopes of the localization errors. This means, that sounds were perceived further away from the centre (0°) than from where they were originally presented. In the back, the baseline performance was much less accurate and comparable to the aided BTE, ITE and Jackrabbit conditions, having errors up to 10° in the ±30° range in the back. There, the beam-forming algorithms STA BF and BF performed very poorly. For sounds presented in the rear, a compression of the perceived auditory space was observed with hearing aids in the form of negative slopes of the localization errors. Sounds in the back were perceived much closer to the midline (180°) than their

original presentation angle. Numerical analysis of binaural cues from HRTF measurements revealed differences in ITDs and ILDs that account very well for the observed localization errors. For frontal sounds, ILD values of HA recordings were larger in the frequency regions between 1.5 – 5.6 kHz than the ones of normal HRTFs and ITDs were generally larger as well. In contrast, for sounds at the back, both the ITD and ILD values were smaller in HA recordings. These differences in binaural cues depending on the microphone position may also explain the localization results from the spatial perception experiment in Chapter 3. The largest amount of front-back confusions occurred in the BTE and BF conditions, while only the BTE condition suffered from a significant amount of back-front confusions. Results from this experiment demonstrate that localization with hearing aids is affected by the altered binaural cues, and that localization with beam-formers is very poor. The Jackrabbit method (Chapter 4), which combines the preservation of pinna cues and enables a high SNR by attenuating disturbers, performed equally well to the ITE condition and did not exhibit any of the drawbacks of the beam-forming conditions. It could be considered as a method that preserves correct spatial perception, while maintaining the SNR-gain capabilities of beam-forming algorithms. The good results of the ITE and Jackrabbit conditions demonstrate the importance of preserving pinna cues for the correct localization of spatial sounds, at least for those who can exploit these cues. Additional testing is needed to verify these findings with HI subjects. The individual weightings of altered ITDs and ILDs, and of specific frequency bands on localization accuracy should also be investigated further. In conclusion, results of this study bolster the idea that in the design of novel hearing aid algorithms, one should consider the preservation of correct cues to avoid a degradation in the spatial perception of sounds.

5.2 Validation of Pinna Cues Preserving Algorithms on Speech Understanding

5.2.1 Summary

Chapter 4 presented two novel methods for noise reduction in hearing aids, methods which combine the benefits of a higher SNR with the preservation of pinna cues important for spatial hearing. This section presents a validation study that investigated the performance of different hearing aid conditions on speech understanding of eight normal-hearing participants wearing hearing aids with linear amplification [Lu, 2017]. The two novel methods were compared to the ITE, BTE and BF hearing aid conditions. Two stimulus conditions were tested. The first was a one-noise (1N) condition with target OLSA sentences coming from 0° in the front and OLSA noise from 180° in the back. The second two-noise (2N) condition comprised target OLSA sentences being presented from 0° in the front together with babble noise (from speakers of the same gender as the target speech) being presented from 165° and a 10-ms delayed undamped reflection of the noise originating from 7.5°. The participant group was split into two groups. Four German native speakers conducted the German OLSA test with male speaker sentences, while the remaining four English natives or fluent speakers conducted the English OLSA test with female speaker sentences. Results show a benefit of microphone directionality. The ITE condition showed a 3.6-dB lower threshold than the BTE condition, and the three beam-forming methods yielded in 11-dB and 6.4-dB lower thresholds than in the BTE in the 1N and 2N conditions, respectively. In addition to having low speech reception thresholds, the Jackrabbit method retained the spatial quality benefits of the ITE (pinna cues) without annoying low-frequency noise enhancement due to roll-off compensation.

5.2.2 Methods

The present study examined the effects of different hearing-aid microphone positions and of beam-forming algorithms on speech understanding of normal-hearing (NH) participants. The advantage of testing NH is that the effect of the hearing aid devices, i.e. the microphone position or effect of specific algorithms, can be tested without hearing loss and different devices affecting the results, like it would inevitably be the case when testing with hearing-impaired participants. In this study, two novel algorithms that combine noise reduction with preservation of pinna cues, i.e. the STA BF and Jackrabbit algorithms, are compared to a standard delay-and-subtract beam-former and the BTE and ITE signals. We chose the OLSA test to assess the speech understanding because its implementation in MATLAB enables automatic recognition of correct sentence words, because it is frequently used in scientific studies, and because both the German and English versions were available for our German and non-German participants. The test was modified in respect to the interferer stimulus. As the aim was to test beam-forming algorithms with attenuation towards the back, a speech shaped noise (SSN) masker was presented from 180° in the 1N condition. Better ear listening is not possible in this condition as the interferer affects both ears in a similar manner. To test a more challenging

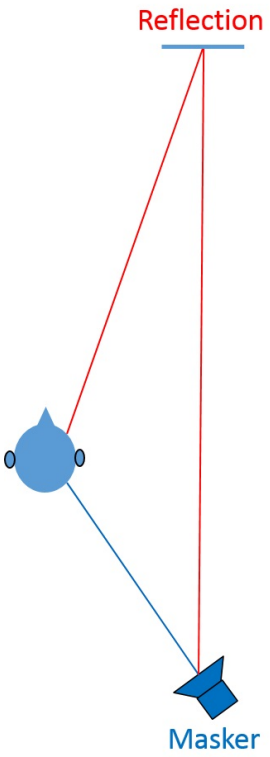

Figure 5.7
Schematic diagram of the 2N masker condition with a masker in the back at 165°*, and an ideal reflection from the front at* 7.5°*. The reflection being delayed by 10 ms.*

situation, we simulated the 2N condition that consisted of a different interferer type, a speech babble masker, located at 165° and a simulated ideal reflection of that masker coming from 7.5°, as shown in Fig. 5.7. The reflection of the masker was delayed by 10 ms, the time of the sound distance difference of 3.4 meters.

The following research questions were formulated: (a) What are the performance differences between the ITE and BTE microphone position? (b) What is the benefit of the beam-forming algorithms compared to ITE and BTE? (c) Is there an advantage of the STA BF, which includes pinna cue information, over the static BF? (d) Does the Jackrabbit method perform similar to the STA BF and BF, despite it selectively filtering out spectral energy of the disturber sounds from the ITE signal and not attenuating certain directions like in traditional beamforming?

5.2.2.1 Hearing Aid Sound Presentation

The hearing aids used in the experiment were custom-made ITE and BTE prototypes by Phonak. The prototypes were connected over cables to a PC on which ran a real-time Simulink model of HA signal processing with a total delay of only 7.8 ms. Frequency-responses of the microphones and the receivers were equalized in the frequency domain.

In addition, a 5-dB gain was applied to the signals after loudness compensation to mask direct sound leakage through the hearing aid shells. Loudness compensation was done by recursively comparing the loudness of stimuli in the ITE aided condition to the unaided condition both bilaterally and unilaterally fitted, and adjusting the gain in the Simulink model to match to the same loudness. The hearing-aid conditions used in this experiment were the processed ITE and BTE signals, a BF with maximum attenuation towards 180° in the back, and two novel beam-forming algorithms: the STA BF, a beam-forming signal that is dynamically filtered with short-time averaged (STA) pinna cues, and the Jackrabbit method, which preserves natural pinna cues and attenuates disturber energy efficiently.

5.2.2.2 Participants

Eight normal-hearing participants (8 male, average age 22 - 35 years) took part in the experiment. Four were native German speakers, the other four were either English native speakers or had excellent English language skills. Everyone had normal hearing thresholds as verified with a calibrated Békésy tracking audiometer [Von Békésy and Wever, 1960, Seeber et al., 2003] in a sound isolated listening booth [Frank, 2000]. All participants had previously taken part in hearing experiments and had used their hearing aid prototypes previously. The TUM ethics committee approved this study.

5.2.2.3 Stimuli and Experimental Procedure

For the German OLSA test, a male speaker was used as target and presented from the frontal loudspeaker of the SOFE apparatus [Seeber et al., 2010]. Target and noise were of the same gender and started at 67-dB SPL instead of the usual 65-dB SPL. These measures were taken to minimize floor effects, which would result in very low SMR thresholds for the beam-forming algorithms, with a potential masking of the target by the hearing aid microphone noise. The OLSA noise in the 1N condition was a scaled superposition of 100 OLSA sentences randomly shifted, with the same long-term spectrum as the target speech and no modulation that would have allowed glimpsing. The noise was presented from 180°, i.e. directly behind the listener. For the 2N condition, a speech babble masker was presented from 165° and a delayed version of the masker from 7.5°. The delayed masker acted as an ideal reflection delayed by 10 ms, the time of the sound distance difference of 3.4 meters (Fig. 5.7). The speech babble masker was a randomly selected section of a longer signal that was constructed by concatenating and super positioning speech sentences of the same sex as the target, yet of a different speaker. Since the OLSA masker was defined to be presented at 67-dB SPL, the theoretical 3-dB intensity sum of two uncorrelated sources was applied and each of the maskers in the 2N condition was presented at 64 dB SPL. The English OLSA test differed only in the speech material for the target and maskers being spoken by female English speakers. A different OLSA list was randomly assigned for each condition of the experiment, such that each participant was tested with ten lists for all combinations of

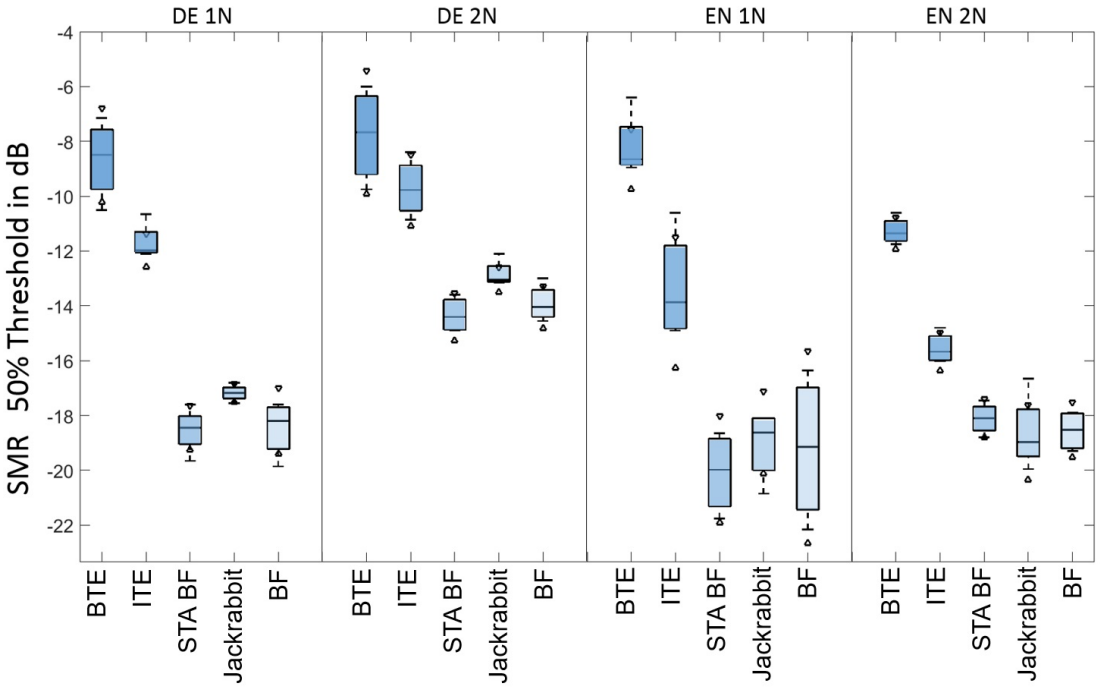

Figure 5.8
Signal to Masker ratios for the OLSA test for 50% correct word recognition. Boxplots represent the distribution of results for individual conditions for 4 subjects each. The squares represent the interquartile range (IQR) including the 25^{th} to 75^{th} quartiles. Whiskers extend to $1.5 * IQR$ representing 2.7 standard deviations from the mean for normally distributed data. The horizontal line in each boxplot shows the median result, while the triangles show the 5% significance intervals, such that overlapping intervals are not significantly different.

the five hearing-aid conditions and the two masker conditions. The presentation order of the experiment was completely randomized across trials and conditions. The experiment took place in complete darkness, participants were seated in the middle of the SOFE loudspeaker ring wearing their hearing-aid shells. Each OLSA sentence was presented only once, after which the results had to be given on the GUI using a touchscreen, by pressing the buttons corresponding to the words thought to be in the presented sentence.

5.2.3 Results

Fig. 5.8 shows the results of the speech understanding experiment for all tested conditions, defined as the signal-to-masker ratio (SMR) for 50% correct speech recognition. What can be seen for all conditions and was expected is the higher (worse) threshold for the BTE and ITE conditions compared to the beamformer conditions. This difference is on average 8.23 dB in the 1N condition and 4.86 dB in the 2N condition. Also remarkable is the better performance of the ITE compared to the BTE in all cases, with an average lower threshold of 3.6 dB.

While there are big SMR differences between the German and the English results, the relative differences between the algorithms are similar. To better compare only the relative differences between algorithms, we normalized the data of the German and English results to a common maximal threshold of 0 dB, shifting all results up. Statistical analysis of the normalized data with a multifactorial ANOVA with algorithms, language and noise as main factors and subjects as random factor show that there are statistically significant differences in all main factors and 2-way interactions except for *subject*algorithms* and *language*noise*. A post-hoc analysis after Tukey for the factor algorithms yielded significant differences between the BTE and ITE conditions ($p < 0.0001$) and between BTE or ITE and any of the beamforming algorithms ($p < 0.0001$), but no significant differences between the three beamforming algorithms. The same is true for the interaction term of *algorithm*language* when looking at the languages separately ($p < 0.0001$ for all significant differences). Also, for the interaction term of *algorithm*noise* there are significant differences between ITE and BTE conditions, but no differences between the beamformer algorithms. When comparing conditions across noises, there are no differences between the conditions BTE 1N and BTE 2N, no differences between ITE 1N and ITE 2N, but differences between the beamformer algorithms for the 1N and 2N conditions of 4 dB, with lower (better) thresholds in the 1N condition.

5.2.4 Discussion

For the English version beamformer algorithms we obtained very large SMR levels of up to -22 dB. We tried to avoid floor effects by increasing the masker and target level from 65 dB SPL to 67 dB SPL and using same sex maskers. While these effects could have been avoided by additionally increasing the level of the masker instead of only lowering the target level, this would have had two disadvantages. First, the hearing aid gain was already high to mask any direct sounds in the low frequencies, where the attenuation of the hearing aid shells is low. Increasing the level of the masker would have resulted in uncomfortable loudness for the participants. Additionally, the masking pattern of the noise would have significantly changed between levels and conditions, i.e. the spectral masking of the masker sounds, which is level dependent, would have resulted in additional level dependent influences on the speech understanding test. We could verify the normal distribution of the data even at high SMR values by plotting histograms of the different conditions. There was no skewness in the results which would have shown in case of floor or ceiling effects.

The better performance of the ITE compared to the BTE clearly shows the benefit of using pinna cues for speech understanding in complex acoustic scenes. In all tested conditions the ITE has a lower threshold than the BTE, on average 3.6 dB. This value is higher than the 2.39 dB reported by ([Pumford et al., 2000]). This is also higher than the just noticeable difference in speech-to-noise ratio of 3 dB found by McShefferty et al. (2015) [McShefferty et al., 2015], which is independent of hearing impairment. The threshold difference between ITE and BTE may have different reasons. The acoustic shadow of the pinna can attenuate disturbing sounds from the back at higher frequen-

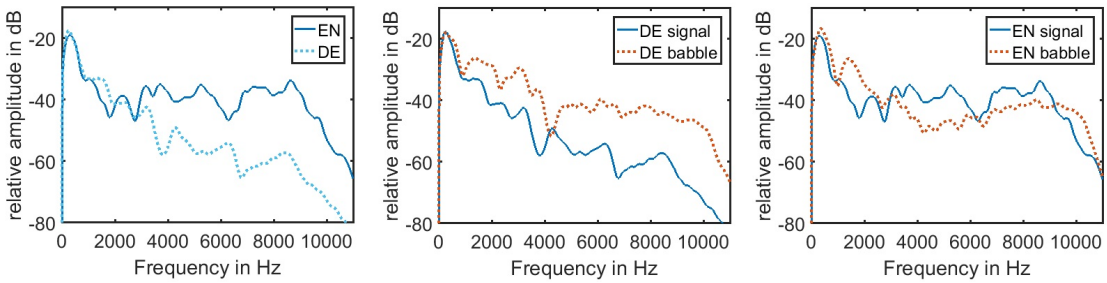

Figure 5.9
Relative amplitude spectra of the German and the English OLSA speech target sentences (left), and of the target speech sentences to the noise babble.

cies, while the BTE has a general directivity gain of a few dB towards the back. This directivity difference between ITE and BTE could partly explain the threshold differences in the 1N condition (4.18 dB) with the masker at 180°. In the 2N condition, the threshold difference is somewhat lower, yet still on average 3 dB. The simultaneous presentation of a same sex masker from the front and back lowered directivity advantages of the ITE, yet there is still a benefit from ITE directivity, since high frequency consonant information, which is important for speech understanding, is still higher weighted energetically for the ITE than for the BTE. Additionally, the easier spatial separability of sources between the front and back in the ITE case due to natural pinna cues can also have improved the ITE results. For comparison, piloting sessions with two subjects showed about -11 dB SMR in the unaided case for the 1N condition. In the BTE case, results may have been lowered due to a loss of spatial release from masking in the presence of front-back confusions or worse sound object separation. In the 2N condition, the advantages of spatial release from masking from opposite hemispheres is reduced, since the lateral offset of the frontal maskers relative to the target was small. Here, the better ear listening starts to influence the speech understanding results, since the left ear experiences lower masker levels than the right ear. While in the English version there is a significant threshold reduction in the 2N condition compared to the 1N condition, we do not see this effect in the German version, rather a slight threshold increase. The differences between the German and the English test were the different talkers, male for the German version and female for the English version, and the corresponding same sex maskers. Why these differences resulted in significantly different thresholds between both languages could possibly be explained when analyzing the sentences. Fig. 5.9 shows the average amplitude spectra of the OLSA sentences. From about 2.8 kHz onwards the English female speech has a higher energy density than the German male speech, which is the most important frequency region for consonant recognition.

Additionally, we analyzed the ratio of energetic parts of the sentence to quite parts, by setting a threshold at -20 dB from the maximum of the speech time signal envelope for the ratio. There is no difference in ratios between the German and the English version, with respective mean ratios of 0.8 and 0.82. Since the stimuli were levelled using RMS levels, it becomes clear that the English speech stimuli must have had spectral (Fig. 5.9)

or temporal content helpful for speech understanding, especially noticeable in the ITE condition, resulting in lower thresholds in the experiment. The noise in both languages was generated in the same manner, differing only in its spectrum. Temporal glimpsing was not possible since both the OLSA noise and the babble noise in the 2N condition was a superposition of several random speech sections, filling the temporal gaps of normal speech modulation. As seen in Fig. 5.9 (middle and right) the noise babble of the German male condition was spectrally more energetic throughout the entire spectrum than the target male speech, leading to a higher degree of masking of spectral components important for speech understanding. In the English condition with female speech babble we see an energetic dominance of the babble only at lower frequencies. While the level of the target speech decreased in the experiment, the German condition had always higher spectral masking levels than in the English condition. Note that in the 1N condition, the long-term spectrum of the noise and of the target speech was identical, since the noise was generated using a high number of superposed target speech sentences. For the analysis of relative differences between algorithms the results of both languages are still comparable and the spectral differences between both languages only led to shifted levels in absolute terms and to differences due to microphone directivities. The three beamforming algorithms STA BF, Jackrabbit and the standard BF performed equally well in all conditions. The average difference between the STA BF and the BF conditions is only 0.15 dB, and the Jackrabbit difference to the mean of the latter two only 0.67 dB, which is statistically insignificant and would not be discriminated in level differences or in speech-to-masker ratio differences, lying well below the corresponding JNDs. The STA BF and BF algorithms physically attenuate the maskers from the back by a subtraction process. In the 1N condition, the masker gets attenuated by 11 dB when comparing the BTE and BF conditions, leading to a very low SMR threshold. In comparison, a disturber from 90° instead of 180° would have only resulted in 3-4 dB lower thresholds ([Wouters et al., 1999]). In the 2N condition, the masker in the back gets attenuated, yet the masker in the front is not affected by the beamforming. In practice, this results in a smaller beamforming benefit than is reflected in the SMR results of 6.4 dB, since each individual masker has a 64 dB SPL level to sum up to 67 at the ears, and only the rear one is attenuated. Thus, the beamforming benefit in the 2N conditions is only about 3.4 dB (when considering level addition of 64 dB from the front and about 50-55dB from the attenuated back). The Jackrabbit algorithm, on the other hand, uses a spatial spectral filter to attenuate masker components out of the target speech signal. It is remarkable that even though the masker in the 1N condition has the same long-term spectrum as the target speech, the filter combined with the ITE benefit can still attenuate the masker to reach the low thresholds of traditional beamforming methods, while retaining natural pinna cues from the ITE microphone position and the underlying benefits thereof. In the 2N condition, the Jackrabbit also attenuates the masker sounds to the same SMR as the STA BF and BF algorithms. In this experiment, the spectrum of the masker and that of the target were similar, since they were of the same sex, although as shown in Fig. 5.9 there were spectral differences at higher frequencies. It is expected that an even better performance could be achieved with the Jackrabbit method when the masker has a different spectrum than the target. The average benefit of the Jackrabbit filter in this

same sex masker experiment is 4.3 dB compared to the ITE condition, with an average 5.5 dB difference benefit in the 1N condition and 3.1 dB in the 2N condition. Since the STA BF and BF algorithms had an integrated roll-off compensation which enhances the low frequencies and thus also the microphone noise to audible levels, this could have partially masked the low-level target. Yet we did not observe any floor effects in the data, as discussed previously, and the high frequency consonant information critical for speech understanding should not have been spectrally masked by the lower frequencies. Other than being annoying and potentially tiresome in the long run, we do not expect the audible low frequency microphone noise to have affected speech understanding results. In the Jackrabbit algorithm, there is no enhancement of the low frequencies caused by the roll-off compensation, since the ratio of cardioid to anti-cardioid is used for the filter calculation and therefore the roll-off compensation would cancel out, making it superfluous. Thus, this method is very advantageous for hearing aids. One last aspect that might have played a role in the speech understanding scores is the fact that sounds from the front are often perceived internalized when the STA BF and BF algorithms are active and the head is held still, as was the case here. Perceiving the target internalized while sounds from outside a $\pm 45°$ region are perceived more externalized, can potentially help to get a clearer object separation and might even be beneficial for speech understanding. This potential influence should be further investigated in future studies. In the 2N condition, where the target and the frontal masker are only separated by 7.5° it is not likely that internalization would have been helpful, since both the target and masker would have been perceived internalized, and only a small lateralization difference might have resulted that would probably not result in significant *spatial unmasking*.

5.2.5 Conclusions

This section presented the performance of two novel methods for increasing the SNR while attempting to preserve pinna cues on speech understanding, the STA BF and the Jackrabbit methods presented in Chapter 4. A validation experiment investigated differences in speech understanding thresholds between the two novel algorithms and the ITE, BTE and static beamforming conditions. A German and an English version of the OLSA test were used in combination with two different masker conditions. The BTE condition got the highest (worst) thresholds. The ITE condition led to an average of 3.6 dB better SMR thresholds than the BTE, showing clear benefits of pinna cues and natural directivity for speech understanding, additional to spatial quality benefits found in our previous studies ([Gomez and Seeber, 2015b, Gomez, 2016]). The two beamforming algorithms STA BF and the standard delay-and-subtract beamformer performed equally well, achieving very low thresholds. The beamforming led to 11 dB lower thresholds than in the BTE condition in the 1N case, and 6.4 dB in the 2N case. The combination of the ITE benefit and the spectral filtering of the Jackrabbit method yielded equally low thresholds than the STA BF and BF algorithms, while retaining the spatial quality benefits of the ITE and lacking disturbing low frequency noise enhancement due to roll-off compensation of the beamformers.

5.3 Investigating Spatial Sound Quality in Complex Acoustic Environments with Hearing Aid Prototypes

5.3.1 Summary

The present study investigates qualitative aspects of spatial sound perception beyond common localization assessments. The experiment is split in two parts, one that investigates the effect of time constants on the spatial perception of sounds, and the second part investigates the effect of different hearing aid conditions on the spatial perception of sounds. For this, five hearing aid conditions were tested for normal hearing participants wearing custom made ITE and BTE hearing aid dummies in a dark room. Adjective antonyms describing diverse spatial dimensions were rated in a semantic differential procedure for a target and a disturber sound separately, presented concurrently from 15° and 165°, once in a dry scene coming from two loudspeakers of a ring, and once embedded in a complex cafeteria scene using virtual acoustics. Results show significantly better spatial ratings for the ITE microphone position than for the BTE microphone position, and a benefit of beamformer directivity over omnidirectional microphones to attenuate disturbers and thus to achieve a better spatial separation of sound sources. Time constants with attack times of up to 4-5 ms can be used in hearing aids without affecting spatial sound quality. Best ratings of all aided conditions were achieved for the Jackrabbit method presented in Chapter 4, which preserves pinna cues and increases SNR by attenuating disturber energy, especially in the dimensions externalization, source separability, saliency and diffuseness. The unaided baseline condition was rated by far better than all aided conditions, showing large detrimental effects of hearing aids on spatial perception for normal hearing subjects, due to limiting factors such as a reduced bandwidth, altered or missing binaural and monaural cues, background noise, delay and occlusion effects.

5.3.2 Methods

In this study, we asked eight NH participants to rate different hearing-aid conditions using individual custom-made prototypes with linear amplification and no additional hearing-loss compensating measures. The conditions rated were the BTE and ITE microphone positions in omnidirectional mode, a static delay-and-subtract beamformer to the front, a beamformer with dynamically applied pinna cues (STA BF) and the novel Jackrabbit algorithm (Chapter 4) that preserves pinna cues and attenuates unwanted sound sources. In extensive piloting sessions, we noticed that a rating of spatial aspects of a sound is best when directly comparable to a different competing sound presented simultaneously from a different spatial location. Additionally, we were interested in testing for differences between microphone positions, such that the differences in monaural cues were an important aspect of the experiment. Therefore, we placed both competing sources on the same cone of confusion to eliminate effects of ITDs and ILDs. We used two different sound scenes for testing. The first always presented a male target talker either from the front or from the back, with a competing disturber presented from the opposite

hemisphere. The disturber stimuli were either female speech or noise, in a continuous modulated way or intermittent. Both the target and disturber were rated separately. The second, reverberant condition, used virtual acoustics and consisted of a simulation and auralization of a cafeteria, where we chose and simulated five different talker dialogs and one music source located at different positions in the virtual space, using the SOFE [Seeber et al., 2010]. From Kawashima and Sato (2015) [Kawashima and Sato, 2015] and Akeroyd et al. (2016) [Akeroyd and Whitmer, 2016] we know that simulating a larger number of sources than 4-5 is not necessary, since it cannot be discriminated by the listeners whether there are more sources present in the acoustic scene. The only difference in the sound fields, when adding more sources, is a smoothing of the stimulus envelope, and a decrease of the temporal gaps and of informational masking, since adding sources will make individual speech tokens less salient, gradually going over to a background speech babble. Two of four talker and disturber stimulus pairs from the dry condition, at the same virtual angle and distance, were reverberated and embedded in the acoustic scene. In both acoustic conditions, the dry and the cafeteria condition, participants had to rate spatial sound quality dimensions separately for both the talker and disturber, in the front and in the back.

The first part of this experiment was used to find suitable time constants for the smoothing of spectra with the short time average method presented in Chapter 4 [Kolotzek, 2016], and to assess how spatial aspects are affected by increased signal smoothing. Here, the adjective pair *static* − *moved* was included, since we experienced fused and moving sources for the spectro-temporal smoothing with larger time constants. For this first part, only the dry condition was used. In the second part, however, the time constants used were low and no movement was expected nor experienced in piloting sessions. Therefore, we replaced the *static–moved* adjective pair with the *salient* − *background* adjective pair, relating to how prominent or salient, and how audible the target or disturber stimulus was perceived in comparison to the acoustic scene, considering the attenuation of the beamforming algorithms. Especially in the cafeteria condition, we expected a clear difference regarding prominence and audibility for the beamformer algorithms compared to the ITE and BTE case.

5.3.2.1 Apparatus

The experiment was conducted in the SOFE v3 [Seeber et al., 2010], consisting of the apparatus and software to simulate and auralize room acoustics as explained below. The SOFE apparatus used in the experiments was located in the basement of a university building of the Technical University of Munich, with a completely dark test room (dimensions 6.8m x 3.9m x 3.3m) in which experiments took place, and an adjacent control room. The floor was carpeted, and the walls and ceiling were covered with sound absorbing curtains. The average T_{60} reverberation time was 0.16 s. The noise floor was 26 dBA measured with an NTI Audio XL2 sound level meter at the listeners position. A ring consisting of 96 loudspeakers (BOSE Freespace 3) with a diameter of 2.48 m at a height of 1.25 m was used for auralization ([Völk, 2010]). A chair with an adjustable head rest was placed in the center of the ring. A touchscreen with a touch pen and a

keyboard were used as input devices. Sound playback was controlled from a standard PC using 3 cascaded 36-channel RME Raydat soundcards. Digital to analog conversion was performed by six 16-channel 24 Bit AD/DA converters (Sonic Core A16 Ultra) which were connected to 48 Samson Servo 120a 2-channel amplifiers. The SOFE was calibrated to a linear frequency response between 200-10000 Hz in amplitude and phase, using a measurement microphone (G.R.A.S. 46 AF), powered and amplified (G.R.A.S. 12 AK) and connected to one of the six AD converters. The BTE and ITE microphones were equalized to compensate for the free field frequency response from the front from 200-8000 Hz with 65 tap FIR digital filters applied during experimental runtime.

5.3.2.2 Hearing Aid Devices and Sound Presentation

The hearing aids used in the experiment were custom-made ITE and BTE prototypes by Phonak, connected over cables to a PC on which ran a real time Simulink model with a total delay of only 7.8 ms. We used linear amplification and no additional hearing-loss compensating measures. We applied a frequency response equalization of the microphones and receivers in the frequency domain. In addition, we applied a 5-dB gain to the signals after loudness compensation to mask direct sound leaked through the hearing aid shells. Loudness compensation was done by recursively comparing the loudness of stimuli in the ITE aided condition to the unaided condition both bilaterally and unilaterally fitted, and adjusting the gain in the Simulink model to match to the same loudness. We also measured the ITE hearing aids' attenuation of external sounds by measuring the binaural hearing threshold with an external loudspeaker once with and once without ITE shells in place. The mean attenuation is greater than 10 dB for frequencies above 600Hz and greater than 20 dB above 1.2 kHz (Fig. 5.10).

We also measured the crosstalk (i.e. the feedback path) between the ITE-receiver playback and the pickup of that signal by the ITE microphones. Using loudness comparisons between the calibrated playback over the SOFE and the playback over the ITE-receivers with broadband noise (0.2-6 kHz) we calibrated the receiver level and subsequently measured the crosstalk given by the ITE-satellites. The average attenuation for the highest measured playback level (90 dB SPL) was greater than 40 dB, such that any crosstalk effects can be ruled out. The hearing aid conditions used in the first part of the experiment were the omnidirectional microphone signals of the ITE and BTE devices, and five different versions of a short time averaging (STA) of magnitude spectra using IIR filters, where the magnitude of the STA ITE was divided by the magnitude of the STA BTE using equal time constants, and taking that division term for a multiplication of the instantaneous BTE signal (eq. 4.6). The five different smoothing constants were chosen for consecutive doubling of the attack time between $2 - 35$ ms for the spectro-temporal smoothing (see Chapter 4). This applies different degrees of smoothed dynamic pinna cues to the instantaneous BTE signal.

The conditions used in the second part of the experiment were the ITE and BTE signals, a standard delay-and-subtract beamformer (BF), the STA method with attack time of 2 ms that applies dynamic pinna cues to the beamformer signal (eq. 4.5), and the Jackrabbit method that preserves pinna cues and attenuates unwanted sound sources

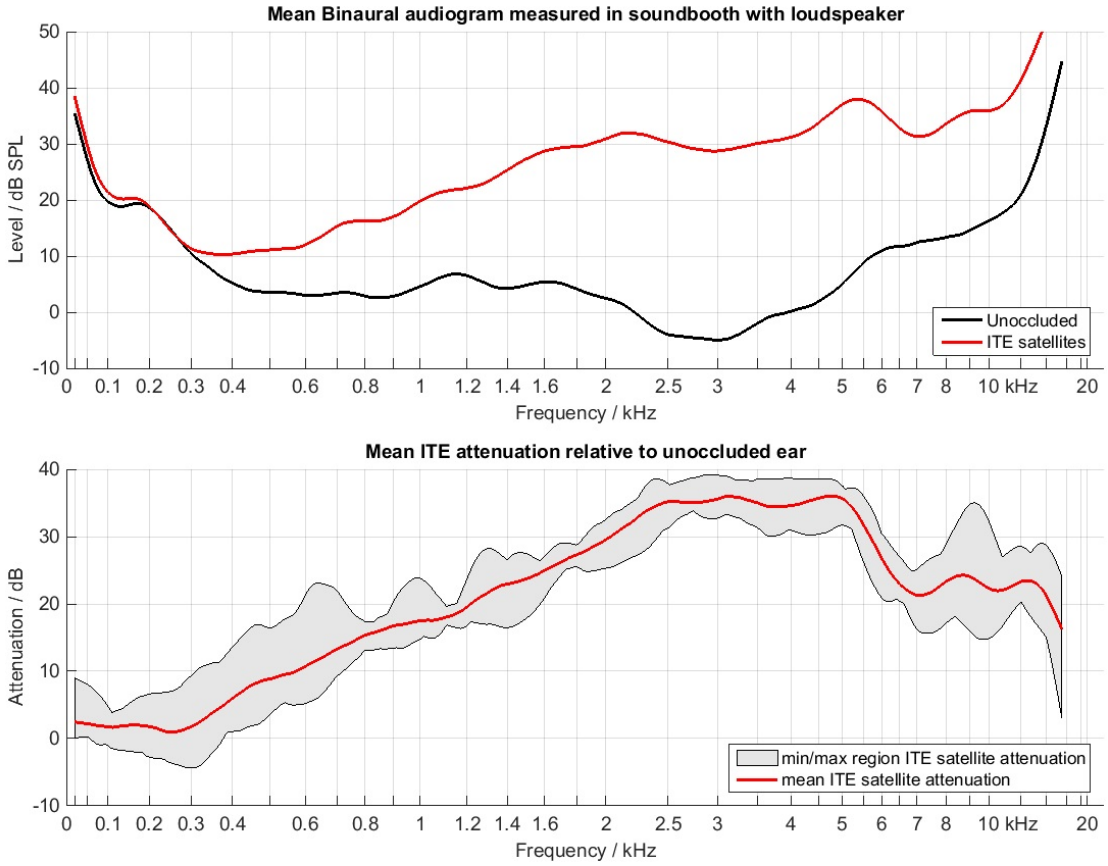

Figure 5.10
Mean binaural audiogram for seven participants (top) for the unoccluded case in black and wearing ITE-satellites in red. The frequency dependent mean attenuation is shown on the bottom.

(Chapter 4). These aided conditions can be directly compared with results from an unaided baseline, as all participants performed the test in the aided and unaided case.

5.3.2.3 Participants

Eight participants (male, age range 23 – 35 years) took part in the experiment, seven of which also took part in the first part of the experiment. All participants were experienced normal hearing listeners (<20 dB HL) as assessed with a Békésy tracking procedure [Von Békésy and Wever, 1960, Seeber et al., 2003] in a sound-isolated listening booth ([Frank, 2000] using Sennheiser HDA-200 closed headphones. All participants had previously taken part in hearing experiments and had used their hearing aid prototypes before. They participated voluntarily after giving written consent and received no compensation for their participation. Subject's well-being was monitored through an intercom during the experiments. The TUM ethics committee approved this study.

5.3.2.4 Stimuli

Four different stimulus-pairs of a target and a competing disturber were used. These were carefully designed to present an alternating energetic dominance of either the target or the disturber. The target sounds were four different speech sentences spoken by male speakers of length 2.2 to 3 seconds. The first disturber consisted of a speech sentence by a female voice, simulating a natural situation of two speakers talking simultaneously from different directions. The second disturber was a 50 ms noise burst train of white Gaussian noise with 3 ms Gaussian on- and offsets, presented at a rate of 5 Hz to test for effects of impulsive sounds on spatial sound quality. The third disturber was an amplitude modulated white noise, whereby the modulation was set to the inverse of the positive target envelope, such that moments of high target energy would have a corresponding low disturber energy, while low target energy would lead to high disturber energy. This stimulus was designed to allow for a continuous alternating energy balance between the front and back. The fourth stimulus consisted of male speaker speech as target, and female speaker speech as disturber. These were created of sections of 250 ms length such that either only the target or the disturber would be active in an alternating manner. The 250 ms sections were cut from speech sentences so that the moment of change between the front and back sections did not fall into a speech pause, thus creating an energetically semi-continuous signal when combined, with 3 ms Gaussian on- and offsets and no overlap. This stimulus was selected to test effects of rapid direction changes on spatial sound quality. Fig. 5.11 shows the time signals of the four target-disturber pairs. All stimuli were presented at levels between 46 dB SPL and 50 dB SPL, after a loudness matching.

For the second part of the experiment, the stimulus pairs 1 and 3 were selected out of the four stimuli used in the first part of the experiment. These two stimuli were chosen because they contain voice and noise as disturbers, respectively, and since they showed significant differences between each other (see results of the first part of the experiment), while not showing significant differences whether presented in the front or back (fig. 5.13) In the dry condition, the stimulus pairs were presented only from two loudspeakers of the ring at 15° and 165° and 1.3 meters distance. For the reverberant cafeteria condition, we simulated a room of size 10.85 x 13 x 7 m ([Cox et al., 2004]). 48 of the 96 loudspeakers were used for the auralization. The target and disturber stimuli were convolved with precomputed room impulse responses at the same angle and distance as the dry stimuli. Additionally, we simulated six additional disturbers of dialog talk or music (Fig. 5.12) to function as background noise. The background noise had a duration of 72 seconds. For each trial, a random section of 4.5 seconds length of the background noise file was selected, and one of the reverberated stimulus conditions of target-disturber pairs, with 1 second of pause at the beginning, was added. A raised cosine ramp function was used to smooth the onset and offset of each trial. The level of the disturber dialog talkers was 58 dB SPL each, and that of the music 48 dB SPL before convolution. The overall reverberated cafeteria noise was reduced in level by 2 dB and the reverberated target and disturber stimuli amplified by 3 dB to ensure audibility in the cafeteria scene, while remaining at an overall comfortable level when considering the

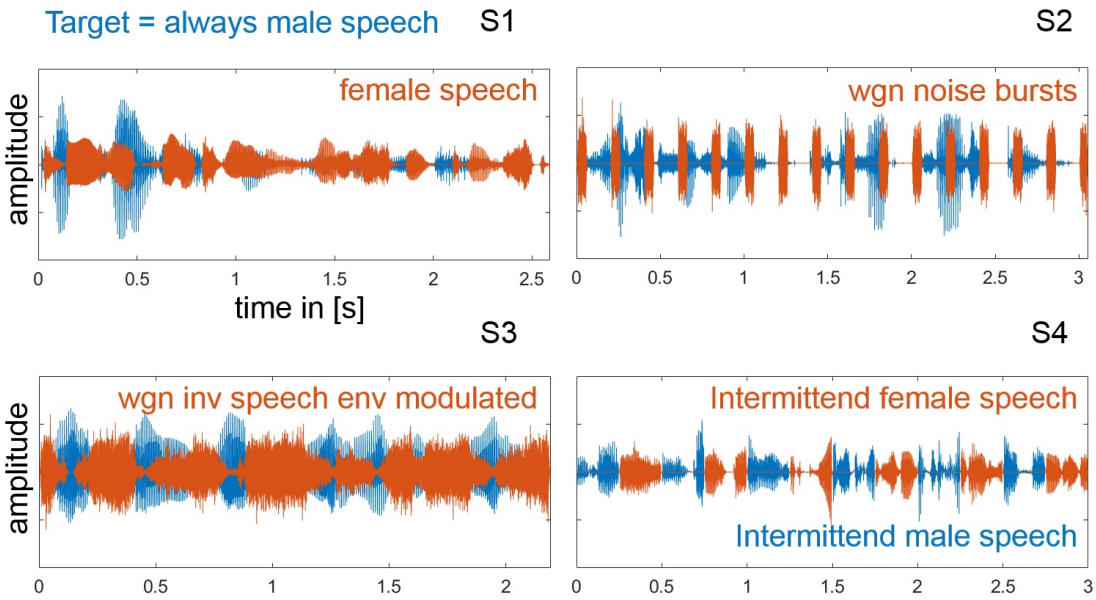

Figure 5.11
Stimulus target-disturber pairs used in the experiment. The blue signal shows the target which was different speech by male speakers. The disturber signals are shown in red and consist of either speech by female speakers, noise bursts at 5 Hz rate or white noise amplitude modulated with the inverse target envelope. All stimulus pairs were presented in the front and in the back at 15° and 165°, participants could repeatedly listen to the stimuli before rating.

hearing aid gain. Since both the talker and disturber were rated once in the front and once in the back for each condition, the room was flipped for those conditions where the target male talker had to be rated in the back. The flipping of the entire room, i.e. of the reverberated cafeteria stimuli, was implemented by mirroring the room at the plane crossing the virtual listener's ears and ensured that identical reverberation patterns were present in the front and back for target and disturber stimuli.

5.3.2.5 Response Measure

Inspired by the work of Wiggins and Seeber (2012) [Wiggins and Seeber, 2012] and as outcome of extensive piloting sessions where the whole range of time constants and several descriptive adjectives were tested, nine adjective pairs were selected that best cover the dimensions of perceived spatial sound quality in this experiment (Table 5.2). The adjectives selected are shown in Table 5.2.

For each of the randomly presented experimental conditions of algorithm (BTE, ITE or STA compensation with different time constants), sound-pairs and position (front or back), one adjective pair was rated. For the second part of the experiment, the adjective pairs used were the same as in the first part of the experiment, except for the fourth adjective pair (*moved–static*) which was replaced by *salient – background*. For each of the randomly presented experimental conditions of algorithm (BTE, ITE, BF, STA BF,

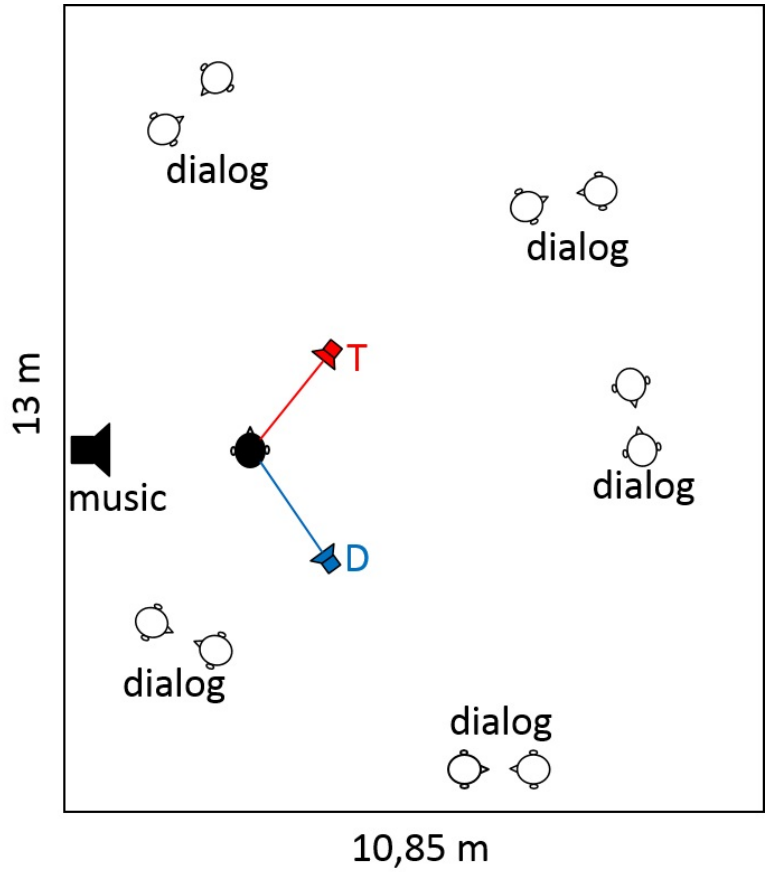

Figure 5.12
Schematic diagram of the simulated cafeteria with one music source and five dialog sources around the virtual listener (black). T and D represent the stimulus target and disturber pairs to be rated by the participants.

Table 5.2
*Antonym adjective pairs as spatial dimensions tested in the first part of the experiment for the semantic differential. * Adjective 4 was replaced in the second part of the experiment for salient–background*

1) Externalized - internalized	2) Wide - pointlike	3) Diffuse - focused
4) Moved - static (Salient - background *)	5) Natural - unnatural	6) Fused - separated
7) Near - far	8) Sure – unsure locatable	9) Front – back

Jackrabbit and unaided baseline), stimulus sound-pairs, position (front at 15° or back at 165°) and acoustic scene (dry or cafeteria), one of the adjective pairs was rated at a time. The rating was done separately for the target and the disturber on the same screen of the GUI. Participants were encouraged to repeat the playback of stimuli as often as needed without moving their heads, to be certain of their ratings. The ratings were done on a touchscreen by setting the position of a slider on a continuous scale between the adjective pair presented for each condition. Separate sliders were used for the target and disturber, with seven markers at each slider between -3 and 3 without labels. The order within an antonym adjective pair was randomly set once, such that positive aspects were either at the left or at the right of the slider, but consistent for all participants. Half of the participants started the experiment with the unaided baseline, while the other half started with the aided part of the test. All trials were completely randomized.

5.3.3 Results and Discussion for the First Part of the Experiment

The results for the seven NH participants showed high interindividual variances for almost all tested conditions and adjectives. All main factors and almost all interactions showed significant differences in a multifactorial ANOVA. Therefore, the results shown here will be post-hoc analysis results of multiple comparisons after Tukey, using marginal means. For the analysis, the positive adjective of the pair was used for the positive values between 0 and 3, while the negative adjective was used for negative values. Overlapping error bars correspond to statistically insignificant differences at a p-level of 0.05, while non-overlapping error bars show significant differences between the conditions. Fig. 5.13 shows average ratings of all algorithms and adjectives combined, for the interaction term of $sound * position$. Ratings for the target and disturber are shown separately. The red lines are shown for better comparability between the sounds and denote a threshold for significant differences.

No differences were found between sounds S1, S2 and S4 for the target nor the disturber. Also, no differences between the sound presentation from the front or the back were found for any sound pair. The target sound S3 in the front was rated significantly worse to sounds S1, S2 and also S4 in the back. For the disturber, both S3 sound presentations in the front and back were rated significantly worse than the rest. Since all target stimuli were male speaker speech sentences, we expected no differences between the target sentences. Yet S3 target ratings were worse than the rest, especially in the front. Since the ratings of S3 for the disturber were also much worse than for the rest, it appears that there is an influence of the disturber in S3 on the target, even while presented on different hemispheres. The design of the S3 stimulus pair of alternating, continuously changing energy between the front and back appears to have a strong detrimental effect on how spatial sound quality is perceived for this stimulus pair with hearing aids.

Fig. 5.14 shows average ratings of all sounds and positions combined, for the interaction term of $algorithms * adjectives$. ITE and BTE ratings are shown in red for each adjective pair for better comparability. Only for the adjective pair back-front, separate marginal means ratings are shown for sounds presented from the front and the back.

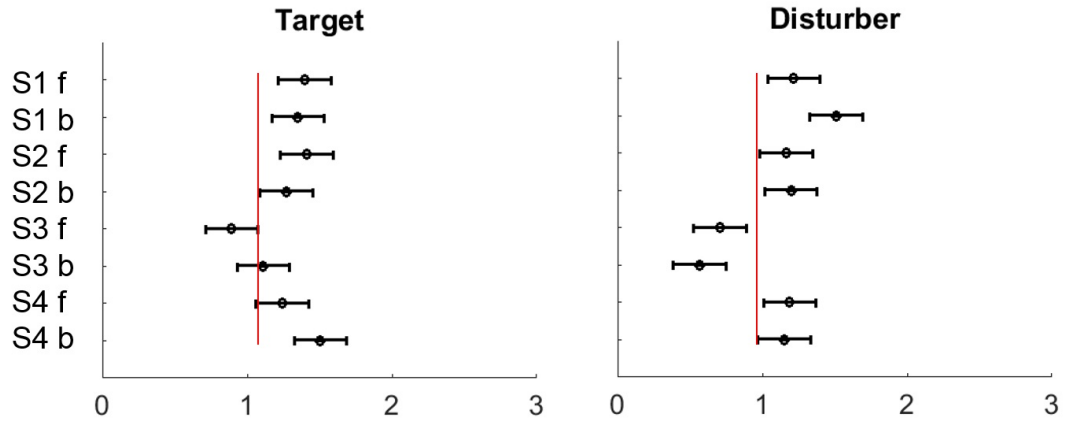

Figure 5.13
*Post-hoc multiple comparison test of position*sound interaction terms, separately for the target and disturber. The error bars show the significance range, while the red lines denote a reference threshold for better visualization of significant differences between sounds (S1 – S4, front (f) of back (b)).*

For all adjective pairs, spatial sound quality ratings were higher for ITEs than for BTEs. Especially for the pairs describing externalization, diffuseness, separation and locatability, the ITE ratings were significantly better than for BTEs. Regarding the back-front adjective ratings, it is noticeable that the BTE ratings are in the back (around -1), regardless of the actual sound presentation, meaning that most of the participants perceived the majority of sounds as coming from the back. For the ITE ratings, sounds were perceived from the correct location with mean values at around 1-2 in the back or the front. The main difference between the BTE and ITE signals is the spectral filtering of the pinna, which is known to be important for front-back discrimination and elevation perception. The data here shows that several aspects of spatial hearing are also affected by the lack of pinna cues, at least in a complex acoustic scene as used in this study, even without reverberation.

Regarding the STA compensation with different time constants, Fig. 5.14 shows similar results of the STA with low time constants with attack times of 2 ms and 4 ms to the ITE ratings. This is because the STA compensation from equation 4.6 takes the current BTE amplitude spectrum (framewise), divides it by the STA averaged BTE amplitude spectrum, and multiplies by the STA averaged ITE amplitude spectrum. At low averaging time constants, the BTE division term results in a vector with values close to 1, while the STA averaged ITE still resembles the original ITE amplitude spectrum. A single frame of 128 samples, processed by the SIMULINK model at a sampling rate of 22050 Hz takes 5.8 ms processing delay time. With an update rate of 32 samples ($\frac{3}{4}$ overlap between consecutive frames) the model updates the amplitude spectra every 1.45 ms. Thus, even at low time constants the STA averages the spectra of consecutive frames, and this smoothing is equally influenced by the model processing, update rate and weighted overlap add reconstruction for all time constants tested. The similarity of

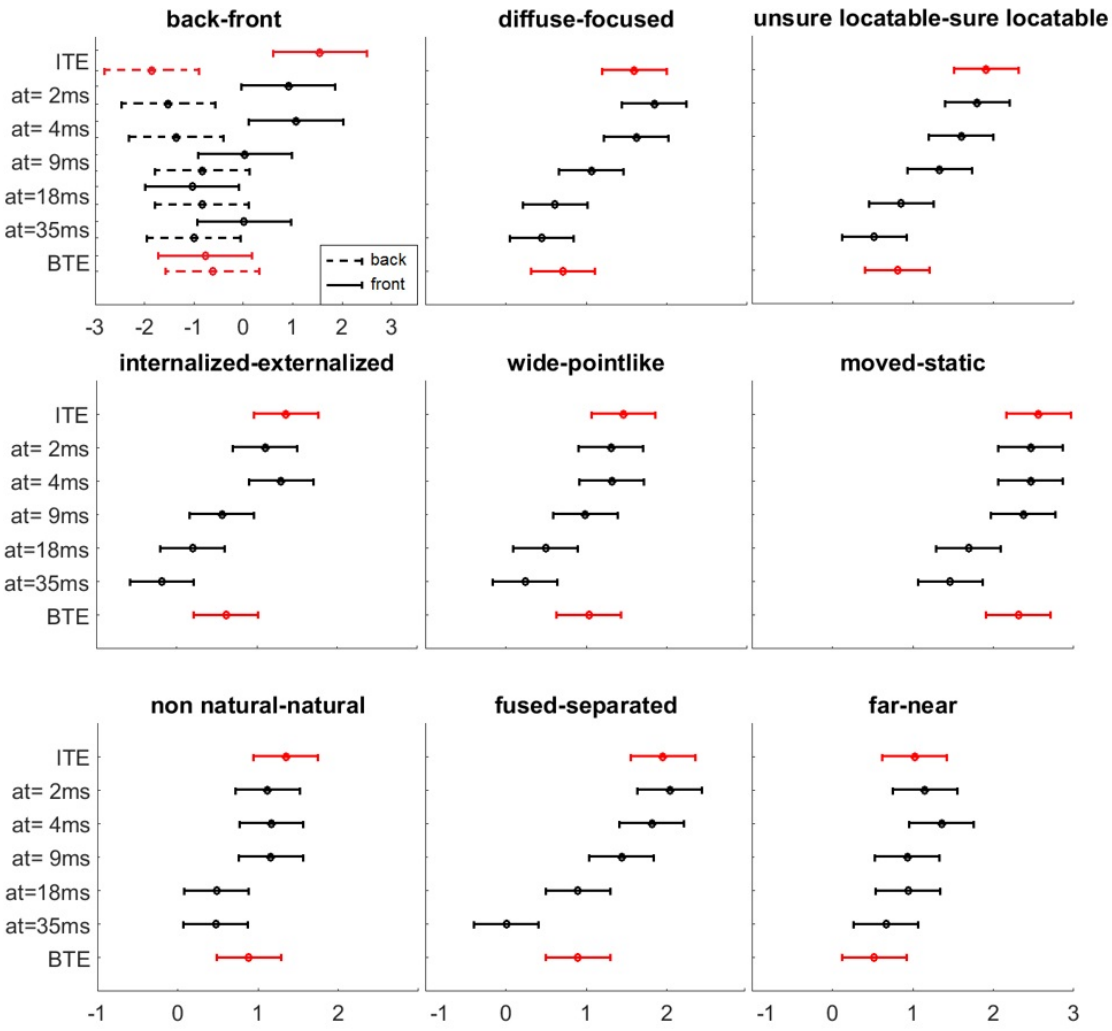

Figure 5.14
*Mean results of semantic differential ratings (range [-3 to 3]) from a post-hoc multiple comparison test of algorithms*adjectives interaction terms, separately for each adjective pair. The error bars show the significance range for differences between conditions with different smoothing constants for attack times (at) between 2-35 ms at a* $p < 0.05$ *level.*

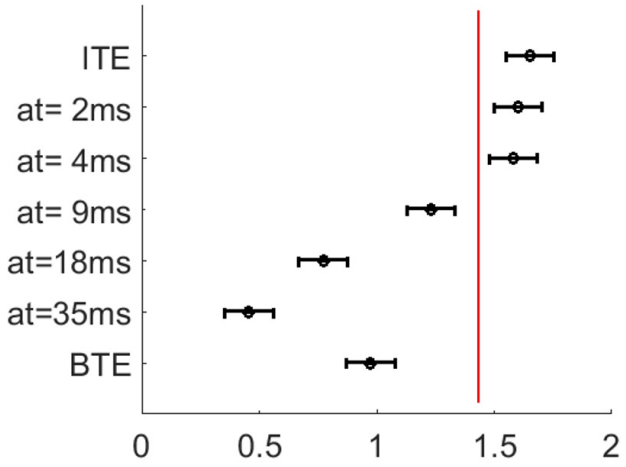

Figure 5.15
Marginal means post-hoc comparison after Tukey of the main factor algorithms. The error bars show the significance range for differences between conditions with different smoothing constants for attack times (at) between 2-35 ms at a $p < 0.05$ level, for all adjectives, positions and sounds combined.

the low time constant's ratings to the ITE rating is therefore caused by the similarity of the spectra, and the good spatial quality of the ITE is preserved. For higher time constants, the smoothing of the amplitude spectra affects both the BTE division term and the ITE multiplication term of eq. 4.6. So, every increase in the smoothing affects the signal twice, such that the effects of the smoothing on the signal increase exponentially rather than linearly. For most of the adjective pairs in Figure 5.14 we observe a decline in spatial sound quality for the attack time of 9 ms. The externalization ratings for 9 ms attack time are even significantly different to the ITE ratings. For the two highest time constants all ratings show a significant deterioration of spatial perception, and for most adjective pairs these ratings become even worse than the BTE ratings. It is especially noticeable for the adjective pair $moved - static$, where we start to observe movement for the two highest time constants due to the alternating spectral dominance of the stimulus target-disturber pairs. Figure 5.15 summarizes the previous findings in a very distinct way. It shows the post-hoc analysis for the factor algorithms, with marginal means of adjectives, positions and sounds combined. It is evident from the figure that the ratings for the ITE and the STA compensation for the two lowest time constants are the best. For 9 ms attack time, the ratings lie between the ITE and BTE ratings, while for the two highest time constants the ratings reach lower values than for the BTE in average.

5.3.4 Conclusions for the First Part of the Experiment

This experiment tested for differences in spatial sound quality between the ITE and BTE microphone position, and additionally compared the perceived spatial sound quality for five different time constants of the STA compensation from eq. 4.6. First, it

becomes clear that the monaural cues present in the ITE signals are very important for spatial sound quality, as can be seen by the difference of ITE ratings to the BTE ratings. Especially the spatial dimensions of externalization, diffuseness, source separation and locatability are significantly affected by the lack of pinna cues. For the STA compensation method, we conclude that the use of low averaging time constants with attack times of up to 4-5 ms can be used in hearing aids without affecting spatial sound quality, but higher time constants should not be used when trying to preserve good spatial sound quality.

5.3.5 Results for the Second Part of the Experiment

The results for the eight participants showed high interindividual variances for almost all tested conditions and adjectives. After ensuring ANOVA conditions were met, we performed a multifactorial ANOVA with the main factors $sounds, conditions, position, adjectives$ and $scene$, taking the mean results over the 8 participants as input. All main factors and almost all interactions showed significant differences in the multifactorial ANOVA. Therefore, the results shown here will be post-hoc analysis results of multiple comparisons after Tukey, using marginal means. For the analysis, the positive adjective of the pair was used for the positive values between 0 and 3, while the negative adjective was used for negative values. Overlapping error bars correspond to statistically insignificant differences at a p-level of 0.05, while non-overlapping error bars show significant differences between the conditions. For the adjective pairs $background - salient$ and $back–front$, we inverted the results for sound presentations from the back, such that for the analysis, these would also have greater positive values the more they were perceived as background or in the back, being a positive (beneficial) spatial rating. The unaided baseline condition is included into the analysis for direct comparison with the aided conditions.

Sounds presented in the front were significantly rated better than sounds in the back, with a mean difference of 0.53 rating points. No significant differences were seen in the ratings of the target compared to the disturber, neither in the front nor the back. The results from the dry scene were significantly rated better than in the reverberated scene, with a mean difference of 0.24 rating points. Sound 1 with the female voice as a disturber was rated significantly better than sound 2 with the speech shaped noise as a disturber, consistent with the observations of the first part of the experiment (time-constant evaluation), with a mean difference of 0.11 rating points. The average rating of each algorithm was not significantly different between the dry and reverb cafeteria condition. The same holds between sound 1 and sound2, where there were no significant differences between sound 1 and sound 2 in the front nor back. Sound 1 was significantly rated better than sound 2 in the dry condition, while no difference was seen in the reverberated condition between the sounds. Mean ratings between front and back were significantly different for both the dry and reverberated condition, with higher ratings for sounds presented in the front. This difference was more prominent in the dry condition, with a mean difference between front and back ratings of 0.74 rating points, while in the reverberated cafeteria condition the difference was only 0.4 rating points.

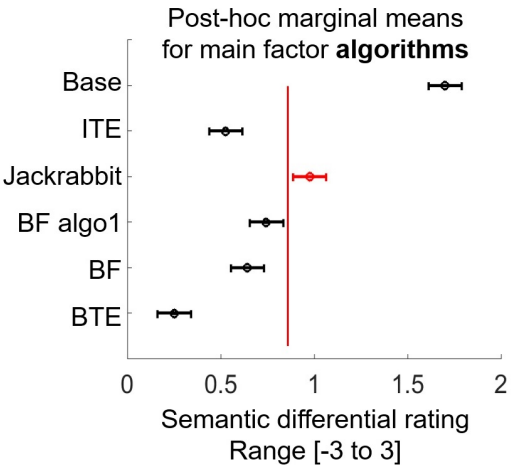

Figure 5.16
Marginal means post-hoc comparison after Tukey of the main factor algorithms. The error bars show the significance range for differences between conditions with different smoothing constants for attack times (at) between 2-35 ms at a $p < 0.05$ level, for all adjectives, positions and sounds combined.

From figure 5.16 it is apparent that the baseline condition was always rated best, and the beamformer conditions were rated better than the ITE and BTE conditions. The STA-BF and the BF conditions did not differ significantly, while the Jackrabbit was significantly rated best of all aided conditions. The ITE was significantly better rated than the BTE, consistent with our previous results in the first part of the experiment. The different mean scores for the ITE and BTE compared to the first part of the experiment are related to having different experimental conditions, with a dry scene and a cafeteria simulation, different algorithms and for half of the participants a comparison to the unaided baseline before the aided part (on different days). The mean ratings for all algorithms were positive, in the range between 0.2 and 1, while the unaided baseline had a mean average rating of close to 1.7, which is about 1 unit more than the aided average. This clearly shows that, despite of having powerful algorithms for noise reduction and in some cases pinna cues preservation, normal hearing participants experienced a significantly reduced spatial sound quality due to the hearing aid devices and signal processing.

Fig. 5.17 shows how the algorithms perform differently for different spatial aspects tested here as adjective antonym pairs. The Jackrabbit algorithm was rated best for the pairs *back-front, unsure-sure locatable, internalized-externalized, background-salient* and *near-far*. This algorithm was never significantly worse than any other algorithm, and even significantly better than the ITE in the dimensions *background-salient, fused-separated* and *near-far*. No significant differences between the aided conditions were seen for the adjective pairs *diffuse-focused, wide-pointlike, unsure-sure locatable* and *unnatural-natural*.

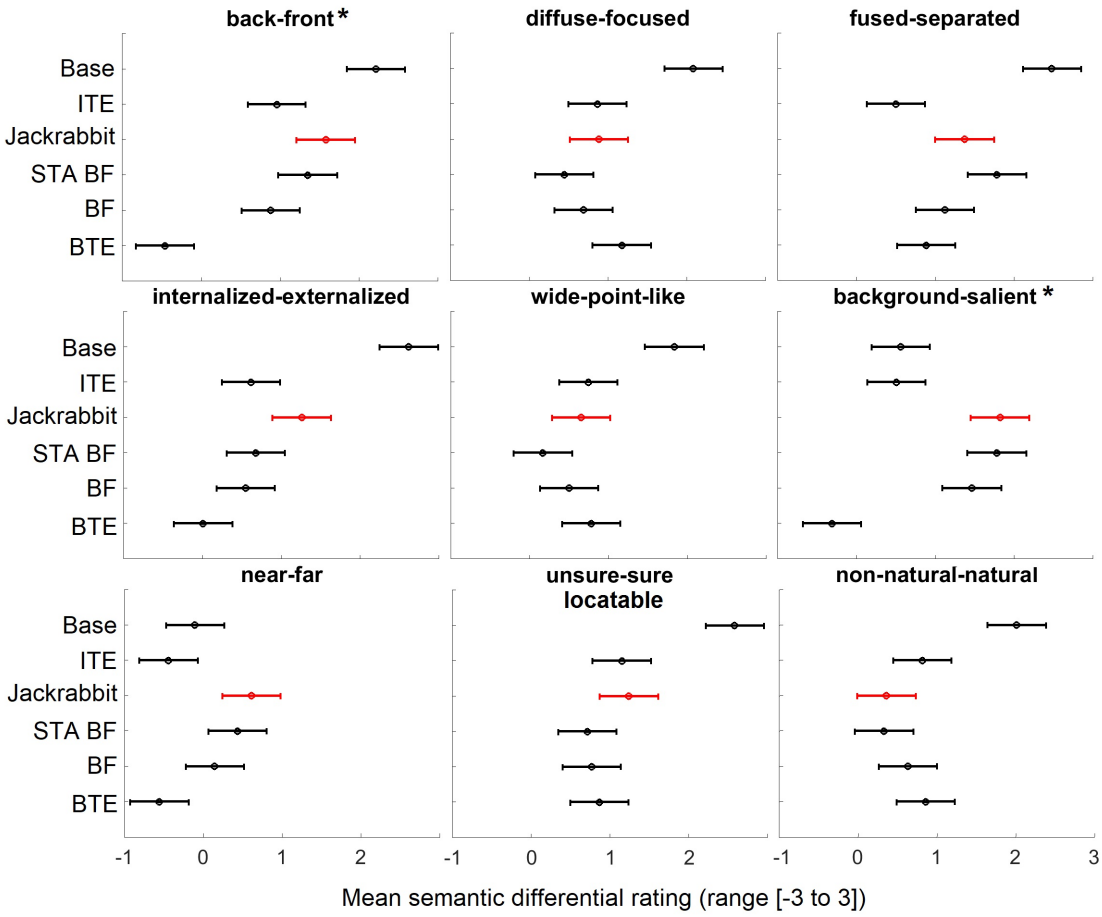

Figure 5.17
*Average ratings of the interaction term of algorithms*adjectives for all sounds and positions combined. Adjective pairs are ordered in the two left columns by related spatial dimensions. The more to the right the ratings are placed, the better the perceived spatial sound quality of an algorithm for a specific adjective pair. Non-overlapping error bars show significant differences at* $p < 0.05$ *level. The "back-front" and "background-salient" adjective pairs were flipped for sounds from the back for the analysis.*

It was expected that the beamforming algorithms BF and STA BF would perform very good in the saliency-background dimension. Interestingly, the Jackrabbit method was rated even slightly better than the beamforming methods, despite taking a different approach to noise reduction (Chapter 4). Since spectral subtraction is used to filter out the energetic difference between the target and disturber energy for each time-frequency bin from the ITE signal, the Jackrabbit does not require a low frequency roll-off compensation as with the beamformers. It is still able to attenuate unwanted noise equally well as the beamformer approaches (achieving equal speech understanding benefit as shown in section 5.2), while still maintaining correct pinna cues and correct binaural cues.

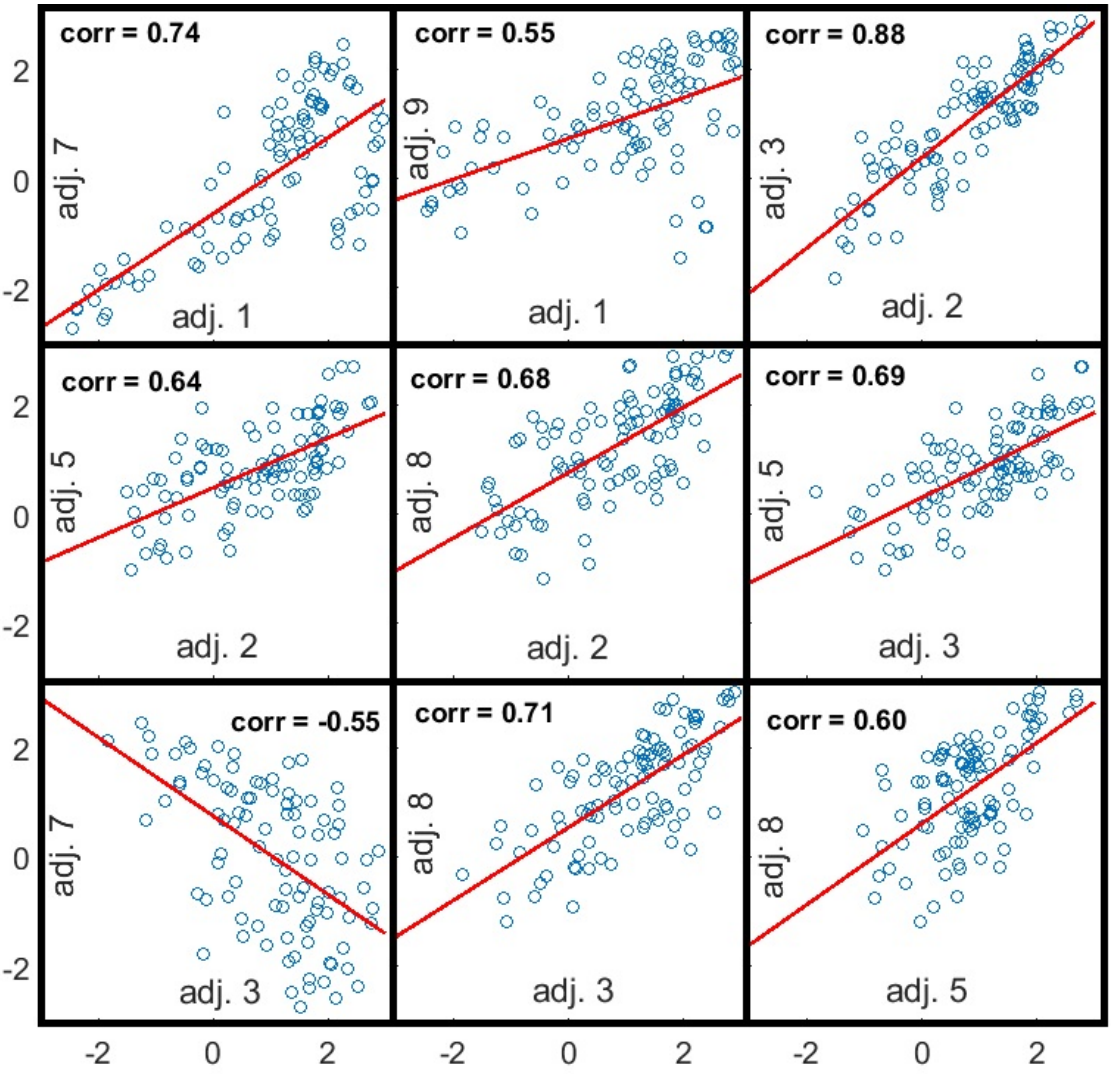

Figure 5.18
Adjective pair relationships with a correlation coefficient higher than 0.5

5.3.6 Correlation Analysis

The correlations shown in Fig. 5.18 clearly exhibit strong relationships and redundancy between some adjective pairs. Externalization (adj. 1) correlated well with distance (adj. 7) and font-back (adj. 9). Apparent source width (adj. 2) correlated well with diffuseness (adj. 3), naturalness (adj. 5) and locatability (adj. 8), and the more focused (adj. 3) a sound was perceived, the closer it was perceived as well (adj. 7), while sounds further away were perceived as more diffuse.

A better, more intuitive representation to summarize these relationships is in form of a dendrogram, as shown in Fig. 5.19, where the data of subject means was clustered into

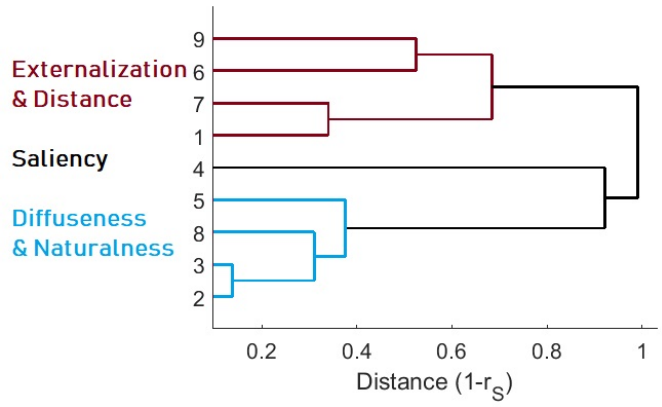

Figure 5.19
Dendrogram from a cluster analysis using the spearman rank correlation as distance between adjective pairs, numbered as in Table 5.2. Different colors represent different groups. The smaller the distance, the closer related the adjectives or clusters are.

groups using the spearman rank correlation coefficient. Adjective pairs 2 and 3 (diffuseness dimension) are clustered strongly together in group 1, showing high redundancy. Also included in group 1 is adjective 8 (locatability), and adjective 5 (naturalness). Interestingly, naturalness was associated closely with narrow, focused sound percepts and good locatability. As a separate group adjective 4 (background - saliency) stands out from the rest. In group 3 we find adjectives 1 and 7 (externalization and distance), combined with adjectives 6 and 9 (separation and front/back). Thus, the original 9 adjective pairs can be categorized into the perception of sources in three spatial sound quality dimensions:

1) Externalization and distance

2) Saliency

3) Diffuseness and naturalness

The relationship between saliency (adj. 4) and some of the other adjective pairs is interesting (not shown graphically). The more externalized a sound was perceived, the more the saliency rating was close to 0. The more internalized a sound was perceived, the more extreme the stimuli were rated as either background or salient. Similarly, the further away a sound was perceived, the more neutral (close to 0) it was perceived in terms of saliency. When sounds were perceived near, they were also perceived either very salient or very much in the background. Also, the more focused or pointlike a sound was perceived, the more background or salient its ratings were set, while they were rated neutral in saliency (0 rating) the more diffuse or wide a sound was perceived.

5.3.7 Conclusions

This section examined the implications of hearing aid devices on the spatial perception of sounds. Normal hearing participants compared different aided conditions to rate a variety of spatial aspects using the semantic differential approach. A target sound and competing disturber sound were rated separately for each trial. In the first part of the experiment, different microphone positions behind the ear and at the ear canal were compared to short time averaged magnitude spectra with different time constants, to assess the amount of smoothing that is tolerable without affecting spatial sound quality in hearing aids. We found that using averaging with attack times greater than 4 ms started to deteriorate spatial sound quality. The ITE microphone position (preserving pinna cues) led to significantly better results than for the BTE microphone position overall, and especially in the externalization, locatability, separability and diffuseness dimensions.

In the second part of the experiment, different hearing aid algorithms for noise reduction based on beamforming were compared to the ITE and BTE microphone positions regarding spatial sound quality. Also, additionally to the dry scene from the first part, where stimuli were presented from loudspeakers at 15° in the front and 165° in the back, a reverberated cafeteria simulation was used as testing environment. In these cocktail party scenes, hearing impaired with state of the art hearing aid devices still encounter great problems communicating. Results showed a benefit from preserving pinna cues (using the ITE microphone position) compared to the BTE microphone position. Also, noise reduction with beamforming showed an overall improvement compared to the ITE and BTE results, since attenuating disturbers in the back allowed for better results especially in the separability and saliency dimensions. The Jackrabbit method, which combines the benefit of preserving pinna cues with noise reduction, performed significantly better than all the other aided conditions, showing that combining an increased SNR by directional noise reduction and pinna cues preservation at the same time is beneficial for spatial perception, and should be considered in the design of future hearing aid devices.

5.4 On the Internalization and Externalization Percept with Hearing Aids

5.4.1 Summary

The present section investigates the perception of externalization for stimuli coming from 15° in the front with different hearing aid conditions and a static head. Eight normal hearing participants wearing custom made ITE and BTE hearing aid dummies rated the perceived distance of sounds in a MUSHRA like test. The hearing aid conditions included the ITE and BTE omnidirectional signals, three degrees of static beamforming as well as the two novel algorithms for noise reduction with preservation of pinna cues (STA BF and Jackrabbit) from Chapter 4. Additionally, an internalized anchor condition was used. 18 different stimuli with different band energy levels were used. Results show that the ITE and Jackrabbit conditions were fully externalized, while increasing degrees of beamforming led to greater internalization [von Unold, 2017]. Higher energy levels in the frequency band between 2 – 4.5 kHz were beneficial for externalization for the hearing aid conditions using the BTE microphone position. No effect of level roving was found on the perceived externalization. These findings demonstrate that preserving correct pinna cues is important for the externalization percept, and beamforming can lead to strongly internalized sound images for static head positions. A cue analysis showed a strong relationship between the standard deviation of the ILDs and of the interaural coherence with perceived externalization on that same frequency band between 2 – 4.5 kHz.

5.4.2 Introduction

The present experiment investigates the externalization of sounds for different hearing aid conditions with NH participants, and how different spectral weights of sound stimuli can affect the externalization perception. Delay-and-subtract beamformers with attenuation in the back normally do not distort sounds coming from the front. In previous piloting sessions, however, we had noticed that with static beamformers the sounds from the front were perceived significantly more internalized than with pure BTE or ITE omnidirectional microphone signals. Thus, we conducted this experiment to quantify the degree by which this beamforming process affects externalization. Also, we were interested in finding out whether a gradual increase in beamforming would result in an increased internalization. Therefore, we designed this study such that the externalization perception of different hearing aid conditions could be easily rated and compared to each other for each presented sound stimulus from a frontal loudspeaker at 15° and for a static head orientation. We chose the *Multi-Stimulus Test with Hidden Reference and Anchor (MUSHRA)* [ITU4R, 2003] to test the externalization ratings in a continuous way between completely internalized in the head, up to a sound perception further away than the loudspeaker. We conducted the experiment with NH participants to selectively study the influence of the different hearing aid conditions, without having additional effects of hearing loss or additional signal processing of commercial hearing aids on the

perceived externalization. For each individual subject, we recorded all stimuli with their own hearing aids to further analyze the corresponding binaural cues and correlate the perceived externalization ratings with these cues. In pilot listening sessions we found that different stimuli were externalized differently, and more specifically, that emphasizing the frequency region between 2 – 4.5 kHz led to higher externalization. Therefore, we designed the stimuli to include different spectral weights by dividing broadband white noise into four frequency bands of 200 - 1300 Hz, 1300 - 2000 Hz, 2000 - 4500 Hz and 4500 - 8000 Hz, and selectively dampening one or more bands by 15 dB relative to the other bands. In total, we presented 18 different stimuli consisting of male talker speech, female talker speech, broadband noise bursts, speech envelope modulated white noise and 14 variations of dampened noise. We applied roving between ±2 dB in 1 dB discrete steps to reduce effects of absolute level on the externalization ratings. We compared eight different hearing aid conditions on perceived externalization. We were especially interested in the externalization ratings of the newly developed STA BF and Jackrabbit algorithms, since these include dynamic pinna cues and should, if pinna cues play a role in externalization, give different results than the static beamformer or the BTE conditions.

5.4.3 Methods

5.4.3.1 Hearing Aid Sound Presentation

The hearing aids used in the experiment were custom-made ITE and BTE prototypes by Phonak, connected over cables to a PC on which we ran a real time Simulink model with a total delay of only 7.8 ms. We applied a frequency response equalization of the microphones and receivers in the frequency domain. In addition, we applied a 5-dB gain to the signals after loudness compensation to mask direct sound leaked through the hearing aid shells. Loudness compensation was done by recursively comparing the loudness of stimuli in the ITE aided condition to the unaided condition both bilaterally and unilaterally fitted, and adjusting the gain in the Simulink model to match to the same loudness. The signals were bandlimited from 172 Hz to 8 kHz in the model by setting all FFT-bins to zero outside that frequency range. The eight hearing aid conditions used in this experiment were the processed ITE and BTE signals, 3 static delay-and-subtract beamformers (BF) with maximum attenuation at 180° in the back, with different degrees of attenuation (30%, 70% and 100%) implemented in the frequency domain according to eq. 5.1.

$$BF_{delay\&subtr.}(f) = BTE_{micfront}(f) - \alpha \cdot H_{sig} \cdot BTE_{micback}(f) \tag{5.1}$$

With H_{sig} being the transfer function to compensate for amplitude and phase differences between the front and back microphone of the BTE device, such that a subtraction leads to maximum attenuation in the back, and α a weighting factor with values of 0, 0.3, 0.7 and 1 respectively for the different degrees of beamforming, as shown in Fig. 5.20. The roll-off compensation was reduced by that factor accordingly.

Additionally, both algorithms from Chapter 4, which reduce noise and attempt to preserve pinna cues (the STA BF and Jackrabbit conditions), were tested. As a reference

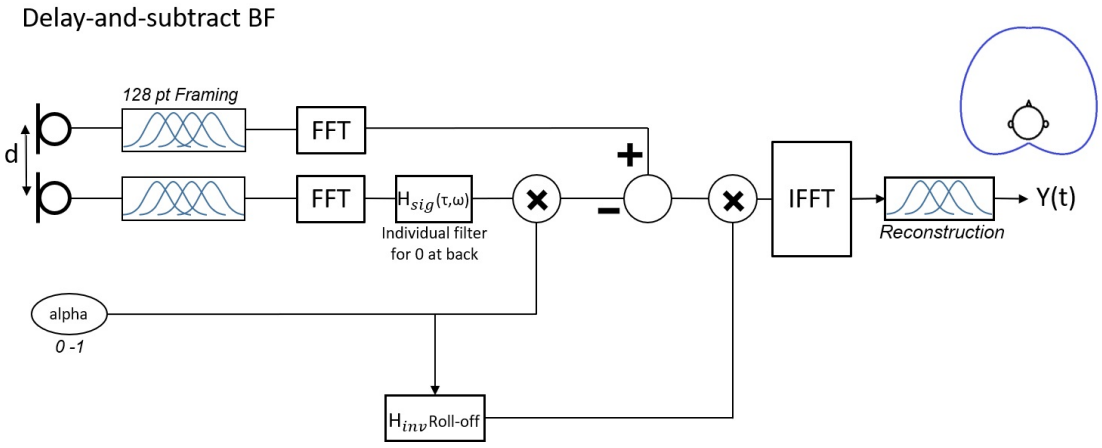

Figure 5.20
Schematic diagram of the signal path for adapting the beamformer strength by setting the factor α between 0 and 1.

for internalization, we designed an additional eighth condition that took the right BTE signal for the right ear, and a mixdown of BTE (20%) and ITE (80%) signals from the right ear applied to the left ear, such that the sound stimuli were always perceived internalized, yet less pointlike than in a diotic condition and slightly lateralized to match the 15° loudspeaker azimuth.

5.4.3.2 Participants

Nine normal hearing participants (male, aged 22 – 35 years) took part in the experiment. All had normal hearing thresholds as verified with a calibrated Békésy tracking audiometer[Von Békésy and Wever, 1960, Seeber et al., 2003] in a sound-isolated listening booth ([Frank, 2000]. All participants except for one had taken part in hearing experiments before and had used their hearing aid prototypes previously. The TUM ethics committee approved this study.

5.4.3.3 Stimuli

18 different sound stimuli were used in the experiment, with long-term amplitude spectra as shown in Fig. 5.21. All stimuli were presented from a loudspeaker at 1.3 m distance from 15° in the front. The angle was chosen since in piloting sessions we experienced the perceived externalization greater than from 0° in some hearing aid conditions. At 15°, the BF conditions still lead to a strong internalization percept. All stimuli were loudness equalized using the Zwicker loudness model [Zwicker and Scharf, 1965]. Final loudness adjustments were done on the two speech stimuli. A discrete roving between ±2 dB was applied to all stimuli in 1 dB steps to reduce effects of absolute level on the externalization ratings when pooling the data. The roving was applied for every stimulus, but not changed between hearing aid conditions, since we wanted to maintain

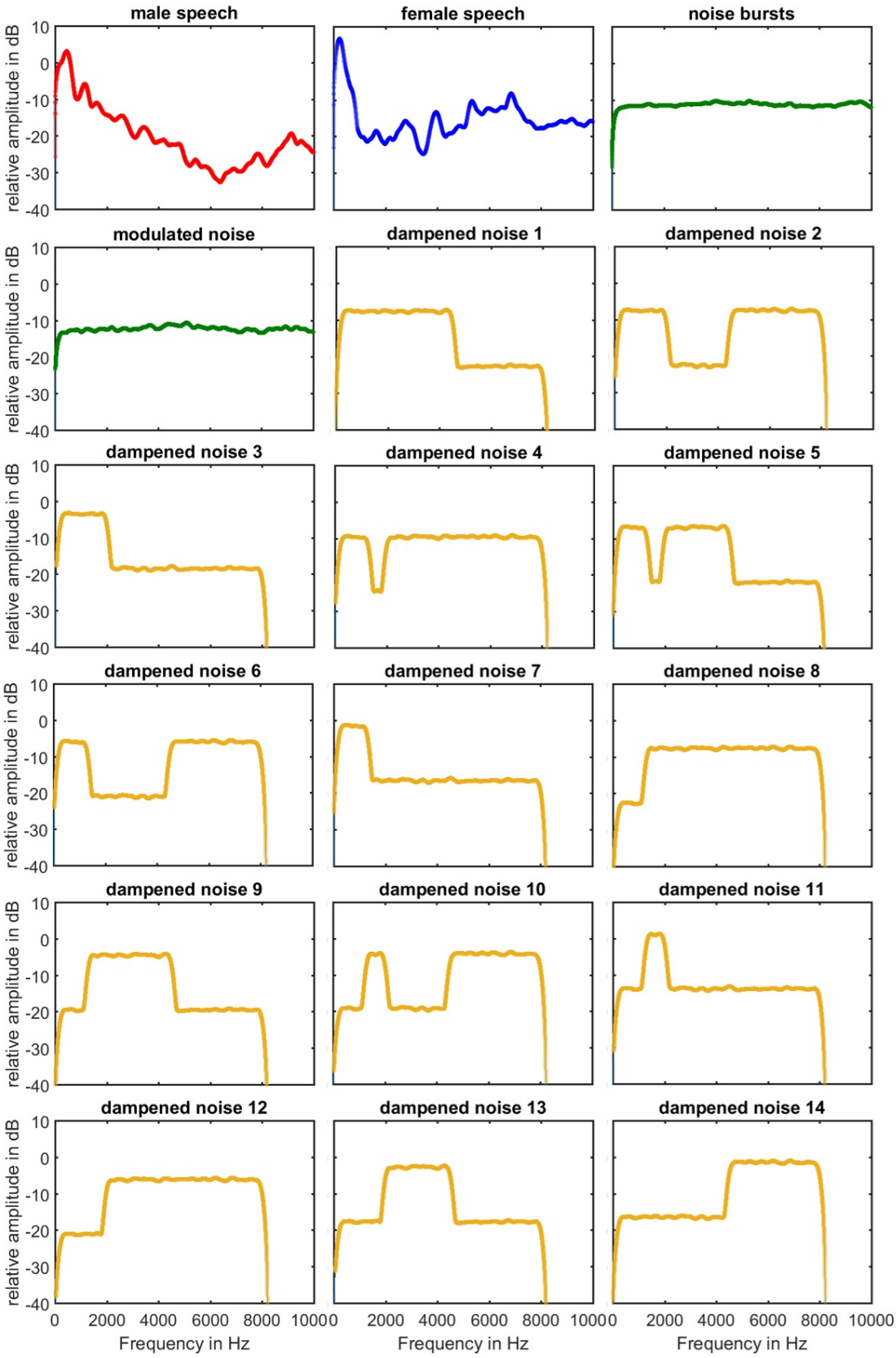

Figure 5.21

Long term amplitude spectra of the 18 stimuli used in the experiment. Stimuli were loudness equalized using a Zwicker loudness model ([Zwicker and Scharf, 1965]). The speech stimuli were manually adjusted to the same loudness as the noise stimuli. The dampened bands in the noise stimuli were attenuated by 15 dB relative to the undamped bands.

relative differences between the algorithms. All stimuli were bandlimited between 200 Hz and 8 kHz prior to playback.

5.4.3.4 Experimental Procedure and Response Method

Each participant sat in the middle of the loudspeaker ring of the SOFE v3 [Seeber et al., 2010] in the dark room, with his head resting on a head rest and wearing the BTE and custom made ITE hearing aid prototypes. After being instructed on the experiment, each participant absolved a short familiarization session to get used to the MUSHRA GUI using a touchscreen and listen to different hearing aid conditions with different degrees of externalization. The GUI consisted of a text field with explanations and a playback button for the ITE reference, which was the most externalized condition. For each of the eight hearing aid conditions, unmarked playback buttons and continuous sliders were aligned next to each other. The markers at the sliders were labelled as *In the head, At the head, In the room* and *At the speaker*, in accordance with previous externalization studies ([Ohl et al., 2010, Catic et al., 2013]). The slider could be set further than the "At the speaker" label to allow for externalization ratings further away. While the internalization is only perceived with a static head, for which we had a headrest mounted to the seat, we encouraged participants to move their heads when listening to the reference (ITE condition) in order to reset the externalization perception which is automatically given with head movements. All 18 stimuli were presented in random order with a random roving applied to. For each stimulus, all eight hearing aid conditions were randomly assigned to the GUI buttons and the externalization rating had to be set using a slider for each condition. Participants were also encouraged to do a final comparison between the algorithms in ascending or descending externalization rating to make fine adjustments before submitting the answer and continuing with the next stimulus. Each stimulus was repeated three times and the externalization ratings were averaged over the three repetitions for the analysis. One subject was only able to do one repetition of the trials.

5.4.3.5 Cue Analysis

From individual recordings of the ITE and BTE microphones we extracted binaural cues for each of the stimuli and hearing aid conditions using the auditory model by Dietz et al. (2011) [Dietz et al., 2011] as implemented in the Auditory Modelling Toolbox (AMT 2013) in MATLAB. In a first step, the ITE and BTE recordings were run in an offline Simulink model through the same signal processing as in the experiment in real-time, such that the audio signals of all hearing aid conditions were available for the analysis for each stimulus and subject. These audio files were subsequently used as inputs for the Dietz et al. model [Dietz et al., 2011], which works as follows:

First, a bandpass filter of 0.5 - 2 kHz is used to model filtering of sounds by the middle ear pathway. The basilar membrane is modelled further by splitting the signal into equivalent rectangular bandwidth (ERB) [Moore and Glasberg, 1996] wide frequency bands using a fourth-order gammatone filterbank. Signals were compressed with a power

of c = 0.4 representing the cochlea compression. Inner hair-cells were modelled by first half-wave rectifying each ERB wide band passed signal and low-pass filtering with 770 Hz cut-off frequency. These band-passed signals were filtered further with three different filters for the extraction of cues. Either with a real valued lowpass filter for the ILDs (see below), or with an additional complex-valued gammatone filter of second order used to obtain amplitude and phase information for each band of the left and right ear signals. These complex filters are centered at the same frequency as the band's center frequency for the temporal fine structure, and for the envelope cues the filters are centred at a modulation frequency of 135 Hz. ITD related cues are then extracted from the low-pass filtered interaural transfer function (ITF) obtained by multiplying the left and conjugate right complex functions for the left and right ear signals, correspondingly, for each time instant t (eq. 5.4.2).

$$ITF(t) = a_l(t) \cdot a_r(t) \cdot e^{j(\phi_l(t) - \phi_r(t))} \tag{5.2}$$

The ITDs from the temporal fine structure are extracted by dividing the interaural phase difference (IPD) by the mean instantaneous frequency of the left and right signals (eq. 5.4.3)

$$ITD(t) = \frac{IPD(t)}{f_{inst}(t)} = \frac{arg(ITF_{lp}(t))}{\frac{1}{2 \cdot 2\pi} \cdot (\frac{d\phi_l(t)}{dt} + \frac{d\phi_r(t)}{dt})} \tag{5.3}$$

The real valued ILDs are extracted from filtering the left and right signals with a 30 Hz lowpass filter set in parallel to the fine structure and modulation gammatone band-pass filters. Here, the instantaneous energy ratio is used (eq. 5.4.4)

$$ILD(t) = \frac{20}{c} \cdot log_{10} \cdot (\frac{h_r(t)}{h_l(t)}) \tag{5.4}$$

Where $c = 0.4$ denotes the fixed value representing the cochlea compression power and is used here to reverse the compression used at the basilar membrane stage for natural ILDs occurring at the outer ears. The interaural coherence (IC) is obtained by the so called interaural vector strength (IVS), given by equation 5.4.5.

$$IC(t) = IVS(t) = \frac{\| \int_0^\infty ITF(t-\tau) \cdot e^{\frac{-\tau}{\tau_s}} \, d\tau \|}{\int_0^\infty \|ITF(t-\tau)\| \cdot e^{\frac{-\tau}{\tau_s}} \, d\tau} \tag{5.5}$$

With τ_s being a frequency dependent time constant for the temporal integration, set to five times the cycle duration T_c corresponding to the center frequency of the band-pass filter.

5.4.4 Results

As can be seen from Fig. 5.22, the highest externalization was achieved with the ITE and Jackrabbit algorithms, with a median externalization at the loudspeaker. The STA BF algorithm, which applies pinna cues to the beamformer signal was rated much lower, as it was perceived closer to the listener. The different degrees of beamforming starting with the BTE (BF 0%), BF 30%, BF 70% and BF 100% show a decreasing externalization

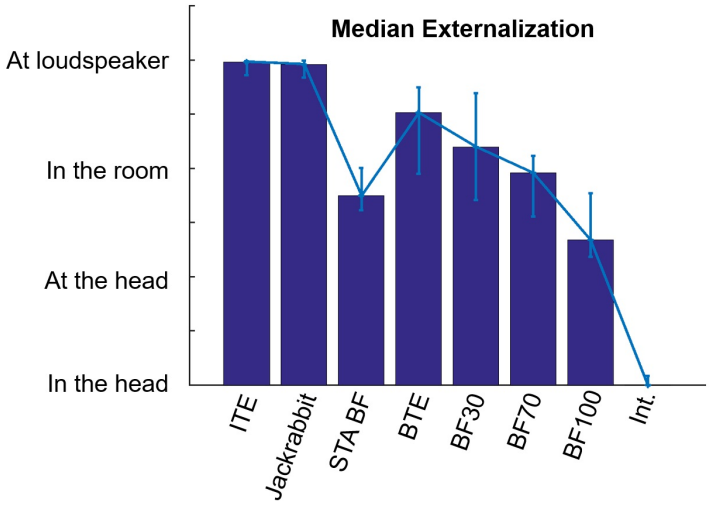

Figure 5.22
Median externalization ratings (ER) over all subjects and all stimuli for the different hearing aid conditions. The ER correspond to 0: "In the head", 1: "at the head", 2: " In the room" and 3: "At the Speaker". Errorbars represent the 25^{th} *and* 75^{th} *percentiles from the median.*

with increasing beamforming strength from about 2.5 decreasing to 1.5. The hidden anchor *Internalized* was consistently rated in the head by all participants.

The horizontal lines in Fig. 5.23 show the mean and median externalization ratings (ER) for all subjects, conditions and stimuli, separated into the applied roving levels. If there was an influence of the roving on the externalization, one should be able to see the lines connecting the mean or median with a certain negative slope, since louder sounds should be perceived closer than softer sounds. Yet here, the roving does not influence the ER, as seen by the slope close to 0, and can be further discarded from the analysis.

Fig. 5.24 shows the weighting of different frequency bands on the ER. They were calculated by taking the mean over all 14 band dampened stimuli for each run and condition, and subtracting the mean ER from each individual stimulus ER. We weighted each of the four bands by that difference, with a factor of 0.25 if the band was dampened and 1 if it was undamped. The final weighting for each condition and each run was the sum of the individual weights. Each band therefore has 25 weightings for each condition (8 subjects x 3 runs + 1 subject x 1 run), and the mean and the 25th and 75th percentiles were calculated and plotted. It becomes apparent that the BTE and BF conditions all have higher weightings for the 3rd band, corresponding to the frequency region of 2 – 4.5 kHz. For the ITE and Jackrabbit conditions, on the other hand, there is no emphasis on any band, the sounds were generally externalized and perceived at the loudspeaker, independent of any dampening of individual bands.

Since the scale used for the ER was a nonlinear one, and we observed ceiling effects for the conditions ITE and Jackrabbit, we used the non-parametric Friedman test for the analysis of the median ER over all stimuli. We found significant rank differences at

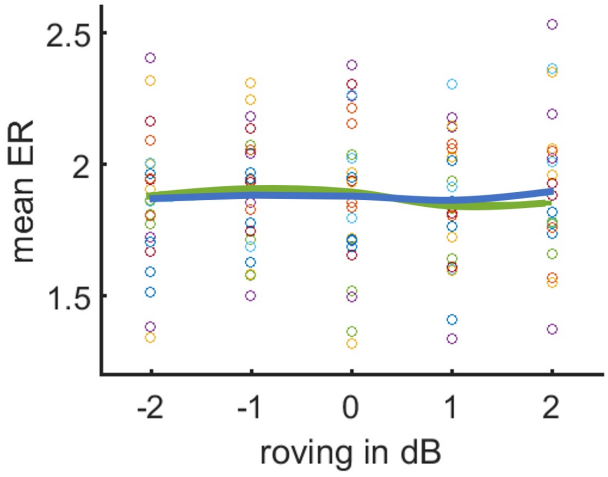

Figure 5.23
Mean externalization rating over all conditions are shown as circles for each individual run of each subject. The blue and green lines represent the median and mean ER for each of the roving levels, respectively.

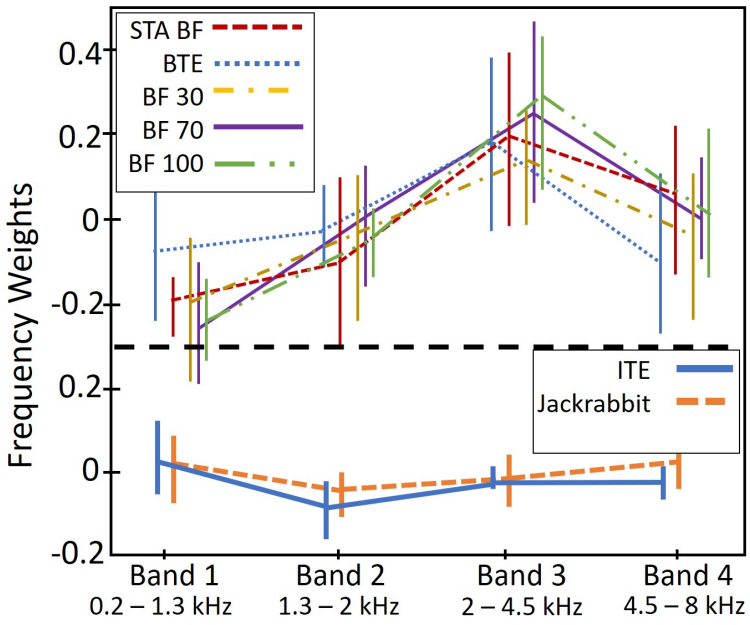

Figure 5.24
Weights for different frequency bands based on the 14 stimuli with dampened bands. Weights for the ITE (blue continuous line) and Jackrabbit (dashed red line) conditions are shown in the bottom. Weights for the BTE (blue), STA BF (orange), BF 30% (yellow), BF 70% (purple) and BF 100% (green) conditions. All conditions in the upper graph show an increased externalization weighting of the 3rd frequency band, compared to the other bands. These graphs reflect a direct change in the ER by the value given on the y-axis.

$p < 0.05$ between the ITE and the STA BF, BF70 and BF100 and internalized conditions, between the Jackrabbit and the BF70 and BF100 and internalized conditions, between the BTE and the BF100 and internalized conditions, and between the BF30 and the internalized condition.

For the analysis of the frequency band weights we used a mixed model multifactorial ANOVA with the fixed factors *band* and *algorithm* and random factor *subjects*, for which all assumptions were fulfilled. No significant differences were found for the factor algorithm, while highly significant differences were found for the factor *band* ($F = 15.34, df = 3, p < 0.0001$), where all bands significantly differed from each other except for band 2 and 3. For the interaction term $band * algorithm$ we found highly significant differences ($F = 6.11, df = 21, p < 0.0001$), but no differences in any bands between conditions ITE, Jackrabbit and internalized in a post-hoc analysis. For all remaining conditions (BTE, STA BF, BF30, BF70 and BF100) band 3 was significantly different from band 1. For the conditions STA BF, BF70 and BF100 band 3 also significantly differed from band 2, and for the conditions BTE, BF70 and BF100 band 3 significantly differed from band 4.

Since in theory the beamforming algorithms should not distort a signal originating from the front, we calculated the polar patterns for all algorithms using HRTF measurements with a 5° resolution using the ITE and both BTE microphones at each ear. Fig. 5.25 shows the normalized polar patterns for the BTE and BF100 conditions for the four frequency bands. No marked differences can be seen for the frontal region which could explain the differences in the externalization ratings, except for band 1 where the beamformer shows a less smooth polar pattern, probably caused by the low frequency roll-off compensation.

5.4.4.1 Binaural Cue Analysis for Externalization Perception

From the individual recordings of the three most experienced participants we analized several cues for different frequency bands. For the fine structure and envelope of the signals we analyzed cues such as ITDs, ILDs, interaural coherence (IC), their standard deviation, minimum and maximum values, percentiles and ratios. As shown in Fig. 5.22, the full BF had the lowest ER and BTE the highest ER from the beamforming strengths tested. The most promising cues correlating with the ER for the different beamforming strengths are shown in Fig. 5.26 color coded. For the frequency band 2 – 4.5 kHz we had the most unambiguous gradings in the cues correlating with the ER, such that the mean cue value of those bands is shown. Eight stimuli of the 18 tested, ranging from bad to good externalization, as can be seen in the spread of each color, were representatively chosen (stimuli 1, 2, 6, 7, 8, 12, 16, 17). The best cue correlations were a decrease of the 75^{th} percentile of the mean envelope ILD magnitude in dB with increasing ER, a decrease in the standard deviation of the envelope ILDs and an increase of the mean envelope coherence values with increasing ER. It should be noted that binaural cue analysis for the Jackrabbit method were not included here, since the Jackrabbit signals here were equal to the ITE signals, as only a target stimulus was presented from the front and thus the Jackrabbit behaves identical to the ITE with it's filter values equal to one.

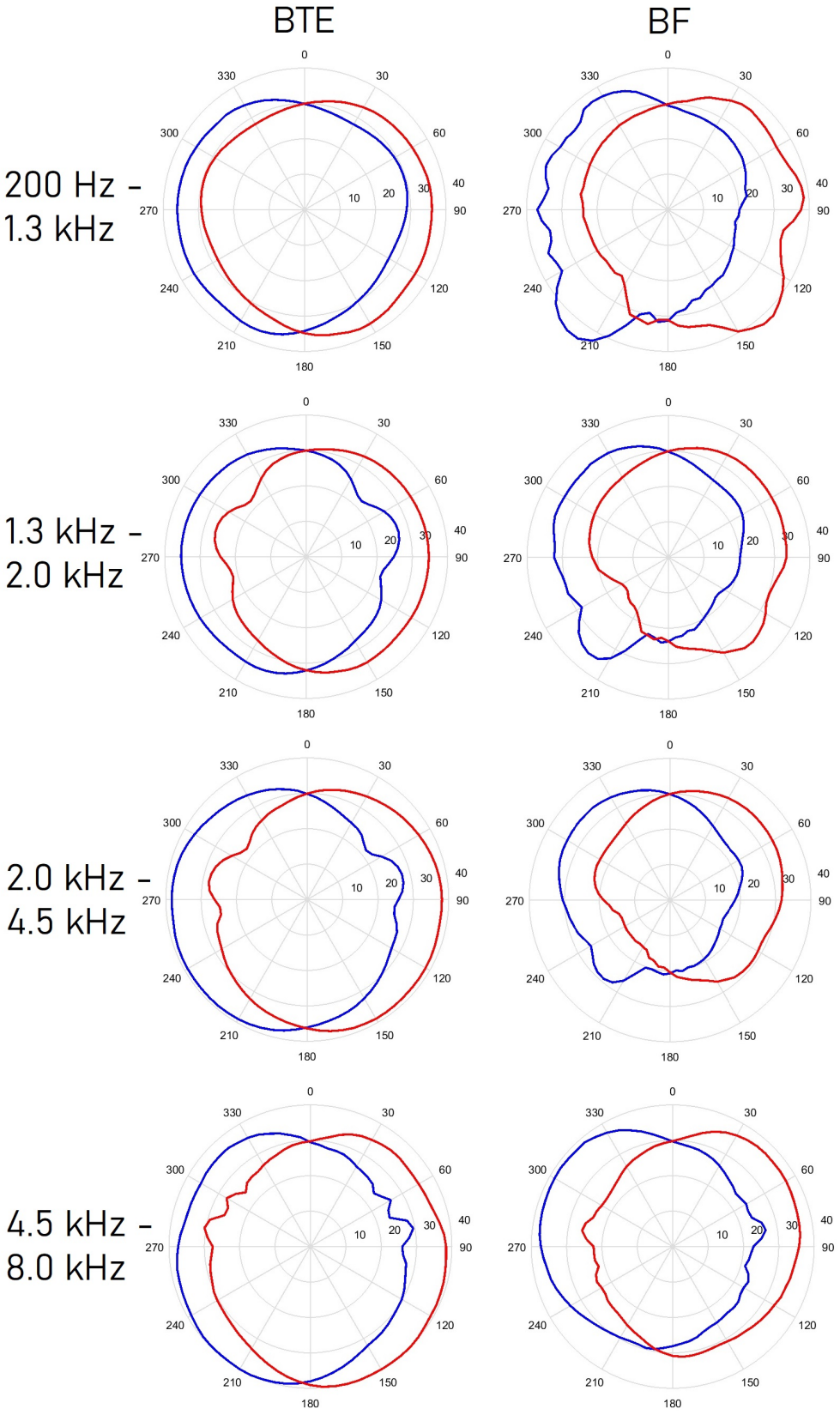

Figure 5.25
Normalized polar patterns of the BTE (left) and BF (right) conditions for the four frequency bands 1 (top) to 4 (bottom). The blue curves represent measured patterns for the left ear and the red lines for the right ear. The rings represent 10 dB steps of attenuation each.

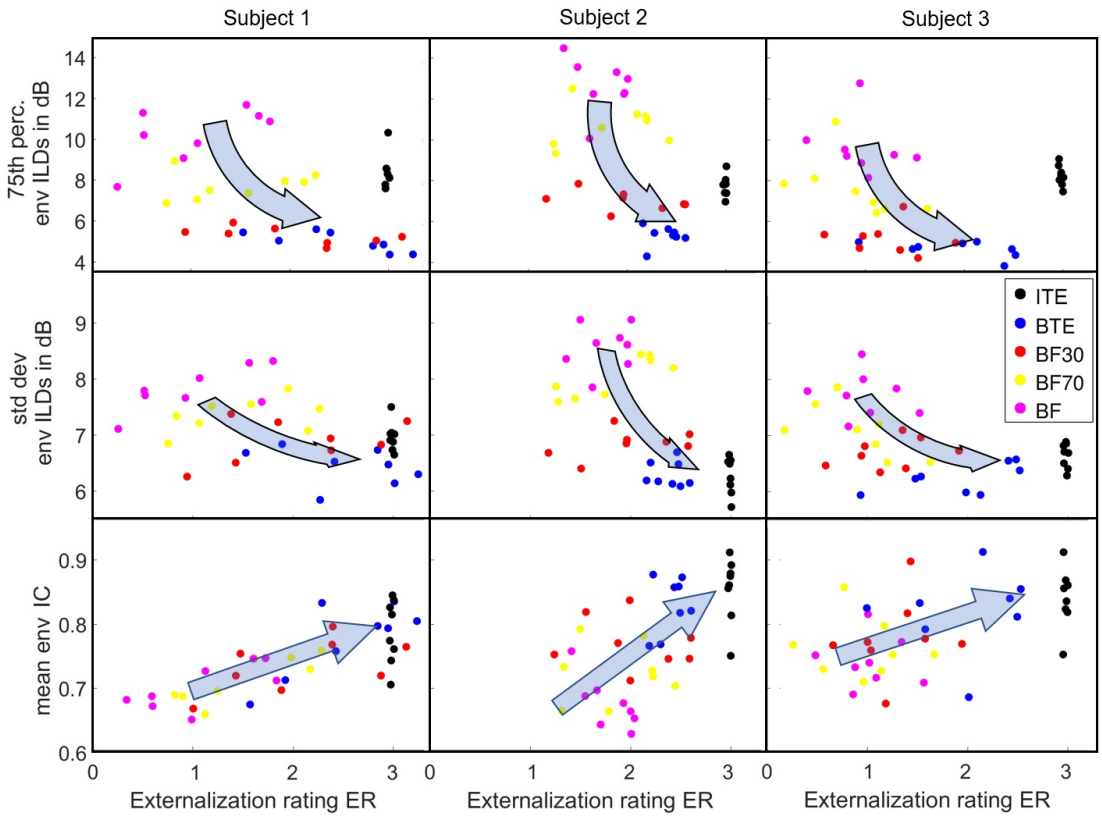

Figure 5.26
Binaural cue analysis for the three participants most experienced with their hearing aids (columns). The upper row shows the 75th percentile of the envelope ILDs in dB, the second row shows the standard deviation of the envelope ILDs in dB and the third row the mean interaural coherence of the frequency bands between 2 – 4.5 kHz for eight selected stimuli against the perceived externalization ratings.

5.4.5 Discussion

The results from Fig. 5.22, showing median externalization ratings over all stimuli and all subjects, clearly show big differences in the perception of externalization of frontal sounds for the different algorithms tested. Especially, it is noticeable that beamforming results in significantly lower ER, perceived close the head, while the ITE condition was perceived at the loudspeaker and the BTE condition close to the loudspeaker. In some situations, participants even perceived the sounds as coming from further away than the loudspeaker in the BTE condition. This remarkably lower ER for the beamformer is unexpected, since from the signal processing aspect and from the measured polar patterns there are no noticeable differences or distortions at 15° in the front that would lead to these results. As we had intended with the design of the experiment, the results also show a gradual decrease in ER with increasing beamforming strength, projecting the perceived sounds from very externalized close to the loudspeaker in the BTE conditions (BF0) to

close to the head in the full BF condition (BF100). This means, that some aspect in the subtraction process of the delay-and-subtract beamformer leads to a distortion of auditory cues relevant for externalization, and the greater the subtraction is, the more these cues are affected. This was further analysed with individual recordings to extract the cues for different hearing aid conditions with help of the auditory model by Dietz et al. (2011) [Dietz et al., 2011]. To achieve a subtraction of sounds coming from the back in the beamforming conditions, we applied a complex filter with amplitude and phase to be able to subtract a subsample-delayed signal of the rear BTE microphone from the frontal microphone. While this leads to a strong attenuation of rear sounds for all frequencies, it is obvious that also sounds from the front are affected by the subtraction of both microphones. Here, different frequencies will be affected differently by the subtraction. While the low frequencies are also attenuated in the front, due to long wavelengths and only small amplitude differences between the BTE microphones, even considering the applied delay, this effect is compensated for by a roll-off compensation filter. At high frequencies, the subtraction will be more noticeable for small wavelengths, which could result in effective gains when both signals are out of phase. Yet, for the microphone distance of the hearing aid of roughly 1 cm, this spatial aliasing does not affect frequencies below 8 kHz, such that this cannot be the reason for perceived internalization. Sounds originating in the front are not perceptually distorted in a way that the sound quality gets affected, as was confirmed in extensive piloting sessions. Stimuli in the BF condition sound similar to the BTE signals, except that the low frequency microphone noise is clearly audible. Yet, apparently, there seems to be a distortion of binaural cues, enough to significantly influence the spatial perception of sounds, like in the externalization perception tested in this experiment.

We conducted a separate analysis of binaural cues using individual recordings of the participants wearing their devices for all stimuli. As seen in Fig. 5.26, we found three cues that represent very well the scaling found in the externalization perception results. These cues show, with increasing ER, first a decrease of the 75^{th} percentile of the mean envelope ILD magnitude in dB, second a decrease in the standard deviation of the envelope ILDs and third an increase of the mean envelope coherence values. Catic et al. ([Catic et al., 2013, Catic et al., 2015]) found that ILD fluctuations decrease with the distance of a source in a reverberant environment, using a range of distances between 0.3 m and 3 m at a 30° angle. They used binaural room impulse responses with microphones placed at the ear canals. In this study, however, we investigated the reduction of externalization with beamforming, using the BTE microphone positions for a fixed distance of 1.3 m at 15° angle. Due to the different microphone position and the different environments tested, with a reverberated environment in the Catic et al. (2013) study [Catic et al., 2013], and a dry one in our case, the cue analysis results are not directly comparable. In our case, the ILD fluctuations, represented by the standard deviation, are reduced with increasing externalization, with the maximum externalization and lowest ILD standard deviation being for the BTE condition (0% beamforming) at similar values to the ITE condition.

Applying beamforming independently for the left and right devices can lead to altered binaural cues that possibly explain the low externalization ratings with beamformers.

From Fig. 5.24 it is apparent that the band from 2 kHz – 4.5 kHz produces a more externalized percept of sounds from the front than the other bands for the conditions that use the BTE microphones. This is the frequency region that is also most prominent in the ITE condition, due to the directional and frequency dependent filtering of the outer ear. Thus, applying a gain in that band compared to the other bands, as is already done in commercial hearing aids, simulates the natural filtering and seems to help in the percept of externalization. In the ITE and Jackrabbit condition, we do not see any significant weightings of individual bands and reach already high externalization ratings independent of the stimulus. The STA BF condition uses the BTE microphones for the beamforming, and applies pinna cues onto that signal. While the pinna cues naturally emphasize the 2 kHz - 4.5 kHz frequency region, the high frequency distortion due to the microphone signal subtraction still seems to dominate, causing much reduced externalization compared to the ITE and Jackrabbit conditions, and only marginally higher ER than the pure BF signals (no significant differences in the Friedman-test).

This experiment tested for externalization perception of frontal sounds with a static head. Not reflected in the results are additional spatial perceptual aspects of the sounds, such as front-back confusions, increased apparent source width or elevated sounds, as reported informally by some participants after conducting the experiment. These aspects might have played an additional role in the responses, since some of the spatial dimensions are linked and could not be completely disentangled by the experimental design. To give an example, one of the participants reported hearing some of the presented sounds at a very large elevation between 60° and 80°. It was up to the individual subjects to project their perception onto the single dimension of externalization. Yet, whether that projection was taken as the Euclidean distance from the head to the perceived sound location, or to the projection onto the horizontal plane (with a reduced externalization/distance range) or something in between cannot be accurately determined. For the BTE condition, some subjects reported hearing the sounds further away than the loudspeaker. While this is a good result if considered only in the externalization dimension, a very externalized sound that is perceived as very wide or diffuse, and thus not easy to accurately localize, is possibly worse than a less externalized but very focused sound percept, if considered in the overall spatial quality.

Regarding the frequency weights, we only considered four frequency bands. The lowest band was chosen to only include the frequency regions where ITDs dominate (200 – 1300 Hz). The third band was chosen from piloting sessions, where that frequency region between 2 kHz – 4.5 kHz led to more externalized sounds when energetically dominant. The remaining two bands were the remaining regions to cover the entire 200 Hz – 8 kHz bandwidth. These bands do not consider the critical band widths of the human auditory system, and it is possible that spectral masking of high energy bands at lower frequencies might have reduced the 15-dB attenuation effect of adjacent higher frequency bands. This question was considered beforehand, but discarded for the experiment since considering critical bands, ITD and ILD dominance regions and spectral masking would have led to a significantly higher experimental complexity and duration, due to the increased number of combinations for attenuating individual bands. A related aspect of higher energy in selected bands influencing spatial perception are the so called Blauert-

bands [Blauert, 1997]. Specific frequency bands lead, when energetically dominant, to a higher probability of sounds being perceived in the front (0.3 – 0.6 kHz and 2.5 – 6 kHz), back (0.8 – 2 kHz and 9 – 15 kHz) or elevated (7 – 9 kHz). This effect is also used in stereo loudspeaker high-fidelity playback to give a presence of closeness and compactness when applying more gain to the frontal Blauert-bands, while the rear Blauert-bands lead to a more distant and diffuse impression ([Sengpiel, 2017]). This effect seems to be the opposite from what we see in our results, with the 2 – 4.5 kHz region leading to a more externalized and distant sound percept. Yet, in our experiment we only see this higher externalization weights for the BTE and BF conditions, while for the ITE and Jackrabbit conditions including pinna cues, which would be the closest setting when listening to stereo loudspeaker hi-fi, there is no effect of any bands on the externalization perception. Also, we only tested for one single source in the front. Thus, the Blauert bands cannot explain or be specifically related to our obtained results.

5.4.6 Conclusions

This experiment considered perceptual differences in externalization for different hearing aid algorithms for a sound source coming from the front and a static head. 18 different stimuli were presented thrice to each participant, with a roving of ±2 dB applied to. 14 of those stimuli were white noise with combinations of dampened frequency bands between 200- 1300 Hz, 1300 – 2000 Hz, 2 – 4.5 kHz and 4.5 – 8 kHz, all with the same overall level, to investigate the effect of individual frequency bands on sound externalization. Externalization ratings were collected using a modified MUSHRA method which allowed for direct comparisons between the 8 hearing aid conditions. The experimental data showed a very good externalization rating for the ITE and Jackrabbit conditions, regardless of the stimulus type. Also, for the internalized anchor condition, ratings were consistently close to 0 (in the head), regardless of the stimulus. Varying degrees of delay-and-subtract beamforming showed a decreased externalization with increasing beamforming strength, ranging from 2.5 ER for the BTE (BF0) condition to 1.5 ER for the full beamformer (BF100). The STA BF condition performed only slightly better than the full beamformer, but the improvement was not significant. The applied roving with a 5dB range did not have any effect on the median externalization ratings, contrary to what might have been expected from a distance cue perspective where level plays an important role. All conditions which used the microphones of the BTE devices showed an increased frequency weight in the 2-4.5 kHz region, meaning that stimuli with higher energy in that frequency band compared to other bands are perceived as more externalized with BTE hearing aid devices for static heads. The importance of pinna cues on externalization becomes evident from the results. While the BTE ratings were only little lower than those of the ITE, participants reported a difference in the compactness between these conditions, with the BTE condition often perceived externalized but broad and more diffuse than in the ITE condition. To better understand the underlying cause of reduced externalization in the beamformer condition, a separate analysis of the binaural cues was conducted. Three cues were found that showed a good correlation with the externalization ratings. These cues show, with increasing ER, first a decrease

of the 75^{th} percentile of the mean envelope ILD magnitude in dB, second a decrease in the standard deviation of the envelope ILDs and third an increase of the mean envelope coherence values. From these findings we can conclude that beamformers in hearing aids reduce the coherence and show increased ILD amplitudes and fluctuations, leading to a reduced perception of externalization, affecting at least those acoustic situations when the head remains still.

6 Summary and Outcomes of the Spatial Perception of Sounds with Hearing Aids

This dissertation and the work underlying addressed several questions on how hearing aid devices affect the perception of sounds in the spatial domain. Firstly, a literature overview was given in chapter 2 that describes the previous findings on localization, distance perception, externalization, apparent source width, front-back confusions, speech understanding and spatial sound quality for normal hearing and hearing-impaired listeners. In chapter 3 a spatial perception study with hearing aids was presented, that simulated a large reverberated room with speech sentences coming from different distances in the front and back of the listener. For each stimulus, different spatial aspects were rated by the test subjects. Results showed a strong influence of microphone directivity, i.e. relative level differences between the front and back, on distance and elevation perception. Overall, a deterioration of spatial sound perception with hearing-aids compared to the unaided baseline was observed, with the BF condition having the biggest negative effect on spatial quality.

Chapter 4 presented two novel methods that reduce noise by static beamforming using both microphones of the BTE devices, and which apply or preserve spatial cues. The STA BF method uses a short time averaging IIR filter to process both the ITE and BTE signals, using the ratio as a spectral cue filter applied to the beamforming signal. The Jackrabbit method takes an inverse approach, using the ITE signal as input signal and reducing disturber energy with a selective noise reduction algorithm based on energy differences between different directions. Chapter 5 presented four different validation studies that compared the new methods to the ITE, BTE and static beamformer signal, in addition to an unaided baseline. The first study on localization showed that in the front, an expansion of the auditory space due to larger binaural cues than at the ear canal leads to large localization errors for the BTE microphone position and the beamformers, while a contraction of the auditory space in the back is caused by smaller binaural cues than in the ear canal, as verified with an HRTF cue analysis. While the beamformer with applied pinna cues (STA BF) showed large localization errors and a high rate of front-back confusions, the Jackrabbit method performed equally well as the ITE condition, with small localization errors and the lowest confusion rate of all aided conditions.

The second study on speech understanding showed a benefit in microphone directionality, with the ITE condition having a 3.6 dB lower (better) threshold than the BTE, and the three beamforming methods led to 11 dB lower thresholds than in the BTE condition in the 1 disturber (1N) case, and 6.4 dB in the disturber with reflection (2N) case. Despite reducing noise not by directional microphone signal subtraction but rather

by spectral subtraction using a time-frequency mask, the Jackrabbit method was able to achieve similar thresholds to the beamforming conditions.

The third study on spatial sound quality firstly investigated the amount of smearing that can be applied with short time averaging of the signals without deteriorating the spatial sound quality. Consecutively, concurrent target and disturber stimuli were rated in the quality of spatial dimensions in a dry and cafeteria like scene for all hearing aid conditions and for an unaided baseline. Results show significantly better spatial ratings for the ITE microphone position than for the BTE microphone position, a benefit of beamformer directivity over omnidirectional microphones to attenuate disturbers and thus to achieve a better spatial separation of sound sources. Best ratings of all aided conditions were achieved with the novel Jackrabbit method that preserves pinna cues and increases SNR by attenuating disturber energy, especially in the dimensions externalization, source separability, saliency and (reduced) diffuseness. The unaided baseline condition was rated by far better than all aided conditions, showing large detrimental effects of hearing aids on spatial perception for normal hearing subjects, due to limiting factors such as a reduced bandwidth, altered or missing binaural and monaural cues, background noise, occlusion effects or some influence of the hearing aid delay.

The final validation study addressed the effect of internalization of sounds with hearing aids. Here, the hearing aid conditions ITE, BTE, STA BF, Jackrabbit and a static beamformer in 30%, 70% and 100% attenuation strength were tested for sounds coming from the front. Results show that the ITE and Jackrabbit conditions were fully externalized, while increasing degrees of beamforming led to greater internalization. Higher energy levels in the frequency band between 2 – 4.5 kHz were beneficial for externalization for the hearing aid conditions using the BTE microphone position. No effect of level roving was found on the perceived externalization. These findings demonstrate that preserving correct pinna cues is important for the externalization percept, and beamforming can lead to strongly internalized sound images for static head positions. A cue analysis showed a strong relationship between both the standard deviation and 75th percentile of the ILDs and of the interaural coherence with perceived externalization in the frequency band between 2 – 4.5 kHz.

In conclusion, the findings of this work clearly show the benefits of preserving pinna cues when using hearing aids for best localization and naturalness. Also, a benefit of beamforming for attenuating disturber sounds, especially beneficial for speech understanding was shown. Combining the advantages of preserving spatial cues and using a beamforming based noise reduction approach led to even better results in all tested scenarios. All experiments were carefully designed to model realistic, demanding sound scenarios with modern technical equipment, allowing participants to wear their dummy hearing aid devices in the sound field with real-time signal processing, thus avoiding many drawbacks of headphone presentation in a sound booth. While the STA BF method showed a slight improvement compared to a static beamformer, the Jackrabbit method clearly outperformed all other hearing aid conditions, demonstrating that noise reduction and the preservation of naturalness with correct spectral cues is useful and can be combined. This method should be considered in the design of future generations of modern hearing aid devices.

Bibliography

[Akeroyd, 2014] Akeroyd, M. A. (2014). An overview of the major phenomena of the localization of sound sources by normal-hearing, hearing-impaired, and aided listeners. *Trends in hearing*, 18:2331216514560442.

[Akeroyd et al., 2007] Akeroyd, M. A., Gatehouse, S., and Blaschke, J. (2007). The detection of differences in the cues to distance by elderly hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 121(2):1077–1089.

[Akeroyd and Whitmer, 2016] Akeroyd, M. A. and Whitmer, W. M. (2016). Spatial hearing and hearing aids. In *Hearing Aids*, pages 181–215. Springer.

[ANSI, 2003] ANSI (2003). S3. 22-2003, specification of hearing aid characteristics. *New York: American National Standards Institute*.

[Bader, 2014] Bader, R. (2014). Microphone array. In *Springer Handbook of Acoustics*, pages 1179–1207. Springer.

[Blauert, 1997] Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization*. MIT press.

[Borish, 1984] Borish, J. (1984). Extension of the image model to arbitrary polyhedra. *The Journal of the Acoustical Society of America*, 75(6):1827–1836.

[Boyd et al., 2012] Boyd, A. W., Whitmer, W. M., Soraghan, J. J., and Akeroyd, M. A. (2012). Auditory externalization in hearing-impaired listeners: The effect of pinna cues and number of talkers. *The Journal of the Acoustical Society of America*, 131(3):EL268–EL274.

[Brandewie and Zahorik, 2010] Brandewie, E. and Zahorik, P. (2010). Prior listening in rooms improves speech intelligibility. *The Journal of the Acoustical Society of America*, 128(1):291–299.

[Bregman, 1994] Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT press.

[Bregman and Pinker, 1978] Bregman, A. S. and Pinker, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 32(1):19.

[Brimijoin and Akeroyd, 2012] Brimijoin, W. O. and Akeroyd, M. A. (2012). The role of head movements and signal spectrum in an auditory front/back illusion. *i-Perception*, 3(3):179.

[Brimijoin and Akeroyd, 2014] Brimijoin, W. O. and Akeroyd, M. A. (2014). The moving minimum audible angle is smaller during self motion than during source motion. *Frontiers in neuroscience*, 8.

[Brimijoin and Akeroyd, 2016] Brimijoin, W. O. and Akeroyd, M. A. (2016). The effects of hearing impairment, age, and hearing aids on the use of self-motion for determining front/back location. *Journal of the American Academy of Audiology*, 27(7):588–600.

[Brimijoin et al., 2013] Brimijoin, W. O., Boyd, A. W., and Akeroyd, M. A. (2013). The contribution of head movement to the externalization and internalization of sounds. *PloS one*, 8(12):e83068.

[Brimijoin et al., 2010] Brimijoin, W. O., McShefferty, D., and Akeroyd, M. A. (2010). Auditory and visual orienting responses in listeners with and without hearing-impairment. *The Journal of the Acoustical Society of America*, 127(6):3678–3688.

[Bronkhorst and Plomp, 1988] Bronkhorst, A. and Plomp, R. (1988). The effect of head-induced interaural time and level differences on speech intelligibility in noise. *The Journal of the Acoustical Society of America*, 83(4):1508–1516.

[Bronkhorst and Houtgast, 1999] Bronkhorst, A. W. and Houtgast, T. (1999). Auditory distance perception in rooms. *Nature*, 397(6719):517–520.

[Byrne and Noble, 1998] Byrne, D. and Noble, W. (1998). Optimizing sound localization with hearing aids. *Trends in Amplification*, 3(2):51–73.

[Calcagno et al., 2012] Calcagno, E., Abregú, E., Eguía, M. C., and Vergara, R. (2012). The role of vision in auditory distance perception. *Perception*, 41(2):175–192.

[Catic et al., 2013] Catic, J., Santurette, S., Buchholz, J. M., Gran, F., and Dau, T. (2013). The effect of interaural-level-difference fluctuations on the externalization of sound. *The Journal of the Acoustical Society of America*, 134(2):1232–1241.

[Catic et al., 2015] Catic, J., Santurette, S., and Dau, T. (2015). The role of reverberation-related binaural cues in the externalization of speech. *The Journal of the Acoustical Society of America*, 138(2):1154–1167.

[Chabot-Leclerc and Dau, 2014] Chabot-Leclerc, A. and Dau, T. (2014). Predicting speech release from masking through spatial separation in distance. In *7th Forum Acusticum*.

[Colburn et al., 2006] Colburn, H. S., Shinn-Cunningham, B., Kidd, Jr, G., and Durlach, N. (2006). The perceptual consequences of binaural hearing. *International Journal of Audiology*, 45(sup1):34–44.

[Colsman et al., 2016] Colsman, A., Aspöck, L., Kohnen, M., and Vorländer, M. (2016). Development of a questionnaire to investigate immersion of virtual acoustic environments. In *Fortschritte der Akustik – DAGA '16*, Aachen, Germany. Dt. Ges. f. Akustik e.V. (DEGA).

[Cox et al., 2004] Cox, T. J., D'Antonio, P., and Avis, M. R. (2004). Room sizing and optimization at low frequencies. *Journal of the Audio Engineering Society*, 52(6):640–651.

[Cubick et al., 2014] Cubick, J., Santurette, S., Laugesen, S., and Dau, T. (2014). Influence of high-frequency audibility on distance perception. *DAGA 2014*, pages 576–577.

[Cubick et al., 2015] Cubick, J., Santurette, S., Laugesen, S., and Dau, T. (2015). The influence of visual cues on auditory distance perception. *DAGA 2015*.

[Dau et al., 2009] Dau, T., Ewert, S., and Oxenham, A. J. (2009). Auditory stream formation affects comodulation masking release retroactively. *The Journal of the Acoustical Society of America*, 125(4):2182–2188.

[Devore et al., 2009] Devore, S., Ihlefeld, A., Hancock, K., Shinn-Cunningham, B., and Delgutte, B. (2009). Accurate sound localization in reverberant environments is mediated by robust encoding of spatial cues in the auditory midbrain. *Neuron*, 62(1):123–134.

[Dietz et al., 2011] Dietz, M., Ewert, S. D., and Hohmann, V. (2011). Auditory model based direction estimation of concurrent speakers from binaural signals. *Speech Communication*, 53(5):592–605.

[Dillon, 2001] Dillon, H. (2001). *Hearing aids*. Thieme.

[Durlach, 1972] Durlach, N. I. (1972). Binaural signal detection-equalization and cancellation theory. *Foundations of Modern Auditory Theory*.

[Ernst et al., 2013] Ernst, S., Grimm, G., and Kollmeier, B. (2013). Evaluation of binaurally-synchronized dynamic-range compression algorithms for hearing aids. In *Proceedings of Meetings on Acoustics ICA2013*, volume 19, page 050085. ASA.

[Faller and Merimaa, 2004] Faller, C. and Merimaa, J. (2004). Source localization in complex listening situations: Selection of binaural cues based on interaural coherence. *The Journal of the Acoustical Society of America*, 116(5):3075–3089.

[Fastl, 2002] Fastl, H. (2002). Psychoacoustics and sound quality. In Jekosch, U., editor, *Tagungsband Fortschritte der Akustik - DAGA 2002, Bochum*, pages 765–766.

[Frank, 2000] Frank, T. (2000). ANSI update: maximum permissible ambient noise levels for audiometric test rooms. *American Journal of Audiology*, 9(1):3–8.

[Freyman et al., 2001] Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *The Journal of the Acoustical Society of America*, 109(5):2112–2122.

[Gil-Carvajal et al., 2016] Gil-Carvajal, J. C., Cubick, J., Santurette, S., and Dau, T. (2016). Spatial hearing with incongruent visual or auditory room cues. *Scientific reports*, 6:37342.

[Gomez, 2016] Gomez, G.: Seeber, B. (2016). Influence of the low-frequency acoustic bypass on distance perception with hearing-aids. In *Proc. 19. Jahrestagung Deutsche Gesellschaft für Audiologie e.V. (DGA)*, pages 1–3. Dt. Ges. f. Audiologie.

[Gomez et al., 2016] Gomez, G., Hoening, V. M., and Seeber, B. U. (2016). Spatial sound perception with hearing aids – examining localization, distance, width, elevation and internalization. In *Proc. International Hearing Aid Research Conference (IHCON)*, page 20, Tahoe City, CA, USA.

[Gomez and Seeber, 2015a] Gomez, G. and Seeber, B. U. (2015a). Distanzwahrnehmung in virtuellen Räumen für Schalle von vorne und hinten. In *DAGA International Conference on Acoustics*.

[Gomez and Seeber, 2015b] Gomez, G. and Seeber, B. U. (2015b). Influence of the hearing aid microphone position on distance perception and front-back confusions with a static head. In *Proc. 18. Jahrestagung Deutsche Gesellschaft für Audiologie e.V. (DGA)*, pages 1–5, Bochum. DGA.

[Good and Gilkey, 1996] Good, M. D. and Gilkey, R. H. (1996). Sound localization in noise: The effect of signal-to-noise ratio. *The Journal of the Acoustical Society of America*, 99(2):1108–1117.

[Hagerman, 1982] Hagerman, B. (1982). Sentences for testing speech intelligibility in noise. *Scandinavian audiology*, 11(2):79–87.

[Hartmann, 1983] Hartmann, W. M. (1983). Localization of sound in rooms. *The Journal of the Acoustical Society of America*, 74(5):1380–1391.

[Hartmann and Constan, 2002] Hartmann, W. M. and Constan, Z. A. (2002). Interaural level differences and the level-meter model. *The Journal of the Acoustical Society of America*, 112(3):1037–1045.

[Hartmann et al., 2016] Hartmann, W. M., Rakerd, B., Crawford, Z. D., and Zhang, P. X. (2016). Transaural experiments and a revised duplex theory for the localization of low-frequency tones. *The Journal of the Acoustical Society of America*, 139(2):968–985.

[Hartmann and Wittenberg, 1996] Hartmann, W. M. and Wittenberg, A. (1996). On the externalization of sound images. *The Journal of the Acoustical Society of America*, 99(6):3678–3688.

[Hassager et al., 2017a] Hassager, H. G., May, T., Wiinberg, A., and Dau, T. (2017a). Preserving spatial perception in rooms using direct-sound driven dynamic range compression. *The Journal of the Acoustical Society of America*.

[Hassager et al., 2017b] Hassager, H. G., Wiinberg, A., and Dau, T. (2017b). Effects of hearing-aid dynamic range compression on spatial perception in a reverberant environment. *The Journal of the Acoustical Society of America*, 141(4):2556–2568.

[Hoening, 2016] Hoening, V. (2016). Untersuchung qualitativer Aspekte der Räumlichkeit mit ITE und BTE Mikrofonen. Bachelor's Thesis, TU München, Germany.

[Hofman et al., 1998] Hofman, P. M., Van Riswick, J., Van Opstal, A. J., et al. (1998). Relearning sound localization with new ears. *Nat Neurosci*, 1(5):417–421.

[Howard and Angus, 2017] Howard, D. M. and Angus, J. (2017). *Acoustics and psychoacoustics*. Focal press.

[Humes et al., 1980] Humes, L. E., Allen, S. K., and Bess, F. H. (1980). Horizontal sound localization skills of unilaterally hearing-impaired children. *Audiology*, 19(6):508–518.

[ITU4R, 2003] ITU4R, R. (2003). Bs. 153441. *Method for the Subjective Assessment of Intermediate Quality Level of Coding Systems, International Telecommunications Union, Geneva, Switzerland*.

[Jensen et al., 2013] Jensen, N. S., Neher, T., Laugesen, S., Johannesson, R. B., and Kragelund, L. (2013). Laboratory and field study of the potential benefits of pinna cue-preserving hearing aids. *Trends in Amplification*, 17(3):171–188.

[Kawashima and Sato, 2015] Kawashima, T. and Sato, T. (2015). Perceptual limits in a simulated "cocktail party". *Attention, Perception, & Psychophysics*, 77(6):2108–2120.

[Keidser et al., 2009] Keidser, G., O'Brien, A., Hain, J.-U., McLelland, M., and Yeend, I. (2009). The effect of frequency-dependent microphone directionality on horizontal localization performance in hearing-aid users. *International journal of audiology*, 48(11):789–803.

[Keidser et al., 2006] Keidser, G., Rohrseitz, K., Dillon, H., Hamacher, V., Carter, L., Rass, U., and Convery, E. (2006). The effect of multi-channel wide dynamic range compression, noise reduction, and the directional microphone on horizontal localization performance in hearing aid wearers. *International Journal of Audiology*, 45(10):563–579.

[Köbler and Rosenhall, 2002] Köbler, S. and Rosenhall, U. (2002). Horizontal localization and speech intelligibility with bilateral and unilateral hearing aid amplification. *International journal of audiology*, 41(7):395–400.

[Kolarik et al., 2013] Kolarik, A. J., Cirstea, S., and Pardhan, S. (2013). Evidence for enhanced discrimination of virtual auditory distance among blind listeners using level and direct-to-reverberant cues. *Experimental brain research*, 224(4):623–633.

[Kolotzek, 2016] Kolotzek, N. (2016). Klangqualität und räumliche Abbildung mit Hörgeräten. Forschungspraxis, TU München, Germany.

[Kolotzek, 2017] Kolotzek, N. (2017). Auswirkung der Kopfdrehung auf die Lokalisation in der Horizontalebene mit Hörgerätesatelliten. Master's Thesis, TU München, Germany.

[Kolotzek et al., 2018] Kolotzek, N., Gomez, G., and Seeber, B. (2018). The effect of head turning on sound localization with hearing-aid satellites. In *Fortschritte der Akustik – DAGA '18*, pages 917–918, München, Germany.

[Korhonen et al., 2015] Korhonen, P., Lau, C., Kuk, F., Keenan, D., and Schumacher, J. (2015). Effects of coordinated compression and pinna compensation features on horizontal localization performance in hearing aid users. *Journal of the American Academy of Audiology*, 26(1):80–92.

[Kuhn, 1987] Kuhn, G. F. (1987). Physical acoustics and measurements pertaining to directional hearing. In *Directional hearing*, pages 3–25. Springer.

[Lindau et al., 2014] Lindau, A., Erbes, V., Lepa, S., Maempel, H.-J., Brinkman, F., and Weinzierl, S. (2014). A spatial audio quality inventory (saqi). *Acta Acustica united with Acustica*, 100(5):984–994.

[Litovsky, 2012] Litovsky, R. Y. (2012). Spatial release from masking. *Acoustics today*, 8(2):18–25.

[Loomis et al., 1998] Loomis, J. M., Klatzky, R. L., Philbeck, J. W., and Golledge, R. G. (1998). Assessing auditory distance perception using perceptually directed action. *Perception & Psychophysics*, 60(6):966–980.

[Lorenzi et al., 1999] Lorenzi, C., Gatehouse, S., and Lever, C. (1999). Sound localization in noise in normal-hearing listeners. *The Journal of the Acoustical Society of America*, 105(3):1810–1820.

[Lu, 2017] Lu, B. (2017). Sprachverständlichkeit für neuartige Hörgerätealgorithmen. Bachelor's Thesis, TU München, Germany.

[Macpherson and Middlebrooks, 2002] Macpherson, E. A. and Middlebrooks, J. C. (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):2219–2236.

[Makous and Middlebrooks, 1990] Makous, J. C. and Middlebrooks, J. C. (1990). Two-dimensional sound localization by human listeners. *The journal of the Acoustical Society of America*, 87(5):2188–2200.

[McNemar, 1947] McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12(2):153–157.

[McShefferty et al., 2015] McShefferty, D., Whitmer, W. M., and Akeroyd, M. A. (2015). The just-noticeable difference in speech-to-noise ratio. *Trends in hearing*, 19:2331216515572316.

[Meier, 1964] Meier, H. (1964). *Deutsche Sprachstatistik*, volume 1. G. Olms Verlagsbuchhandlung.

[Mershon, 1997] Mershon, D. H. (1997). Phenomenal geometry and the measurement of perceived auditory distance. *Binaural and spatial hearing in real and virtual environments*, pages 257–274.

[Mershon et al., 1989] Mershon, D. H., Ballenger, W. L., Little, A. D., McMurtry, P. L., and Buchanan, J. L. (1989). Effects of room reflectance and background noise on perceived auditory distance. *Perception*, 18(3):403–416.

[Mershon and Bowers, 1979] Mershon, D. H. and Bowers, J. N. (1979). Absolute and relative cues for the auditory perception of egocentric distance. *Perception*, 8(3):311–322.

[Mershon and Hutson, 1991] Mershon, D. H. and Hutson, W. E. (1991). Toward the indirect measurement of perceived auditory distance. *Bulletin of the Psychonomic Society*, 29(2):109–112.

[Middlebrooks et al., 1989] Middlebrooks, J. C., Makous, J. C., and Green, D. M. (1989). Directional sensitivity of sound-pressure levels in the human ear canal. *The Journal of the Acoustical Society of America*, 86(1):89–108.

[Min and Mershon, 2005] Min, Y.-K. and Mershon, D. H. (2005). An adjacency effect in auditory distance perception. *Acta acustica united with acustica*, 91(3):480–489.

[Mine, 1993] Mine, M. R. (1993). Characterization of end-to-end delays in head-mounted display systems. *The University of North Carolina at Chapel Hill, TR93-001*.

[Monaghan et al., 2013] Monaghan, J. J., Krumbholz, K., and Seeber, B. U. (2013). Factors affecting the use of envelope interaural time differences in reverberation. *The Journal of the Acoustical Society of America*, 133(4):2288–2300.

[Moore and Glasberg, 1996] Moore, B. C. and Glasberg, B. R. (1996). A revision of Zwicker's loudness model. *Acta Acustica united with Acustica*, 82(2):335–345.

[Mueller et al., 2010] Mueller, H. G., Johnson, E., and Weber, J. (2010). Fitting hearing aids: A comparison of three pre-fitting speech tests. *AudiologyOnline, Article*, 2332.

[Musa-Shufani et al., 2006] Musa-Shufani, S., Walger, M., von Wedel, H., and Meister, H. (2006). Influence of dynamic compression on directional hearing in the horizontal plane. *Ear and hearing*, 27(3):279–285.

[Musicant and Butler, 1984] Musicant, A. D. and Butler, R. A. (1984). The influence of pinnae-based spectral cues on sound localization. *The Journal of the Acoustical Society of America*, 75(4):1195–1200.

[Noble and Byrne, 1990] Noble, W. and Byrne, D. (1990). A comparison of different binaural hearing aid systems for sound localization in the horizontal and vertical planes. *British Journal of Audiology*, 24(5):335–346.

[Noble et al., 1994] Noble, W., Byrne, D., and Lepage, B. (1994). Effects on sound localization of configuration and type of hearing impairment. *The Journal of the Acoustical Society of America*, 95(2):992–1005.

[Noble et al., 1998] Noble, W., Sinclair, S., and Byrne, D. (1998). Improvement in aided sound localization with open earmolds: observations in people with high-frequency hearing loss. *Journal-American Academy of Audiology*, 9:25–34.

[Ohl et al., 2010] Ohl, B., Laugesen, S., Buchholz, J., and Dau, T. (2010). Externalization versus internalization of sound in normal-hearing and hearing-impaired listeners. In *Jahrestagung der Deutschen Gesellschaft für Akustik*. Deutsche Gesellschaft für Akustik.

[Perrett and Noble, 1997] Perrett, S. and Noble, W. (1997). The contribution of head motion cues to localization of low-pass noise. *Attention, Perception, & Psychophysics*, 59(7):1018–1026.

[Petzschner and Glasauer, 2011] Petzschner, F. H. and Glasauer, S. (2011). Iterative bayesian estimation as an explanation for range and regression effects: a study on human path integration. *The Journal of Neuroscience*, 31(47):17220–17229.

[Petzschner et al., 2015] Petzschner, F. H., Glasauer, S., and Stephan, K. E. (2015). A bayesian perspective on magnitude estimation. *Trends in cognitive sciences*, 19(5):285–293.

[Pickles, 1988] Pickles, J. O. (1988). *An introduction to the physiology of hearing*, volume 2. Academic press London.

[Picou et al., 2014] Picou, E. M., Aspell, E., and Ricketts, T. A. (2014). Potential benefits and limitations of three types of directional processing in hearing aids. *Ear and hearing*, 35(3):339–352.

[Pumford et al., 2000] Pumford, J. M., Seewald, R. C., Scollie, S. D., and Jenstad, L. M. (2000). Speech recognition with in-the-ear and behind-the-ear dual-microphone hearing instruments. *Journal of the American Academy of Audiology*, 11(1):23–35.

[Rayleigh, 1907] Rayleigh, L. (1907). On our perception of sound direction. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 13(74):214–232.

[Schwartz and Shinn-Cunningham, 2013] Schwartz, A. H. and Shinn-Cunningham, B. G. (2013). Effects of dynamic range compression on spatial selective auditory attention in normal-hearing listeners. *The Journal of the Acoustical Society of America*, 133(4):2329–2339.

[Seeber, 2002] Seeber, B. (2002). A new method for localization studies. *Acustica-Stuttgart*, 88(3):446–449.

[Seeber et al., 2003] Seeber, B., Fastl, H., and Koci, V. (2003). Ein PC-basiertes Békésy-Audiometer mit Bark-Skalierung (A PC-based Békésy-audiometer using Bark frequency scaling). In *Fortschritte der Akustik–DAGA'03*.

[Seeber et al., 2010] Seeber, B. U., Kerber, S., and Hafter, E. R. (2010). A system to simulate and reproduce audiovisual environments for spatial hearing research. *Hearing research*, 260(1):1–10.

[Sengpiel, 2017] Sengpiel, E. (last visited oct. 2017). Die Bedeutung der Blauertschen Bänder für die Tonaufnahme'. *UdK Berlin http://www.sengpielaudio.com/DieBedeutungDerBlauertschenBaender.pdf*.

[Shinn-Cunningham, 2008] Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in cognitive sciences*, 12(5):182–186.

[Slattery and Middlebrooks, 1994] Slattery, W. H. and Middlebrooks, J. C. (1994). Monaural sound localization: acute versus chronic unilateral impairment. *Hearing research*, 75(1):38–46.

[Thompson, 1882] Thompson, S. P. (1882). On the function of the two ears in the perception of space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 13(83):406–416.

[Udesen et al., 2013] Udesen, J., Piechowiak, T., Gran, F., and Dittberner, A. B. (2013). Degradation of spatial sound by the hearing aid. In *Proceedings of the International Symposium on Auditory and Audiological Research*, volume 4, pages 271–278.

[Van den Bogaert et al., 2011] Van den Bogaert, T., Carette, E., and Wouters, J. (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):164–176.

[Van den Bogaert et al., 2006] Van den Bogaert, T., Klasen, T. J., Moonen, M., Van Deun, L., and Wouters, J. (2006). Horizontal localization with bilateral hearing aids: Without is better than with. *The Journal of the Acoustical Society of America*, 119(1):515–526.

[Völk, 2010] Völk, F. (2010). Psychoakustische Experimente zur Distanz mittels Wellenfeldsynthese erzeugter Hörereignisse. *DAGA 2010*, pages 1065–1066.

[Von Békésy and Wever, 1960] Von Békésy, G. and Wever, E. G. (1960). *Experiments in hearing*, volume 8. McGraw-Hill New York.

[von Unold, 2017] von Unold, P. (2017). Externalisierung bei Hörgeräten. Bachelor's Thesis, TU München, Germany.

[Vorländer and Summers, 2008] Vorländer, M. and Summers, J. E. (2008). Auralization: Fundamentals of acoustics, modelling, simulation, algorithms, and acoustic virtual reality. *Acoustical Society of America Journal*, 123:4028.

[Wagener et al., 1999] Wagener, K., Brand, T., and Kollmeier, B. (1999). Entwicklung und Evaluation eines Satztests für die deutsche Sprache iii: Evaluation des Oldenburger Satztests. *Zeitschrift für Audiologie/Audiological Acoustics*, 38:8695.

[Wallach, 1940] Wallach, H. (1940). The role of head movements and vestibular and visual cues in sound localization. *Journal of Experimental Psychology*, 27(4):339.

[Wenzel et al., 1993] Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (1993). Localization using nonindividualized head-related transfer functions. *The Journal of the Acoustical Society of America*, 94(1):111–123.

[Westermann and Buchholz, 2013] Westermann, A. and Buchholz, J. M. (2013). Release from masking through spatial separation in distance in hearing impaired listeners. In *Proceedings of Meetings on Acoustics*, volume 19, page 050156. Acoustical Society of America.

[Whitmer et al., 2012] Whitmer, W. M., Seeber, B. U., and Akeroyd, M. A. (2012). Apparent auditory source width insensitivity in older hearing-impaired individuals. *The Journal of the Acoustical Society of America*, 132(1):369–379.

[Wiggins and Seeber, 2011] Wiggins, I. M. and Seeber, B. U. (2011). Dynamic-range compression affects the lateral position of sounds. *The Journal of the Acoustical Society of America*, 130(6):3939–3953.

[Wiggins and Seeber, 2012] Wiggins, I. M. and Seeber, B. U. (2012). Effects of dynamic-range compression on the spatial attributes of sounds in normal-hearing listeners. *Ear and hearing*, 33(3):399–410.

[Wiggins and Seeber, 2013] Wiggins, I. M. and Seeber, B. U. (2013). Linking dynamic-range compression across the ears can improve speech intelligibility in spatially separated noise. *The Journal of the Acoustical Society of America*, 133(2):1004–1016.

[Wightman and Kistler, 1997] Wightman, F. L. and Kistler, D. J. (1997). Monaural sound localization revisited. *The Journal of the Acoustical Society of America*, 101(2):1050–1063.

[Wouters et al., 1999] Wouters, J., Litière, L., and Van Wieringen, A. (1999). Speech intelligibility in noisy environments with one-and two-microphone hearing aids. *Audiology*, 38(2):91–98.

[Yost, 2001] Yost, W. A. (2001). *Fundamentals of hearing: an introduction*. Academic Press, San Diego, CA.

[Zahorik, 2002] Zahorik, P. (2002). Assessing auditory distance perception using virtual acoustics. *The Journal of the Acoustical Society of America*, 111(4):1832–1846.

[Zahorik et al., 2005] Zahorik, P., Brungart, D. S., and Bronkhorst, A. W. (2005). Auditory distance perception in humans: A summary of past and present research. *Acta Acustica united with Acustica*, 91(3):409–420.

[Zhang and Hartmann, 2010] Zhang, P. X. and Hartmann, W. M. (2010). On the ability of human listeners to distinguish between front and back. *Hearing research*, 260(1-2):30–46.

[Zwicker, 1961] Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands (Frequenzgruppen). *The Journal of the Acoustical Society of America*, 33(2):248–248.

[Zwicker and Fastl, 2013] Zwicker, E. and Fastl, H. (2013). *Psychoacoustics: Facts and models*, volume 22. Springer Science & Business Media.

[Zwicker and Scharf, 1965] Zwicker, E. and Scharf, B. (1965). A model of loudness summation. *Psychological review*, 72(1):3.