Technical University of Berlin
Faculty I - Humanities
Audio Communication Group

Master's Thesis

Static and Dynamic Localization Cues of Head-Related and Spherical Head Transfer Functions

Submitted by: Silke Bögelein
[redacted]

Submition Date: September 11, 2019

1. Examiner: Prof. Dr. Stefan Weinzierl

2. Examiner: Fabian Brinkmann

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt gegenüber der Fakultät I der Technischen Universität Berlin, dass die vorliegende, dieser Erklärung angefügte Arbeit selbstständig und nur unter Zuhilfenahme der im Literaturverzeichnis genannten Quellen und Hilfsmittel angefertigt wurde. Alle Stellen der Arbeit, die anderen Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind kenntlich gemacht. Ich reiche die Arbeit erstmals als Prüfungsleistung ein. Ich versichere, dass diese Arbeit oder wesentliche Teile dieser Arbeit nicht bereits dem Leistungserwerb in einer anderen Lehrveranstaltung zugrunde lagen.

Mit meiner Unterschrift bestätige ich, dass ich über fachübliche Zitierregeln unterrichtet worden bin und diese verstanden habe. Die im betroffenen Fachgebiet üblichen Zitiervorschriften sind eingehalten worden. Eine Überprüfung der Arbeit auf Plagiate mithilfe elektronischer Hilfsmittel darf vorgenommen werden.

Berlin, September 11, 2019
Ort, Datum

Silke Bögelein

Abstract

Spherical head models have been used for many years in various applications as an alternative to individual Head-related Transfer Functions (HRTFs). Using these models, sound sources can be localized very accurate in the horizontal plane. In the median plane, however, sound source localization normally is based on so-called monaural cues. Those cues are missing in spherical head models. Furthermore, the influence of motion induced cues has not yet been investigated for spherical head models.

In this thesis, static and dynamic localization cues of HRTFs and Spherical Head Transfer Functions (SHTFs) are investigated in detail. A theoretical consideration of the cues for three geometries (KEMAR's artificial head, sphere, ellipsoid) is carried out. Therefore, transfer functions of KEMAR's head and the ellipsoid are calculated using the Boundary Element Method (BEM), whereas SHTFs are analytically generated. In addition, the influence of off-set ears is examined and declared advantageous. A model of localization mismatch between simplified geometries and human head using dynamic cues is presented.

In addition to the theoretical examination, a perceptual evaluation is carried out. A listening test is performed in Virtual Reality (VR) that compares individual representations of the spherical head model and individual HRTFs. In addition, different head movements (dynamic, static) and transfer functions (magnitude, original, phase) are used to analyze the influence of spectral and temporal cues. The participants must locate sound sources in the median plane. Two measures of error (quadrant error and local polar bias) are considered when assessing localization ability. The statistical analysis is performed after a three factorial repeated measures Multivariat Analysis of Varianz (MANOVA). A significant influence can only be determined on the quadrant error. Key outcome is that the quadrant error is reduced by using head movements. In contrast, the geometry itself plays a minor role.

Zusammenfassung

Kugelkopfmodelle werden seit vielen Jahren in verschiedenen Anwendungen als Alternative zu individuellen Head-related Transfer Functions (HRTFs) eingesetzt. Mit diesen Modellen können Schallquellen sehr genau in der horizontalen Ebene lokalisiert werden. In der Medianebene basiert die Lokalisierung von Schallquellen jedoch in der Regel auf sogenannten monauralen Cues. Diese Cues fehlen bei Kugelkopfmodellen. Der Einfluss von bewegungsinduzierten Cues für Kugelkopfmodelle wurde noch nicht untersucht.

In dieser Arbeit werden statische und dynamische Lokalistions Cues von HRTFs und SHTFs im Detail analysiert. Eine theoretische Betrachtung der Cues für drei Geometrien (Kunstkopf, Kugel, Ellipsoid) wird durchgeführt. Die Übertragungsfunktionen des KEMAR Kunstkopfes und des Ellipsoids werden mit Hilfe der Boundary Element Method (BEM) berechnet, während SHTFs analytisch generiert werden. Darüber hinaus wird der Einfluss von versetzten Ohren untersucht und für vorteilhaft befunden. Ein Modell, das Lokalisations-Mismatchs zwischen vereinfachten Geometrien und dem menschlichen Kopf unter Verwendung dynamischer Cues darstellt, wird vorgestellt.

Zusätzlich zur theoretischen Betrachtung wird eine perzeptuelle Analyse durchgeführt. Ein Lokalisationstest wird in Virtual Reality (VR) ausgeführt, der individuelle Darstellungen des Kugelkopfmodells und individueller HRTFs vergleicht. Darüber hinaus werden verschiedene Kopfbewegungen (dynamisch, statisch) und Übertragungsfunktionen (Magnitude, Original, Phase) verwendet, um den Einfluss von spektralen und zeitlichen Cues zu analysieren. Die Teilnehmer müssen Schallquellen in der Medianebene lokalisieren. Bei der Beurteilung der Lokalisierbarkeit werden zwei Fehlermaße (Quadrantenfehler und lokaler Polar Bias) berücksichtigt. Die statistische Analyse erfolgt nach einer drei faktoriellen MANOVA mit Messwiederholung. Ein signifikanter Einfluss kann nur auf den Quadrantenfehler festgestellt werden. Die wichtigste Aussage ergibt, dass der Quadrantenfehler durch Kopfbewegungen reduziert wird. Im Gegensatz dazu spielt die Geometrie an sich eine untergeordnete Rolle.

Table of Contents

Nomenclature

Physics

λ_{max}	Maximum wave length
μ	Normalized frequency
ϕ	Altitude angle
Ψ	Infinite series expansion
ρ_0	Density of air
Θ	Angle between two points on great circle
θ	Polar angle
φ	Azimuth angle
ϱ	Normalized distance to sound source
ϑ	Elevation angle
a	Radius of sphere
c	Speed of Sound
d	Depth of head
f	Frequency
H	General transfer function
h	Height of head
$h_m^{(2)}$	m-th order spherical Hankel function of second kind
j	Complex number
k	Wave number
L	Sound pressure level
$L_{Aeq,T}$	A-weighted energy equivalent continuous sound pressure level
m	Order index
p	Sound pressure
p_0	Free field sound pressure

P_m	Legendre polynomial of degree m
Q	Strength of point source
Q_m	m-th order modified spherical Hankel function of second kind
r	Radius
w	Width of head
x	X-coordinate
y	Y-coordinate
z	Z-coordinate

Statistics

χ^2	Chi-Square
η_p^2	Effect size
μ_θ	Mean polar bias
θ_{error}	Local polar bias
p	Significance level
df	Degree of freedom
F	F-ratio
L_n	n-th level of a factor
M	Mean value
std	Standard deviation

List of Abbreviations

BEM	Boundary Element Method
BIE	Boundary Integral Equation
EHIR	Ellipsoid Head Impulse Response
EHTF	Ellipsoid Head Transfer Function
FABIAN	Fast and Automatic Binaural Impulse response AcquisitoN
HMD	Head Mounted Device
HP	High-Pass
HRIR	Head-related Impulse Response
HRTF	Head-related Transfer Function
ILD	Interaural Level Difference
ITD	Interaural Time Difference
JND	Just Noticeable Difference
LP	Low-Pass
MANOVA	Multivariat Analysis of Varianz
ML-FMM	Multi-Level Fast Multipole Method
MTB	Motion Tracked Binaural
SHIR	Spherical Head Impulse Response
SHTF	Spherical Head Transfer Function
SOFA	Spatially Oriented Format for Acoustics
TF	Transfer Function
TOA	Time Of Arrival
VR	Virtual Reality

1 Introduction

The human auditory system, amongst many other physiological functions, allows the listener to locate sound sources in three-dimensional space. The localization of sound sources is permanently encountered in everyday life. In addition to visual perception, it plays an important role in orientation in rooms or the surroundings. The localization ability helps in cases such as assessing dangerous situations in road traffic (approaching car, horn, siren) or focusing acoustic objects (speech, music). However, the spatial hearing does not necessarily have to take place in a real environment. The constantly increasing number of virtual reality applications requires an equally realistic sound experience. With the increasing popularity of virtual reality, the 3D sound becomes more and more important for stimulating the acoustic sense (Cuevas-Rodríguez et al., 2019). This means that the realistic impression is strengthened by different sound sources that can be precisely localized. One possibility to achieve an authentic three dimensional virtual sound impression is by using binaural synthesis (Brinkmann et al., 2017b).

Three dimensional sound source localization takes place on two spatial dimensions. These dimensions are the horizontal direction and the median direction. Depending on the respective dimension, different characteristic features, so-called localization cues, exist that are evaluated for the perception of a spatial orientation of a sound source. These cues are either determined by temporal and spectral relations between the two ears (binaural cues) or by general spectral characteristics of the received sound that are caused by the fine structure of the pinnae (monaural cues). It is widely accepted that the accurate localization in the median plane relies mainly on the latter. Nevertheless, weaker elevation cues originate from reflections and shadowing at the human shoulder and torso, head-induced spectral cues, and Interaural Time Difference (ITD) fine structure (Algazi et al., 2001a; Algazi and Duda, 2002; Benichoux et al., 2016).

The previously mentioned localization cues are static cues. They exist without any motion of the source or the listener and are included in the individual Head-related Transfer Function (HRTF). However, additional motion cues originate from movements of the source or listener, inducing temporal changes of the static localization cues. Various studies have shown that these can lead to an improvement in the localization of sound sources, especially in median plane (McAnally and Martin, 2014; Jiang et al., 2018).

The processing of user specific head-related characteristics for virtual reality applications can be technically demanding. The same applies for the application of sound field reproduction methods (Algazi et al., 2004). A generally accepted simplification is therefore the use of so-called spherical head models resulting in Spherical Head Transfer Functions (SHTFs) (Duda and Martens, 1998; Cuevas-Rodríguez et al., 2019). By approximating

the head as a sphere, sound source localization in the horizontal plane works almost as well as with an individual HRTF (Xie, 2013). In contrast, localization in the median plane might be particularly difficult, due to the missing pinnae. However, this topic has not been further investigated. Only very few studies have produced weak evidence that a localization in the median plane is still possible without a pinnae (Algazi et al., 2001a; Fiedler et al., 2017).

1.1 Objective and Outline

This work is devoted to the question of to what extent a localization of sound sources using a spherical head model is possible and, above all, which localization cues serve as the basis for it. The main focus here is on the motion cues, since the assumption is that these contribute significantly to localization with spherical head models in the median plane. In addition, the influence of the ear positions on a spherical head model is analyzed. The transfer function of an ellipsoid is also compared with that of the sphere, since the ellipsoid is also a simple model and corresponds more to a human's head shape.
Up to now, the relevance and effect of motion cues has not been studied for the case of a spherical head model. Sound source localization using transfer functions by spherical head models is poor in static scenes, but might improve as soon as motion cues come into play. The current study thus investigates the similarity of static and motion cues found in HRTFs and SHTFs. This is done within a physical and a perceptual evaluation of the research aim.

Chapter 2 first describes all the basics, as well as related work, which contribute to the overall understanding of the study.
The following Chapter 3 deals with the physical evaluation of the question. Comparisons of simulated transfer functions of three different geometries (human head, sphere, ellipsoid) were made. The influence of off-set ears on the spherical head is analyzed. In addition, a simple model for estimating the localization error in the median plane based on dynamic Interaural Time Difference (ITD) and Interaural Level Difference (ILD) cues is introduced.
Chapter 4 evaluates theoretical considerations with a perceptual listening test. The listening test took place in a Virtual Reality (VR) environment. Additionally, the feasibility of localization experiments in VR is also discussed.
Chapter 5 then compares the in Chapter 3 introduced model for estimating the localization error with data from the listening test.
Finally, Chapter 6 gives an overall summary and discussion and additionally deals with possible future work approaches.

2 Fundamentals and Related Work

This Chapter serves as a more detailed introduction to the topic of the thesis. It also includes all necessary basics, in order to be able to understand the further expiration of this work.

2.1 Coordinate System

In the further context it is important to define coordinate systems in order to be able to indicate positions of sound sources relative to the position of the observer. Blauert (1997) defines the structure of the coordinate systems as follows: The origin of the coordinate systems is determined by the intersection of the three planes median plane, horizontal plane and frontal plane. The centre of the frontal plane is obtained by connecting the upper points of the entrance of the auditory canal and the centre between these two points. The horizontal plane is at the level of the lower end of the eye socket with a right angle to the frontal plane. The median plane is at right angles to the other two planes. Together they form the center of the coordinate system at the intersection.

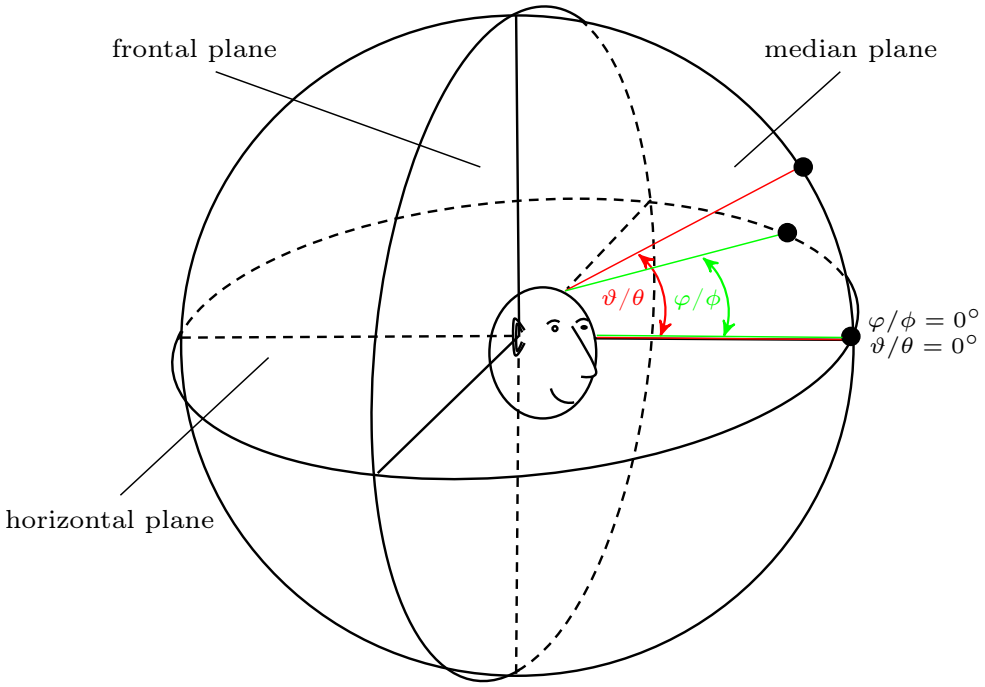


Figure 2.1: Structure of a head-related coordinate system (similar to Blauert (1997) Figure 1.4) where φ/ϕ represent angles in the horizontal plane (Azimuth/Altitude) and ϑ/θ representing angles in the median plane (Elevation/Polar).

In this thesis two coordinate systems were used. A scheme of head-related coordinate systems is shown in Figure 2.1. On the one hand a vertical polar coordinate system was used, where the horizontal plane is the main plane of the coordinate system (Blauert, 1997). The horizontal angle is called Azimuth (φ) and the angle related to the median plane Elevation (ϑ). Azimuth angles are defined from 0° to 360° starting from 0° frontal (see Figure 2.1). Elevation angles are defined from -90° at the lower pole, over 0° frontal, to +90° at the upper pole.
On the other hand a interaural polar coordinate system was used (Baumgartner et al., 2014; Ziegelwanger, 2012). In this case, angles on the median plane are called Polar angles (θ). These represent the main plane. Angles on the horizontal plane are called Latitudes (ϕ). Polar angles are defined from -90° (lower pole) over 0° frontal and from there over +90° at the upper pole to +270° to the lower pole. Latitude angles go from -90° over 0° frontal to +90°. Azimuth respectively Altitude angles are both used clockwise in this work.

2.2 Localization

Sound source localization describes the ability of a person to orientate himself in his acoustic environment. They are an important part of hearing.
Localization cues depend on the angle of incidence of sound and can be broken down into subsequent cues.
First there is a subdivision within the different planes of the coordinate system. This results in the sound source localization in the horizontal plane, which concentrates on altitude angles as well as a localization of the sound source in the median plane, respectively in the vertical direction, the polar angle. A combination of altitude and polar angle leads to a combination of different cues.
In the horizontal plane the cues consist of the interaction of both ears, therefore they are called binaural cues. In the median plane, however, they refer to one ear and are called monaural cues. All cues are individual and depend on the size and shape of the head as well as the position and shape of the pinnae. Additionally, all cues are frequency dependent. Every individual has learned to use their specific cues for locating a sound source.

2.2.1 Binaural Cues

As already mentioned, binaural cues develop in the horizontal direction and describe the differences between sound pressure signals at the two ears. They can be divided

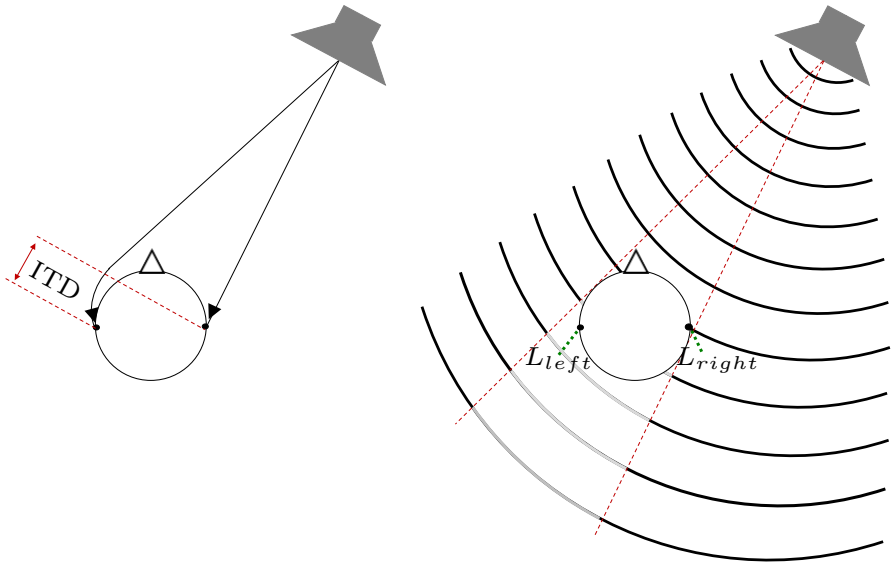


Figure 2.2: Scheme of binaural cues for sound source localization. The left side shows the different time of arrivals for both ears, thus ITD cues. The right side shows shadowing by the head and therefore different sound pressures or sound pressure levels (L_{left} and L_{right}) at both ears, which lead to ILD cues.

into Interaural Time Difference (ITD) of the incoming sound wave between both ears and Interaural Level Difference (ILD) between both ears. Figure 2.2 shows these two subdivisions. The left side of Figure 2.2 shows the principle of interaural time differences. The sound event reaches the respective ear at different times. The arrival time of the propagating sound wave from one source to one ear is called Time Of Arrival (TOA). The interaural time difference is calculated from the subtraction of the TOA of one ear and the TOA of the second ear. This results in time and phase differences between both ears. The resulting ITD depends on the position of the sound source.

The maximum ITD for humans is between $\pm 600\,\mu s$ and $\pm 800\,\mu s$ and occur at an angle of incidence of 90° (Middlebrooks, 1999). The perceptive threshold is about $10\,\mu s$ (Zwislocki and Feldman, 1956). Interaural time differences are the preeminent cues for localization in low frequencies up to 1.5 kHz (Mills, 1972; Irvine, 1992). In the wider frequency range up to 4 kHz, the interaural time differences overlap with the cues of the Interaural Level Difference (ILD). In the frequency range from 4 kHz, the ILDs play a dominant role (Xie, 2013).

The right side of Figure 2.2 shows the principle of the interaural level difference. Due to shadowing by the head, the level L_{left} at the contralateral side is lower than the level L_{right} at the ipsilateral side by the sound wave. This applies to the frequency range where the wavelengths are smaller than the head diameter. Thus, sound source localization by means of ILD takes place in the range of high frequencies from 4 kHz (Mills, 1972;

Irvine, 1992). If the head diameter is about the same as the wavelength, the sound field is diffracting around the head. If the wavelength is above the head diameter, this has no influence on the interaural level difference.
Interaural level differences can exceed 20 dB and are highest for sound sources at 90° (Gulick et al., 1989). The perceptive threshold is about 1 dB (Hall, 1968) and is called Just Noticeable Difference (JND). However, this value depends on the volume of the sound source. For sound events below 40 dB, this value may deviate and increase to up to 3 dB (Hershkowitz and Durlach, 1969).
For conflicting ITD and ILD cues, the sound source localization is performed based on the ITD cues (Wightman and Kistler, 1992; Macpherson and Middlebrooks, 2002).
The localization blur in the horizontal plane is between ±3.6° and ±10°, depending on the angle of incidence. It is lowest for a sound incidence angle of 0° and highest for a lateral sound incidence angle of 90° (Blauert, 1997).

2.2.2 Monaural Cues

Monaural cues are used for the localization process in the median plane. They refer to spectral influences and are therefore also called spectral cues.
Since the distances of the sound source to both ears are the same on the median plane, only little information are available through binaural cues (ITD and ILD) for sound source localization (Algazi et al., 1999; Benichoux et al., 2016). The assignment of the location of the sound source in median plane is much more based on the spectral change caused by the filtering effect of the pinnae. An approximate symmetric head and ear shape is assumed. Due to those assumptions they are often called monaural cues. The filtering effect for a sound source to be localized is the same for both ears in the median plane. There is almost no difference in the localization of elevated sources between people who can hear in both or in only one ear (Hebrank and Wright, 1974; Oldfield and Parker, 1986), which confirms a monaural evaluation of the cues. The filter effect is mainly caused by the combination of direct and reflected sound in front of the eardrum. In addition, spectral colorations are caused by the resonances of the different pinnae caves. However, these are largely independent of the angle of incidence. Figure 2.3 shows the idea of direct and reflected sound. The reflection (here at the Antihelix) depends on the angle of incidence. The spectral changes that are caused, differ for each angle of incidence due to different reflections on the fine structure of the pinnae. The addition of direct and reflected similar sounds results in a comb filter-like pattern in the spectral dimension. In Figure 2.3 the reflection is at the Antihelix, but the location of

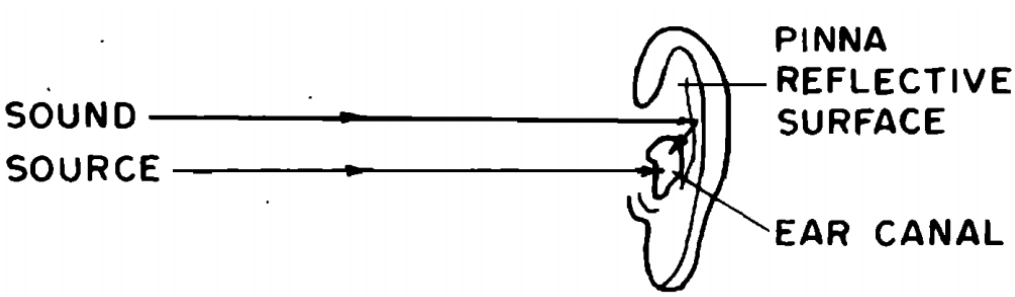


Figure 2.3: Principle of the reflection of an incoming sound wave caused by the pinnae (Figure 1 from Wright et al. (1974)).

reflection on the fine structure of the pinnae defers, depending on the angle of incidence. The notches caused by the addition of direct and reflected sounds are thus defined as pinnae notches. These occur at frequencies between 6 kHz and 13 kHz (Xie, 2013).
In addition to the spectral differences due to the fine structure of the pinnae per angle of incidence, there are also spectral influences due to the body parts torso and shoulders. They start occurring at frequencies around 4 kHz and higher (Xie, 2013). However, these are significantly weaker. Nevertheless, they help in the localization of strongly elevated sources by reflections on the torso and in low elevated sources by shading (Algazi and Duda, 2002; Benichoux et al., 2016).
Sound sources in the median plane can be localized less accurately than in the horizontal plane. Accuracy is higher in the anterior region, as in the posterior region in the median plane (Gardner and Gardner, 1973).
The localization blur varies depending on the elevation angle (see Figure 2.4) (Blauert,

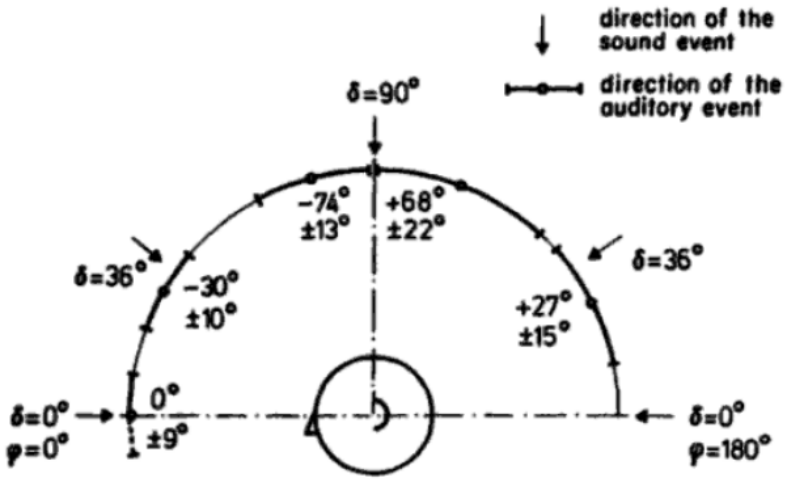


Figure 2.4: Localization blur in the median plane for continues, familiar speech (from Blauert (1997) Figure 2.5 based on data from Damaske and Wagener (1969)).

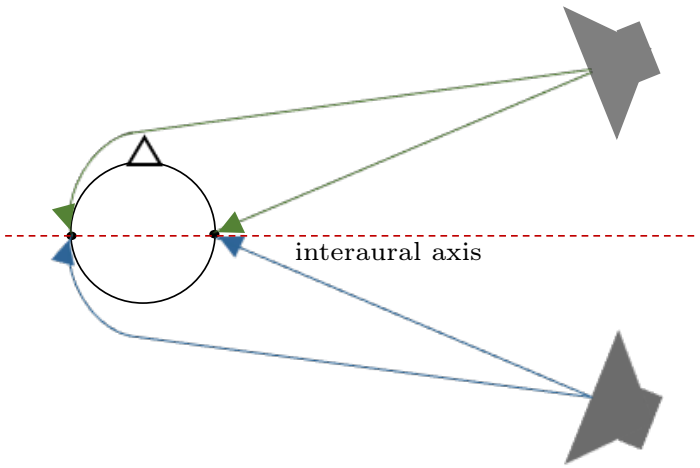


Figure 2.5: Scheme of the principle of the cone of confusion. Identical paths of a propagating sound wave, created by two different sound source positions.

1997). The lowest localization blur can be achieved with a sound source placed in the front of the listener ($\varphi = 0°, \vartheta = 0°$). There the localization blur is ±9°. It is highest at 90° elevation. The accuracy (mean of the perceived value) can be in the anterior or posterior hemisphere. The blur is ±13° (mean anterior: 74°) respectively ±22° (mean posterior: 68°) for a sound source at 90°(Damaske and Wagener, 1969). This results in a total response range of 73° for the upper pole (measured here using a speaker known to the persons). For noisy signals this value presumably decreases (Wettschureck, 1970).
With the pinnae blocked, the localization ability in the median plane deteriorates drastically (Gardner and Gardner, 1973).

2.2.3 Cone Of Confusion

The so-called cone of confusion complicates the localization of sound sources. They are cones of identical binaural cues produced by identical sound propagation paths (Xie, 2013). Figure 2.5 gives an example for sound sources on the cone of confusion. The two paths of the propagating sound waves of the sound sources are exactly the same length. The sound source is mirrored on the interaural axis and thus offers the same cues in the rear hemisphere as the sound source in the front hemisphere. The same applies to shading by the head and the resulting interaural level differences.
A perceived reversal of the sound source on the cone of confusion is called front/back confusion and is measured by the front/back error. In addition, there is the same case for a top and bottom confusion. Usually both types of confusions are combined and are called quadrant errors.

2.2.4 Dynamic Cues

All previously mentioned cues are mostly described in static state of the head. Whereby in this case static means that no head movement takes place. As soon as a movement of the head or source is added, further cues result, which are called dynamic or motion cues. Cues caused by a head movement are created by evaluating the sound source properties at both ears (Carlile, 1996, 2014).
Talking about motion cues, it is usually meant a movement of the head in the horizontal plane. This means a movement of the head during the localization process from left to right or vice versa, rotating around the center of the head. This results in a change of the interaural time and level difference during this movement (Xie, 2013). Left/Right movements in the range of 16° - 32° and wider enhance the localization accuracy and almost eliminate the front/back confusion (McAnally and Martin, 2014). The more the sound source is elevated, the smaller the changes in the ITD due to head movements become, therefore the more the head has to be moved to achieve noticeable changes (Shaw, 1974).
It has long been known that the movement of the head during the localization process leads to a distinction between sources coming from the front and the back. This means that the front/back error/quadrant error described in paragraph **Cone Of Confusion** can be minimized (Klensch, 1948; Koenig, 1950; Jongkees and Van der Veer, 1958).
However, the natural movement of the head during sound source localization (with a stimulus of 5 seconds) is not limited to the horizontal plane (Thurlow et al., 1967). The natural movement in the horizontal plane is 29.2° for High-Pass (HP) signals and 42° for Low-Pass (LP) signals. In the median plane, the mean value of natural motion is 15.2° (HP) and 13.1° (LP), for a diagonal lateral movement (pivot), the values are 11.6° (HP) and 10.2° (LP), respectively.
Jiang et al. (2018) found that using head movement, but without spectral components in the high frequency range, the error measures front/back error, up/down error, and accuracy in median plane were as good as in the static case using full spectral cues. However, there was no improvement in the combination of the two cues.

2.3 Head-Related Transfer Function

Head-related Transfer Function (HRTF) describe changes in sound from a source to the eardrum (Blauert, 1997; Xie, 2013). The corresponding impulse response in the time domain is called Head-related Impulse Response (HRIR). They contain all the previously mentioned main cues for sound source localization, whereby motion cues are only partially

apparent.
HRTF describe all individual scattering, reflection, absorption and diffraction properties caused by the body parts head (including pinnae), shoulder and torso. Their description depends on the angle of incidence of the sound.
HRTF can be determined by measurement or simulation. Still, the most common way is to measure them.
When measuring individual HRTFs, small microphones are placed at the ear canal entrances. An array of loudspeakers is placed in a circular arrangement around the subject in an anechoic environment. The sitting test person is then rotated by means of a turntable and each measurement is performed with a broadband signal. Both the loudspeakers and the microphones are equalized so that they have no influence on the transfer function.
By definition, the head-related transfer function is determined by forming the sound pressure at the entrance to the auditory canal in relation to the sound pressure at the center of the head, but without the presence of the head (Xie, 2013). For one ear (in this case the right ear) it applies:

$$H_r(f, r, \varphi, \vartheta) = \frac{p_r(f, r, \varphi, \vartheta)}{p_0(r, f)} \tag{1}$$

This results in two transfer functions, $H_r(f, r, \varphi, \vartheta)$ (right ear) and $H_l(f, r, \varphi, \vartheta)$ (left ear). These are both frequency-dependent (f in [Hz]), distance-dependent (r in [m]), and dependent on the angle of incidence of Azimuth (φ in [deg]) and Elevation (ϑ in [deg]). Together they form the binaural head-related transfer function. The function p_r in [Pa] is the determined sound pressure at the right ear canal and p_0 in [Pa] the free field sound pressure at the center of the head with the head absent. It is defined as

$$p_0(r, f) = j\frac{k\rho_0 c Q_0}{4\pi r} e^{(-jkr)} \tag{2}$$

where ρ_0 in $[\frac{kg}{m^3}]$ is the density of air, c in $[\frac{m}{s}]$ the speed of sound, Q_0 the strength of the point source and the wave number $k = \frac{2\pi f}{c}$. This is called Green's function and represents the sound pressure of a monopole sound source in the free field and frequency domain (Morse and Ingard, 1971). If far-field conditions are given ($r \gg a$ (head radius)), resulting cues are approximately the same for both ears (Xie, 2013) .
Equations 1 - 2 apply generally, therefore also to simulated HRTFs. These types of transfer functions can be stored in the so-called standardized Spatially Oriented Format for Acoustics (SOFA). This was specially designed to store spatial data that is angle-dependent (AES Standards Comittee, 2015; Majdak et al., 2013). In addition to the actual HRTF data, specifications such as source position, listener position, and distance can be defined.

The SOFA toolbox[1] is used to load, edit or save SOFA formats in Matlab.
All transfer functions used in this work were transferred to this format and loaded, transformed or saved with the functions provided in the SOFA toolbox.

As already mentioned, head-related transfer function can not only be measured, but also simulated. Simulation methods have been used and refined for more than two decades (Bronkhorst, 1995; Ziegelwanger et al., 2015b). Until then, the possibility of creating HRTF was exclusively in measurements. However, this can be very time-consuming and costly. Measurements require a lot of expensive equipment and a lot of space.
In order to be able to simulate transfer functions, however, high-resolution scans of the heads to be simulated must be made. The most common way to achieve these scans is with a high resolution 3D scanner. However, there are also approaches in which a 3D image of the head is generated from many different individual images (Dellepiane et al., 2008). Since the fine structure of the pinnae is especially important, there are approaches to improve it by means of neural networks (Kaneko et al., 2016). The resulting 3D models are displayed in high-resolution triangle meshes and used as input for numerical simulations. A BEM is the most common approach for the numerical calculation of HRTF (Ziegelwanger et al., 2015b). In the field of acoustics, this is based on the Kirchoff-Helmholtz integral (Xie, 2013). This can be used due to the simplification that any body structure is assumed to be reverberant and there is no propagation through the head, since it has been shown that the skin is acoustically rigid. Only the hair has a slight absorbing character, but is neglected in this case (Katz, 2001a,b). The maximum calculable frequency is based on the node distances of the triangular mesh. It is defined as $\lambda_{max} = 6 * internode$. In general, HRTF simulated by BEM can be mapped up to a frequency of 22 kHz (Katz, 2001a). To achieve this maximum frequency without artifacts, a maximum edge length of around 2.5 mm is required.
Ziegelwanger et al. (2013, 2015a) showed in comparisons between measured and calculated HRTF that calculated HRTF are at least as accurate as measured ones. Spectral and temporal structures of measured and simulated transfer functions are largely identical. A major difference that leads to slight changes in simulated HRTF is that simulations are usually performed without the body parts shoulder and torso. This results in an abrupt break-off at the bottom of the grid surface. In the performance of the actual sound source localization, however, no significant differences between the two methods could be found.

[1] `https://github.com/sofacoustics/API_Cpp` (last downloaded: Oktober 22, 2018)

2.4 Spherical Head Model

Spherical head transfer functions (SHTF) are often used in 3D audio applications and recording methods (Algazi et al., 2004; Cuevas-Rodríguez et al., 2019). The analytical representation of sound propagation on a sphere was already described in the 19th century by Lord Rayleigh (Rayleigh, 1894).
Since analytical solutions to calculate the sound propagation on a sphere exist, the numerical calculation of the SHTF by means of BEM was omitted.
To calculate the SHTF the model of Duda and Martens (1998) was used. However, this model is limited exclusively to ear shifts within the horizontal plane. Since an ear off-set in the median plane is essential for a more precise definition of the SHTF and thus for greater comparability with original head-related transfer functions, Duda and Martens (1998) approach has been extended.
First, the transfer function is generally defined as

$$H = \frac{p_s}{p_{ff}} \tag{3}$$

where p_s in [Pa] represents the sound field on the sphere surface and p_{ff} in [Pa] the sound field at the location of the original center of the sphere, but with the sphere absent.
With the normalized frequency $\mu = ka = f\frac{2\pi a}{c}$, where $\frac{2\pi a}{c}$ is the time a wave needs to once travel around a sphere and the normalized distance to the source $\varrho = \frac{r}{a}$, respectively $\varrho_0 = \frac{r_0}{a}$ as the normalized distance to the source starting from the coordinate system, the transfer function results in

$$H(\varrho_0, \varrho, \mu, \Theta) = -\frac{\varrho_0}{\mu} e^{-j\varrho_0\mu} \Psi(\varrho, \mu, \Theta) \tag{4}$$

with the infinit series expansion

$$\Psi(\varrho, \mu, \Theta) = \sum_{m=0}^{\infty} (2m+1) P_m(cos(\Theta)) \frac{h_m^{(2)}(\varrho, \mu)}{h_m^{(2)\prime}(\mu)} \tag{5}$$

and the angles (calculated with the definition of a great circle on a sphere)

$$\Theta = arccos(sin(\vartheta)sin(\vartheta_{l,r}) + cos(\vartheta)cos(\vartheta_{l,r})cos(\varphi - \varphi_{l,r})) \tag{6}$$

as the angle of incidence between the source and the location of the ear. P_m is the Legendre polynomial of degree m and the m-th order spherical Hankel function of second kind (outgoing wave - from centre of head to sound source) is described by $h_m^{(2)}$ in equation 5.

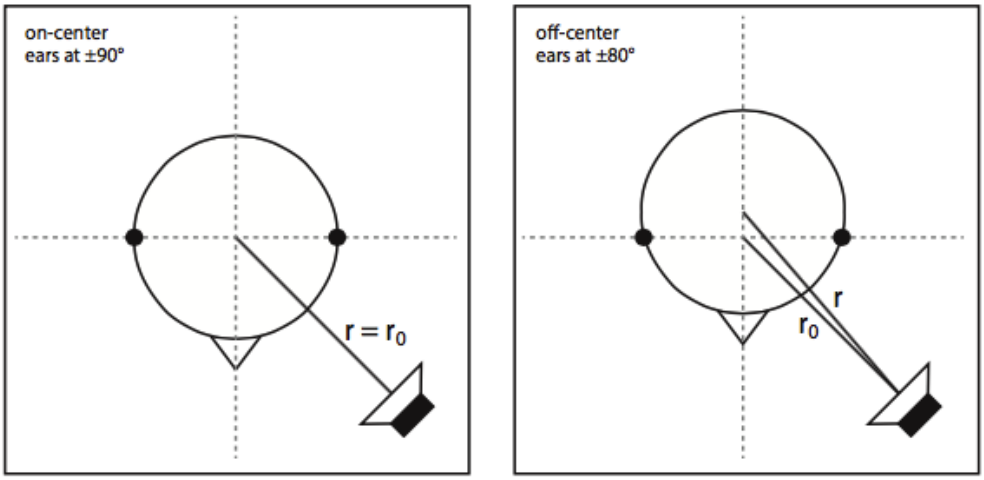


Figure 2.6: Spherical head model with on-set ears (left) and off-set ear positions (right) (Figure used from AKtools[2] AKspherical Head documentation). Left side shows ear positons at $\pm 90°$ and $r = r_0$. Right side shows ear positions at $\pm 80°$ and $r \neq r_0$.

For all equations it applies, that the angular frequency is $w = 2\pi f$, f the frequency in [Hz], c the speed of sound in [m/s], a the radius of the sphere in [m], r the distance from the sound source to the centre of the sphere in [m] and r_0 the distance from the sound source to the centre of the coordinate system in [m]. In the case of an off-set of the ears, $r \neq r_0$. Equations 3 to 6 correspond to the specifications of Duda and Martens (1998), with the exception of the extension to radius r_0 and the additional possibility to shift the ears on the horizontal and median plane with equation 6.
Θ is the angle between two points on the great circle. In equation 6, φ and ϑ are the angles of the sampling grid and $\varphi_{l,r}$ and $\vartheta_{l,r}$ are the ear positions for the left and the right ear. Thus, Θ indicates the angle between each point on the sampling grid and the ear positions by means of a great circle. For equation 3 or rather 4 to continue to apply, the sphere or coordinate system must be shifted so that the interaural center is again the center of the coordinate system. This is achieved by shifting the spherical head by

$$x = a * sin(\varphi_{l,r} - \frac{\pi}{2}) \tag{7}$$

in x-direction and by

$$z = a * sin(\vartheta_{l,r}) \tag{8}$$

in z-direction (Møller, 1992).
Figure 2.6 shows the principle of the spherical head mode. The left side of Figure 2.6 refers to the principle of Duda and Martens (1998), the right side to the new approach

with off-set ears. Here the the ear position shifts from ±90° to ±80°, so that the original radius r_0 and the radius r resulting from the shifting serve as a description.
The Duda and Martens (1998) algorithm including the extension was implemented in Matlab and integrated into AKtools[2] (`AKsphericalHead.m` including documentation) (Brinkmann and Weinzierl, 2017a). For the implementation the iterative solution of equation 4 provided by Duda and Martens (1998) was used for a faster calculation (see equation 9).

$$H(\varrho_0, \varrho, \mu, \Theta) = \frac{\varrho_0}{j\mu} e^{-j(\mu\varrho - \mu\varrho_0 - \mu)} \sum_{m=0}^{\infty} (2m+1) P_m(cos(\Theta)) \frac{Q_m(\frac{1}{j\varrho\mu})}{\frac{m+1}{j\mu} Q_m(\frac{1}{j\mu}) - Q_{m-1}(\frac{1}{j\mu})} \tag{9}$$

Equation 9 varies slightly compared to the equation from Duda and Martens (1998), due to the extension of the algorithm.
The here explained, by Fabian Brinkmann extended, analytical approach of a spherical head model with off-set ears can be found throughout the AKtools Matlab Toolbox[2] (Brinkmann and Weinzierl, 2017a) again under AKsphericalHead.

2.5 Anthropometric Data

All analytically or numerically calculated transfer functions are based on anthropometric data of the human head. They describe the different dimensions of individual structures of the body, in this case the head. For the determination of the dimensioning of different geometries the anthropometric data of the head width, the head height, as well as the head depth are usually necessary.
In the case of SHTF, there are two ways to calculate optimal head radii for spherical heads. One possibility is the least squares minimization method, which uses standard anthropometric data to start the optimization process. It also requires transfer functions of a human head to calculate the corresponding radius with best fit to the ITDs of the HRTF. This approach is explained in more detail in Section 3.1.3.
A more basic approach to calculate the radius of a spherical head was introduced by Algazi et al. (2001b). It is a general linear model for estimating the radii using the weighted sum of anthropometric head dimensions. His approach was taken up and optimized by Ziegelwanger (2012). The radius $\tilde{r}$ of the corresponding sphere can be calculated

$$\tilde{r} = \frac{0.53w + 0.22h + 0.24d}{2} \tag{10}$$

[2] `https://www.ak.tu-berlin.de/menue/publications/open_research_tools/aktools/` (last downloaded: January 10, 2019)

on the basis of the weighted head dimensions height (h), width (w) and depth (d). The width of the head is weighted with more than 50% to the overall model. The height, as well as the depth lie with in each case somewhat over 20%.
Since there are often no individual head dimensions in applications with spherical head models, common values are taken. If anthropometric data from DIN33402-2 (2005) are used (see Table A.1 in **Appendix**) and the radii are calculated using equation 10, the average radius is 8.87 cm for male heads and 8.51 cm for female heads. This results in an average head radius of 8.69 cm (see Table A.2 in **Appendix** for more details). In comparison, the radius of a KEMAR artificial head calculated according to this method (w=15.2 cm, d=19.1 cm, h =22.4cm) is 8.78 cm. Both values are close to the a standard value of 8.75 cm for calculating spherical head models in many studies (Algazi et al., 2001a; Fiedler et al., 2017).

However, if elliptical geometries are used, there is no optimization procedure for dimensioning this geometry. The calculation of the Ellipsoid Head Transfer Function (EHTF), whether analytical or numerical, is based exclusively on anthropometric data. Usually the head width and the head height of an average human or an artificial head are used (Lins et al., 2016).

The position of the ear on the human head is not recorded in any standard. Burkhard and Sachs (1975) found that these are on average 87° in the horizontal plane and -7° in the median plane. An off-set of -7° in the median plane means that the ear deviates from the original head axis by -7°. This information relates to the entrance to the ear canal.

3 Physical Evaluation

In this chapter, the different geometries used are presented and their properties with regard to sound source localization are considered. In particular, different localization cues, with a main focus on median plane sound source localization, are discussed. Static binaural and spectral cues are considered. Furthermore, motion cues are analyzed and a model for the consideration of possible localization errors by using HRTF by non-human heads for sound source localization is introduced. Advantages of off-set ears with non-human geometries are experienced, as well as advantages and disadvantages of the introduced geometries.

3.1 Method

This sections explains the numerical simulation of transfer functions for corresponding geometries. The dimensioning of the different geometries and their subsequent generation of meshes is also discussed.

3.1.1 Numerical Simulation of Transfer Functions

Transfer functions of two of the geometries more precisely defined in Section 3.1.3 were calculated numerically using the boundary element method. For this purpose, the open source software Mesh2HRTF[3] was used. The following steps also apply to every simulated HRTF. The approach is based on the transformation of the Helmholtz equation into the Boundary Integral Equation (BIE) and then offers different calculation methods for the BEM (e.g. Multi-Level Fast Multipole Method (ML-FMM), 3-dimensional Burton-Miller collocation BEM).
For calculating transfer functions by BEM, a discrete sampling grid of the boundary object was required. Therefore, closed triangular polygon meshes were used. They had to be imported or created in the open source software Blender[4] (Flavell, 2010). Ziegelwanger et al. (2015a) evaluated that a high mesh resolution of 1 mm to 2 mm around the pinnae leads to a natural localization ability. Thus, a polar root mean square error deterioration on average by 2.2°/mm and the quadrant error by 4.6 %/mm can be expected. The steps for creating a HRTF are explained in the following. The file `export_mesh2hrtf.py` had to be imported into Blender. The skin surface as well as the left and right ear canal entrance had to be marked on the mesh and the coordinate system defined on the interaural axis. After defining the mesh some further specifications for the BEM were

[3] `http://mesh2hrtf.sourceforge.net/` (last downloaded: August 07, 2018)
[4] `https://www.blender.org` (Version 2.79b.; last downloaded: August 07, 2018)

Table 1: Settings used when exporting geometries from Blender using Mesh2HRTF.

Setting	Input
Ear	Left
Pictures	✗
Source (y)	5
Reciprocal	✓
c ($\frac{m}{s}$)	346.18
rho ($\frac{kg}{m^3}$)	1.1839
Unit	m
Evaluation Grid 1	Lebedev
Near Field calculation	✗
Frequency Step (Hz)	100
Frequency max (Hz)	22000
Frequency dependency	✗
Method	BEM
CPU First	1
CPU Last	1
Number used cores	3

selected for the calculation. All settings selected for the calculation of the HRTFs are listed in Table 1. In this work, all HRTFs were calculated with the same specifications. In this case a 3-dimensional Burton-Miller collocation BEM (reverberant Neumann boundary conditions) with a 10 m source distance was used (Ziegelwanger et al., 2015b). The faster calculation with the ML-FMM was omitted due to distortions in the high frequency range, starting at 17 kHz (see Pelzer (2018)). HRTFs were calculated reciprocally in a frequency range from 100 Hz to 22 kHz with a step size of 100 Hz. A Lebedev-Laikov-Grid of 35th order was chosen, due to nearly equally spaced nodes on a sphere (Bernschütz, 2016). As it was taken for granted that all geometries are axial symmetrical, HRTFs were simulated for the left ear only.

When exporting, one gets the folder NumCalc, which contains the input files for the BEM. In addition, the discrete points of the evaluation grid and a Matlab file `Output2HRTF.m` are generated. The input files were loaded into the application startNumCalc for BEM calculation. At the end of the calculation, a SOFA file was created using the generated

Matlab script `Output2HRTF.m`. The HRTF was calculated using the sound pressures on the sampling grid of the BEM and equation 1, already described in Chapter 2.3.

3.1.2 Post-Processing

For the further use and analysis of the calculated SOFA files, a post-processing was necessary. An existing Matlab file for data post-processing of a previous thesis (Pelzer, 2018) was extended to some specifications. The following steps were performed:
In order to be able to use AKtools Matlab toolbox[2], the SOFA file data was first transformed into the AKtools format. Each transfer function was manually extended by the missing 0 Hz coefficient such that the HRTF($f = 0$) = 0 dB. For unexplained reasons, a phase rotation occurred during the previous calculations. This was bypassed by a complex conjugation of the transfer function. Furthermore, a two-sided spectrum was generated and transformed into the time domain. Due to regulations for the calculation of HRTF a shift of 60 samples was made to avoid a non-causal transfer function (see Pelzer (2018)). Subsequently, the length of the impulse response was limited to 256 samples. The transfer function of the left ear was rotated and mirrored to obtain a binaural transfer function, due to symmetries of the geometries. Finally, the impulse response of both ears were filtered low-pass, with a cut-off frequency of 3000 Hz, in order to obtain TOAs and ITDs.

3.1.3 Geometries

This Section presents all geometries used. The main focus is on dimensioning. In addition, the types of representation are explained and related to human hearing.

KEMAR's head

In this study the head of a KEMAR mannequin was used to assess human localization abilities. Therefore, a mesh of KEMAR's head was created as described in Brinkmann et al. (2017c) and a gliding node spacing between 1 mm (ipsilateral ear) and 10 mm (contralateral side) was rendered via the open source tool Gmsh[5] (Geuzaine and Remacle, 2009; Ziegelwanger et al., 2016). The non-uniform discretization was chosen due to reasons of computing capacity. The left side of Figure 3.1 shows the mesh of KEMAR's head. KEMAR's head was simulated without torso and shoulders, and with an incomplete neck (see Figure 3.1). This serves to improve comparability with geometries

[5] `http://gmsh.info/` (last downloaded: August 01, 2018)

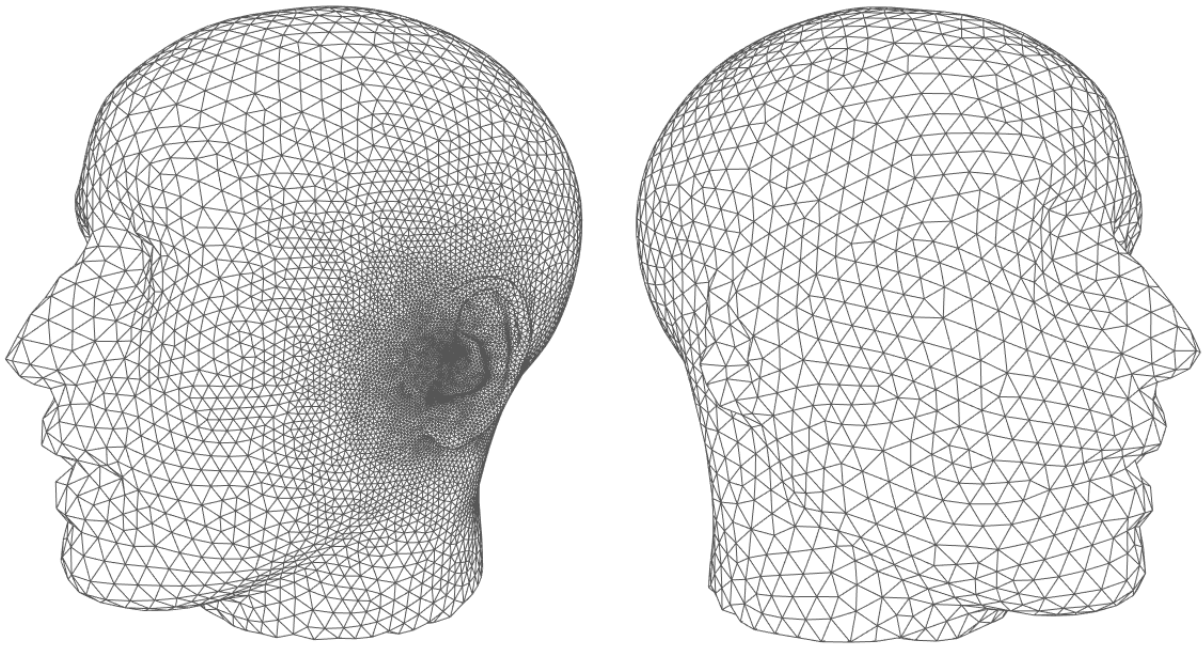


Figure 3.1: Mesh of the KEMAR mannequin with 1 mm grid size on the ipsilateral side, merging to 10 mm on the contralateral side.

later introduced in this section. Spherical head models are often used without torso models. In comparison with measured HRTF/HRIR, there will be slightly less reflections, scattering and shading of the sound wave. The construction of the Motion Tracked Binaural (MTB) also consists exclusively of a sphere with several microphones. KEMAR's HRTF was calculated using the in Section 3.1.1 introduced Boundary Element Method.

As already mentioned, KEMAR's HRTF served as a representative for the properties of human hearing with respect to sound source localization. Figure 3.2 shows the transfer functions of the left ear for KEMAR's head and thus typical transfer functions for the human head. The left two graphs show the functions in the horizontal plane, the right two in the median plane. The two upper graphs are each in the time domain and therefore show the HRIRs, the two lower graphs in the frequency domain and therefore show the HRTFs. Looking at HRIRs in the horizontal plane (see upper left graph in Figure 3.2), it can be seen that the earliest incident of sound, as well as the one with the highest amplitude, is at 90°, that is when the sound source points directly to the ear. The first incidence occurs around 1 ms. In contrast, the latest angle of incidence is 270°, when the sound source points directly to the contralateral ear and the circulating paths around the head from the ipsilateral ear are longest. The correspondent time of arrival is approximately 1.8 ms. This results in an additional weak, so-called x-shape, since the circulating paths around

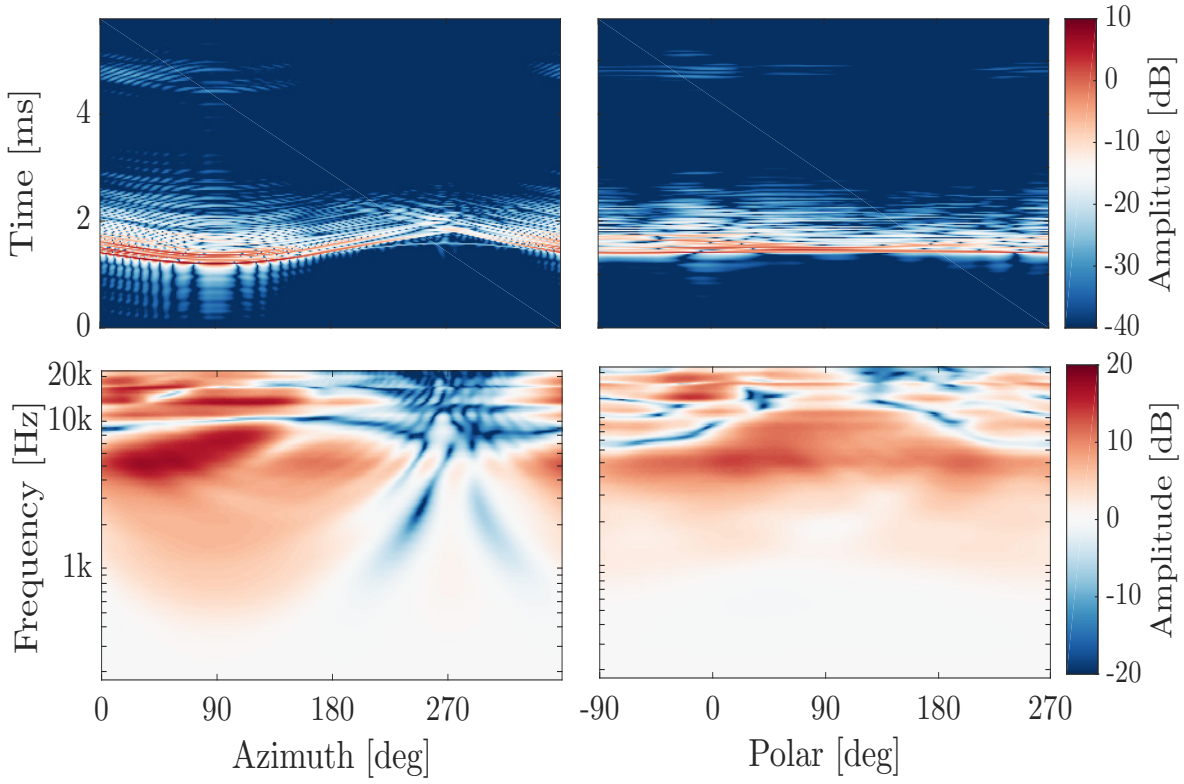


Figure 3.2: HRIRs and HRTFs in horizontal plane (left side) and median plane (right side) for KEMAR's head.

the front of the head and around the back of the head are about the same length.
Considering the HRTF depicted in the lower column of Figure 3.2, an influence of the geometry at large wavelength, thus at low frequencies, can not be observed. In contrast, the influence of the fine structure of the head on high frequencies can be identified over the entire angular range. The greatest influence can be recognized at a frequency approximately 4 kHz with a wavelength of about 8.5 cm. This can be attributed to reflections, shading and scattering on different facial features, such as the pinnae and protrusions like the forehead, chin and occipital bone.
Looking at the HRTFs in the horizontal plane (see lower left graph in Figure 3.2), the partial cancellation at around 270° can be seen. Additionally, the influence of the pinnae can be observed, especially at high frequencies.
In the median plane, the time of sound incidence is approximately the same across all angles, since the distance from the sound source to both ears is approximately the same (see upper right graph in Figure 3.2). Only slight variations of the time of arrival occur, resulting from the off-set of the ear.

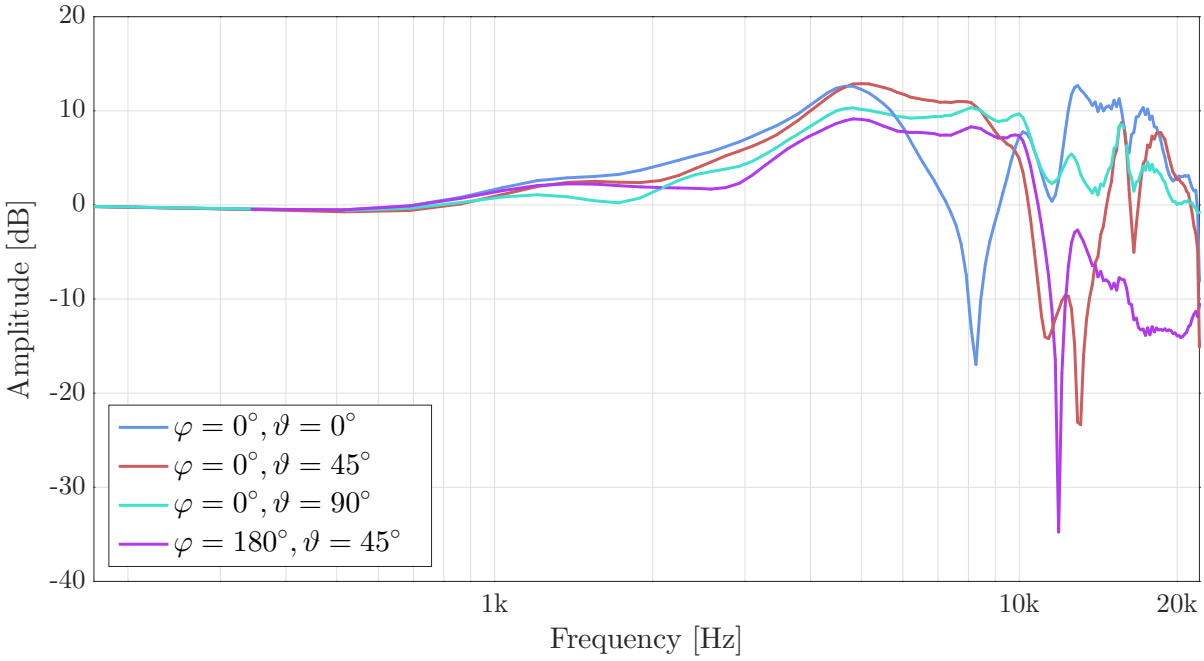


Figure 3.3: Pinnae Notches of KEMAR's head for four different elevation angles ($\vartheta = 0°, 45°, 90°$ with $\varphi = 0$ and $\vartheta = 45°$ with $\varphi = 180°$)

There are further differences in HRTFs (see lower right graph in Figure 3.2), due to the angle of incidence on the pinnae and the resulting different reflections and scattering on the fine structure. Additionally, further shadowing can be observed when the sound source emits from 90° onwards, in the back of the head. This is due to the protruding auricle, which interferes the sound path to the ear canal. It only applies at high frequencies, since there is no diffraction around the auricle. Spectral changes in the high frequency range are between 6 kHz and 13 kHz and are called pinnae notches.
Figure 3.3 shows frequency responses of the HRTF for different sound source elevations. These notches shift with the angle of incidence (Xie, 2013). At the horizontal plane (0° azimuth and 0° elevation), the pinnae notch is at its lowest frequency, in this specific case at about 8.2 kHz. The more elevated the sound source, the more the notch shifts to higher frequencies. At 90°, the centre frequency is at its highest and shifts to lower frequencies in the back of the head. Since these notches are very pronounced, it is considered the main indicator for sound source localization in the median plane (Shaw and Teranishi, 1968; Gardner, 1997). Note that for geometries with absent pinnaes notch patterns are not present.

Spherical Head

For better comparability with the existing literature, a numerical approach for the spherical head was not applied. Instead, the calculation of the SHTF was done analytically (see Section 2.4). Therefore, the radius of the sphere as well as the ear positions had to be determined beforehand. Ziegelwanger's time of arrival (TOA) model (Ziegelwanger and Majdak, 2014), which is included in the Auditory Modeling Toobox[6] (Søndergaard and Majdak, 2013), was used for this purpose. This algorithm was created to estimate TOA of an existing HRTF data set, since calculations from measurement data often lead to artifacts. The most accurate TOA estimates are made using mimial-phase correlation. The parameters radius and ear position were determined by solving a minimization problem with a non-linear least square solver. Ziegelwanger and Majdak (2014) presented two approaches, the on-axis approach and the off-axis approach. In the on-axis approach, the TOA estimates, especially in the maximum for human HRTF, led to an underestimation the TOA. The head position of the subject (during measurement) would have to be exactly on the axis. For this reason, the author introduced the off-axis approach. In addition to estimations of TOA, HRTF are simplified to sphere geometries and shifted on the axis so that the resulting transfer function of the sphere (especially ITDs) fit the data of the HRTF. Subsequently, the estimated TOA are corrected by these geometric data. This approach has been used to obtain the exact geometry needed to calculate SHTF.
The calculated parameters resulted in an ear position of $\varphi_e = 89.62°$ azimuth and $\vartheta_e = -4.87°$ elevation, and a sphere radius of $a = 8.35\,\text{cm}$ for KEMAR's head.
Finally, the corresponding SHTFs were calculated analytically using the in Section 2.4 explained and expanded Spherical Head Model.

Ellipsoidal Head

The third and final geometry used in this work is an ellipsoid. An ellipsoid represents a compromise between a realistic head shape and the very simplified spherical head model. Bomhardt and Fels (2014) found that the height and width of the head are the most important anthropometric data for an accurate reproduction of the ITD for an ellipsoidal head. Therefore, the head depth corresponds to the width. Accordingly, the two parameters radius and head width were used for dimensioning the ellipsoidal geometry. The exact designation of the geometry with two dimensioning parameters corresponds to a rotational ellipsoid, the so-called spheroid.

In order to be able to compare the geometry with the HRTF characteristics of a human

[6] `http://amtoolbox.sourceforge.net/download.php` (last downloaded: January 18, 2019)

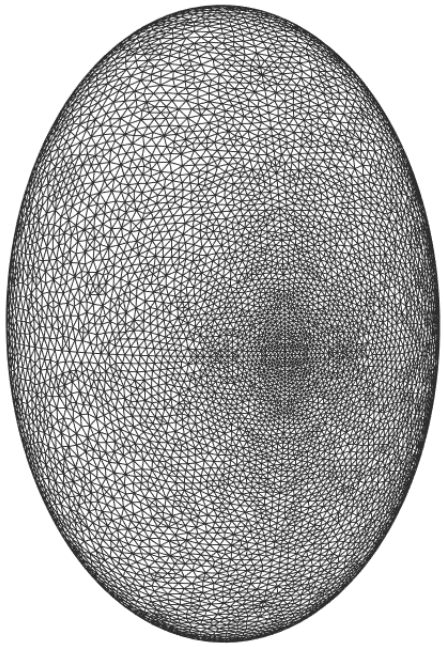


Figure 3.4: Generated mesh of the geometry ellipsoid with a width of 15.2 cm, a height of 22.4 cm and ear positions of $\varphi_e = 81.99°$, $\vartheta_e = -4.28°$.

being, the dimensioning is carried out exemplary on the artificial head KEMAR. The height of the artificial head is 22.4 cm and the width 15.2 cm (Burkhard and Sachs, 1975). The ear positions of the KEMAR are 12.5 cm below the highest point of the head. Based on this data, a mesh was generated in Blender[4] (Flavell, 2010).
Since the KEMAR and the resulting ellipsoid did not have the same depth and the ear positions were only indicated in their elevations, both meshes of the geometries were superimposed as exactly as possible (minimum overhang on all axes). The ear positions of the KEMAR's head were transferred to the ellipsoid. This procedure revealed additional information on the ear position, especially in the horizontal plane. The marked ear positions resulted in 81.99° in the horizontal plane and -4.28° elevation.
Figure 3.4 shows the constructed mesh of the ellipsoid. For reasons of symmetry and computing time only the transfer function for one ear (left) was calculated. The node size was selected from 1 mm (ipsilateral) to 10 mm (contralateral) using Gmsh (Geuzaine and Remacle, 2009). In the following ellipsoidal head transfer function are revert to as EHTF and Ellipsoid Head Impulse Response (EHIR). The resulting EHTFs were calculated numerically, using the same method and settings explained in Section 3.1.1.

3.2 Evaluation and Results

The main focus of this section is on the evaluation of the sound source localization in the median plane, depending on the geometry. KEMAR's HRTF serves as the reference representing cues for sound source localization of the human head. Static cues (binaural cues and spectral cues) as well as motion cues are investigated in detail. A model for calculating theoretical mismatches between the human ear and non-human HRTF is introduced and explained. The influence of off-set ears in non-human geometries is considered, as well as advantages and disadvantages of ellipsoidal heads.

3.2.1 Offset-Ears

A person's ear positions are usually shifted out of the centre of the head. Burkhard and Sachs (1975) identified an average ear position of 87° in the horizontal plane and -7° in the median plane. This means that the ears are slightly shifted forward and downward on average. In the horizontal plane there sometimes may also be shifts to the rear instead of to the front. These shifts lead to two paths of different lengths around the head for each plane. In the case of the calculated ear off-set for the SHTF matching KEMAR's HRTF, the path in horizontal plane around the front of the head

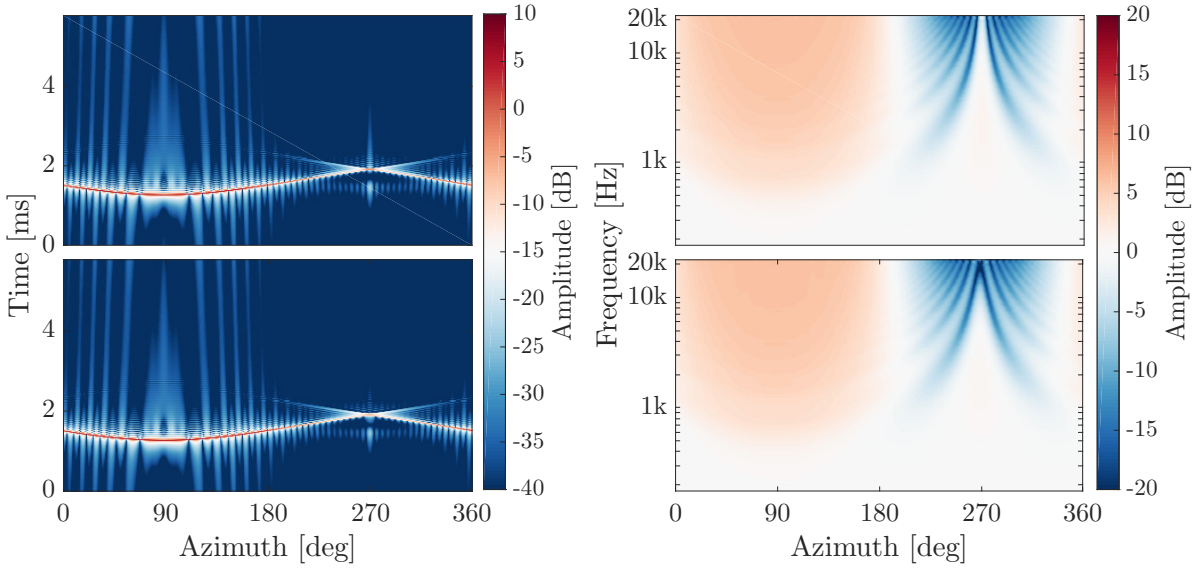


Figure 3.5: SHIR (left side) and SHTF (right side) of two spherical head models in horizontal plane. Upper two graphs: sphere with on-set ears $\varphi_e = 90°$, $\vartheta_e = 0°$; Lower two graphs: sphere with off-set ears ($\varphi_e = 89.62°$, $\vartheta_e = -4.87°$)

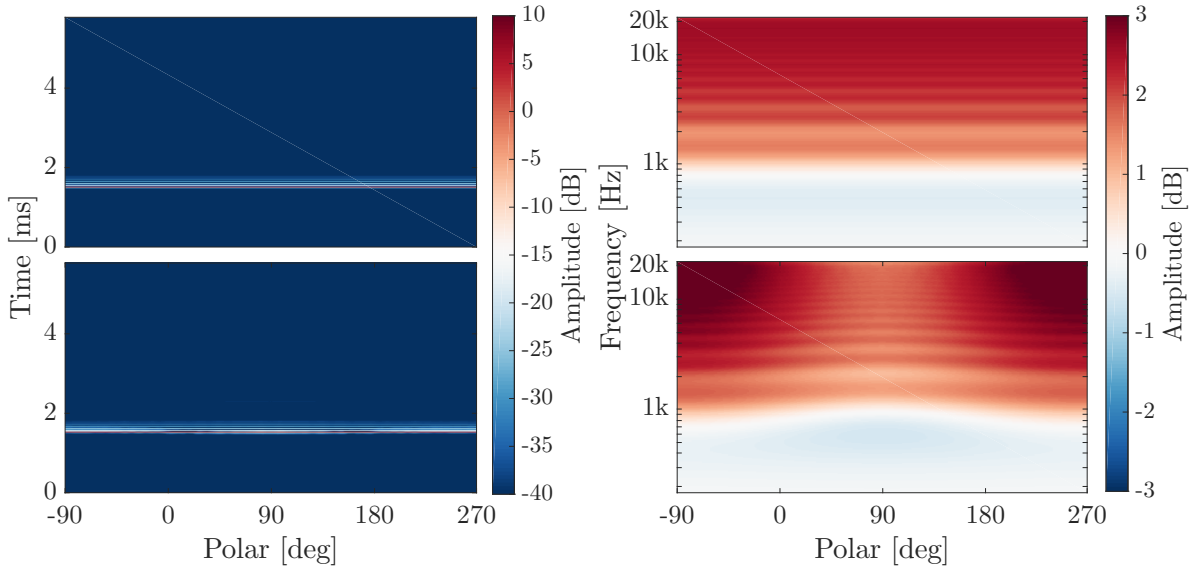


Figure 3.6: SHIR (left side) and SHTF (right side) of two spherical head models in median plane. Upper two graphs: sphere with on-set ears $\varphi_e = 90°$, $\vartheta_e = 0°$; Lower two graphs: sphere with off-set ears ($\varphi_e = 89.62°$, $\vartheta_e = -4.87°$)

is shorter from one ear to the other than around the back. For the median plane, the path around the bottom of the spherical head is shorter than the path around the top. This leads to slight changes in the ITD, ILD, as well as slight changes in the spectral amplitude range, due to creating an asymmetry within an all-axis-symmetric geometry. Figure 3.5 and Figure 3.6 show the effects of the ear off-set on the localization cues in the horizontal plane and on the median plane. In each Figure the upper two graphs show the SHIR (left side) and the SHTF (right side) of a spherical head model without ear off-set ($\varphi_e = 90°$, $\vartheta_e = 0°$). In contrast, the lower two graphs show SHIR and SHTF with off-set ears ($\varphi_e = 89.62°$, $\vartheta_e = -4.87°$).

In the horizontal plane in Figure 3.5, it can be seen that the impulse response is shortened at an azimuth angle of about 270° (lower left side) for the SHIR with off-set ears. Since the off-set in the horizontal plane deviates from the axis only by 0.38°, the influence is therefore very weak. The greater the ear off-set, the greater the influence on the different cues. Looking at the SHTF, the same loss of amplitude at around 270° can be observed at high frequencies (lower right side).

Figure 3.6 shows the different influences in median plane (on-set ears upper row, off-set ears lower row). The difference caused by the ears off-set downwards (elevation) is clearly visible. Looking at the SHIR for the case with centred ears, no change in the length or amplitude of the impulse response over the entire angular range can be observed. However,

compared to the SHIR with off-set ears, a slight change in the amplitude can be identified. This change can be divided into three angular sections, each of which corresponds to about the same pattern of amplitude over time. In the range from -90° to about 30°, as well as from about 150° to 270°, the same pattern results and thus deviates from the middle range (0° to 180°). This becomes more clear when looking at the corresponding SHTF. For the case of on-set ears, the SHTF is constant over the entire median plane. However, observing the SHTF with off-set ear, one can recognize changing characteristics in the transfer function at certain angular sections. For frequencies above 1 kHz, level differences of up to 3 dB result between the individual sections. These slight level differences could make a difference in sound source localization in the median plane and contribute to a more natural behavior of the SHTF.
In the following, the spherical head model is used and considered exclusively with off-set ears.

3.2.2 Static Cues

Binaural Cues

Binaural cues result in differences between both ears. They are represented by Interaural Time Difference (ITD) and Interaural Level Difference (ILD). Figure 3.7 shows the ILD, Figure 3.8 the ITD in the horizontal plane for different sound source elevations. In both figures the left graph shows the cues for the KEMAR, the middle one for the spherical head and the right one for the ellipsoid. The broadband ITDs were estimated from low-

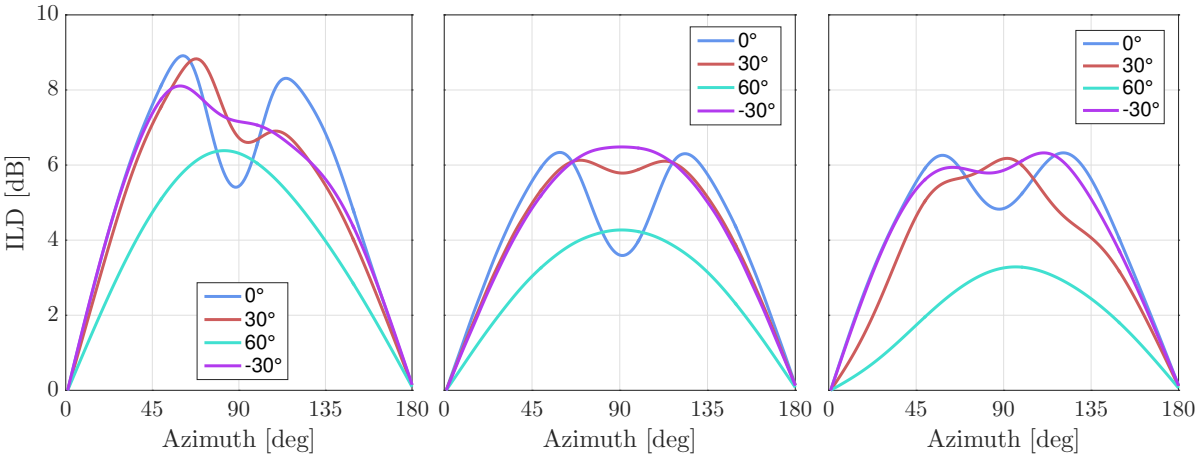


Figure 3.7: Interaural level differences in horizontal plane for four different elevations and different geometries. Left side: ILD of KEMAR's head; Middle one: ILD of spherical head; Right side: ILD of ellipsoidal head.

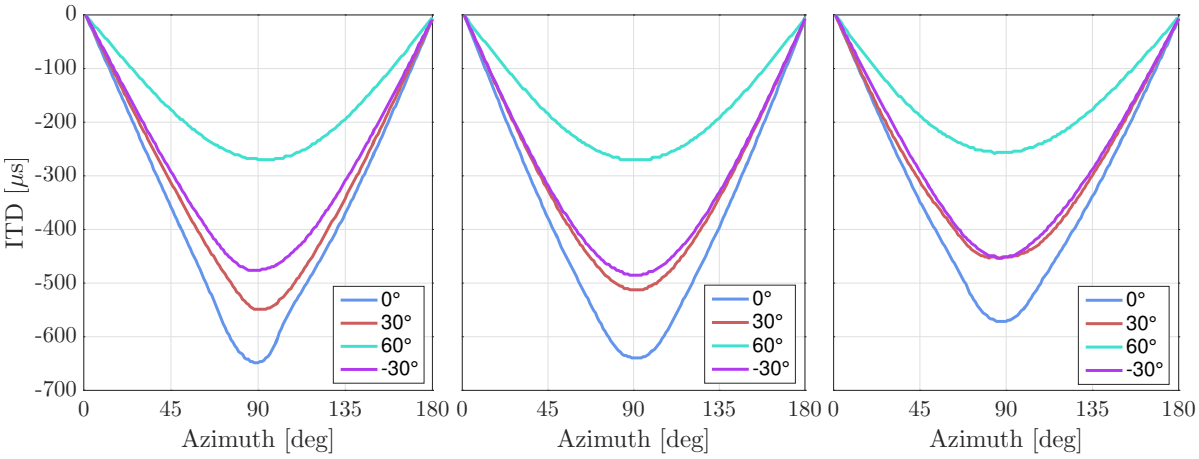


Figure 3.8: Interaural time differences in horizontal plane for four different elevations and different geometries. Left side: ITD of KEMAR's head; Middle one: ITD of spherical head; Right side: ITD of ellipsoidal head.

passed ($f_c = 3$ kHz) and ten times up-sampled impulse responses using a threshold based onset detection (threshold -20 dB) (Andreopoulou and Katz, 2017). The broadband ILDs were estimated from the logarithmic RMS level differences between left and right ears.

Comparing the ILDs of KEMAR, the spherical head and the ellipsoid, slight differences can be seen throughout horizontal plane (see Figure 3.7) and different elevations. The ILD of the sphere and the ellipsoid underestimates that of the KEMAR, as the influence of the pinnae on the ILDs in the case of the sphere and the ellipsoid is not present. They underestimate the ILD cues of KEMAR's head by about 2.5 dB for all elevations. Additionally, the ILD of the ellipsoid is less symmetrical than the ILD of the spherical head geometry. This is reflected in an increased number of peaks and turning points with an asymmetric distribution over different azimuth angles.

Figure 3.8 shows the ITD in the horizontal plane for the different geometries and elevations. The ITD of the ellipsoid is about 100 μs lower than that of the KEMAR or spherical head. Nevertheless, the temporal characteristics (rise, peaks) are very similar for all three geometries. This can be explained by matching the sphere's geometry and the ellipsoid's geometry to KEMAR's TOAs.

Spectral Cues

Spectral cues are most important for localization in median plane. Nevertheless, they also give information about left/right localization in horizontal plane. Figure 3.9 shows HRTFs of KEMAR (left graph), SHTF (middle graph) and EHTF (right graph) in the horizontal plane. Compared to KEMAR's transfer function, the spectral fine structure of

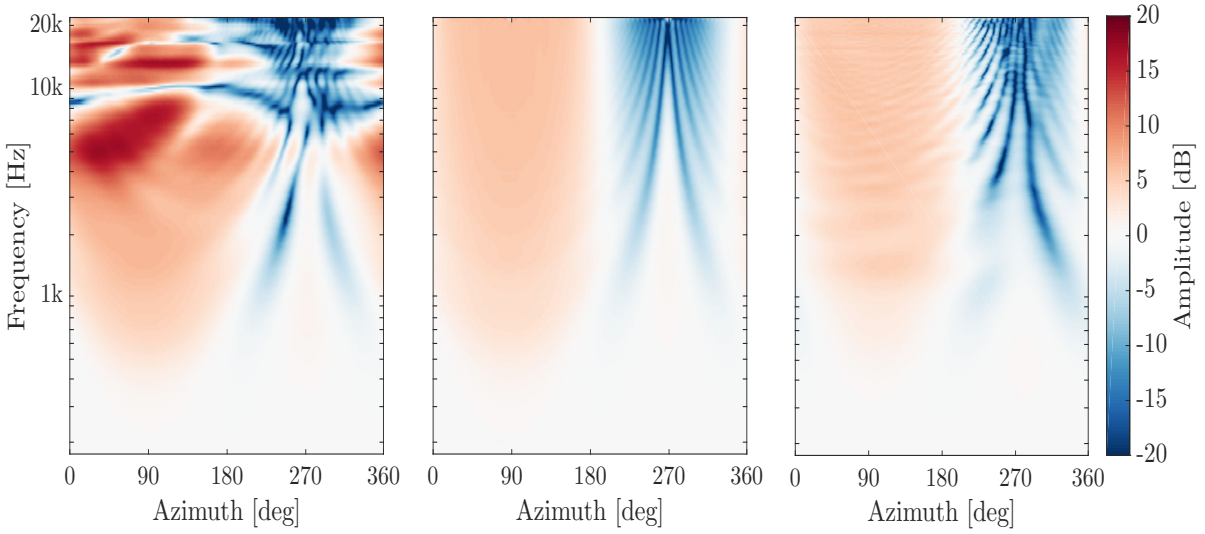


Figure 3.9: Transfer functions in horizontal plane for different geometries. Left side: HRTF of KEMAR's head; Middle one: SHTF of spherical head; Right side: EHTF of ellipsoidal head;

the spherical head and the ellipsoidal head transfer function is absent at high frequencies due to the absence of the pinnae. Additionally, there is a slight loss in the maximal amplitude for the ipsilateral side of the spherical head and the ellipsoidal head. The structure of the EHTF is more irregular than that of the SHTF and is therefore more similar to the fine structure of the HRTF. In general, all three geometries show similar characteristics in horizontal plane. They all have higher amplitudes at the ipsilateral side, and lower amplitudes at the contralateral side. Moreover, the inverse x-shape amplitude pattern at contralateral angles, due to shading of the (spherical/ellipsoidal) head at high frequencies, are existing in all three cases. Thus, left/right localization should also be possible with spectral cues from a spherical head and a ellipsoidal head. As already mentioned, the influence of the pinnae on the localization ability in median plane is said to be crucial for human localization. The left graph in Figure 3.10 shows HRTFs of KEMAR's head, the middle graph shows SHTFs and the right graph EHTFs in median plane. Note, that the amplitude scale differs from the one used in Section 3.2.1. In order to be able to draw better comparisons, an equal amplitude range for all three geometries in the illustration of the median plane was selected in this section. This makes the structure mentioned in Section 3.2.1 less visible in the case of the spherical head. Due to the missing pinnae, spectral features in median plane of the sphere's SHTF with off-set ears and the ellipsoid's EHTF are a lot weaker than the well known spectral cues of KEMAR's head. However, weak spectral cues of the SHTF and EHTF are the result of shifting the ears out of the principal axis, as it creates an asymmetry within an all-axis-symmetric geometry. Those weak spectral features, are in a range of 3 dB at high

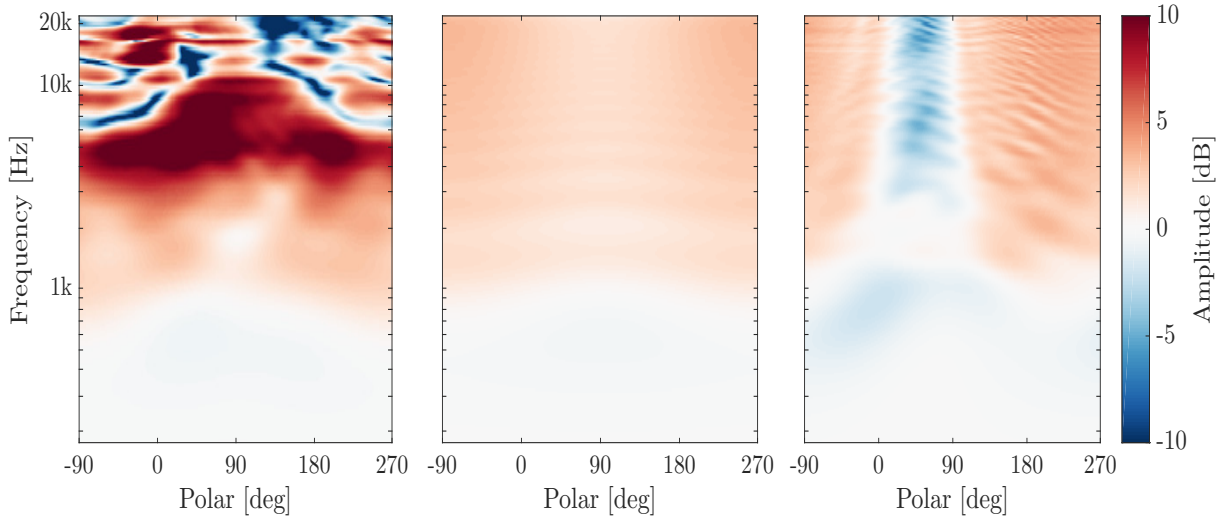


Figure 3.10: Transfer functions in median plane for different geometries. Left side: HRTF of KEMAR's head; Middle one: SHTF of spherical head; Right side: EHTF of ellipsoidal head;

frequencies for SHTF and in a range of about 10 dB for EHTF and might aid the elevation localization. However, these cues are almost identical for sources above and below the horizontal plane.

3.2.3 Motion Cues

Motion cues are localization cues induced through head or source movements that result in changes of the ITD, ILD and spectral cues. Motion cues were calculated in the horizontal plane independently for ILD and ITD, and analyzed depending on the median plane. The change of the values per degree head movement of the two static cues ILD and ITD were calculated over an horizontal angular range from 0° to 10°. This angular interval was chosen because an almost constant behavior of the interaural level difference (see Figure 3.7) and interaural time difference (see Figure 3.8 can be observed in this range. For the binaural cue ILD, the derivative with regard to $|\varphi|$ results in the dynamic ILD cue ($\overline{\Delta \text{ILD}}$ in [dB/deg]). For the static ITD, the derivative with regard to $|\varphi|$ results in the dynamic ITD cue ($\overline{\Delta \text{ITD}}$ in [μs/deg]).
Dynamic ILD cues and dynamic ITD cues in the median plane are shown in Figure 3.12 and in Figure 3.11. In general, the dynamic cues are largest for sources on the horizontal plane ($\vartheta = 0°$), and decrease with increasing distance to the horizontal plane. Comparing KEMAR's head and the spherical head in terms of the dynamic ITD cues, it can be observed that the magnitude of the dynamic ITD is larger for the latter geometry. One possible cause is that the sphere's diameter overestimates KEMAR's head width. In contrast, the

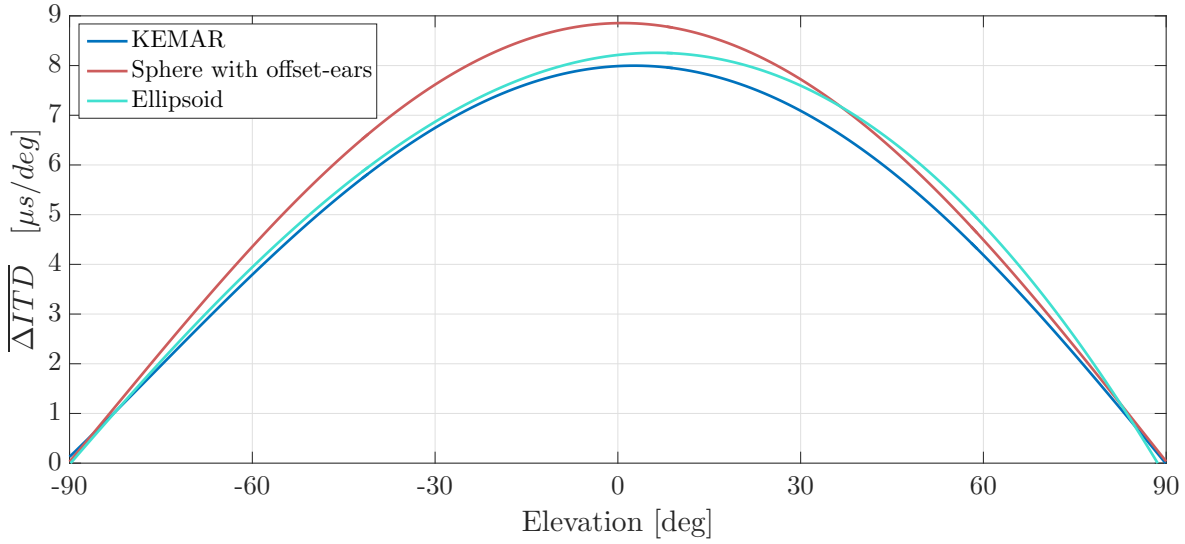


Figure 3.11: Change in ITD per degree for sources in the median plane and head/source movement in the range of $|\varphi| \leq 10°$.

dynamic ILD cues are smaller for the spherical head, and do not follow the asymmetry observed in the KEMAR data, which is caused by the missing pinnae.
In the case of the ellipsoidal head, the dimensioning was carried out according to the anthropometric data of the KEMAR in two dimensions. Accordingly, the dynamic ITD cues of these two geometries correspond very close (see Figure 3.8). Slight differences may be due to the fact that the head depth is not equal. Considering Figure 3.12, the dynamic ILD cues, in the case of the ellipsoidal head, also show an asymmetry similar to that of

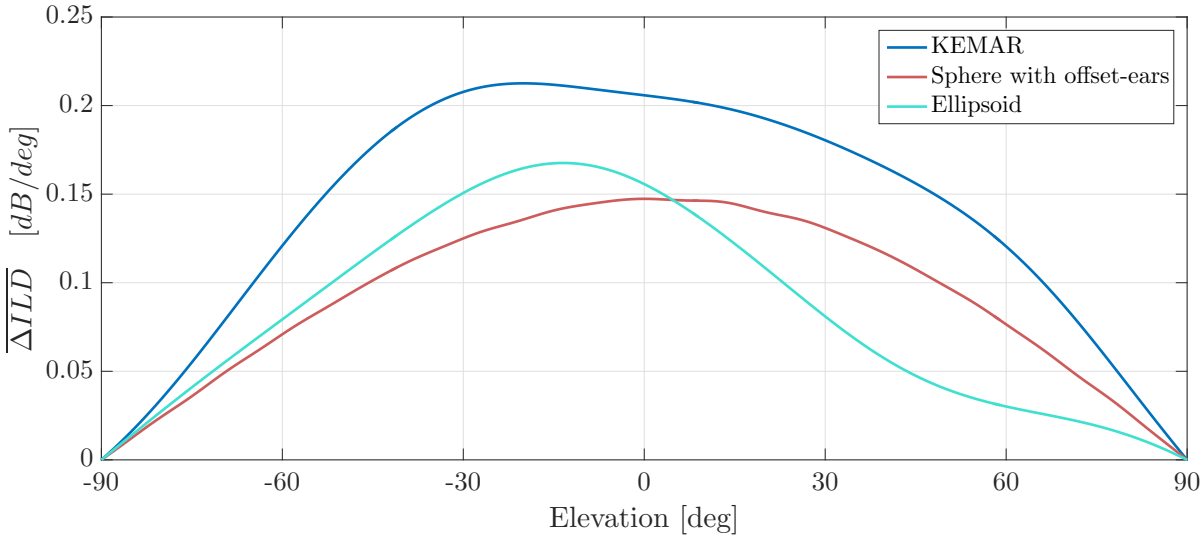


Figure 3.12: Change in ILD per degree for sources in the median plane and head/source movement in the range of $|\varphi| \leq 10°$.

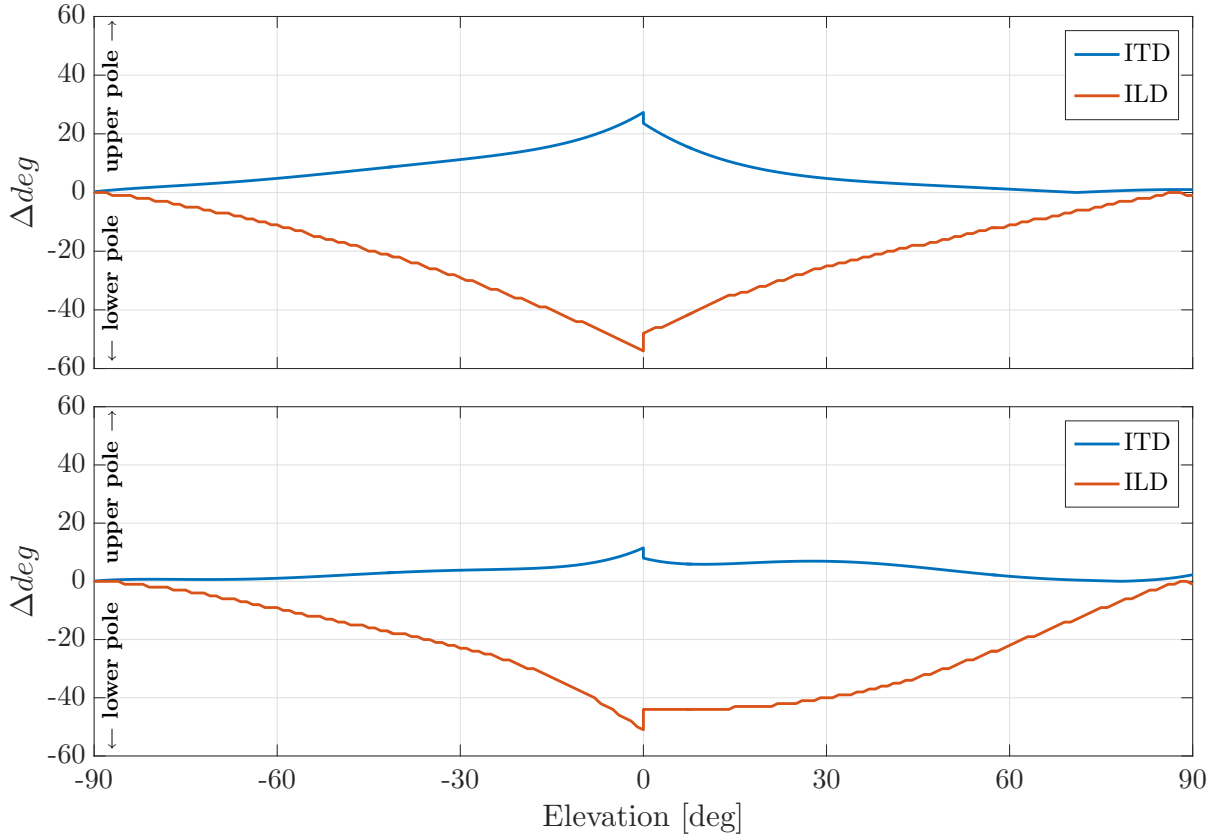


Figure 3.13: Estimated localization error caused by dynamic ITD and ILD cues of the spherical head model and the ellipsoidal head. Upper graph: Mismatch between spherical head and KEMAR's head; Lower graph: Mismatch between ellipsoidal head and KEMAR's head;

the KEMAR, at least in the lower region ($\vartheta < 0°$). Like the ILD of the spherical head, the ILD of the ellipsoidal head is underestimated to KEMAR's ILD.

It is likely to assume that the human auditory system learned these personal dynamic localization cues, and that a deviations from the learned features results in a localization error. For example, KEMAR expects a ΔITD of 6 μs/° at -39° elevation, but the spherical head delivers 6 μs/° at -47°, which results in a localization error of 8°. These expected localization errors were evaluated over the entire median plane. Upper graph of Figure 3.13 shows the estimated localization error between the spherical head and KEMAR's head, the lower graph shows it between the ellipsoidal head and KEMAR's head. Motion cues related to ITD result in a localization shift of up to 30° for the spherical head and up to 10° for the ellipsoidal head. ILD related motion cues cause errors of up to 50° for both geometries. Both cues cause a shift towards the upper pole or the lower pole, which means away from the horizontal plane.

Base on this mismatch model, it can be assumed that as soon as the decision to localize a sound source is made on the basis of dynamic ITD cues, the perceived sound source shifts to the upper pole, and as soon as the decision is made on the basis of dynamic ILD cues, the perceived sound source shifts to the lower pole.

3.2.4 Ellipsoid vs. Sphere

As already mentioned, the human head shape rather represents an ellipsoid, like a sphere. However, both geometries are relevant models for simplifying head geometry. In this section, both geometries are compared in terms of localization influencing features.
An advantage of using ellipsoidal geometries instead of spherical head models can be found when looking at the EHIR and EHTF in Figure 3.14. The left side of the Figure 3.14 shows EHIR and EHTF in the horizontal plane, the two graphs on the right side in the median plane. EHTFs in the horizontal plane show more pronounced structures and a higher dynamic range compared to SHTF (see Figure 3.5). Especially the fine structure from 180° onwards (sound source closer to the contralateral ear) shows varying spectral

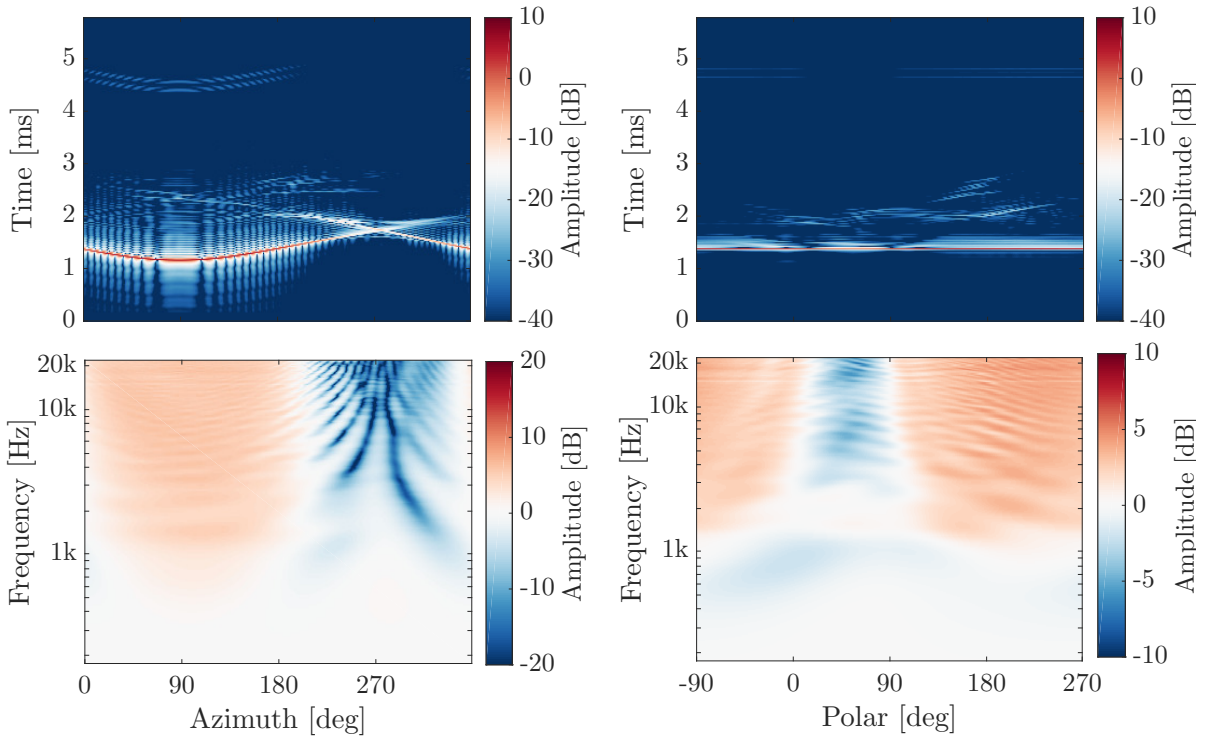


Figure 3.14: EHIR and EHTF in horizontal plane (left side) and median plane (right side) for the ellipsoidal head.

characteristics. Thus, it can be assumed that a better source localization is also enabled by the more diverse characteristics of the transfer function of the ellipsoid.
A larger difference regarding the maxima and minima of the EHTF compared to the SHTF can be also detected in the median plane. The larger amplitude range plays an important role here. If there are maximum differences of 3 dB between individual sections of the SHTF, this difference is around 8 dB in the case of the ellipsoid. Based on the higher dynamic of the EHTF, it can be assumed that a front back confusion is less likely with an ellipsoidal head model.
A more detailed structure can also be confirmed in the time domain. In the case of the spherical head, there are hardly any sound events in the median plane after the arrival of the direct sound. In the case of the ellipsoid, one can see further activities after the arrival of the direct wave front.
The theoretical findings of this section revealed that the use of an ellipsoidal model might result in a better sound source localization in both planes. An ellipsoidal design, such as an MTB microphone where the microphones are not located directly on the geometry axis and the microphone capsules are also adapted to achieve off-set ears in both planes, would be one way to improve the localization ability of sound sources. For applications using an analytical spherical head model, it is possible to use an analytical ellipsoidal head model based on the same approaches as SHTF (Duda et al., 1999; Bomhardt and Fels, 2014; Lins et al., 2016; Bomhardt et al., 2016b). The only disadvantage of this method is that the dimensioning is not as easy as with the spherical head model.

3.3 Summary and Conclusion

First, it could be shown that off-set ears are an advantage when using spherical head models. Shifted ears lead to different path lengths and thus different cues depending on the angle of incidence of a sound wave. Second, a spherical and ellipsoidal head with offset-ears has been compared to the KEMAR dummy head regarding static and dynamic localization cues. While static cues of the spherical head model provide sufficient information about the left/right position, the included localization cues in the median plane are rather weak. Fortunately, movement induced dynamic cues are a promising candidate for providing the missing information. In the case of left/right localization, the static ILD cues are underestimated by the spherical head model, but it might be assumed that the more realistic ITD cues dominate the perceived source location in this case (Wightman and Kistler, 1992). Localization cues caused by an ellipsoid head are generally are stronger and more detailed pronounced than those of the spherical head. This is true for the median and the horizontal plane.

Finally, a model for estimating the localization bias, induced by using a simplified geometry instead of a realistic head model, was given in relation to dynamic cues. Based on this model it could be shown that dynamic cues in the median plane are conflicting, as they cause a shift of the perceived source position in opposite direction. Further investigations are needed to find out whether the conflicting cues balance each other, or if the auditory system prefers one cue over the other. Nevertheless, if the dynamic ITD cues are used to identify the location of the sound source, the mismatch is much smaller using cues of an ellipsoidal head as with the ones of a spherical head.

Based on the previous findings, an ellipsoid model can be an alternative to the conventional spherical head model. With regard to practical applicability, it is advantageous that an analytical solution also exists for the calculation of the transfer functions of an ellipsoid (Lins et al., 2016). A disadvantage, however, could be the dimensioning of the ellipsoid, since due to the parameters head width and head height a standardized ellipsoid might deviate more from the actual head of the listener than a standardized sphere. In addition, a method for estimating the dimensions of an ellipsoidal model based on an optimal ITD fit does not yet exist.

The findings of this chapter were all based on numerical simulations or analytical evaluations. To further validate the drawn assumptions, a listening test was designed and carried out (see Chapter 4).

4 Perceptual Evaluation

The considerations in Chapter 3 were all of theoretical origin. In order to be able to transfer these into practice and to support the findings given, those findings will be further investigated in this chapter with a listening test.
In order to examine the localization ability in the median plane under different conditions, a localization listening test was implemented and carried out. An important aspect here is the localization ability with a spherical head model, as well as the influence of head movement on the localization ability.
The listening test was carried out in Virtual Reality (VR). Additionally, the listening test itself is in the direct field of application using SHTF generated by spherical head models. In addition, conventional test setups are associated with considerable effort, since visual components and sound speaker setups are cost-intensive and time-consuming.
The popularity of virtual reality is increasing. For example, in a survey conducted in Germany in 2019[7], 63% of all respondents stated that they had already used VR glasses, were interested in buying one or already owned one. However, virtual reality and its applications does not only play a role in consumer related topics. The use of virtual reality already extends across the education, medical and industrial sectors. Since there are only very few listening tests in a complete virtual environment to date, these chapter also provides information on the feasibility and implementation of such a test setup, in addition to examining the question of sound source localization at the median plane.
In this chapter, the entire test setup, the various conditions to be tested, the structure of the listening test and its implementation, as well as the results and their evaluation are explained.

4.1 Listening Test Design

Test Conditions

An experiment with three different factors (in statistic cases)/conditions was developed to investigate the different influences on localization.
All test conditions are listed in Table 2. Only the individual head and the spherical head of a participant were considered as geometries under test. An ellipsoidal geometry was neglected, to reduce test conditions and avoid fatigue and concentration effects. Therefore,

[7] https://de.statista.com/statistik/daten/studie/438899/umfrage/umfrage-zum-interesse-an-virtual-reality-brillen-in-deutschland/ (last downloaded: August 09, 2018)

Table 2: Conditions for the sound localization test.

Head Movement	Geometry	TF Type
dynamic	Head	Original
		Magnitude
		Phase
	Sphere	Original
		Magnitude
		Phase
static	Head	Original
		Magnitude
		Phase
	Sphere	Original
		Magnitude
		Phase

the individual head of each participant and a corresponding spherical head with individual diameter were used. The choice of a spherical geometry instead of an ellipsoidal geometry was made due to simple analytical models of a spherical head, which are often used in applications.

To later analyze the influence of head movement on localization, two levels were defined. One variant was the possibility of individual unrestricted head movement (dynamic). The head movements were not limited in order to make the listening test as natural as possible and to be able to transform connections into everyday life. The stimulus duration for the dynamic case was chosen to be 5 seconds. This was considered to be a sufficiently long duration to move the head several times in any direction.

In the second case, head movements are not desired (static). In order to achieve this, head tracking was disabled in this condition, so that head movements of the participants are not considered for playback of the stimulus. In addition, in the static case, the test signal was reduced to 100 ms in order to reduce the overall listening test duration. A stimulus duration of 100 ms is sufficient to propagate all relevant localization characteristics (Vliegen and Van Opstal, 2004). Vliegen and Van Opstal (2004) found out that the localization of sound sources in the elevation plane increases with a sound duration of up to 100 ms and

stagnates with a longer duration.
In order to obtain a separable influence of spectral and temporal cues, three Transfer Function (TF) were used for each subject, geometry and movement situation. The original HRTF or SHTF of each participant was provided, as well as one each containing only the phase of the original TF and a third TF containing only the magnitude of the original TF. For Further information see Section 4.1.
The above conditions result in the 12 individual localization scenarios listed in Table 2 that each subject must pass through during the test.

Transfer Function

During the listening test, the individual HRTFs of the participants were used. Those HRTFs are based on a mesh of the participants heads, which are used for caluculating HRTF with the Mesh2HRTF BEM method. The meshes, as a results of a previous study, were adapted for this purpose (Dinakaran et al., 2016).
In that study, the heads of subjects were scanned in high resolution to extract their anthropometric features. A high-resolution Kinect 3-D scanner was used for this purpose. For further information on the generation of those HRTF see Dinakaran et al. (2016) and Pelzer (2018).
With this method, images of the heads of 93 subject could be taken. These 93 participants served as the basis for the selection of the subjects in the experiment of this work. The given subject IDs are based on the IDs of the subjects of the previous study.
All transfer functions used were diffuse field corrected[8].
HRTFs were calculated using the BEM method used in Chapter 3. These were saved with a angular resolution of 2° in SOFA format. A higher resolution could not be used, due to computing capacities in real-time rendering of the graphics card (see Paragraph **Technical Equipment and Software**) a minimum resolution of 2° was possible.
SHTFs were calculated as described in Section 2.4 and Section 3.1.3. They were calculated individually for each participant. This means that for each subject, Ziegelwanger and Majdak (2014) TOA estimation[6] was used to calculate a suitable spherical head radius and ear positions based on the individual ITDs (see Section 3.1.3).
The transfer function types which contain only the magnitude of the original versions, were created by forming a symmetrical absolute magnitude spectrum of the impulse

[8] SHTF were diffuse field corrected using AKtools `AKsphericalHeadDiffuse.m` function. HRTF were diffuse field corrected within the process of Mesh2HRTF SOFA calculation.
A headphone equalization was not possible, due to increasing the number of HRTF bins by convoluting both signals. Thus, exceeding the standard SOFA length of 256 bins valid for the use of the rendering tool in the Game Engine.

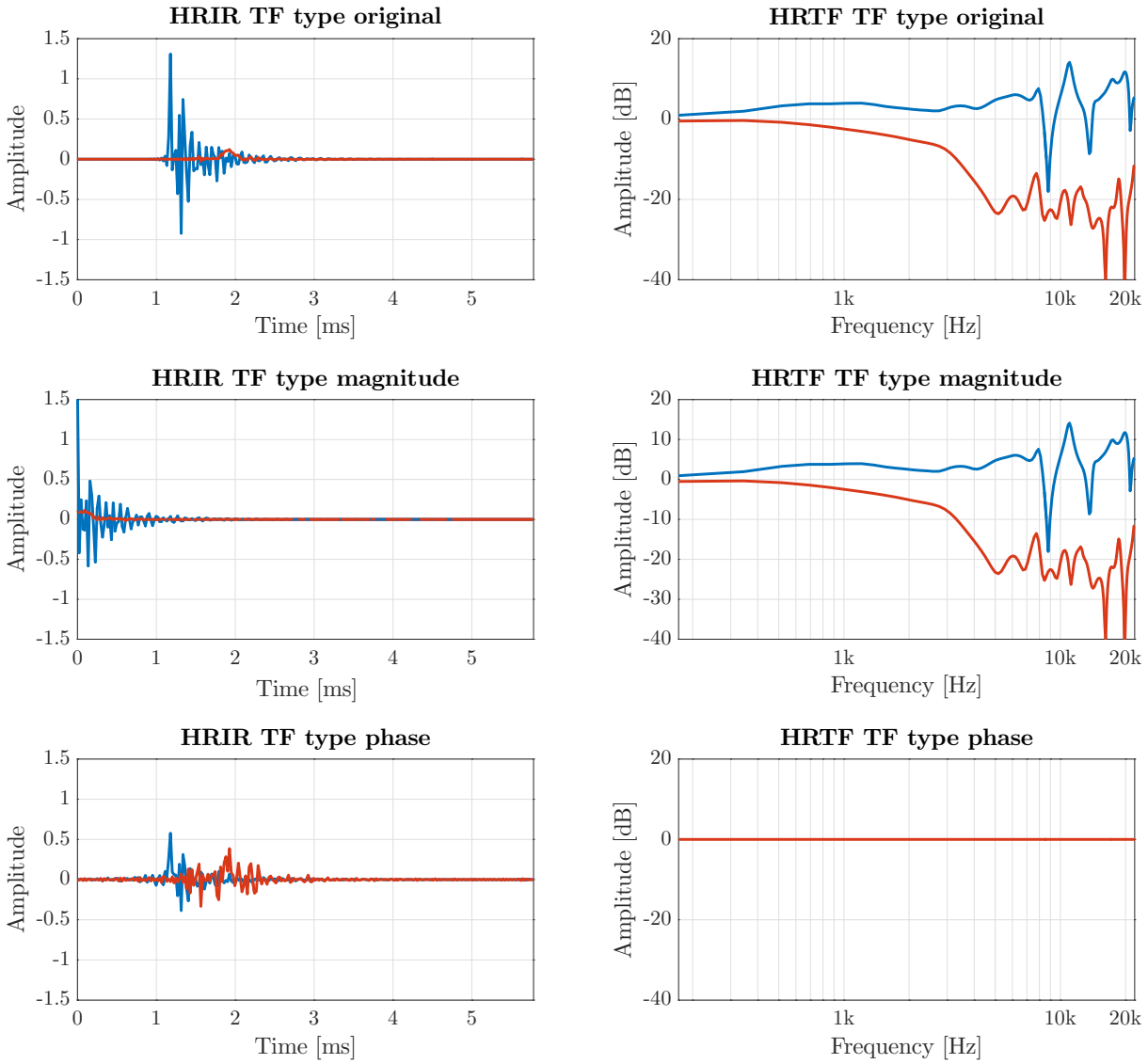


Figure 4.1: HRIR (left side) and HRTF (right side) of three different TF types for participant 31 ($\varphi = 90°$ and $\vartheta = 0°$). Blue indicates the left (ipsilateral) ear and red the right (contralateral) ear. Upper graphs: TF type original; Middle one: TF type magnitude; Low graph: TF type phase;

response, transforming it back into the time domain, and generating a minimum phase using `AKphaseManipulation.m`[2] function (Brinkmann and Weinzierl, 2017a).
In the case of the transfer functions which contain the phase, an absolute spectrum of 0 dB over all frequencies was calculated and the original phase information was retained . Figure 4.1 gives an example of all three TF types for individual HRTF in time and spectral domain for $\varphi = 90°$ and $\vartheta = 0°$. Figure 4.2 shows the same TF types for the spherical head model. The blue lines represent the function of the left ear and

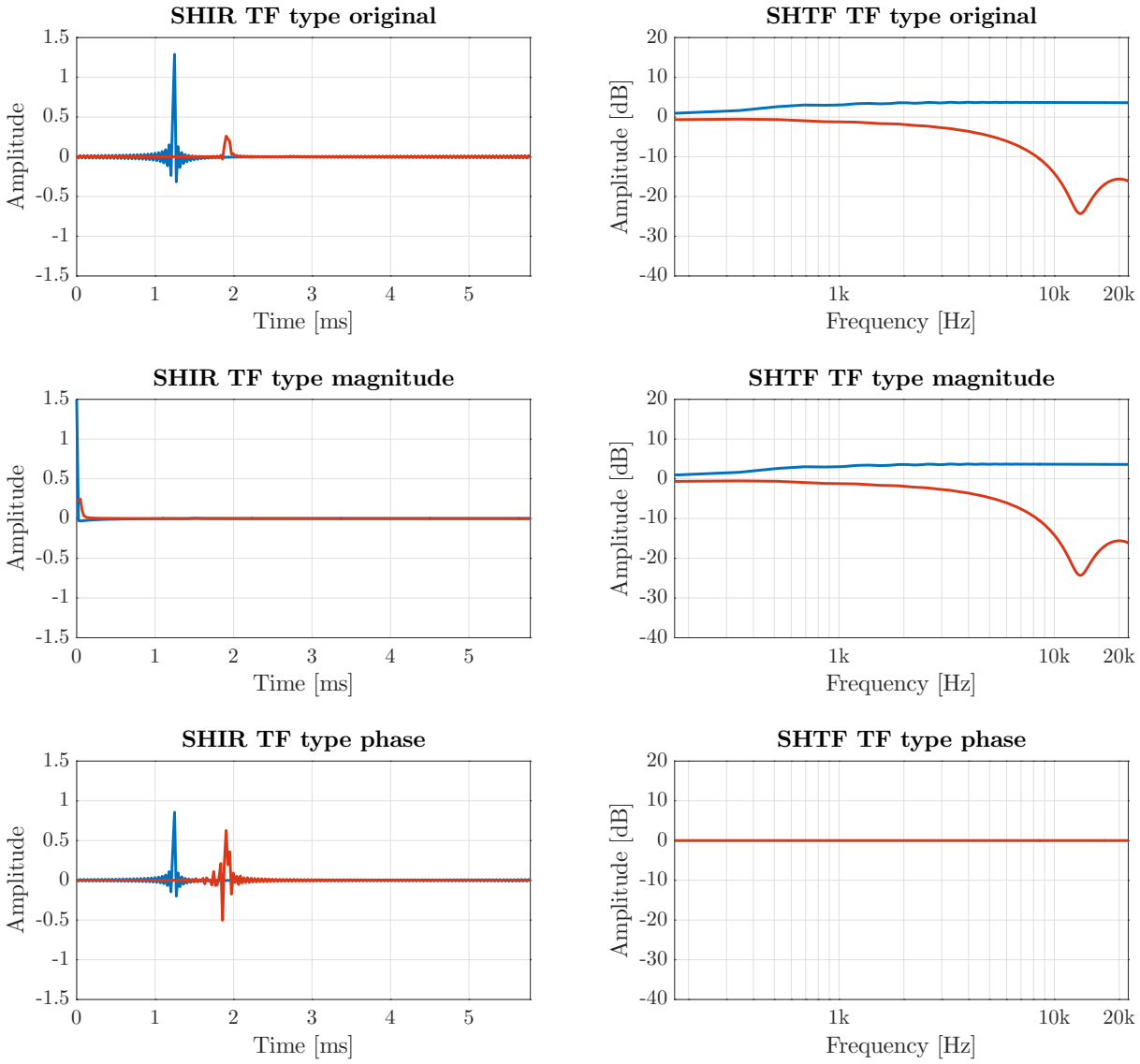


Figure 4.2: SHIR (left side) and SHTF (right side) of three different TF types for participant 31 ($\varphi = 90°$ and $\vartheta = 0°$). Blue indicates the left (ipsilateral) ear and red the right (contralateral) ear. Upper graphs: TF type original; Middle one: TF type magnitude; Low graph: TF type phase;

the red lines the functions of the right ear. The left ear is the ipsilateral side, the right the contralateral one. Left graphics show the functions in the time domain, the rights in the frequency domain. The two upper graphics (HRIR and HRTF or SHIR and SHTF) each show the original function in the time as well as in the spectral range. The middle columns show the graphics for the transfer functions with minimum phase, the lower ones show the functions with linear spectrum and the original phase information. Spectra of the original HRTF/SHTF and the spectra of the condition magnitude are identical.

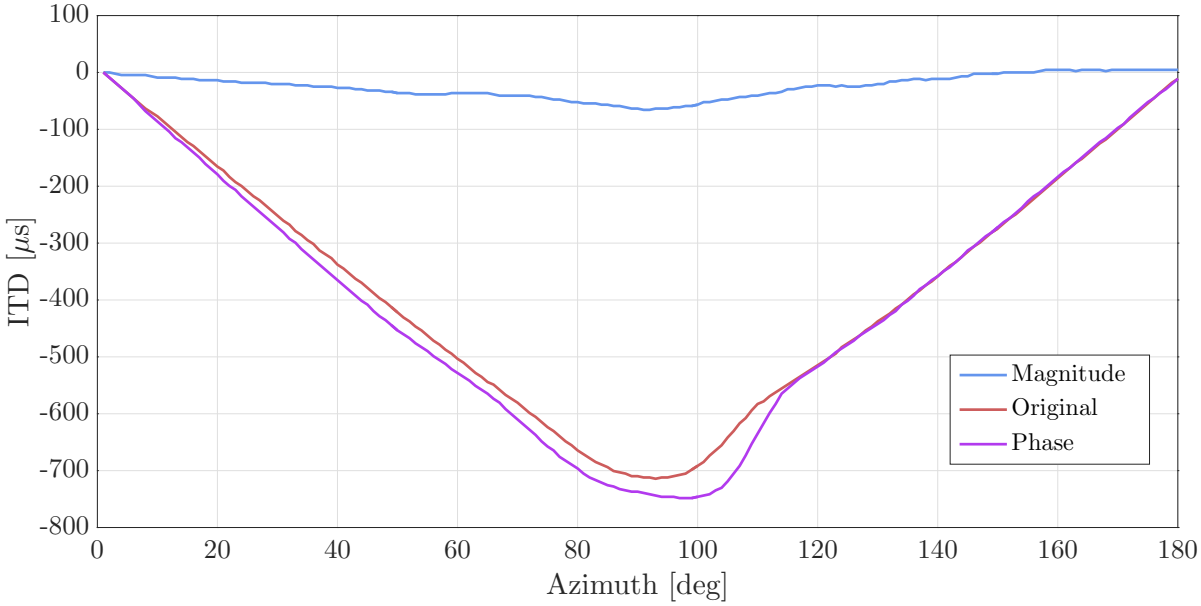


Figure 4.3: ITD Cues for the TF types original (red line), magnitude (blue line) and phase (purple line) for participant 31.

The condition phase shows that there are no spectral cues for sound source localization. Considering the HRIR/SHIR, one can see that in the TF type magnitude there is no information about the time of arrival and therefore about phase ratios. In the case of the HRIR/SHIR phase, there is a slight loss of amplitude of the left ear signal compared to the original impulse responses. However, the signal of the contralateral ear increases in amplitude so that the ratios of the two signals are the same. This is due to the equalization of the signal in the spectral range, which is approximately equal to a lack of shading and a lack of different distance between the sound source and the ears.
In Figure 4.2 it can be stated that in general spectral cues are minimal due to the spherical head shape.
By manipulating the original transfer functions (to TF type magnitude and phase), slight deviations from the characteristics of the original HRTF may have occurred. These differences, as well as the influences of the different TF types, can be described by the localization cues ITD and ILD. Figure 4.3 shows the ITD and Figure 4.4 the ILD of HRTF for all three TF types (original (red line), magnitude (blue line), phase (purple line)) in horizontal plane with $\vartheta = 0°$.

In the case of TF type magnitude (blue line), it can be seen that almost no temporal information is available for localization over the entire horizontal plane.
If one compares the original TF (red line), as well as the TF which represent the TF type phase (purple line), one can only see a minimal overestimation of ITD cues. The slight

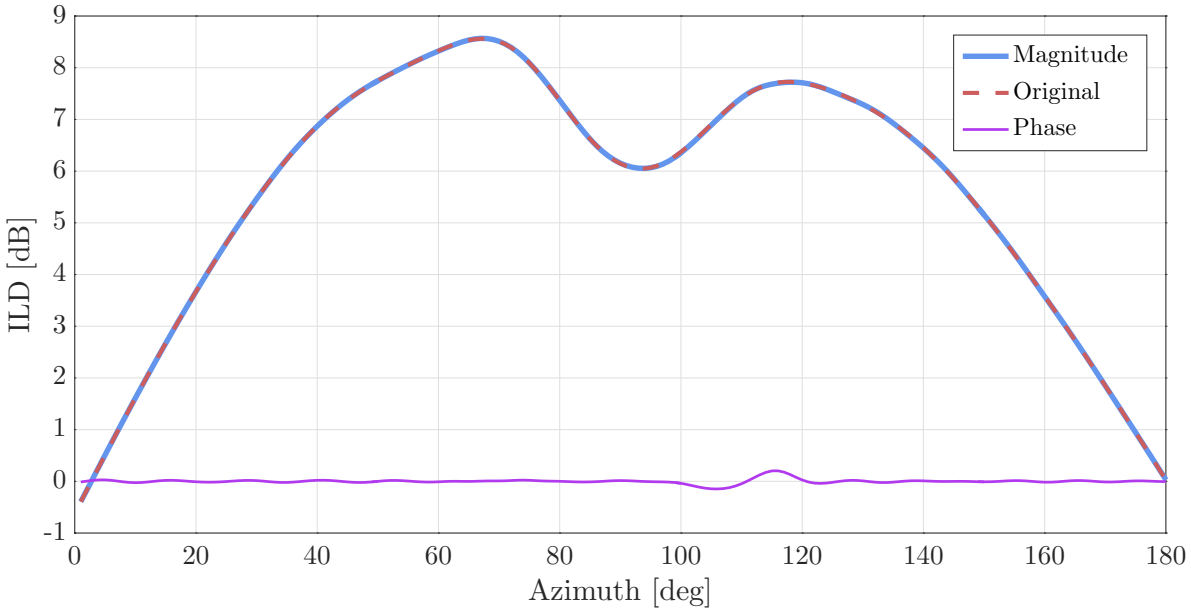


Figure 4.4: ILD Cues for the TF types original (red dashes), magnitude (blue line) and phase (purple line) for participant 31.

discrepancy may be due to a simplified detection of the ITD.
Figure 4.4 shows the same TF types as in Figure 4.3, this time for the ILD cues of the different transfer functions. In this case, the TF type phase (purple line) does not contain any localization information in ILD cues. The ILD cues of the TF type original (red dashes) and the TF type magnitude (blue line) match exactly.

Stimuli

The signal the virtual sound source emits in the listening test was chosen to be pink noise. It is considered pleasant and less shrill than white noise and has similar properties to speech and music. Studies found that the pink noise is one of the few signals which, without being known by the listener, leads to the correct assignment of the sound source position during localization attempts (Weinzierl, 2008, p. 95).
Two different stimuli were used for the listening test, one stimulus with a duration of 5 seconds, one with a duration of 100 milliseconds. Figure 4.5 shows both stimuli types[9]. In the dynamic case, the test subjects heard a stimulus of pulsed pink noise over a total length of 5 s (see left image of Figure 4.5). The stimulus consisted of a train of pulses with a length of 100 ms. The respective pauses were also 100 ms and the individual pulses were

[9] The stimuli were generated via the AKtools `AKpulsedNoise.m` Matlab function (Brinkmann and Weinzierl, 2017a)

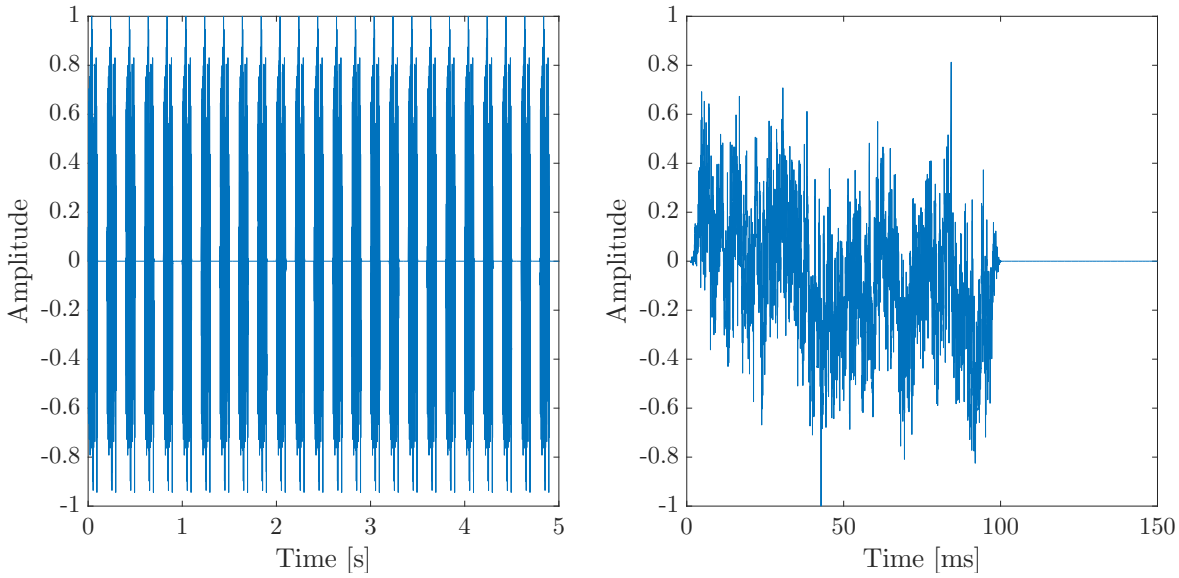


Figure 4.5: Different stimuli used in the listening test. Left side: 5 s train of pink noise; Right side: 100 ms Pink noise pulse;

faded for 5 ms each. For the whole signal one and the same randomly generated pulse was used. The pink noise and the 5 ms fade also applies to the second stimulus, with the exception that it consists of a single 100 ms pulse. The amplitude of both signals was normalized to a range of 1 to -1.

Technical Equipment and Software

An desktop computer with a Windows 2010 operating system was used to perform the listening test. The VR Head Mounted Device (HMD) used was an Oculus Rift Virtual Reality Headset with Touch Motion Controller Bundle and 3 sensors.The use of a HMD requires a fast, high-resolution graphics card. To achieve this, a Nvidia GeForce GTX 1060 was used and in our case resulted in an update rate of 60 frames per second of the HMD. The headphones mounted on the Oculus Rift were removed and replaced by Sennheiser HD 650 headphones with a Lake People Phone-AMP G109 headphone amplifier.
To use the Oculus Rift, the Oculus App[10] is generally required. The entire setup of the listening test was created using the game engine Unity 3D [11]. The HMD was connected to the game engine via the application Steam[12].
The creation of environments and tasks in Unity 3D is based on the C# programming

[10] `https://www.oculus.com/setup/?locale=de_DE` (last downloaded: December 03, 2018)
[11] `https://store.unity.com/download-nuo` (64 bit version Unity 2018.2.14f1; last downloaded: December 03, 2018)
[12] `https://store.steampowered.com/?l=german` (last downloaded: November 30, 2018)

language. The C# scripts included in Unity were created with Microsoft Visual Studio 2015. Predefined objects are existent that can be loaded as plugins into Unity 3D. For this listening test the open source plugin SteamVR[13] was used to mainly control the controller functions and the head tracking. In addition, for rendering the HRTF and the associated sound output per tracking position, the plugin SteamAudio[14] was used.
All visual components (meshes, objects) were created with the open source graphic software Blender[4].

Room Set-up

The listening test was carried out in the Media Lab in the main building of the TU Berlin. Experimenter and Participant were placed in different rooms. The walls of the room where the listening test took place were covered with a black curtain and had a dimmed lighting. Figure 4.6 (upper Photo) shows the placement of the three motion tracking sensors of the Oculus rift, indicated by red circles. Two sensors were at a distance of 1.2 m from each other (left circles), at a height of 1.1 m, slightly inclined to each other to focus the center of the playing field. Another one was opposite, slightly lower (0.9 m), also directed to the center.

The following preparations were done, in order to ensure the freedom of movement of the participants during the test: In the center of the play area, participants were placed on a swivel chair with no armrest and no backrest. A cable holder was used to hold the cables of the HMD and of the headphone above the heads of the participants. This was to avoid getting tangled up in the cables. Nevertheless, the participants were asked to rotate the swivel chair backwards by a maximum of 180° and to start the rotation again from the other side for further rotation.
Scenes were started subsequently by the experimenter in the control room. The experimenter was able to communicate with the participant via microphone and loudspeaker, but could not see him. Only movements on the playing field in Unity could be observed on the screen.
The A-weighted energy-equivalent continuous sound pressure level $L_{\mathrm{Aeq,T}}$ in the test room, measured over 5 minutes was $L_{\mathrm{Aeq,5\,min}} = 42.7\,\mathrm{dB}$.

[13] `https://github.com/ValveSoftware/steamvr_unity_plugin` (Version 1.0.12; last downloaded: December 03, 2018)
[14] `https://github.com/ValveSoftware/steam-audio` (Version v2.0-beta.16; last downloaded: October 25, 2018)

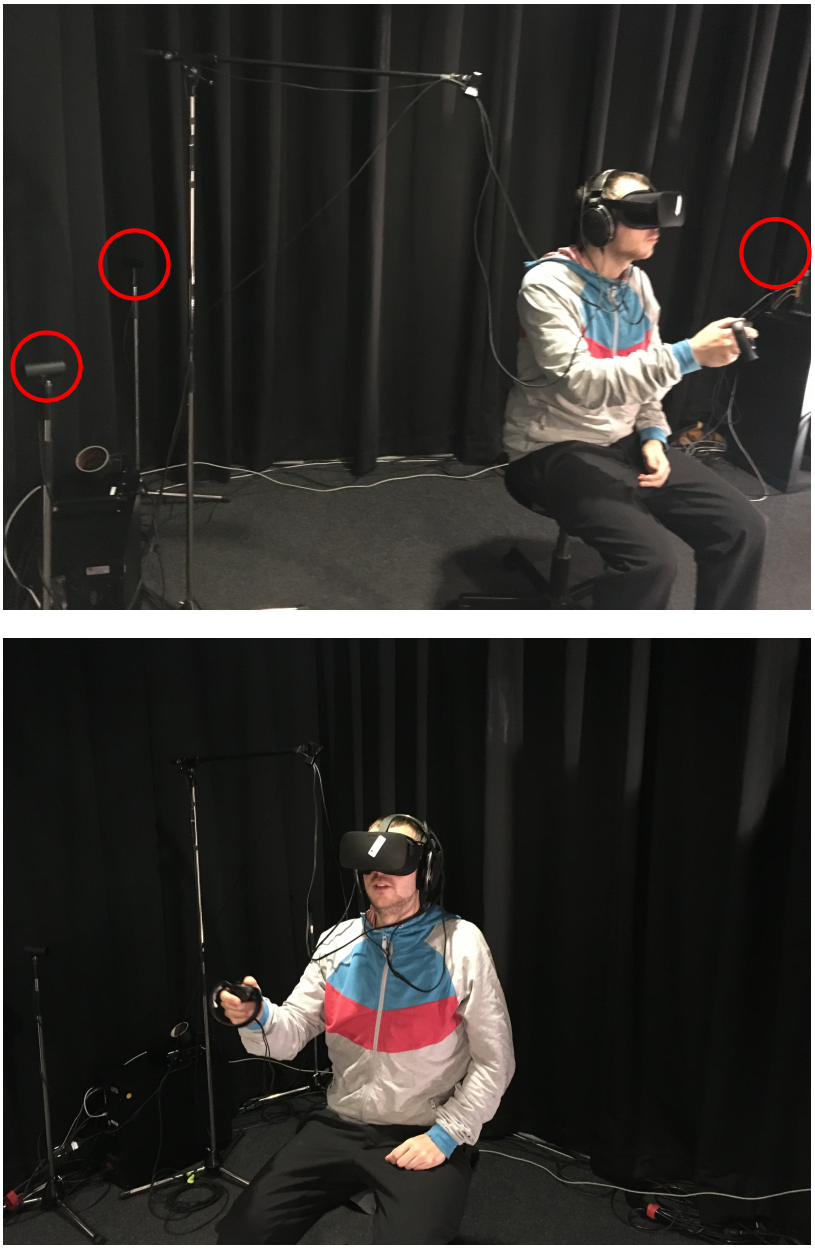


Figure 4.6: Listening test set-up at Media Lab TU Berlin. Red circles indicate the positions of the sensors. Exemplary for a participant sitting on the swivel chair and performing the listening test wearing the HMD, headphones and using the controller.

Pointer Method

The pointer method is the method by which the participant indicates the location of his perceived sound source.

The choice of the pointer method is important to get an intuitive, accurate response and to be easy to use. Two different methods were used in previous studies to give answers

during the listening test via VR device. It is possible to indicate the perceived location by the viewing direction or nose tipping (Middlebrooks, 1999). Furthermore, a detected laser beam coming from the VR controller to indicate the location can also be used (Seeber, 1997).
Majdak et al. (2010) showed that there is no significant difference in the localization accuracy between these two pointer methods in VR, especially without long-term training. However, a comfortable use for the participant is necessary.
In this listening test the sound sources were only placed in the median plane. Sources near the upper and lower poles can lead to unpleasant positions when using the nose tipping method. Thus, the laser pointer method was chosen. The participant does not have to take unpleasant head positions and can therefore reach all positions more easily and unhindered with the laser pointer.

Visual Information

Montello et al. (1999) showed that subjects who were in a visual, spatial environment without visual stimuli leading them to the sound source had a higher localization precision and accuracy than subjects whose eyes were blind folded during the experiment. This means that the localization ability improves as soon as a plausible spatial environment is provided. Majdak et al. (2010) found that the localization bias in blindfolded experiments is significantly lower in the anterior region than in the posterior region. For head mounted devices, however, there is no difference between the bias in the anterior and posterior hemispheres. In addition, in contrast to a rectangular room geometry, Winter et al. (2017) discovered that in localization listening tests the accuracy of sound source localization increases as soon as the room is spherical. Since the sound sources usually originate from a spherical loudspeaker array in order to achieve a constant distance from the sound source to the listener, this distance is also given visually in a spherical room. This is particularly advantageous in virtual reality, where the sound sources are not visually present in the form of loudspeakers.
Redon and Hay (2005) found that a visual constant grid structure in a room reduces the pointing bias. Accordingly, the spherical surface was provided with a grid structure in 5° steps. This serves as an orientation, especially for static localization cases with short stimuli, in which the participants cannot move in the direction of the sound source, but have to remember from where it came from. The grid structure was highlighted on the median plane, as well as on 90° and 270° azimuth to provide more information on the orientation (see Figure 4.7 and Figure 4.8 in Section 4.2).

Spatial Accuracy of Headtracking Device

The accuracy of head tracking can be divided into two components. The first component is the temporal resolution, which is the transmission rate between the head tracking and the associated optical and acoustic update. On the other hand, the spatial accuracy is expressed.
In the first case, the entire transmission chain (depending on computer capacity, load, etc.) offered an update rate of 60 frames per second in this case, which corresponds to a very good temporal resolution [15].
No precise data provided by the company Oculus could be found for the spatial accuracy [16].
Xu et al. (2015) conducted a study on the accuracy of the head tracking of the Oculus Rift compared to a high accuracy tracking sensor. The subjects had to move their heads to certain points. The difference between the tracking data of the on the Oculus Rift mounted sensor and the internal tracking of the HMD was observed. The authors found a lateral averaged deviation of 0.8° (std: 2.2°) and a axial rotation of 0.8° (std: 2.4°) to the original head position.
However, no information on the number and positioning of the sensors was provided as well as no exact information how they ensured the exact head position. To ensure that the accuracy of the internal tracking system was sufficient for a localization test, a small experiment was performed. The Fast and Automatic Binaural Impulse response AcquisitoN (FABIAN) artificial head of the Audio Communication Group at TU Berlin was used for this purpose (Lindau and Weinzierl, 2007). The head of the FABIAN can be accurately adjusted to certain positions (azimuth: [0,360], elevation: [-30,+30]).
The artificial head was placed in the middle of a room, wearing the Oculus Rift HMD. Three sensors were used for this setup, as well as for the actual localization test. Two of these sensors were placed at the level of the head at the front, a third sensor diagonally behind and slightly below the artificial head. The HMD was adjusted so that at a position of 0° azimuth and 0° elevation of the rotation joint, no deviation was indicated by the tracking sensors. A combination of 19 azimuth angles (from 0° to 180° in 10° steps) with 7 angles in elevation (from ±30° in 10° steps) were examined. Accordingly, a total of 133 angles were considered.
The evaluation of the considered angles showed an average deviation of 0.53° with a standard deviation of 0.35° in the horizontal plane and an average deviation of 0.34° with a standard deviation of 0.23° in the elevation plane. The maximum deviation in the azimuth

[15] Frame rates using an Oculus Rift are between 45 and 90 frames per second. `https://developer.oculus.com/documentation/pcsdk/latest/concepts/dg-performance-guidelines/`
[16] One reason may be, that the spatial accuracy depends strongly on the environment, the number of sensors, as well as their positioning.

plane was 1.45°, in the elevation plane only 0.89°.
The observed deviation of the sampled positions are similar to the deviations examined by Xu et al. (2015). Further, with a spatial uncertainty of the head tracking below the spatial resolution of the HRTF, the data acquisition should not affect the listening test results.

4.2 Test Procedure

A questionnaire was handed out to the test person before the experiment. This included personal information (age, gender), the subject's sense of hearing (impairment, tinnitus, cold) and their expertise in acoustics, listening tests and explicit localization tests. In addition, a declaration was signed to the effect that the participant could stop the test at any time. Subsequently, the participant received a 20 to 30 minute briefing on the localization test and the equipment. They were informed about the respective tasks in all three training steps, as well as the structure of the localization test. The participants were told beforehand that there are two different stimuli lengths. n addition, participants were encouraged to move their head and upper body freely for localization purposes.
However, the participants were not informed about the exact conditions of the experiment. During the localization test, the subjects wore a HMD and held a controller in their right hand[17]. In addition to the HMD, the participants were equipped with headphones.
The participants were virtually placed inside a sphere with a grid structure for better orientation (see schematic representation in Figure 4.7[18]). The median plane, as well as 90° and 270° azimuth were color marked to avoid disorientation (see green line in Figure 4.8). In steps of 5° there were orientation lines over the whole sphere. At 0° azimuth and 0° elevation there was a green reference point. At the beginning of each trial, the participants always had to navigate to this point. The point was additionally used for orientation and as a reset point for head tracking. For each participant, the HMD had to be calibrated, which is briefly explained in the following. The tracking of the Oculus Rift can, after consultation with the company Oculus, only be reset in horizontal shifts. It is not possible to influence the initial position of the elevation angle. Due to the nature of the HMD and different head shapes, there may be an off-set in the elevation. For a sharp view the glasses must be pressed firmly to the head. For this purpose, the HMD is equipped with kink points which the operator can adjust to achieve an optimum, individual fit. An elevation of 0° would only be possible if

[17] Each participant was asked if they could handle the operation of the controller in their right hand. Everyone stated that they are comfortable with it. Otherwise it would have been possible to use the left controller
[18] Note, that the grid on the sphere is not the exact grid used in the listening test.

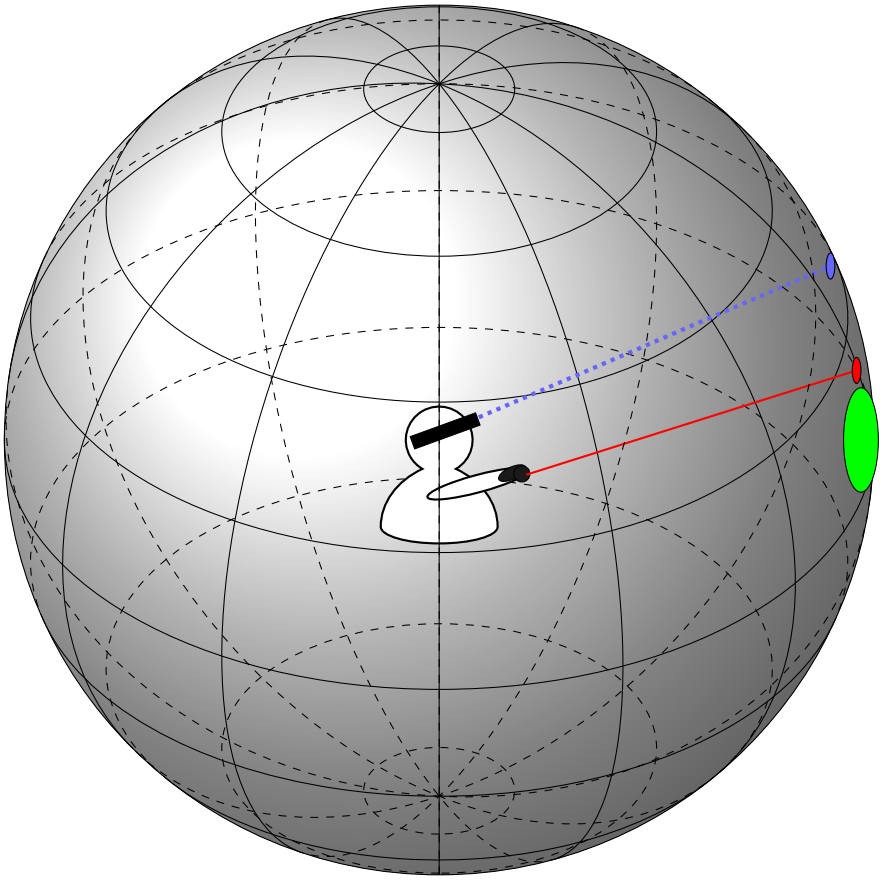
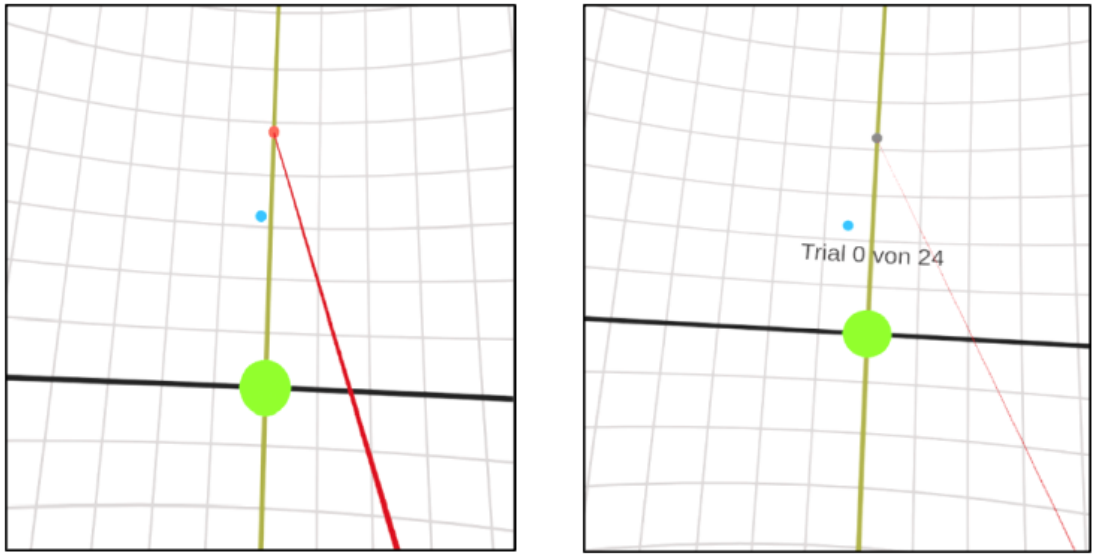


Figure 4.7: Scheme of the set-up in virtual reality. Person indicates the view positioning of the participants. Blue dot indicates the head position (to reset after each trial); Red line and dot indicate the laser beam, coming from the controller, and the response dot; Green dot is the reference point;

the glasses were seated exactly horizontally at the head of the participant. The off-set varies depending on the depth of the eye socket and the cheek bones. Deviations of up to 8° were determined. To avoid these deviations, a manual reset was performed at the beginning of each trial. This was realized by displaying a dot on the sphere indicating the viewing direction (see blue dot in Figure 4.8). This dot had to be navigated by rotating the head to the reference point at the beginning of each trial (green point in Figure 4.8). In order to start a new trial, the deviation of the reference point and the head orientation dot must not exceed 1° in all directions. Thus the deviation of the elevation angle was reduced to below 1°, when starting each trial. If the range of 1° was exceeded, the participant was notified by the entire VR image getting a red cast for one second. The dot of the viewing direction also indicates the recorded tracking data during a trial. Only a rotational movement was stored, but not a translational movement of the head.
A permanently displayed laser beam was emitted from the controller to the observation plane. When the so-called trigger button was pressed, the laser beam became thicker. A



(a) Active laser beam (thick red line) within a trial to position the response dot (red dot).

(b) After a trial is finished, the progress display appears and the answer is grayed out.

Figure 4.8: Perspective through the HMD of the test participants in virtual reality.

red dot was placed at the end of the laser beam, on the surface of the sphere, which could be moved by pressing the trigger button. This red dot was used to indicate the perceived sound source position.
Within each trial, the source location had to be confirmed as soon as the participant placed the red dot of the laser at the perceived position. For this purpose, the B button of the controller had to be pressed. In the following, the red dot was grayed out. A moving of the dot by the laser was no longer possible. If the button was pressed accidentally, the respondent could release the point by pressing the B button again. The position of the perceived sound source was not stored until a new trial was started.
The listening test consisted of five parts. The first four of which were dedicated to handling the equipment and were called 'training' and the fourth one was the listening test itself.

Training

The first part of the training was used to learn how to handle the controller. In order to learn the handling and the precision, a gray dot was shown on the surface of the sphere for each trail. The test person now had to try as precisely as possible to place the red dot at the end of the laser beam onto the gray dot on the surface of the sphere. After confirming the red dot, it was shown whether the indication was precise enough (the sphere turned green for a sufficiently precise indication or red for an insufficiently precise indication).

Deviations below 1° were considered sufficient. One trial consisted of one gray dot on the spheres surface. A dot was randomly displayed in each octant. The order of the octants was also randomly selected. The controller training was finished as soon as the respondent had set the dot precisely enough in all eight octants. No acoustic stimuli or confirmation was used in this task. On average, the participants had about 2 failed attempts in this part of the training. The handling of the controller was learned very fast by the participants and did not cause any problems during the listening test.

The second part of the training served to adjust the volume for all subsequent signals of the training, as well as the localization test. In order to avoid a previous listening test[19] and still achieve the same volume for all subjects, the participants should adjust their own volume. For this purpose, a continuous pink noise was presented. The participants had to adjust the signal level with the controller so that they could no longer hear it. They could turn the signal level in 0.5 dB steps up and down to determine their hearing threshold. The volume change took place in 0.5 dB steps.

In the further listening test, the stimuli volume was adjusted to an over all 50 dB level (± 2.5) individually for each test person on the basis of their threshold. Thus a previous hearing test of the hearing threshold could be avoided. In order to avoid loudness effects, a volume of 50 dB based on Vliegen and Van Opstal (2004) was used. In addition, the signal level was randomly increased or decreased by ±2.5 dB to avoid localization based on trainings effects based on the total level (Majdak et al., 2010).

The third part of the training is dedicated to the localization of sound sources. Each trial, the subject was presented with a pulsed pink noise for an unlimited time[20]. The used transfer function was the original HRTF of the participant. During stimulus playback, the participants were allowed to move freely to identify the location of the sound source. The stimulus length was not limited so that the participants could get a feeling for the change of cues by moving the head. The subjects had to decide on the perceived sound source location. They were then asked to place the dot at the end of the laser beam at perceived position. The participants received no feedback on the actual location of the sound source. The sound sources presented were, at the median level only. One stimulus per median plane quadrant was presented in four trials. The subject had to hit the correct quadrant to complete the exercise.

The last part of the training was the same as part three, with the exception that the subjects were now presented with a limited stimulus. This had a length of 5 s, just like in

[19] A pronounced, undiscovered hearing loss was not to be expected due to the age of the test persons. The proportion of hearing impaired people in the age group 25 to 34 years is 5.2 %. (https://de.statista.com/statistik/daten/studie/913261/umfrage/anteil-der-schwerhoerigen-nach-altersgruppen-in-deutschland/)

[20] same conditions as the in Paragraph **Stimuli** described 5 s pulsed noise, but looped

one part of the main experiment.
The localization training was intentionally kept short and without feedback to avoid any training effects and to keep the listening test intuitive and natural. The interest of the listening test is in the discussion of the different influences from the investigated conditions. An improvement or absolute localization performance is not of relevance here. It is assumed that the localization ability does not change over the entire listening test without feedback on the actual sound sources.

Listeningtest

In the main experiment, the test subjects had to locate the sound source in twelve different conditions (see Table 2), each with 24 angles. The angles of the sound source to be localized were distributed in 15° steps over the entire median plane. However, since the resolution of the HRTF was 2°, angles that were not divisible by 2 were rounded to the nearest value divisible by 2 (depending on the sign). Due to a programming error, there were six angles which lost their sign in the course of the angle determination. Accordingly, six angles in the lower median plane were not presented to the participants, but six angles in the upper median plane were presented twice. This is true for all conditions and all subjects to the same extent.
The conditions were grouped into four blocks. A block consisted of a head movement condition, a geometry and the three TF types (resulting in three laps per block). Within the blocks, the order of the TF types was randomized. The four blocks were also presented in a randomized order.
The subjects were informed that only angles on the median plane were considered. Therefore, they only had to be precise in the elevation but not in the azimuth angle. Angles in the horizontal plane were set to 0° or 180° respectively.
All perceived and actual sound sources were stored in a text file. Additionally the data of the head tracker was stored with 60 entries per second.

4.3 Results

This section evaluates all data collected during preprocessing as well as during the listening test. The statistical method for evaluating the data from the listening test is declared and applied. The results obtained are explained and discussed.

Subjects

The participants were selected from an existing pool of subjects. In this study, there were 13 participants in total, 2 women and 11 men, whereas three of them had to be excluded due to a programming error in the HRTF creation. The remaining subjects had an average age of 34.5 years (men ∅ 35.9 years / women ∅ 33 years).
One of them indicated maybe having a hearing loss on both sides and a second one had a little cold. However, the evaluation did not reveal any outliers attributable to these participants.
All participants, with one exception, stated that they were active in the field of acoustics in their private or professional lives and had already participated in several listening tests, each of them having taken part in at least two, usually more than ten. However, none had special expertise in sound source localization.
Two of the participants were spectacle wearers who did not have the opportunity to use contact lenses. However, their frames were narrow enough to be worn under the VR HMD, as suggested by the manufacturer[21]. They stated that they had not noticed any limitations or disadvantages.
All participants needed between 1.5 hours and 2 hours for the listening test including filling the questionair, preliminary discussion, explanation of the equipment, training, the actual listening test and a break of approximately 15 minutes. None of the participants needed more than 50 minutes for the task with VR HMD.
The number of participants in this localization test corresponds to similar studies on sound source localization (Majdak et al., 2010; Jiang et al., 2018).

SHTF Model Parameters

As already explained in paragraph **Test Conditions**, the dimensioning of the SHTF for each participant individually, was calculated with the TOA estimate algorithm, invented by Ziegelwanger and Majdak (2014). This led to a radius of the sphere, an ear position in the azimuth plane and an ear position in the elevation plane for each participant. The calculated dimensions of the spherical heads led to the smallest possible difference between the ITDs of the SHTF and the original HRTF.
The position of the left and right ear were optimized independently. Thus, positional differences of the left and right ear, as well es the radius of the sphere, depending on the side and the corresponding participant, occurred. Figure 4.9 shows the differences for the

[21] Suggested by Oculus, the glasses may have the maximum dimensions: 142 mm wide, 50 mm high. (`https://support.oculus.com/191247164573652/`)

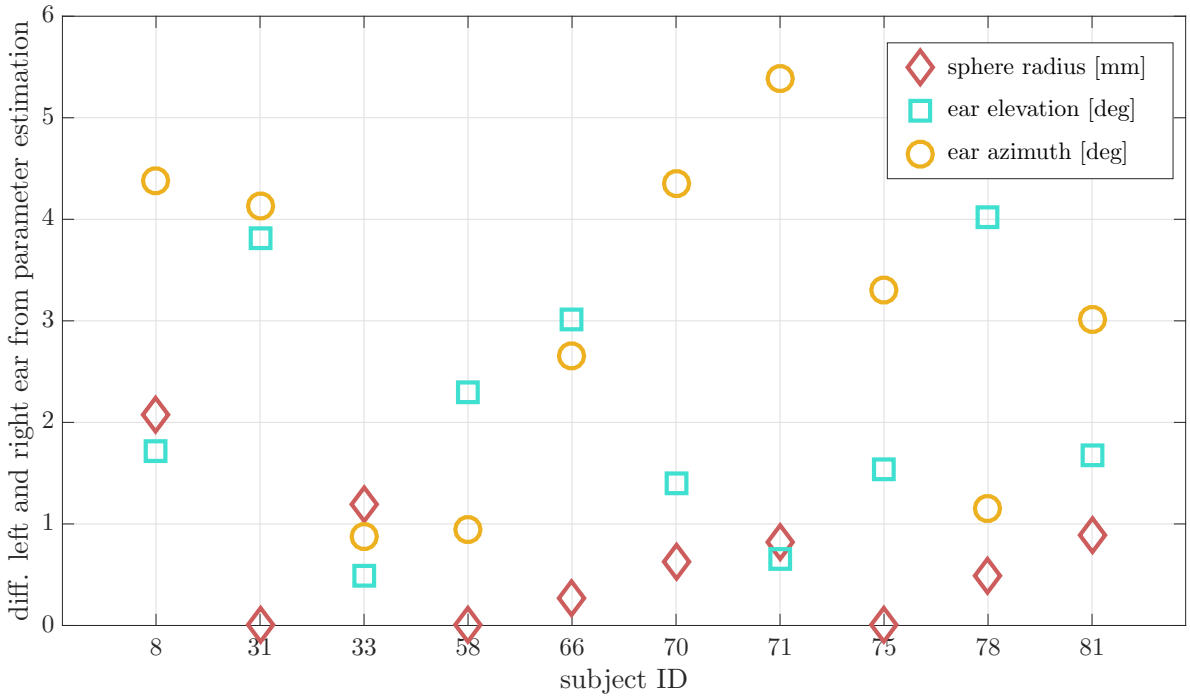


Figure 4.9: Difference between the left and right ear parameters from the TOA estimation algorithm. Parameters are the radius of the sphere (diamond), elevation of the ear (rectangle) and azimuth of the ear (circle).

azimuth and elevation angle between the left and the right ear and the differences between both radii for the two optimization processes per subject. By observing the depicted positional difference in the elevation plane, a minimum of 0.49° and a maximum of 4.02° can be found. In horizontal plane the range is between 0.88° and 5.38°. The differences in ear position between the two ears vary greatly depending on the subject. For example, subject 33 has a very low estimated asymmetry between both ears, whereas subject 31 has a high estimated asymmetry in both planes.

The calculated deviation for the spherical head radius between the left and right ear is relatively small. For three participants, the same radius was calculated for both sides, so the difference is 0 mm. The largest difference is 2.1 mm.

The differences between the estimated parameters between the two ears are most likely due to the head shape and its asymmetries, which lead to different cues on both sides of the original HRTF. The asymmetry does not necessarily have to be in the off-set of the ears, but are estimated as that due to the simplification of the spherical shape.

When calculating transfer functions, symmetry is always assumed. Since the spherical head model allows only symmetrical ear positions, the values of the left and right ear estimations were averaged per person to calculate the SHTF.

Table 3: Estimated parameters averaged over all participants.

parameters	mean	std	min	max
radius	8.7 cm	0.46 cm	8.2 cm	9.48 cm
ear azimuth	88.6°	1.89°	86.59°	91.66°
ear elevation	-6.64°	2.59°	-13.18°	-4.13°

The result for each subject was an ear position consisting of azimuth and elevation angle and a radius for the spherical head model. To be able to compare the estimated parameters with values from the literature and to be able to generalize subsequent results, the ear positions and the radii were averaged over all participants and shown in Table 3. In the elevation plane, the difference between the minimum and maximum values was 9.05°. However, all shifts were below 0° elevation. In azimuth plane the difference was 5.07°. In this case, however, the off-set lay between a shift to the front and a shift to the rear.
To asses the individual parameter values, a comparison to other studies will be done in the following.
Comparing the values of Table 3 to values in the literature, the mean pinnae off-set in elevation of -6.64° (std. 2.59°) applies very well to the value of -7° mentioned by Burkhard and Sachs (1975). This is also true for the mean pinnae off-set in the azimuth plane. In this work the average off-set was 88.6° (std: 1.89°) and 87° for Burkhard and Sachs (1975). However, three test persons in this work have an off-set in the azimuth plane to the rear (>90°), whereby the maximum value was 91.66°. A backward off-set is more unusual and explains the slight difference to values in literature.
In total, the calculated sphere radii had a range from 8.2 cm to 9.48 cm with a mean value of 8.7 cm (sd: 0.46 cm) . According to Hartley and Fry (1921), an average measured head radius is 8.75 cm. Ziegelwanger and Majdak (2014) calculated with their model an average head radius of 8.695 cm. In comparison, the mean value of 8.7 cm is exactly in this range. The dimensioning of the MTB microphones of Algazi et al. (2004) (8.75 cm) and Fiedler et al. (2017) (8.8 cm) also apply to these results.

Data Post Processing

The participants were informed that only sound sources in the median plane would be examined and that no exact information in the horizontal plane was required. This was in order to avoid unnecessary time expenditure. For responses that deviate from the median plane the azimuth angle information was corrected. Accordingly, all azimuth angles were

set to $\varphi = 0°$ or $\varphi = 180°$ on whether the sound source was perceived in the of the head or in the back of the head.
The Matlab function `localization.m` of the Auditory Modeling Toolbox[6] was used to calculate the error measures for all test persons and all conditions (Søndergaard and Majdak, 2013).
All data from the localization experiment were used to calculate the following error measures in spherical coordinates (azimuth (φ), elevation (ϑ)) and polar coordinates (altitude (ϕ), polar (θ)). This is a prerequisite for using the AM Toolbox.

Error Measures

Two different error measures were used to describe and analyze perceived localization of sound sources of the listening test.
The first error measure is the *quadrant error*. It describes the percentage of cases in each condition in which participants report perceiving the sound source in a quadrant where it did not actually occur. The confusion in terms of sound source localization can be between the front of the head ($\varphi = 0°$) and the back of the head ($\varphi = 180°$). They can also occur between sources from above the head ($\vartheta = 0°$ to $90°$) and below the head ($\vartheta = 0°$ to $-90°$). Both of these confusions can be present at the same time. For something to be a quadrant error the actual sound source and the perceived location of the sound source are in different quadrants. In addition, however, these two positions must be at least 45° apart.
For statistical evaluation the quadrant error should be corrected for all other error measures in order to keep error measures separate and not receive a statistical double rating of the quadrant error (Carlile, 1996; Majdak et al., 2010)
The second error is the so called *local polar bias* $\overline{e}_\theta$. A synonym that is often used in the literature is *mean polar accuracy with quadrant error removed*. It measures the accuracy of the perceived position of a sound source in polar dimension averaged over a set of trails. Letowski and Letowski (2011) define, that "Accuracy [...] is the measure of the degree to which the measured quantity is the same as its actual value.". This means, in the terms of averaged accuracy, it is the mean value of the differences between actual and target value, calculated including circular/spherical statistics" (Letowski and Letowski, 2011). In the following, it is briefly explained how this error measure was calculated.
Frist, the angle θ, corresponding to a specific participant rating, was quadrant error corrected.[22] Second, the deviation from the source position was yielded by subtracting the

[22] The quadrant error is compensated by transforming the data of the perceived sound source into the same quadrants as the actual sound source.

target angle $\hat{\theta}$ so that

$$\theta_{\text{error}} = |\theta - \hat{\theta}|. \tag{11}$$

Further, to calculate the mean angular deviation $\overline{\theta}_{\text{error}}$ a participant achieved over a set of N trials of a certain condition, circular statistics was used. This is necessary since conventional statistical methods are subject to the prerequisite of linear infinite distributions. However, in this case the distribution of values is circular (Letowski and Letowski, 2011).
Considering that each angle θ_{error} belongs to a point on the unit circle, a correct averaging can be achieved by calculating the mean over each of its cartesian coordiantes with

$$x = \frac{1}{N}\sum_{n=1}^{N}\sin(\theta_{\text{error}}^{(n)}) \quad \text{and} \quad y = \frac{1}{N}\sum_{n=1}^{N}\cos(\theta_{\text{error}}^{(n)}). \tag{12}$$

The local polar bias is therefore the angle corresponding to the resulting point $\mathbf{x} = [x, y]$

$$\overline{\theta}_{\text{error}} = \begin{cases} \arctan(\frac{y}{x}) & x > 0 \\ \pi + \arctan(\frac{y}{x}) & x > 0, y \geq 0 \\ -\pi + \arctan(\frac{y}{x}) & x < 0, y < 0 \\ \frac{\pi}{2} & x = 0, y \geq 0 \\ -\frac{\pi}{2} & x = 0, y < 0 \end{cases}. \tag{13}$$

For the statistical evaluation in this thesis, the local polar bias was additionally averaged over the M subjects of the listening test. The resulting *mean polar bias* μ_θ

$$\mu_\theta = \frac{1}{M}\sum_{m=1}^{M}\overline{\theta}_{\text{error}}^{(m)} \tag{14}$$

provides a general statement about the accuracy of the position statements of the test persons under a certain condition. When averaging local polar biases across participants, an arithmetic mean is sufficient and a transformation to circular statistics is not required. The two error measures used describe the localization ability of a participant in the most important characteristics (Majdak et al., 2010).
As described above the *local polar bias* is quadrant error corrected (to avoid double rating of the quadrant error). The same metric without the quadrant error correction will be referred to as *polar bias*.

Statistical Method

The two error measures, the local polar bias and the quadrant error, are the dependent variables used for statistical analysis.
The different conditions (see Table 2) of the localization test form the three different factors used in statistical analysis. These factors are the head movement, geometry and TF type. The factors head movement (dynamic/static) and geometry (head/sphere) each contain two levels. The factor TF type is divided into three levels (original/magnitude/phase).
Since each participant had to go through all the conditions, the answers between the different conditions are not independent of each other. Due to intrinsic motivation, personal localization abilities and possibly slightly different hearing ability, each person has a different response behavior, which is reflected in all conditions in this experimental setup. Additionally, answers to previous trials may influence future responses. Therefore, the statistical model must be based on repeated measures (Bortz, 2005).
Since each test person passed each condition once and three test persons were excluded beforehand, the available data is a balanced design (Field, 2007).
A three-factor repeated measures MANOVA was conducted to assess which factors influenced localization ability. In a multivariate analysis, two or more dependent variables can be included in the evaluation without increasing the alpha error (Bortz, 2005).
Statistical significance was determined based on the Pillai's trace test statistic. Pillai's trace test is considered to be very robust and particularly suitable for small samples (Field, 2007).

A number of post hoc tests were used to examine the influence of the factors in more detail and the effect on the respective error measures individually.
Two univariate variance analysis were used as the first follow up statistic. These consider the influence of the different factors and their combinations on the two error measures independently.
All results of the post hoc tests are Bonferroni corrected to reduce a type I error. A type I error exists, if a null hypothesis is wrongly rejected. Since in the univariate comparison two univariate hypotheses are derived from the MANOVA hypothesis, the significance levels are aligned with the Bonferroni correction so that the probability of committing at least one 1st type error is limited. This Bonferroni correction means that the likelihood to wrongly reject a null hypothesis is reduced to the level of a single statistical test (Field, 2007).
In order to gain a detailed insight into which level of the factors have an influence on the respective error measure, a pairwise comparison was also carried out. This compares the difference between the respective levels of a factor. Since this can only be indicated as

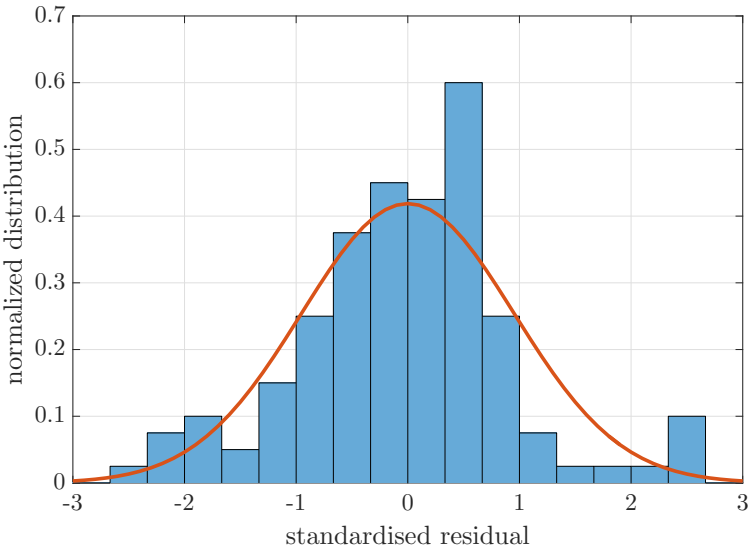


Figure 4.10: Distribution of the residuals of the quadrant error and additional normal distribution curve (solid line).

meaningful for the individual factors by means of significance values, the combinations of factors considered significant in the univariate analysis were compared visually. This is illustrated by the estimated marginal mean. Estimated marginal means represent the mean responses of one factor, controlling for covariates (Field, 2007).
All calculations were carried out using the statistics program IBM SPSS Statistics[23]. For all analysis the alpha level was set to $\alpha = 0.05$. However, p values above 5% and below 10% were determined as marginally significant ($0.05 < p \geq 0.1$)in this work. This choice was made due to the low number of subjects in the listening test, which may result in less statistical power.

Assumptions

One assumption for the MANOVA is the normal distribution of the dependent variables. In this case, a normal distribution of the data of the two error measures (Bortz, 2005). For this purpose, the residuals for each of the error measures were calculated and visually examined. A consideration of the normal distribution of residuals is meaningful in this case, since a large deviation of the distribution around the mean value can occur under 12 different conditions. Due to this, the data mean value is released by means of the residuals. Residuals are determined by calculating the distance between the data and the regression line created by the data. Figure 4.10 shows the distribution of the residuals of

[23] `https://www.ibm.com/de-de/products/spss-statistics`(IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp. Last downloaded: June 19, 2019)

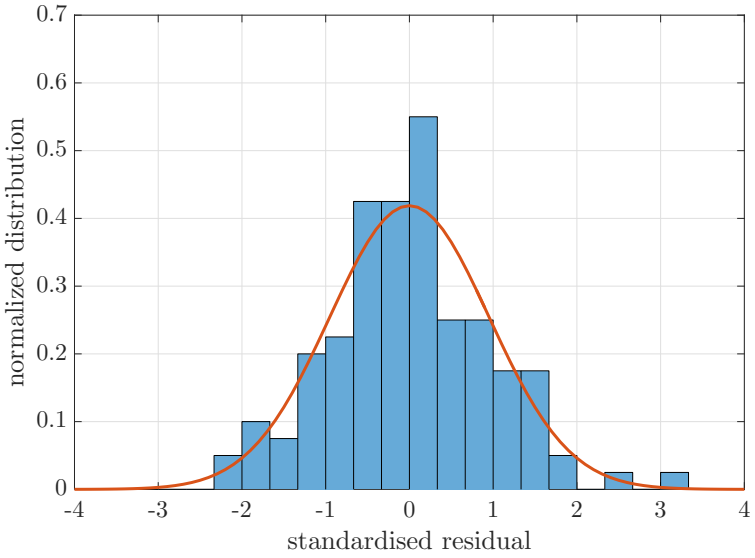


Figure 4.11: Distribution of the residuals of the local polar bias and additional normal distribution curve (solid line).

the quadrant error and the corresponding normal distribution curve. Figure 4.11 shows the residuals of the local polar bias and their corresponding Gaussian distribution curve, based on a sample number of N = 120 (number of persons (10) × number of conditions (12)). In both cases, a normal distribution can be assumed based on the residuals.
For the residuals of local polar bias, the Kolmogorov Smirnov test with Lillefors correction confirms this assumption with $p = 0.4795$. For the normal distribution of the quadrant error, the significance level is $p = 0.029$. A significant Kolmogorov Smirnov test means that these residuals are not normally distributed. However, an analysis of variance with repeated measurements is robust against a violation of this assumption (Vasey and Thayer, 1987). An additional requirement is that the data must be verified for sphericity. "Sphericity refers to the equality of variances of the differences between treatment levels" (Field, 2007).

Table 4: Results of the Mauchly test to verify sphericity.

factor	error measure	χ^2	df	p
TF type	quadrant error	0,161	2	0,923
	local polar bias	3,775	2	0,151
Head Movement*	quadrant error	2,393	2	0,302
TF type	local polar bias	0,971	2	0,615
Geometry*	quadrant error	2,128	2	0,345
TF type	local polar bias	2,344	2	0,310

This condition has to be considered in repeated measures and is only required for factors with three or more levels. The Mauchly test is commonly used to examine sphericity. Sphericity is assumed when the significance of the test is greater than 0.05 ($p > 0.05$). The sphericity test was performed on the factor TF type, as this consists of three levels. In addition, it was also performed for all combinations of the other factors with the factor TF type. Table 4 shows the results of the Mauchly test for the factors described. Since all significance values are $p > 0.05$, sphericity can be assumed for the factor TF type, as well as for the interaction of the factor TF type with other factors.

4.4 Results

This Section presents all relevant statistical results. A detailed overview of the results of the statistical calculation can be found in the **Appendix**.

Descriptive Analysis

First, the quadrant error for each test person and condition will be visually examined in the following. Figure 4.12 shows the quadrant error in percent per participant and

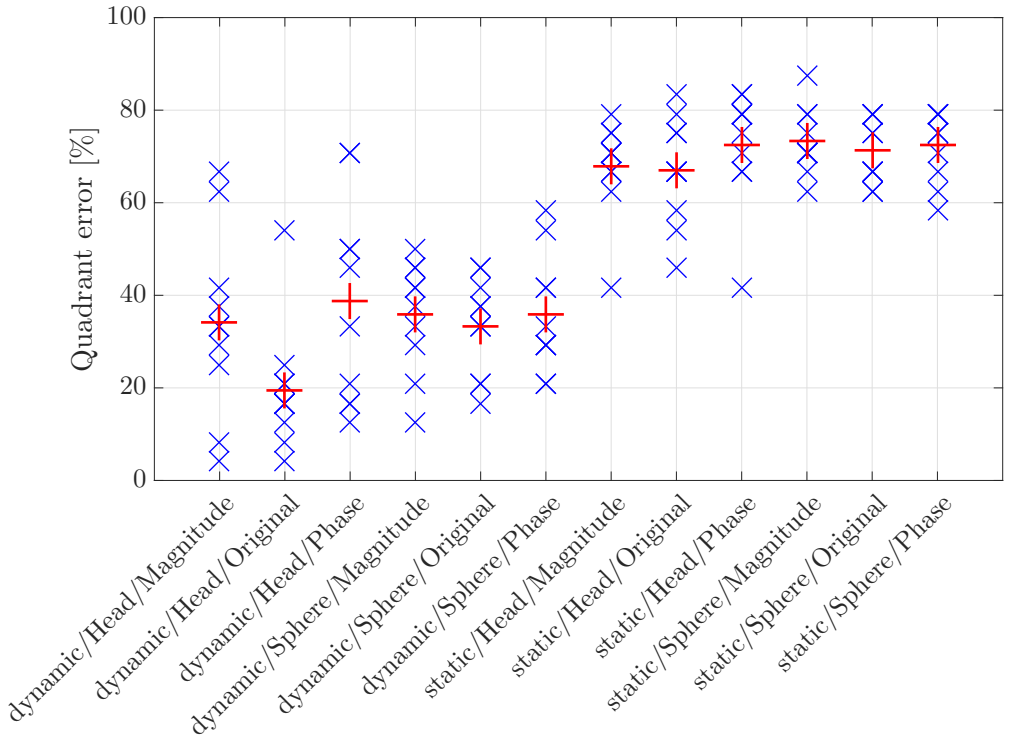


Figure 4.12: Calculated quadrant error per subject and condition (blue crosses) and the respective mean values (red crosses).

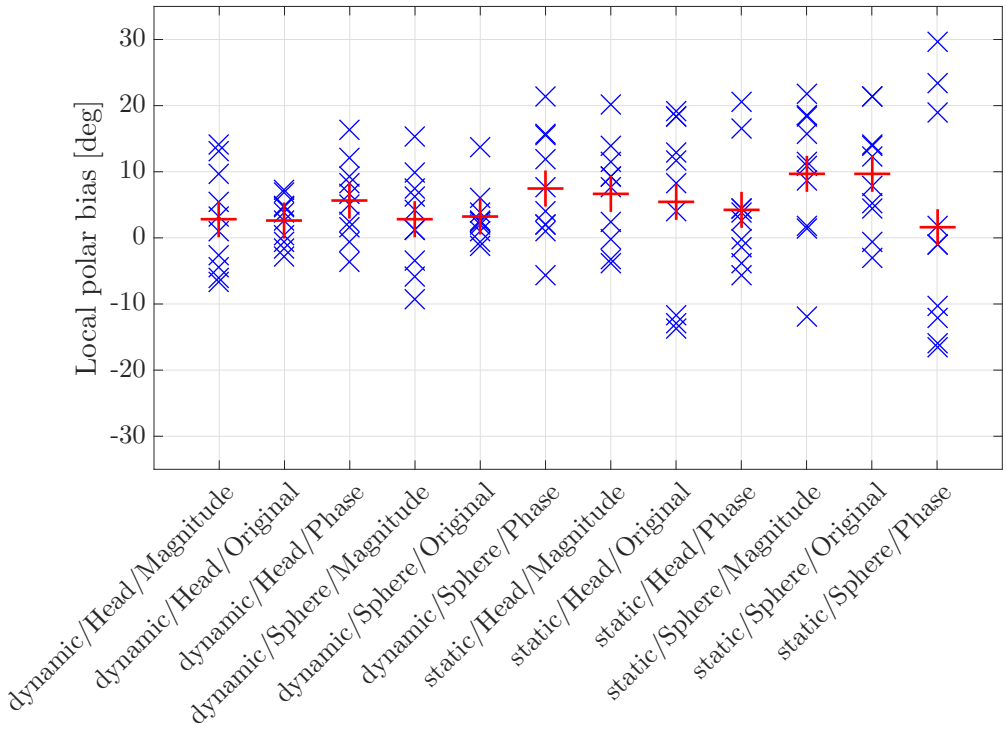


Figure 4.13: Calculated local polar bias per subject and condition (blue crosses) and the respective mean values (red crosses).

condition. Each blue cross represents the quadrant error made by a specific participant. The red crosses indicate the mean value over all test persons for a respective condition. The conditions change from left to right first by the TF type, then by the geometry and finally by the head movement.

What stands out most is that the mean values of the quadrant error are much lower in the dynamic case than in the static case. In addition, however, the scattering over the test participants is particularly high in the dynamic case (left 6 conditions), whereas in the static case (right 6 conditions) it is relatively low. Further, it can be noted that the lowest average quadrant error occurred when a participant used their original, individual HRTF in a dynamic localization scenario. The lowest quadrant error here is approximately 4 %, which is equivalent to one confusion in 24 trials. In total, all but one of the participants were at a maximum of 25 %.

While quadrant errors were lowest for the dynamic head original condition, most of the participants showed relatively low quadrants across all dynamic conditions.

Looking at the second dependent variable (see Figure 4.13), the local polar bias, at first glance none of the conditions appear to result in a comparatively lower or higher local polar bias. Initially, visual observation shows that all mean values of local polar bias are

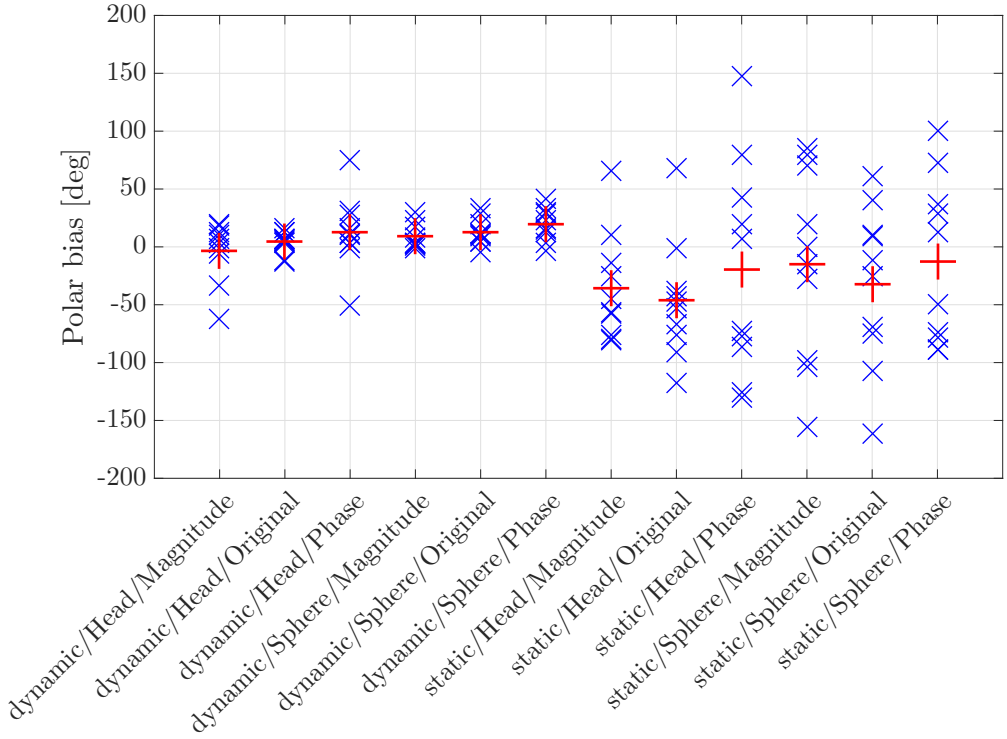


Figure 4.14: Calculated polar bias per subject and condition (blue crosses) and the respective mean values (red crosses).

in a range between 0° and 10°. This means that the sound source is perceived on average at a higher position, which is then indicated above the actual sound source.
The lowest scattering of the local polar bias across subjects is found in the dynamic case of the original individual HRTF (geometry of the head), as well as in the dynamic case of the original SHTF (geometry of the sphere).
If one compares those findings with the polar bias, which has not been corrected for the quadrant error, the division into dynamic and static can be recognized (see Figure 4.14). In this case, there is an overestimation of the angles in the dynamic case and an underestimation in the static case. The overestimation implies a perception of the sound source above the actual sound source, which means the perception is shifted to the upper pole. The underestimation of the sound source results in exactly the opposite case, namely a shift of the perceived sound source to the lower pole. Since both variables of the error measures still must be independent of each other when using MANOVA the local polar bias is used for further analysis.
The findings of this section were generated by a descriptive analysis of the available data. This does not allow for any generalization of these conclusions to a wider population outside the present sample. Thus, the in paragraph **Statistical Method** described

multivariate variance analysis and post hoc tests were carried out. Its results are presented in the subsequent section.

Multivariate Results

The three factorial repeated measures MANOVA was performed to consider the overall influence of the different factors on the error measures. Table 5 gives an overview of the test results using the Pilla's trace test. To describe the results, the following statistical measures apply: The F-ratio, which determines the limits of significance is a measure of the ratio of the variation explained by the model and the variation explained by unsystematic factors (Field, 2007). It is based on the F-distribution. The second statistical measure is the respective significance value (p) and the corresponding partial eta square (η_p^2), which indicates the effect size and is independent of the sample size. The effect is more pronounced the closer η_p^2 is to 1. The significance level here is $\alpha = 0.05$. If one first considers the three factors individually, the factors head movement ($F(2,8)=62.690$, $p<0.01$) and TF type ($F(4,36)=2.769$, $p=0.042$) have a significant influence on the error measures. This cannot be confirmed for the factor geometry ($F(2,8)=1.108$, $p=0.376$), thus it has no significant influence on the overall model (localization ability) containing two dependent variables. The significant influence on the localization ability can be corroborate with a large effect size of $\eta_p^2=0.940$.
In addition, the combination of geometry and TF type has a significant effect on the localization ability ($F(4,36)=2.740$, $p=0.044$). A marginal significant influence could also be observed with the combination of the conditions head movement and TF type ($F(4,36)=2.123$, $p=0.098$).

Table 5: Results of the multivariate analysis using Pillai's trace test for all factors and factor combinations. Showing the values for the F-ratio, the significance level p and the effect strength η_p^2.

Factor	F	p	η_p^2
Head Movement	62.69	<0.01	0.940
Geometry	1.108	0.376	0.217
TF Type	2.769	0.042	0.235
Head Movement * Geometry	0.041	0.960	0.010
Head Movement * TF Type	2.123	0.098	0.191
Geometry * TF Type	2.740	0.044	0.233

Initially, this analysis can uncover significant influences, but not yet a statement about the type of influence, the influence on the different error measures, and the difference between the levels of a factor can be given. Post hoc tests were carried out to provide further information.

Post Hoc Tests

The first post hoc test is an univariate analysis of variance. Here, the influence of the individual factors and their combinations on one dependent variable, was analyzed. The same statistical sizes (F-ratio, p and η_p^2) as in the multivariate analysis were used to describe the results.
All subsequent results are subject to a Bonferroni correction of $\alpha = 0.05$. Since the sphericity was already assumed in the assumption (see paragraph **Assumptions**), the results could be used uncorrected for interpretation.
First the variance analyses for the quadrant error are considered. Table 6 gives an overview of the influence of the individual factors and their combinations on this error measure.
If the three factors are examined individually, a significant influence of the head movement ($F(1,9)=138.761$, $p<0.01$) and the TF type ($F(2,18)=5.946$, $p=0.010$) on the quadrant error can be observed. As the only condition, the head movement has additionally a large effect size ($\eta_p^2=0.939$). In line with the MANOVA results, the univariate case of the quadrant error, the factor geometry has no significant influence ($F(1,9)=1.864$, $p=0.205$). The results displayed in Table 6 also show a significant interaction effect between the geometry and TF type ($F(2,18)=6.611$, $p<0.01$). No effect on the quadrant error could be observed for the interactions of head movement/geometry and head movement/TF type.
The same type of univariate analysis was performed for the second dependent variable, the

Table 6: Results of the univariate analysis of the quadrant error. Showing the values for the F-ratio, the significance level p and the effect size η_p^2.

Factor	F	p	η_p^2
Head Movement	138.761	<0.01	0.939
Geometry	1.864	0.205	0.172
TF Type	5.946	0.010	0.398
Head Movement*Geometry	0.042	0.841	0.005
Head Movement*TF Type	1.805	0.193	0.167
Geometry*TF Type	6.611	<0.01	0.423

Table 7: Results of the univariate analysis of the local polar bias. Showing the values for the F-ratio, the significance level p and the effect size η_p^2.

Factor	F	p	η_p^2
Head Movement	0.992	0.345	0.099
Geometry	0.626	0.449	0.065
TF Type	0.085	0.919	0.009
Head Movement*Geometry	0.026	0.874	0.003
Head Movement*TF Type	2.684	0.095	0.230
Geometry*TF Type	0.665	0.526	0.069

local polar bias. Table 7 gives an overview of the results. No significant main effects of the factors or interaction effects could be found. Only the interaction term head movement and TF type shows marginal significance ($F(2,18)=2.684$, $p=0.095$).
Furthermore, the influences of the different factor levels on the error measures were examined in a pairwise comparison. In this comparison, the respective levels of a factor are examined for differences. A significant result between two levels indicates whether there are different influences between the two factor levels on the error measure. This only needs to be done for factors that have been significantly tested in the univariate analysis. Table 8 shows these pairwise comparisons for the error measure quadrant error. As already confirmed in the univariate analysis, there is a significant difference between the two head movement conditions static and dynamic ($p<0.01$). The difference between the two mean values of the factors is -37,847 %. This means that with the dynamic level, the average value of quadrant errors is 37,857 % less than in the static case.
If one considers the pairwise comparisons of the TF types, one can observe that there are no significant differences between the two levels magnitude and original ($p=0.127$), and magnitude and phase ($p=1.000$). However, there is a significant dif-

Table 8: Results of the pairwise comparisons for the error measure quadrant error.

Factor	Level 1 (L_1)	Level 2 (L_2)	Mean difference [%] (L_1-L_2)	p
Head Movement	dynamic	static	-37.847	<0.01
	magnitude	original	5.000	0.127
TF type	magnitude	phase	-2.083	1.000
	orginal	phase	-7.083	0.018

ference between the TF types original and phase (p=0.018). The mean difference here is -7,083, so the quadrant error is on average significantly lower when using original transfer functions compared to transfer functions with phase information only. In the case of the local polar bias, no significant influence of the factors on this error measure was found during the univariate analysis. Accordingly, no pairwise comparison is made for the individual factors.

Since significant influences of the factor levels can only be described statistically for individual factors, the influences of the factor levels of the interaction terms are described by means of interactions on the basis of estimated marginal means. For this, the significant interaction terms in the univariate analysis are taken into account. In the case of the quadrant error, there is a significant influence of the interaction TF type and geometry. In the case of local polar bias, a marginally significant influence is present for the interaction between head movement and TF type.

Figure 4.15 shows the interaction between TF type and geometry with respect to the quadrant error. In this case the x-axis shows the type of geometry, the y-axis the estimated marginal means and the legend (respectively graphs) the three different TF types.

Figure 4.15 shows that there is hardly any difference between the TF types when looking at the geometry sphere. In contrast when considering the geometry head, participants have the lowest quadrant error in the original condition compared to magnitude and phase. The difference between the two geometries is the largest for the original TF type and quite

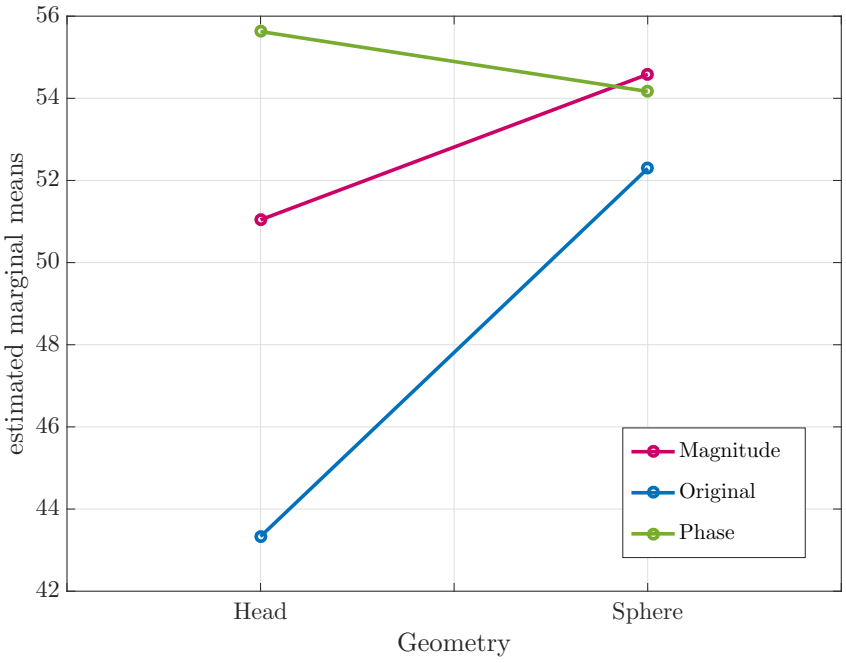


Figure 4.15: Interaction effects for the factors geometry and TF type on the quadrant error.

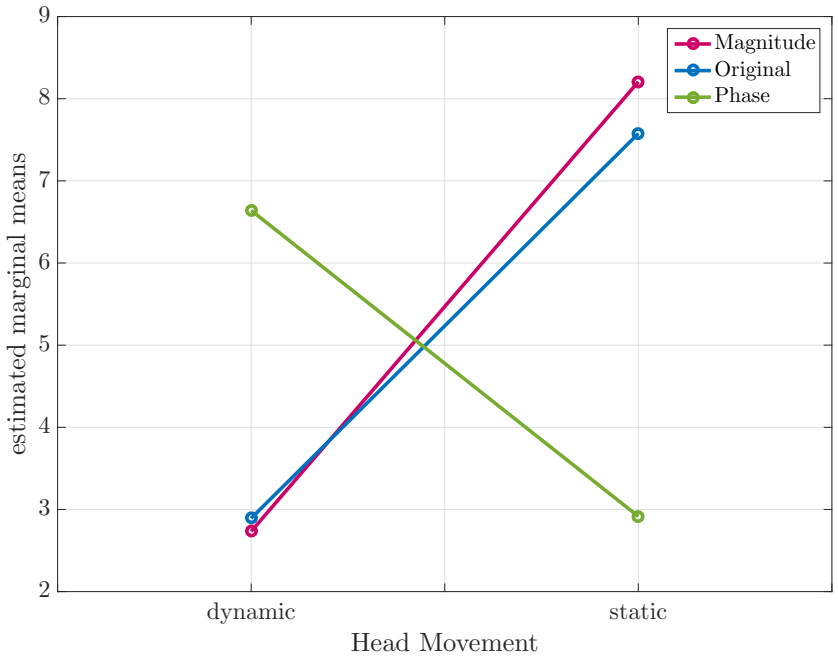


Figure 4.16: Interaction effects for the factors head movement and TF type on the local polar bias.

small for the other two TF types.

Next, the marginal means related to the second error measure local polar bias were considered. The interaction between the head movement and the TF type is illustrated in Figure 4.16.

Figure 4.16 shows the marginal means for different transfer functions depending on whether the localization scenario in the listening test was static or dynamic. In this case the x-axis shows the type of head movement, the y-axis the estimated marginal means and the legend (respectively graphs) the three different TF types.

Figure 4.16 shows that in the dynamic condition participants have higher error rates in the phase type compared to the magnitude and original type (where error rates are similar). In contrast in the static condition participants have lower error rates in the phase type compared to the magnitude and original type (where again error rates are similar).

4.5 Conclusion and Discussion

This section gives an interpretation and summary of all the results of the listening test. In addition, it critically examines possible sources of error in the test setup or procedure.

Summary of the Results

Comparable conditions (untrained subjects with individual HRTF in the dynamic case) show similar values for the quadrant error, as well as for the local polar bias in similar studies (Majdak et al., 2010; Jiang et al., 2018).
The multivariate statistical analysis first showed that the factors head movement and TF type have significant influence on the localization ability of the participants. This means that, on average, the error measures improve as soon as the subjects have the opportunity to move their head during the localization process. In addition, the choice of TF type plays a role in the localization ability.
Geometry as a single factor has no significant influence on the localization ability. Only in interaction between the TF type there is an influence.
However, as the univariate analysis revealed, these factors only have an influence on the quadrant error, but not on the local polar bias. Only a marginally significant effect could be observed in the interaction of the factors head movement and TF type for the local polar bias.
A possible reason for the improvement of the local polar bias with the TF type phase in the static case compared to the dynamic case can be due to the positive as well as the negative range of the error measure. The scattering in both directions is much higher in the static case, which on average a value closer to 0° equals. This indicates that the participants overestimate and underestimate source heights to the same extent in transfer functions that only contain phase information.
The use of a sphere instead of a head as geometry has no influence on the local polar bias. In relation to the quadrant error geometry does not show a significant main effect, but a marginal interaction effect with TF type can be detected. The interaction between the geometries and the TF types shows mainly a increase of the quadrant error in the interaction of the geometry sphere and the TF type original compared to the interaction of the geometry head and the TF type original. This shows that the quadrant error is lowest with the individual, full HRTF. This HRTF is known to the participant and is used to train the most efficient sound source localization possible, so that this result can be expected. Since there is only an interaction effect in both error measures, but no significant effect of the geometry on the localization ability, it can be assumed that the sound source localization in the median plane is not exclusively based on the use of the different cues generated by the pinnae.
The biggest influence on the quadrant error, however, is the head movement. The error rate improves on average by about 38% using motion cues.
The different TF types indicate the localization ability using ILD cues (magnitude),

ITD cues (phase) and the combination of both (original). The quadrant error worsens significantly when using transfer functions that contain only phase information compared to the original transfer functions that represent both cues. Using the type phase, the quadrant error deteriorates on average by about 7% compared to the TF type original. The type magnitude did not differ significantly from the other two types.

In summary it can be said that no changes could be found due to different geometries, TF types or head movements on the local polar bias, only marginal significant interactions between the head movement and the TF types (mainly with the TF type original). Thus, the main focus of the localization ability is on the quadrant error. A head movement (dynamic case) during the localization of the sound source can already considerably reduce the quadrant error. In addition, the quadrant error is reduced as soon as transfer functions have at least spectral components. The presence of the pinnae as a spectral filter is only conditionally necessary (Geometry head compared to sphere). In general, however, the individual original HRTF can best be used for localization. In this listening test the different transfer functions were not trained. A more frequent use of an HRTF/SHTF (as is of course the case with the individual HRTF) would lead to an additional improvement of the quadrant error (Majdak et al., 2010).

Possible Sources of Error

A possible source of error is the selected error measure local polar bias. This contains no absolute values and thus bears the danger that the evaluated mean values of the participants represent a supposedly better result. If participants would over- and underestimate the sound source to the same extent, a local polar bias of 0° would result on average. The results on the local polar bias are therefore not transferable to an absolute bias.
However, there are some factors related to the measurement setup that contribute to the distortion of the results.
If a sound source is coming from the back participants could simply swivel around on the swivel chair, which meant that the sound source would now be located at the front. The localization of a source located at the front is known to be easier and therefore this condition may produce a smaller localization error simply due to the increased localization of a source from the front. The listening test was designed to be as natural as possible. If, for example, a recording of the MTB procedure is used, it would also be possible to freely move one's head to localize.
A slight improvement of the quadrant error can be observed using the geometry sphere

compared to the geometry head in combination with the TF type phase. Spherical head radii have been calculated to match their ITD as closely as possible to their original ITD. The spectral components differ more from their personal HRTF. This could be a possible reason for the better localization with the TF type phase compared to the TF type magnitude in the case of the spherical head.
Some participants mentioned a missing externalization of the presented sound sources in the static listening scenarios. The subjects stated that they did not have the feeling that the sound source was in their head rather than external in the dynamic case. Externalization is more pronounced using the individual HRTF. This was only given in one of six conditions in the static case. Furthermore, a head movement significantly favors externalization (Brimijoin et al., 2013). In the static case, this aspect is therefore omitted. None of the participants had this feeling over the entire static conditions.
One difficulty in the case of static conditions was the structure of the experiment itself. As soon as the test persons pressed the start confirmation, the signal played immediately. Some participants stated, that they needed a couple of trials to focus on the short stimulus duration. They said, in the beginning their focus was on finding the confirmation button as the thump was not displayed in the VR environment. The stimulus therefore came too quickly and suddenly in relation to the stimulus length for some subjects. In addition, the controller produced a sensory as well as a slight acoustic click sound due to the mechanical construction of the button. According to some, this directed the attention of the test persons in a different direction. A delay of the signal, or to avoid all these disadvantages, a slightly longer stimulus duration is therefore suitable for similar experiments.
In addition, missing cues of the torso in transfer functions could have generally led to a perception of the sound source shifted to higher elevation angles.

5 Theoretical vs. Actual Localization Error

This chapter validates the theoretical findings from the physical evaluation in Chapter 3 with the observations from the listening test in Chapter 4. For this purpose, the theoretical localization error resulting from the use of spherical head models was determined for each subject as described in Section 3.2.3. Since the HRTF and the SHTF depends on the subject, the suggested localization mismatch is different for each participant. This theoretical mismatch is compared with the localization errors collected during the listening test. In the following, the localization results of the listening test for the original transfer function of a sphere and dynamically presented source scenarios are used for comparison with the theoretical model.

Figure 5.1 shows the theoretically estimated and the measured localization error of the participant with ID 31. The red line represents the theoretical mismatch on the basis of motion induced ILD cues, the blue line from motion induced ITD cues. In contrast, the localization errors obtained by the listening test for each trial are shown by red circles.As the figure shows, multiple measured errors (red circle) exist for certain angular sections in the elevation plane. This is caused by the mapping of the rear and front hemispheres ($\varphi = 0°$ and $\varphi = 180°$) to the elevation range from -90° to 90°. Additionally, due to the

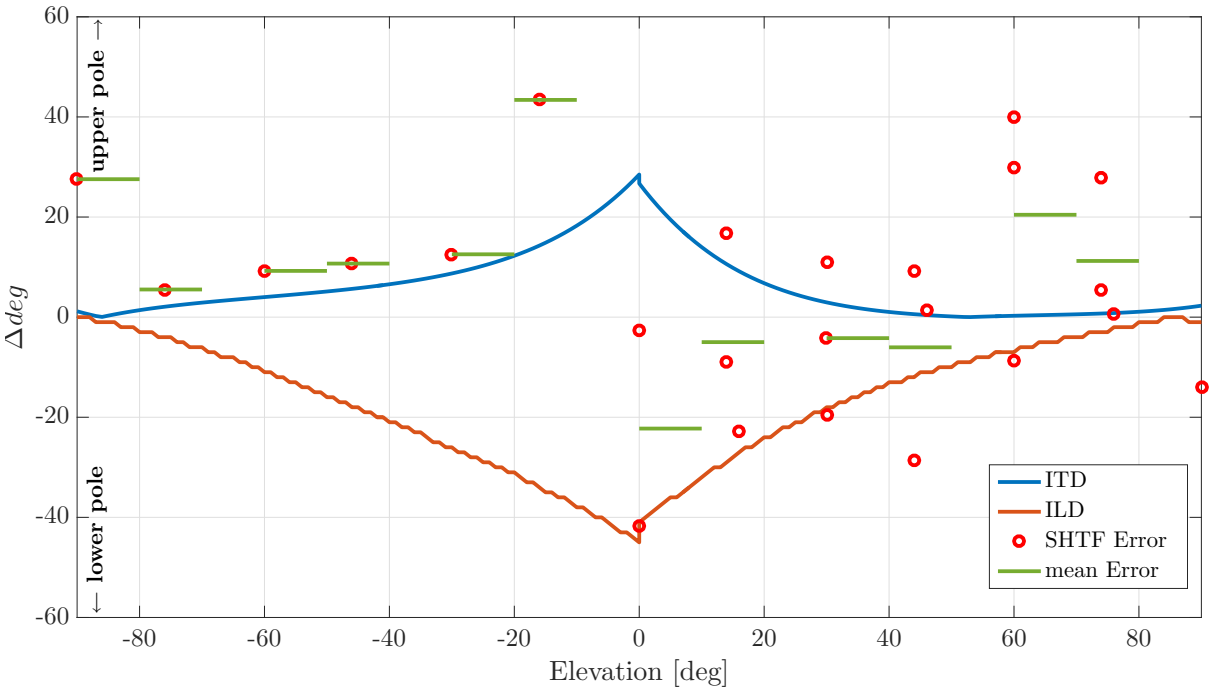


Figure 5.1: Estimated localization error as well as collected data for the condition original TF with geometry sphere in the dynamic case for subject number 31.

programming error mentioned in Chapter 4, up to three angles are in the (approximately) same range. Therefore, mean values were calculated in steps of 10° in the range of -90° to 90°. These are shown as green lines in Figure 5.1.

From the 24 localization tasks 13 mean values could be calculated. Using the example of participant 31 (Figure 5.1), a positive mean value could be calculated in 8 cases, a negative mean value in 5 cases. In most cases, the perceived location of the sound source of this subject shifts to higher elevations, which means a shift to the upper pole. This means that in most localization scenarios, a localization decision might be based on motion induced ITD cues. The mean localization error over the entire angular range is 6.87° for this participant. For the lower elevation range (ϑ = -90° to ϑ = 0°) the calculated difference applies very well to the theoretical mismatch.

In the following, results of all participants are investigated. The number of participants who made a localization decision on the basis of motion-induced ITD cues in the majority of trials is 8 out of 10 participants (see Table 9). On average, 8 out of 13 blocks (13 mean values per participant) are decided in favor of ITD cues. This is congruent with Wightman and Kistler's (1992) finding that in the case of opposing cues, the decision is made primarily after ITD cues.

The opposite shift of the theoretical localization mismatch to the poles is considered separately for ITD and ILD in the following. The different conditions of the transfer functions, phase (ITD) and magnitude (ILD), each represent one of the two cues. For the SHTF condition with only phase information, representing only ITD cues, 7 out of 10 participants opted for elevation angles above the actual angle in the majority of blocks (average in about 9 out of 13 blocks). This equals a decision by ITD cues based on theoretical considerations. In the case of the SHTF condition with only magnitude, however, the participants also opted for higher elevation angles than the actual sound source in a ratio of 7 to 3 participants (here on average in about 8 of 13 blocks), even if only ILD cues are present.

Table 9 summarizes previous observations, whereby *decision based on ITD* or *decision based on ILD* indicates the number of participants who have decided in favor of one of these cues. Further, *overall mean mismatch* is the averaged mismatch over all 10 participants, regardless on which cues the decision was based. Finally, *mean amount of blocks when decided on ITD* describes the averaged number of bocks indicating that a decision was based on ITD cues. While *mean amount of blocks when decided on ILD* stands for the averaged number of blocks that led to a decision based on ILD cues and thus a shift to lower elevation angles.

A general observation of the previous comparison is that sound source locations were mostly assessed higher by the participants of the listening test than the actual source

Table 9: Summary of localization decisions for all participants.

	SHTF original	SHTF phase	SHTF magnitude
Decision based on ITD	8	7	7
Decision based on ILD	2	3	3
overall mean mismatch	5.04°	4.24°	4.4°
mean amount blocks when decided on ITD	8	8	8
mean amount blocks when decided on ILD	7	7	8

positions. One reason might be that none of the transfer functions had characteristics of the torso included. Usually, reflections at the torso introducing further localization cues, especially in the low frequency range. This preferably applies when sound sources are located in the upper range of the median plane (Algazi et al., 2001a). However, since this work is based on considerations of the Motion Tracked Binaural method and other spherical head model applications which do not contain a torso, all transfer functions were used without any influence of the torso. Nevertheless, this might have led to a falsification of the personally learned localization cues. Another possibility is that participants generally overestimate sound sources in the median plane more frequently without referring to any different cues.

All findings of this chapter are based on the assumption that each participant with their individual HRTF can locate without error by means of head movement. However, this is almost impossible in practice. Especially with high elevations, a polar bias occurs even with individual HRTF cues (see Section 2.2.2). The actual localization error would need to be corrected by this bias to get an unbiased representation. However, this would require a large number of repeated measurements to get the true localization mismatch. Therefore, only a first tendency of decisions based on dynamic ITD cues has been observed. For further evaluation, another localization listening test is needed which deals with this question in more detail. This would require repeated trials in all cases. Furthermore, the head movement should be restricted to movements in the horizontal plane, since only those are included in the calculation of the theoretical mismatch. Otherwise, the results could be affected by roll movements that are not included in the current model.

6 Summary and Conclusion

The theoretical evaluation in Chapter 3 concluded that the use of off-set ears is advantageous for spherical head models, or applications where a sphere is used as a replacement for a head. Adapting the spherical head models results in additional cues that can lead to a minimization of the quadrant error (up to 3 dB difference between behind the head and above the head).
In addition, Section 3.2.2 outlines that under static conditions with a spherical head only minor cues are present for sound source localization.
However, since dynamic cues usually play an important role in applications where SHTF are used, these have been analyzed in detail in Section 3.2.3. Dynamic cues result from slight changes in binaural cues due to head movement. Therefore, a model was introduced to compare differences in localization between human dynamic cues and those of other geometries. It turned out that in comparison to the individual, familiar cues, the use of SHTF results in opposing dynamic ITD and ILD cues. This means that using the dynamic ITD cues, the perceived sound source is shifted to the upper pole, whereas evaluating the dynamic ILD cues, the perceived sound source is shifted to the lower pole. According to Wightman and Kistler (1992), participants opt for information based on ITD cues in the case of opposing localization cues. The extent of the deviation between individual dynamic cues and dynamic cues from geometries depends on the individual HRTFs and the dimensions of the geometry.
Since a sphere only resembles a human head to a limited extent, the localization cues of an ellipsoid were also considered. These were compared to the cues of the spherical head. It turned out that in the static case spectral cues are somewhat more pronounced with an ellipsoidal head. They presumably exert a further improvement of the localization ability especially through reducing quadrant error.
The investigation of the dynamic ILD cues found that using an ellipsoidal head does not result in an advantage compared to using spherical heads. However, when subjects use dynamic ITD cues, the perceived sound source is 10° to 20° lower (with a maximum deviation of 10° from individual HRTF) with a ellipsoidal head compared to when subjects use a spherical head.
Using TF of an ellipsoid instead of TF of a sphere as a substitute for individual transfer functions can therefore be advantageous.

A listening test was performed to analyze the differences between dynamic and static cues, as well as monaural and binaural cues for individual heads and individualized spherical heads. The listening test revealed that neither the geometry (head and sphere), nor the TF type (original, magnitude (ITD) and phase (ILD)), nor the head movement (dynamic

and static) had an influence on the local polar bias (polar accuracy). Only a marginally interaction between the TF type and the head movement could be determined for this error measure. In this case it is noticeable that the TF type phase in the static case has just as good values as the TF types magnitude and original in the dynamic case.
Nevertheless, the main focus of the evaluation is on the quadrant error. This can be influenced by using different TF types. It was found that the quadrant error is significantly higher using the TF type phase compared to the original. However, the greatest influence on this error measure results from the use of head movement. In the dynamic case, the quadrant error is on average 38% lower than the error in the static case. An influence of the geometry cannot be determined. Only a marginally significant one in the interaction with the TF type. In the case of the geometry head with the TF type original there was a difference of about 7% compared to the TF type original in combination with the geometry sphere. For the other two TF types, the difference between the geometries was only about 2% and 3%, respectively.

The approach of using a VR environment for localization listening testing is relatively new. Few studies with traditional consumer products such as the Oculus Rift and Unity have been conducted so far. In order to be able to use those products, some considerations have been made. The accuracy of head tracking for such a listening test has been investigated (see Paragraph **Spatial Accuracy of Headtracking Device**). It could be confirmed that the deviations are small and thus an accurate reproduction of the binaural synthesis is guaranteed. For all participants, the experience of performing an experiment in a virtual reality was very exciting. A single disadvantage is that in the case of the Oculus Rift no reset of the tracking position in the elevation can be performed. For this reason, the participants had to perform a manual reset by navigating the head into a certain area (see Section 4.2).

The evaluation of the theoretical model for dynamic localization cues with data from the listening test showed that in most cases the sound sources were perceived in higher elevations. However, this perception also occurred using the TF type Magnitude, where no ITD cues were present. It is still unclear whether this was due to the missing body parts torso and shoulder or whether it is due to the choice of cues. In order to validate this model, a further listening test is needed which is specifically aimed at this question. Since the localization of a sound source with individual HRTF is not at 0° perceived sound source to actual sound source, a general statement cannot be made with a non-repeat of the different trials. Thus an listening test with several repetitions for each condition needs to be performed.

In conclusion, the theoretical and empirical work suggests that head movements exert

the greatest influence on the accuracy of sound source localization. The ability to move the head freely significantly reduces the quadrant error. No significant improvement or deterioration of the local polar bias can be observed under any condition. Likewise, the geometry has no significant influence on both error measures. Therefore, with the inclusion of head movements, sound source localization in the median plane using SHTF is almost as accurate as using individual HRTF.

6.1 Future Work

Future work should further investigate the estimated localization error model presented in Chapter 3 and compared with listening test data in Chapter 5 can be further investigated. In order to be able to fully verify it, a listening test with several repetitions per trial should be carried out that is specially adapted to the model. This is needed to subtract the localization error with individual HRTF (which ideally should be 0°, but in reality has a larger bias) from the localization error of the geometry. This will allow to obtain a real mismatch.

Another approach for future studies is to evaluate the collected tracking data during the listening test. In dynamic cues studies, the effect of head movement was primarily addressed, but not the head movement as such (Algazi et al., 2001a; Jiang et al., 2018; Carlile, 2014). The following questions should be addressed through further research:

- Which head movement do the participants perform (only left/right or also lateral or up and down movements)?
- Are there differences in the number of movements per trial between subjects?
- Are there differences in the size of the head movement between subjects?
- Is it more accurate to localize with a certain type of head movement?

Since the tracking data of head movement is stored 60 times per second and for 5 seconds per trial, this is a large amount of data per trial. In order to be able to evaluate this data, a schema must be found to extract, for example, reversal points (of the head movement), as well as an adequate representation type for this kind of data.

Since a VR environment has proven to be a very good tool for listening tests related to sound source localization, it would be beneficial to create a sophisticated version for subsequent listening tests. This includes the integration of the automatic handling of

different HRTFs. In addition, it is useful to create a higher-level structure in which all processes are automated and no adjustment by the test leader is necessary during the listening test.

Furthermore, the influence of the reflection pattern of reverberant environments on the perceived source position could also be an interesting field of research in the context of spherical head models.

Bibliography

AES Standards Comittee (2015): *AES69-2015: AES standard for file exchange - Spatial acoustic data file format.* Audio Engineering Society, Inc.

Algazi, V. Ralph; Carlos Alvendano; and Richard O. Duda (2001a): "Elevation localization and head-related transfer function analysis at low frequencies." In: *The Journal of the Acoustical Society of America*, **109**(3), pp. 1110–1122.

Algazi, V. Ralph; Carlos Avendano; and Richard O. Duda (1999): "Low-frequency ILD elevation cues." In: *The Journal of the Acoustical Society of America*, **106**(4), pp. 2237–2237.

Algazi, V. Ralph; Carlos Avendano; and Richard O. Duda (2001b): "Estimation of a Spherical-Head Model from Anthropometry." In: *Journal of the Audio Engineering Society*, **49**(6), pp. 472–479.

Algazi, V. Ralph and Richard O. Duda (2002): "Approximating the head-related transfer function using simple geometric models of the head and torso." In: *The Journal of the Acoustical Society of America*, **112**(5), pp. 2053–2064.

Algazi, V. Ralph; Richard O. Duda; and Dennis M. Thompson (2004): "Motion-Tracked Binaural Sound." In: *The Journal of the Acoustical Society of America*, **52**(11), pp. 1142–1156.

Andreopoulou, Areti and Brian F. G. Katz (2017): "Identification of perceptually relevant methods of inter-aural time difference estimation." In: *The Journal of the Acoustical Society of America*, **142**(2), pp. 588–598.

Baumgartner, Robert; Piotr Majdak; and Bernhard Laback (2014): "Modeling sound-source localization in sagittal planes for human listeners." In: *The Journal of the Acoustical Society of America*, **136**(2), pp. 791–802.

Benichoux, Victor; Marc Rébillat; and Romain Brette (2016): "On the variation of interaural time differences with frequency." In: *The Journal of the Acoustical Society of America*, **139**(4), pp. 1810–1821.

Bernschütz, Benjamin (2016): *Microphone arrays and sound field decomposition for dynamic binaural recording.* Ph.D. thesis, Technische Universität Berlin.

Blauert, Jens (1997): *Spatial Hearing: The Psychophysics of Human Sound Localization.* Cambridge, MA, USA: MIT Press.

Bomhardt, Ramona and Janina Fels (2014): "Analytical interaural time difference model for the individualization of arbitrary Head-Related Impulse Responses." In: *137th Audio Engineering Society Convention Convention, Los Angeles, USA*. p. Paper 9131.

Bomhardt, Ramona; Marcia Lins; and Janina Fels (2016b): "Analytical Ellipsoidal Model of

Interaural Time Differences for the Individualization of Head-Related Impulse Responses." In: *Journal of the Audio Engineering Society*, **64**(11), pp. 882–894.

Bortz, Jürgen (2005): *Statistik für Human- und Sozialwissenschaftler*. 6. vollständig überarbeitete und aktualisierte Auflage. Heidelberg: Springer.

Brimijoin, W. Owen; Alan W. Boyd; and Michael A. Akeroyd (2013): "The Contribution of Head Movement to the Externalization and Internalization of Sounds." In: *PLOS ONE*, **8**(12).

Brinkmann, Fabian; Alexander Lindau; and Stefan Weinzierl (2017b): "On the authenticity of individual dynamic binaural synthesis." In: *The Journal of the Acoustical Society of America*, **142**(4), pp. 1784–1795.

Brinkmann, Fabian and Stefan Weinzierl (2017a): "AKtools - An open software toolbox for signal acquisition, processing, and inspection in acoustics." In: *142nd Audio Engineering Society Convention Convention, Berlin, Germany*. pp. e–Brief 309.

Brinkmann, Fabian; et al. (2017c): "A High Resolution and Full-Spherical Head-Related Transfer Function Database for Different Head-Above-Torso Orientations." In: *Journal of the Audio Engineering Society*, **65**(10), pp. 841–848.

Bronkhorst, Adelbert W. (1995): "Localization of real and virtual sound sources." In: *The Journal of the Acoustical Society of America*, **98**(5), pp. 2542–2553.

Burkhard, Mahlon D. and Richard M. Sachs (1975): "Anthropometric manikin for acoustic research." In: *The Journal of the Acoustical Society of America*, **58**(1), pp. 214–222.

Carlile, Simon (1996): *The Physical and Psychophysical Basis of Sound Localization*. Berlin, Heidelberg: Springer.

Carlile, Simon (2014): "The plastic ear and perceptual relearning in auditory spatial perception." In: *Frontiers in Neuroscience*, **8**, pp. 237–250.

Cuevas-Rodríguez, María; et al. (2019): "3D Tune-In Toolkit: An open-source library for real-time binaural spatialisation." In: *PLOS ONE*, **14**, pp. 1–37.

Damaske, Peter and Bernd Wagener (1969): "Directional Hearing Tests by the Aid of an artificial Head." In: *Acustica*, **21**, pp. 30–35.

Dellepiane, Matteo; Nico Pietroni; Nicolas Tsingos; Manuel Asselot; and Roberto Scopigno (2008): "Reconstructing head models from photographs for individualized 3D-audio processing." In: *Computer Graphics Forum*, **27**(7), pp. 1719–1727.

DIN33402-2 (2005): *DIN 33402-2: Ergonomie - Körpermaße des Menschen - Teil 2: Werte*. Berlin: Beuth.

Dinakaran, Manoj; Peter Grosche; Fabian Brinkmann; and Stefan Weinzierl (2016): "Extraction of

Anthropometric Measures from 3D-Meshes for the Individualization of Head-Related Transfer Functions." In: *140th Audio Engineering Society Convention, Paris, France*. p. 9579.

Duda, Richard O.; Carlos Avendaño; and V. Ralph Algazi (1999): "An adaptable ellipsoidal head model for the interaural time difference." In: *IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99*, **2**, pp. 965–968.

Duda, Richard O. and William L. Martens (1998): "Range dependence of the response of a spherical head model." In: *The Journal of the Acoustical Society of America*, **104**(5), pp. 3048–3058.

Fiedler, Felicitas; David Ackermann; Fabian Brinkmann; Stefan Weinzierl; and Martin Schneider (2017): "Development and evaluation of a microphone array for recording spatial sound fields according to the motion-tracked binaural (MTB) algorithm (in German)." In: *Fortschritte der Akustik - DAGA 2017, Kiel, Germany*. pp. 1115–1117.

Field, Andy (2007): *Discovering Statistics Using SPSS*. Introducing Statistical Methods Series. SAGE Publications.

Flavell, Lance (2010): *Beginning Blender: Open Source 3D Modeling, Animation, and Game Design*. Apresspod Series. Apress.

Gardner, Mark B. and Robert S. Gardner (1973): "Problem of localization in the median plane: effect of pinnae cavity occlusion." In: *The Journal of the Acoustical Society of America*, **53**(2), pp. 400–408.

Gardner, William Grant (1997): "Head tracked 3-D audio using loudspeakers." In: *Proceedings of 1997 Workshop on Applications of Signal Processing to Audio and Acoustics*. pp. 4–.

Geuzaine, Christophe and Jean-François Remacle (2009): "Gmsh: a three-dimensional finite element mesh generator with built-in pre- and post-processing facilities." In: *International Journal for Numerical Methods in Engineering*, **79**(11), pp. 1309–1331.

Gulick, W. Lawrence; George A. Gescheider; and Robert D. Frisina (1989): *Hearing: Physiological Acoustics, Neural Coding, and Psychoacoustics*. Oxford University Press.

Hall, Joseph L. (1968): "Maximum-Likelihood Sequential Procedure for Estimation of Psychometric Functions." In: *The Journal of the Acoustical Society of America*, **44**(1), pp. 370–370.

Hartley, Ralph V. L. and Thornton C. Fry (1921): "The Binaural Location of Pure Tones." In: *Physical Review*, **18**, pp. 431–442.

Hebrank, Jack and David Wright (1974): "Are two ears necessary for localization of sound sources on the median plane?" In: *The Journal of the Acoustical Society of America*, **56**(3), pp. 935–938.

Hershkowitz, Ronald M. and Nathaniel I. Durlach (1969): "Interaural Time and Amplitude

jnds for a 500-Hz Tone." In: *The Journal of the Acoustical Society of America*, **46**(6B), pp. 1464–1467.

Irvine, Dexter R. F. (1992): "Physiology of the auditory brainstem." In: Arthur N. Popper and Richard R. Fay (Eds.) *The Mammalian auditory pathway.*, vol. 1, chap. 4. New York: Springer, pp. 153–231.

Jiang, Jianliang; et al. (2018): "The Effects of Dynamic Cue and Spectral Cue on Auditory Vertical Localization." In: *Audio Engineering Society Conference: International Conference on Spatial Reproduction - Aesthetics and Science, Tokyo, Japan.*

Jongkees, Leonard B. W. and R. A. Van der Veer (1958): "On Directional sound Localization in Unilateral Deafness and its Explanation." In: *Acta Oto-Laryngologica*, **49**(1), pp. 119–131.

Kaneko, Shoken; Tsukasa Suenaga; and Satoshi Sekine (2016): "DeepEarNet: Individualizing Spatial Audio with Photography, Ear Shape Modeling, and Neural Networks." In: *Audio Engineering Society Conference: International Conference on Audio for Virtual and Augmented Reality, Los Angeles, USA.*

Katz, Brian F. G. (2001a): "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation." In: *The Journal of the Acoustical Society of America*, **110**(5), pp. 2440–2448.

Katz, Brian F. G. (2001b): "Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparisons to real measurements." In: *The Journal of the Acoustical Society of America*, **110**(5), pp. 2449–2455.

Klensch, Herbert (1948): "Beitrag zur Frage der Lokalisation des Schalles im Raum." In: *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere*, **250**(4), pp. 492–500.

Koenig, Walter Jr. (1950): "Subjective Effects in Binaural Hearing." In: *The Journal of the Acoustical Society of America*, **22**(1), pp. 61–62.

Letowski, Tomasz and Szymon Letowski (2011): "Localization error: Accuracy and precision of auditory localization." In: Pawel Strumillo (Ed.) *Advances in Sound Localization*, vol. 1, chap. 4. Croatia: InTech, pp. 55–87.

Lindau, Alexander and Stefan Weinzierl (2007): "FABIAN - Schnelle Erfassung binauraler Raumimpulsantworten in mehreren Freiheitsgraden." In: *Fortschritte der Akustik - DAGA 2007, Stuttgart, Germany.* pp. 633–634.

Lins, Marsia; Ramona Bomhardt; and Janina Fels (2016): "Individualisierung der HRTF: Ein Ellipsoidmodell zur Anpassung von interauralen Pegeldifferenzen." In: *Fortschritte der Akustik - DAGA 2016, Aachen, Germany.* pp. 78–80.

Macpherson, Ewan A. and John C. Middlebrooks (2002): "Listener weighting of cues for lateral

angle: The duplex theory of sound localization revisited." In: *The Journal of the Acoustical Society of America*, **111**(5), pp. 2219–2236.

Majdak, Piotr; Matthew J. Goupell; and Bernhard Laback (2010): "3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training." In: *Attention, Perception, & Psychophysics*, **72**(2), pp. 454–469.

Majdak, Piotr; et al. (2013): "Spatially Oriented Format for Acoustics: A Data Exchange Format Representing Head-Related Transfer Functions." In: *134th Audio Engineering Society Convention, Rome, Italy.*

McAnally, Ken I. and Russell L. Martin (2014): "Sound localization with head movement: implications for 3-d audio displays." In: *Frontiers in Neuroscience*, **8**, p. 210.

Middlebrooks, John C. (1999): "Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency." In: *The Journal of the Acoustical Society of America*, **106**(3), pp. 1493–1510.

Mills, A. William (1972): "Auditory localization." In: Jerry V. Tobias (Ed.) *Foundations of modern auditory theory.*, vol. 4. New York: Academic Press, pp. 301–345.

Møller, Henrik (1992): "Fundamentals of binaural technology." In: *Applied Acoustics*, **36**, pp. 171–218.

Montello, Daniel R.; Anthony E. Richardson; Mary Hegarty; and Michael Provenza (1999): "A Comparison of Methods for Estimating Directions in Egocentric Space." In: *Perception*, **28**, pp. 981–1000.

Morse, Philip M. and K. Uno Ingard (1971): *Theoretical Acoustics.* McGraw-Hill.

Oldfield, Simon R. and Simon P. A. Parker (1986): "Acuity of Sound Localisation: A Topography of Auditory Space. III. Monaural Hearing Conditions." In: *Perception*, **15**(1), pp. 67–81.

Pelzer, Robert (2018): *Perceptually motivated analysis of head-related transfer functions for individualization.* Master's thesis, Technische Universität, Fakultät 1, Fachgebiet Audiokommunikation und -technologie, Berlin.

Rayleigh, John William Strutt Baron (1894): *The Theory of Sound.* No. 1 in The Theory of Sound. Macmillan.

Redon, Christine and Laurette Hay (2005): "Role of visual context and oculomotor conditions in pointing accuracy." In: *NeuroReport*, **16**(18), pp. 2065–2067.

Seeber, Bernhard (1997): "A New Method for Localization Studies." In: *Acta Acustica united with Acustica*, **83**, pp. 446–450.

Shaw, Edgar A. G. (1974): "The External Ear." In: Wolf D. Keidel and William D. Neff (Eds.)

Auditory System: Anatomy Physiology (Ear). Berlin, Heidelberg: Springer, pp. 455–490.

Shaw, Edgar A. G. and Ryunen Teranishi (1968): "Sound Pressure Generated in an External-Ear Replica and Real Human Ears by a Nearby Point Source." In: *The Journal of the Acoustical Society of America*, **44**(1), pp. 240–249.

Søndergaard, Peter and Piotr Majdak (2013): "The Auditory Modeling Toolbox." In: Jens Blauert (Ed.) *The Technology of Binaural Listening, Modern Acoustics and Signal Processing*, vol. 1, chap. 2. Berlin, Heidelberg: Springer, pp. 33–56.

Thurlow, Willard R.; John W. Mangels; and Philip S. Runge (1967): "Head Movements During Sound Localization." In: *The Journal of the Acoustical Society of America*, **42**(2), pp. 489–493.

Vasey, Michael W. and Julian F. Thayer (1987): "The Continuing Problem of False Positives in Repeated Measures ANOVA in Psychophysiology: A Multivariate Solution." In: *Psychophysiology*, **24**(4), pp. 479–486.

Vliegen, Joyce and John Van Opstal (2004): "The influence of duration and level on human sound localization." In: *The Journal of the Acoustical Society of America*, **115**, pp. 1705–1713.

Weinzierl, Stefan (2008): *Handbuch der Audiotechnik*. VDI-Buch. Berlin, Heidelberg: Springer.

Wettschureck, Rüdiger (1970): "Über Unterschiedsschwellen beim Richtungshören in der Medianebene." In: *Gemeinschaftstagung für Akustik und Schwingungstechnik, Berlin, Germany*. Düsseldorf: VDI-Verlag, pp. 385–388.

Wightman, Frederic L. and Doris J. Kistler (1992): "The dominant role of low-frequency interaural time differences in sound localisation." In: *The Journal of the Acoustical Society of America*, **91**(3), pp. 1648–1661.

Winter, Fiete; Hagen Wierstorf; and Sascha Spors (2017): "Improvement of the Reporting Method for Closed-Loop Human Localization Experiments." In: *142th Audio Engineering Society Convention, Berlin, Germany*.

Wright, Duncan; John H. Hebrank; and Barbara M. Wilson (1974): "Pinna reflections as cues for localization." In: *The Journal of the Acoustical Society of America*, **56 3**, pp. 957–962.

Xie, Bosun (2013): *Head-Related Transfer Function and Virtual Auditory Display: Second Edition*. Plantation, FL, USA: J. Ross Publishing.

Xu, Xu; Karen B. Chen; Jia-Hua Lin; and Robert G. Radwin (2015): "The accuracy of the Oculus Rift virtual reality head-mounted display during cervical spine mobility measurement." In: *Journal of biomechanics*, **48 4**, pp. 721–724.

Ziegelwanger, Harald (2012): *Modell zur effizienten Kodierung von Signallaufzeiten für die binaurale Wiedergabe virtueller Schallquellen*. Master's thesis, Universität für Musik und darstellende Kunst Graz, Institut für Elektronische Musik und Akustik, Wien.

Ziegelwanger, Harald; Wolfgang Kreuzer; and Piotr Majdak (2015b): "MESH2HRTF: An open-source software package for the numerical calculation of head-related transfer functions." In: *Proceedings of the 22nd International Congress on Sound and Vibration, Florence, Italy*.

Ziegelwanger, Harald; Wolfgang Kreuzer; and Piotr Majdak (2016): "A-priori mesh grading for the numerical calculation of the head-related transfer functions." In: *Applied Acoustics*, **114**, pp. 99–110.

Ziegelwanger, Harald; Piotr Majdak; and Wolfgang Kreuzer (2015a): "Numerical calculation of listener-specific head-related transfer functions and sound localization: Microphone model and mesh discretization." In: *The Journal of the Acoustical Society of America*, **138**(1), pp. 208–222.

Ziegelwanger, Harald; Andreas Reichinger; and Piotr Majdak (2013): "Calculation of listener-specific head-related transfer functions: Effect of mesh quality." In: *Proceedings of Meetings on Acoustics*, **19**(1), p. 050017.

Ziegelwanger, Wolfgang and Piotr Majdak (2014): "Modeling the direction-continuous time-of-arrival in head-related transfer functions." In: *The Journal of the Acoustical Society of America*, **135**(3), pp. 1278–1293.

Zwislocki, Jozef J. and Alan S. Feldman (1956): "Just Noticeable Differences in Dichotic Phase." In: *The Journal of the Acoustical Society of America*, **28**(5), pp. 860–864.

List of Figures

List of Tables

Appendix

Table A.1: Anthropometric data from DIN33402-2 (2005).

	percentil	male	female
head depth	50 %	19.5 cm	18.5 cm
	95 %	20.5 cm	19.5 cm
head height	50 %	22.0 cm	21.0 cm
	95 %	23.5 cm	23.5 cm
head width	50 %	15.5 cm	15.0 cm
	95 %	16.5 cm	16.0 cm

Table A.2: Calculated Radii (calculated with Equation 10) for anthropometric data from DIN33402-2 (2005)

	Percentile	$\tilde{r}$
female	50 %	8.51 cm
	95 %	9.17 cm
male	50 %	8.87 cm
	95 %	9.41 cm
KEMAR	w=15.2 cm, d=19.1 cm, h=22,4 cm	8.78 cm

Table A.3: Descriptive statistics quadrant error. Mean value (M), standard deviation (std) and number of participants (N) for all conditions with respect to the quadrant error.

Condition	M	std	N
dynamic/Head/Magnitude	34.17	20.01	10
dynamic/Head/Original	19.58	13.62	10
dynamic/Head/Phase	38.75	22.05	10
dynamic/Sphere/Magnitude	35.83	11.98	10
dynamic/Sphere/Original	33.33	10.58	10
dynamic/Sphere/Phase	35.83	12.91	10
static/Head/Magnitude	67.92	10.40	10
static/Head/Original	67.08	11.69	10
static/Head/Phase	72.50	12.45	10
static/Sphere/Magnitude	73.33	7.14	10
static/Sphere/Original	71.25	6.93	10
static/Sphere/Phase	72.50	7.66	10

Table A.4: Descriptive statistics quadrant error. Mean value (M), standard deviation (std) and number of participants (N) for all conditions with respect to the local polar bias.

Condition	M	std	N
dynamic/Head/Magnitude	2.73	7.77	10
dynamic/Head/Original	2.54	3.55	10
dynamic/Head/Phase	5.73	6.07	10
dynamic/Sphere/Magnitude	2.74	7.52	10
dynamic/Sphere/Original	3.24	4.23	10
dynamic/Sphere/Phase	7.55	8.41	10
static/Head/Magnitude	6.72	7.82	10
static/Head/Original	5.45	13.43	10
static/Head/Phase	4.16	8.43	10
static/Sphere/Magnitude	9.70	10.20	10
static/Sphere/Original	9.70	8.40	10
static/Sphere/Phase	1.67	16.83	10

Table A.5: Mauchly test of sphericity for both error measures, including possible correction methods.

		Mauchly-W	Approx. χ^2	df	p	Greenhouse-Geisser	Huynh-Feldt	lower limit
Head Movement	quadrant error	1.000	0.000	0		1.000	1.000	1.000
	local polar bias	1.000	0.000	0		1.000	1.000	1.000
Geometry	quadrant error	1.000	0.000	0		1.000	1.000	1.000
	local polar bias	1.000	0.000	0		1.000	1.000	1.000
TF type	quadrant error	0.980	0.161	2	0.923	0.980	1.000	0.500
	local polar bias	0.624	3.775	2	0.151	0.727	0.830	0.500
Head Movement* Geometry	quadrant error	1.000	0.000	0		1.000	1.000	1.000
	local polar bias	1.000	0.000	0		1.000	1.000	1.000
Head Movement* TF type	quadrant error	0.741	2.393	2	0.302	0.795	0.937	0.500
	local polar bias	0.886	0.971	2	0.615	0.897	1.000	0.500
Geometry*TF type	quadrant error	0.766	2.128	2	0.345	0.811	0.963	0.500
	local polar bias	0.746	2.344	2	0.310	0.797	0.942	0.500
Head Movement* Geometry*TF type	quadrant error	0.898	0.860	2	0.651	0.908	1.000	0.500
	local polar bias	0.991	0.072	2	0.964	0.991	1.000	0.500

Table A.6: Multivariate Analysis using Pillai's Trace.

	value	F	Hypothesis df	Error df	p	η_p^2	Noncent. Par.	Observed Power
Head Movement	0.940	62.690	2.000	8.000	0.000	0.940	125.380	1.000
Geometry	0.217	1.108	2.000	8.000	0.376	0.217	2.216	0.182
TF type	0.471	2.769	4.000	36.000	0.042	0.235	11.078	0.702
Head Movement* Geometry	0.010	0.041	2.000	8.000	0.960	0.010	0.082	0.054
Head Movement* TF type	0.382	2.123	4.000	36.000	0.098	0.191	8.494	0.571
Geometry*TF type	0.467	2.740	4.000	36.000	0.044	0.233	10.959	0.697
Head Movement* Geometry*TF type	0.256	1.319	4.000	36.000	0.282	0.128	5.275	0.369

Table A.7: Univariat analysis for both error measures (quadrant error (q. e.), local polar bias (l. p. b.)).Sphericity accepted;

	error measure	Type III sum of squares	df	mean square	F	p	η_p^2	Noncent. Par.	Observed of Power
Head Movement	q. e.	42972.367	1	42972.367	138.761	0.000	0.939	138.761	1.000
	l. p. b.	137.791	1	137.791	0.992	0.345	0.099	0.992	0.145
Error(Head Movement)	q. e.	2787.182	9	309.687					
	l. p. b.	1249.705	9	138.856					
Geometry	q. e.	406.395	1	406.395	1.864	0.205	0.172	1.864	0.231
	l. p. b.	44.160	1	44.160	0.626	0.449	0.065	0.626	0.109
Error(Geometry)	q. e.	1961.950	9	217.994					
	l. p. b.	635.374	9	70.597					
TF type	q. e.	1060.185	2	530.093	5.946	0.010	0.398	11.892	0.816
	l. p. b.	9.903	2	4.952	0.085	0.919	0.009	0.170	0.061
Error(TF type)	q. e.	1604.745	18	89.153					
	l. p. b.	1048.199	18	58.233					
Head Movement* Geometry	q. e.	7.089	1	7.089	0.042	0.841	0.005	0.042	0.054
	l. p. b.	4.125	1	4.125	0.026	0.874	0.003	0.026	0.052
Error(Head Movement* Geometry)	q. e.	1504.774	9	167.197					
	l. p. b.	1402.787	9	155.865					
Head Movement* TF type	q. e.	355.324	2	177.662	1.805	0.193	0.167	3.611	0.327
	l. p. b.	519.562	2	259.781	2.684	0.095	0.230	5.368	0.464
Error(Head Movement* TF type)	q. e.	1771.412	18	98.412					
	l. p. b.	1742.069	18	96.782					
Geometry*TF type	q. e.	542.824	2	271.412	6.611	0.007	0.423	13.222	0.857
	l. p. b.	40.782	2	20.391	0.665	0.526	0.069	1.330	0.144
Error(Geometry* TF type)	q. e.	739.005	18	41.056					
	l. p. b.	551.966	18	30.665					
Head Movement* Geometry*TF type	q. e.	278.935	2	139.468	2.406	0.119	0.211	4.812	0.422
	l. p. b.	95.552	2	47.776	0.488	0.622	0.051	0.976	0.118
Error(Head Movement* Geometry*TF type)	q. e.	1043.403	18	57.967					
	l. p. b.	1763.008	18	97.945					

Table A.8: Estimates for pairwise comparison of the error measure quadrant error.

		mean	standard error	95% confidence intervall	
				lower bound	upper bound
Head Movement	dynamic	32.917	3.818	24.280	41.554
	static	70.764	1.910	66.443	75.085
Geometry	head	50.000	3.501	42.079	57.921
	sphere	53.681	2.106	48.915	58.446
TF type	magnitude	52.813	2.851	46.363	59.262
	original	47.813	2.362	42.468	53.157
	phase	54.896	3.216	47.620	62.172

Table A.9: Estimates for pairwise comparison of the error measure local polar bias.

		mean	standard error	95% confidence intervall	
				lower bound	upper bound
Head Movement	dynamic	4.089	0.921	2.006	6.172
	static	6.232	1.735	2.307	10.158
Geometry	head	4.554	1.158	1.934	7.175
	sphere	5.767	1.175	3.110	8.425
TF type	magnitude	5.470	1.619	1.808	9.132
	original	5.234	1.350	2.180	8.288
	phase	4.778	0.887	2.772	6.784

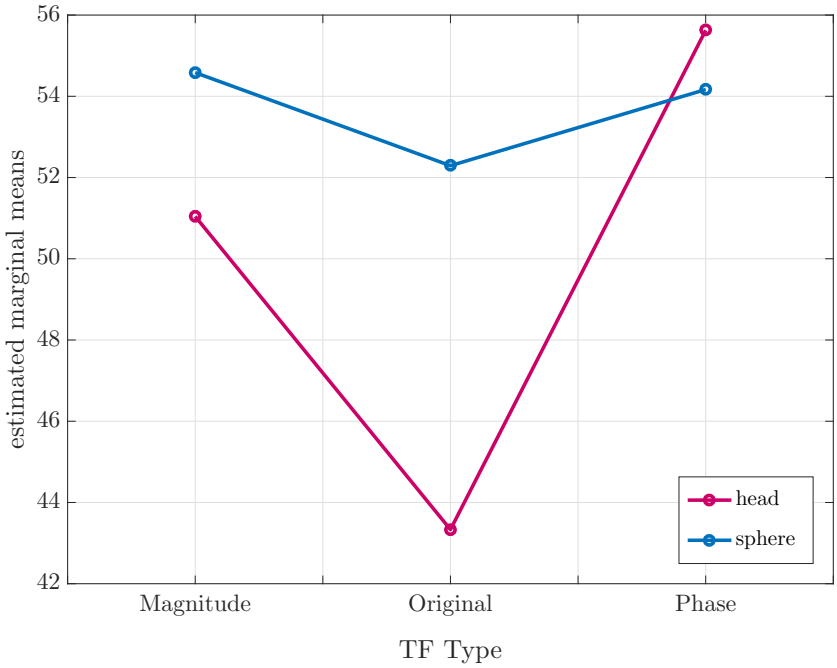


Figure A.1: Interaction effects for the factors TF type and geometry on the quadrant error.

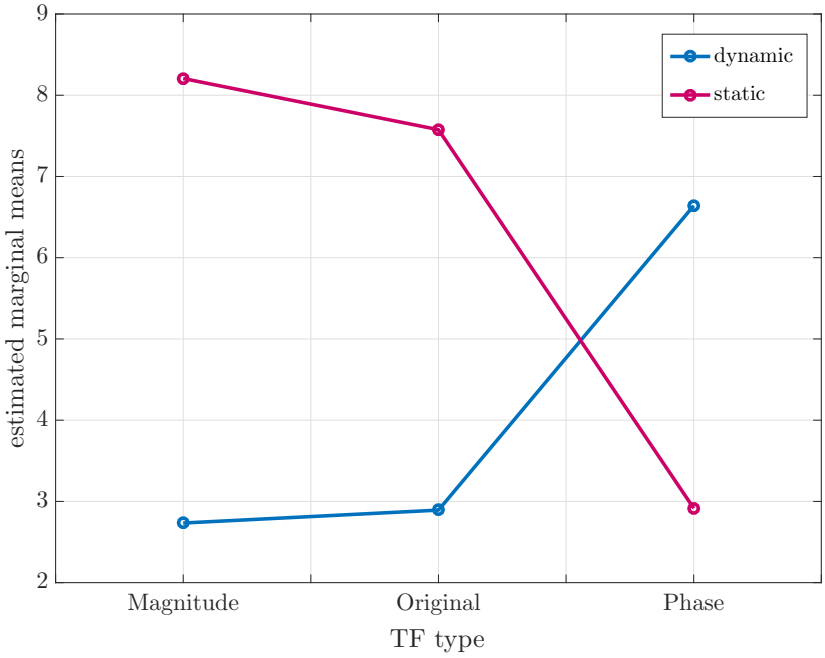


Figure A.2: Interaction effects for the factors TF type and head movement on the local polar bias.

Electronic Device

All folders of the electronic appendix are listed here and the content is briefly described.

- **Audio Signals**: Contains the audio signal of the listening test.
- **BEM Data (Input and Output)**: Contains all input and output data for the BEM for the geometries KEMAR, ellipsoid with on- and off-axis ears, as well as an unused version of the KEMAR without ears and one with exclusively the ears of the KEMAR. Input data are blender files of the mesh and the data and folders generated by the blender export, as well as the results of the BEM calculation.
- **Bibliography**: The PDF Bibliography List gives an overview of all cited papers and books and highlights the ones missing in the folder *Literature*. Additionally, the references.bib file is provided.
- **Conference Paper**: Includes the DAGA conference paper of the physical evaluation of this work
- **Data for Physical Evaluation**: This folder contains all SOFA files (simulated as well as analytically calculated) and a summarized version of the data as `.mat` file.
- **Data Localization Mismatch**: Contains `.mat` files with the calculated dynamic delta ILD and delta ITD and the required slope.
- **Evaluate Tracking**: This folder has an Excel spreadsheet with the target and actual angles to validate the accuracy during tracking of the Oculus Rift.
- **Forms Listening Test**: Includes the information sheet, which was handed to the participants, as well as all scanned questionnaires and statements of agreement for all participants.
- **Head Parameter**: It includes all data generated during TOA estimation (estimated ear positions, radii).
- **HRTF/SHTF (10 analyzed participants)**: Contains all SOFA files for all conditions of the 10 evaluated participants.
- **Listening Test**: Includes the whole Unity project with all assets and scripts. All steps are described in the scripts. With the mentioned versions the project can be loaded into Unity and is ready to use.
- **Literature**: Covers any literature listed in *Literature List* in PDF form.
- **Masters Thesis**: Includes the thesis in PDF version.
- **Matlab Scripts**: This folder contains any relevant Matlab scripts.
- **Raw Data, Results and Tracking Listening Test**: All raw data of the hearing test (participants' answers) and all data stored by the Head Tracker are located in this folder.
- **SPSS Raw Data**: Contains raw unfiltered SPSS output data as well as all command files in SPSS.