CZECH TECHNICAL
UNIVERSITY
IN PRAGUE

Faculty of Electrical Engineering
Department of Circuit Theory in cooperation with the Institute of Hearing Technology and Acoustics of RWTH Aachen University

F3

Master's Thesis

Examining the Interrelation and Perceptual Influence of Head-Related Transfer Functions Distance Metrics

Natálie Brožová
Medical Electronics and Bioinformatics, Signal Processing

August 2021
Supervisor: Shaima'a Doma, M. Sc., Ing. Tereza Tykalová, Ph.D.

Acknowledgement

At first, I would like to thank Shaima'a Doma, under whose supervision I have worked on my thesis within Erasmus+ exchange, for the excellent support and advice I received during my work. She was always ready to help and her door was always open whenever I encountered a problem. Her knowledgeable feedback helped me every time.

I would also like to thank Professor Janina Fels for giving me the opportunity to work with the Institute of Hearing Technology and Acoustics of RWTH Aachen University Medical Acoustics group. And also to Tereza Tykalová for allowing this work officially possible.

MASTER'S THESIS ASSIGNMENT

I. Personal and study details

Student's name: **Brožová Natálie** Personal ID number: **465404**

Faculty / Institute: **Faculty of Electrical Engineering**

Department / Institute: **Department of Circuit Theory**

Study program: **Medical Electronics and Bioinformatics**

Specialisation: **Signal processing**

II. Master's thesis details

Master's thesis title in English:

Examining the Interrelation and Perceptual Influence of Head-Related Transfer Functions Distance Metrics

Master's thesis title in Czech:

Zkoumání vzájemných vztahů a percepčního vlivu distančních metrik Head-Related Transfer Functions

Guidelines:

1) Analyze the interaction and mutual information between different objective distance metrics for HRTF comparison, using tools like correlation analysis or factor analysis.
2) Reduce the given metrics to a smaller set of measures suitable for a listening experiment and covering both spatial and spectral aspects.
3) Design a listening experiment paradigm to examine just noticeable differences and give first insights into various perceptual attributes (e.g. changes in sound coloration, spatial aspects and audible artifacts), which might be predicted by the respective distance metrics.

Bibliography / sources:

[1] K. Lee and S. Lee, 'A Relevant Distance Criterion for Interpolation of Head-Related Transfer Functions,' IEEE Transactions on Audio, Speech, and Language Processing, 2011;19: 1780-1790.
[2] A. Bondu, V. Lemaire, R. Nicol a S. Busson, „Looking for a relevant similarity criterion for HRTF clustering: A comparative study" Audio Engineering Society Convention. 2006;120: 4.
[3] D. Ballabio, A MATLAB toolbox for Principal Component Analysis and unsupervised exploration of datastructure. Chemometrics and Intelligent Laboratory Systems. 2015;149: 1-9.

Name and workplace of master's thesis supervisor:

Ing. Tereza Tykalová, Ph.D., Department of Circuit Theory, FEE

Name and workplace of second master's thesis supervisor or consultant:

Date of master's thesis assignment: **06.01.2021** Deadline for master's thesis submission: **13.08.2021**

Assignment valid until: **30.09.2022**

Ing. Tereza Tykalová, Ph.D. Supervisor's signature	doc. Ing. Radoslav Bortel, Ph.D. Head of department's signature	prof. Mgr. Petr Páta, Ph.D. Dean's signature

III. Assignment receipt

The student acknowledges that the master's thesis is an individual work. The student must produce her thesis without the assistance of others, with the exception of provided consultations. Within the master's thesis, the author must state the names of consultants and include a list of references.

.

Date of assignment receipt Student's signature

/ Declaration

I declare that the presented work was developed independently and that I have listed all sources of information used within it in accordance with the methodical instructions for observing the ethical principles in the preparation of university theses.

In Prague 13. 8. 2021

....

Natálie Brožová

Abstrakt /

Head-Related Transfer Functions (HRTFs) se využívají pro popis virtuálních audio scén, pro fitování naslouchacích pomůcek a v psychologii. Aproximují interakci zvuku s tělem posluchače. Jelikož existují různé stupně individualizace HRTF pro posluchače roste potřeba jejich porovnání. Pro tento účel se využívají distanční metriky, které mohou popisovat různé typy rozdílů mezi HRTFs. Bylo analyzováno pět distančních metrik pro porovnání HRTFs. Cílem této práce je analyzovat interakce a vzájemné informace poskytnuté danými distančními metrikami a zredukovat jejich počet na počet vhodný pro experiment. Dále, navrhnout a provést daný experiment pro prozkoumání slyšitelného prahu (just noticable difference: JND) daných metrik a také pro získání prvních poznatků o predikci percepčních atributů pomocí distančních metrik.

Dané vzájemné interakce a informace poskytnuté distančními metrikami jsou analyzovány pomocí korelační analýzy, analýzy hlavních komponent a faktorové analýzy. Tři distanční metriky (Inter-subject Spectral Difference: ISSD, Mean Squared Error: MSE and Mel-frequency Cepstral Distortion: MFCD) zachovávající nejrozmanitější informace jsou vybrány pro návrh experimentu. 3AFC paradigma je navrženo pro prozkoumání slyšitelného prahu těchto metrik. Test se také zaměřuje na zkoumání vztahu mezi distančními metrikami a subjektivními percepčními atributy (kolorizace a lokalizace). Tyto atributy jsou zkoumány za různých podmínek, což umožnilo pozorovat výsledky shodující se s předpokladem “výhody pravého ucha” (right ear advantage). Na základě této práce byly předloženy návrhy na další kroky ve vývoji distančních metrik pro porovnání HRTFs.

Klíčová slova: HRTF, distanční metriky, 3AFC, JND, percepční atributy.

Abstract /

Head-Related Transfer Functions (HRTFs) become necessary when using headphones to create virtual audio scenes. They approximate the interaction of the sound with the body of the listener. They can be also used in hearing aid fitting and psychology, where individually measured HRTFs are better suited than generic ones. For comparison of HRTF datasets, distance metrics, taking different types of errors into account, can be used. Five distance metrics were examined in the present work. The main objective of this thesis was to examine mutual information and interaction of these metrics and reduce them to a smaller set of measures suitable for a listening experiment. A further objective was to design and conduct a listening experiment paradigm to examine just noticeable differences (JND) as well as gain first insights into various perceptual attributes using respective distance metrics for prediction.

The present mutual information and interactions between different objective distance metrics were analysed and evaluated using tools such as correlation analysis, principal component analysis and factor analysis. Three distance metrics (Inter-subject Spectral Difference: ISSD, Mean Squared Error: MSE and Mel-frequency Cepstral Distortion: MFCD) providing the most diverse information were selected for the listening test. To examine an audible threshold of these metrics, a 3AFC listening test paradigm was proposed. The test also focused on finding a relation of selected distance metrics and subjective perceivable attributes (coloration and localization). These attributes were inspected in different conditions, enabling observations in alignment with the right ear advantage presumption. Suggestions were given for further steps in the metric development, based on the present work.

Keywords: HRTF, distance metrics, 3AFC, JND, perceptual attributes.

Contents /

Tables / Figures

Chapter 1
Introduction

Head-related transfer functions (HRTFs) are often used in virtual audio scenes and for medical purposes such as hearing aid fitting [1]. They are also used in psychological research and psychoacoustics, e. g. regarding auditory selective attention [2]. HRTF describes the head as an LTI system, more precisely as a frequency and direction dependent filter. The incoming soundwaves on the ear are filtered in every direction based on resonance, diffraction at a listener's head and torso as well as based on interference and reflection of the soundwaves. The goal of using HRTFs is to achieve the most realistic impression of the audio scenes. Within this objective, individually measured HRTFs perform better than HRTFs measured for an artificial head. However, the measuring process is very complex and requires special equipment. Hence it might not be possible to measure HRTF for every person individually. Many studies have been conducted in order to individualize it, e.g. based on a person's preference or morphology such as the shape of the head and resonance properties of the pinna [3].

With different ways of obtaining individual or individualized HRTFs the need for their comparison arose. This task is not intuitive as the visualization of HRTFs is challenging due to its high dimensionality. The HRTFs may differ in spectral peaks and notches for every direction. Objective measures (referred to as distance metrics) allow for different ways of comparison. They can be mainly focused on directional differences, frequency differences or they can take both into account [4–10]. So far, no detailed analysis of the information provided by distance metrics has been conducted. In this thesis, the relationship between different distance metrics will be examined by different tools, such as correlation analysis, principal component or factor analysis in order to eliminate a subset due to redundancy. Five direction dependent distance metrics will be inspected in detail.

Based on findings from the analysis of interrelations and mutual information of the distance metrics a few distance metrics providing the most diverse information will be selected. We further hope to look for an audible threshold of these selected distance metrics. For that purpose, a just noticeable difference (JND) listening test is proposed and conducted. The listening experiment also focuses on finding a relation of the few different distance metrics and human auditory perception. Subjectively perceivable attributes of audible differences, coloration and localization, are examined in relationship with the selected distance metrics.

Chapter 2 describes the basics of human auditory perception and binaural hearing. HRTFs are introduced in the chapter, as well as mathematical methods used in this thesis. Chapter 3 discusses the current research on binaural technology relevant to the present work. That includes HRTF measurement and individualization, objective and subjective measures for HRTF comparison and an overview of common listening paradigms. Chapter 4 describes the input data and further is separated into two main sections. The first section describes steps taken to analyse and reduce investigated objective distance metrics. Based on findings from this analysis (described in section 5.1) the listening experiment is proposed in the second section of this chapter. Results of both parts of this thesis are presented and discussed in chapter 5. A summary and suggestions for further work are then given in chapter 6.

Chapter 2
Theoretical Background

This chapter is concerned with describing the basics of human auditory perception as perception of pitch and loudness. The basic principles of binaural hearing are also described. Then, head-related transfer functions (HRTFs) and their directional components directional transfer functions (DTFs) are introduced. Moreover, mathematical principles used in this thesis are explained.

2.1 Human Auditory Percetion

Hearing – our auditory perception provides us with the possibility to communicate with others and it helps us to orient in new events and environments. The acoustic signal is not simply received by the human auditory system but also analyzed by hearing organs and then transmitted to the brain, where it is further analyzed and interpreted. Each sound event is therefore subjective and influenced by factors such as the pitch and loudness leading to an overall perception of sound quality.

2.1.1 Perception of Pitch and Loudness

The human audible frequency range is approximately 20 Hz – 20 kHz. The sensation of a sound wave's frequency is called pitch. Pitch perception of the human auditory system is highly nonlinear, this is thought to have their basis in cochlear filtering [11]. At low frequencies, the frequency is proportional to the perceived sensation. At higher frequencies, the perceived sensation is not proportional. For reference tone $fref = 8$ kHz the tone of 4 kHz is not perceived "half as high" but rather 1.3 kHz is [12].

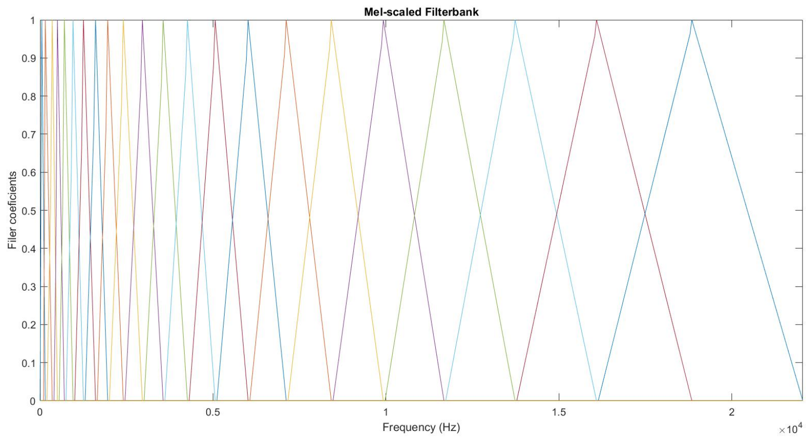

Figure 2.1. Mel scale filter bank with 24 critical bands.

Mel (melody) is a unit of the perceived pitch. It was shown that the human auditory system processes tones in frequency groups, so-called critical bands [12]. Their effect can be described as a band-pass filter (BPF). In Figure 2.1 Mel scaled filter bank is shown with twenty-four critical bands. The critical bandwidth is frequency dependent. It is approximately 100 Hz wide up to an audible frequency of 500 Hz, for higher frequencies it greatly increases.

Similarly, to human pitch perception, there is also a difference between the actual sound pressure level and the perceived loudness level. The relation between the perceived loudness level and actual sound pressure level is described by a unit phon. It is a unit of loudness for a sine wave. The loudness level of one phone matches sound pressure level in decibels of similarly perceived 1 kHz sine wave [13]. Therefore 60 phons means “as loud as 60 dB, 1 kHz wave”. The relation between the loudness level is frequency dependent. Furthermore, the loudness level is a non-linear metric, for that reason the unit sone was introduced. In this scale for each 10-phon increase, the loudness in sones almost exactly doubles [14].

2.1.2 Binaural Hearing

Binaural hearing, the ability to hear with two ears, is very important to humans for sound localization. Temporal and spectral disparities between the signal in the two ears provide cues about the spatial location of the sound. The most important binaural cues are the interaural time difference (ITD) and the interaural level difference (ILD), see Figure 2.2.

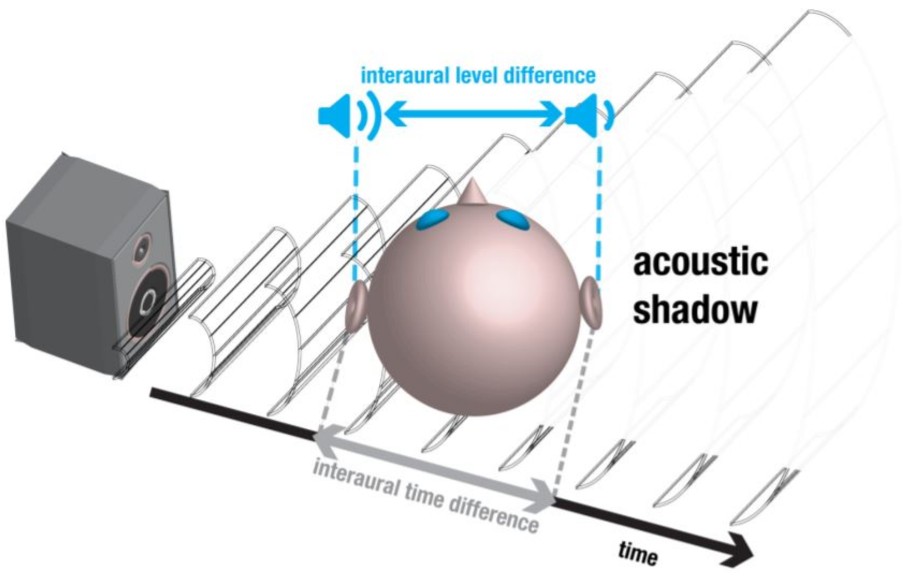

Figure 2.2. Illustration of ITD and ILD occurrence in free-field. Temporal and spectral disparities between the signals in the two ears provide cues about the incidence direction of the incoming sound [15].

ITD refers to the different time delays of sound wave’s arrival to left and right ear. ILD refers to different sound pressure level perceived by each ear, the sound pressure levels drops at the contralateral side ear compared to ipsilateral side ear. The duplex theory states ITD would be more relevant in lower frequencies (originally only below 125 Hz) while ILD would be relevant for higher frequencies (originally from 500 Hz onwards) [16]. ILD is irrelevant for lower frequencies as their long wavelength is not influenced by the body. ITDs should be relevant only for low frequencies as higher wavelength shall allow for direct detection of phase delays between each ear’s signals. Hovewer, similar effect has been shown in higher frequencies due to amplitude modulation. [17] Information of ITD and ILD from the whole spectrum are combined to enable horizontal localization [18].

Only changes in ITD and ILD, which shall be responsible for human localization in the horizontal plane, shall lead to detectable differences in positions. However, localization is also partially enabled by so-called monaural cues which are strongly influenced by the listener's anthropometry. An example of limitation, when describing sound localization, is the; "cone of confusion" [19], see 2.3.

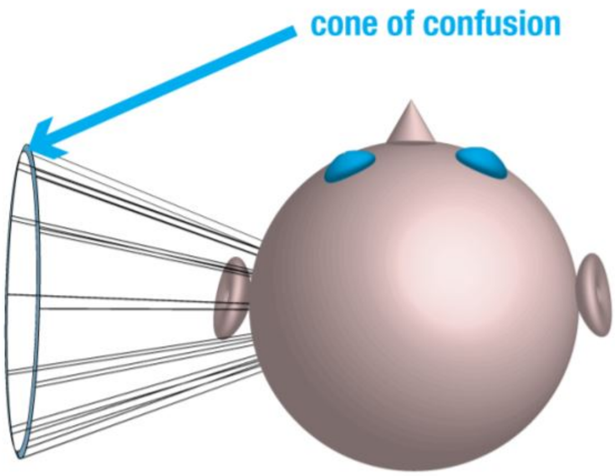

Figure 2.3. Illustration of a "cone of confusion", the ambiguity of localizing of the sound source while using only ITD and ILD. The cone is centered on the interaural axis expanding from each ear entrance, representing on its surface locations of the same interaural differences. Differentiating between these directions is difficult. [15, 19].

2.2 Head Related Transfer Function

The effects described at 2.1.2 can be summarized by the head-related transfer function. Head-related transfer functions (HRTFs) are often used in virtual audio scenes and for medical purposes such as hearing aid fitting. HRTF describes the filter process of the incoming sound on the ear by reflection, diffraction, interference and rezonance [17]. It also describes the head as an LTI system, more precisely as a frequency and direction dependent filter (1).

$$H(f, \theta, \varphi) \tag{1}$$

The HRTF contains prominent spectral features, such as peaks and notches, that vary according to source direction. Playback of a sound that has been filtered with individual's HRTF leads to more accurate localization of the sound source. On the contrary, playback of a sound filtered with HRTF from different subject, the virtual acoustic scene is less realistic and localization less acurate. [20].

Figure 2.4 describes coordinate system in use. HRTFs consist of directional and frequency-discrete complex values, that can be subdivided into a magnitude and a phase spectrum. Hence HRTF has coefficients for each ear, direction and frequency. Meaningful comparison method between HRTFs has been in the interest of research for many years. Different distance metric for comparison of HRTFs are in use. [4–5].

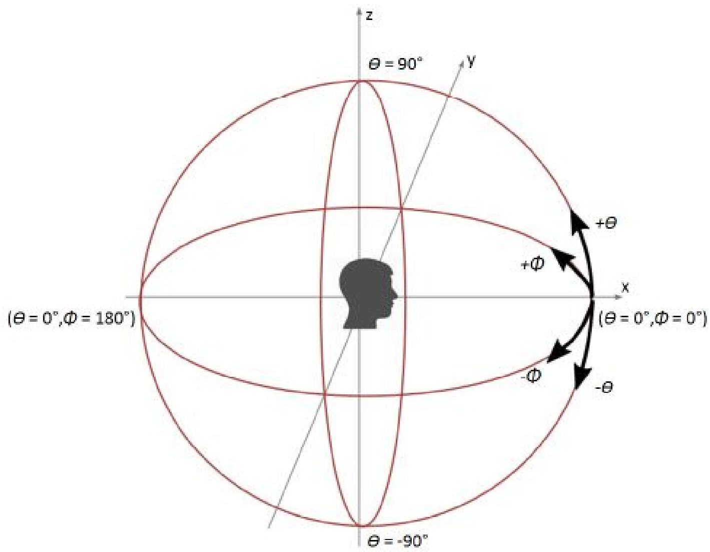

Figure 2.4. Head-related coordinate system. Theta θ describes zenith (elevation), φ the azimuth angle. The origin is located at the center of the head. Both ears lie on the y axis. The head faces towards the positive x direction, where the azimuth angle φ and the elevation angle θ both equal zero. [21].

HRTF can be displayed in freguency as well as time domain. The time-domian counterpart is called head-related impulce response (HRIR). In HRIR binaural cues become evident, see Figure 2.5. In magnitute spectrum of HRTF monoaural cues and ILD are visible, see Figure 2.6.

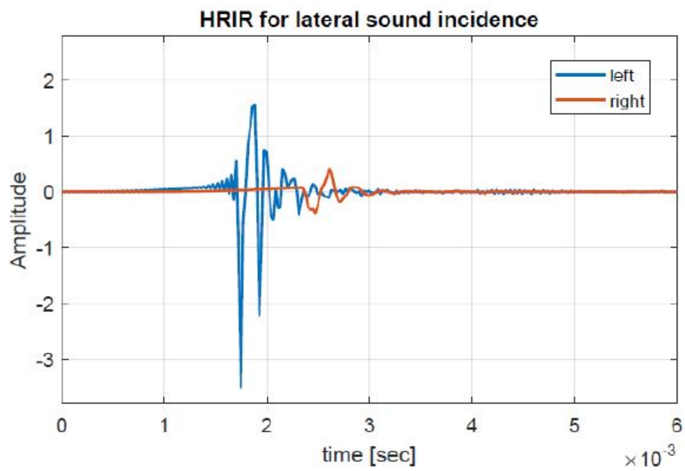

Figure 2.5. Head-related transfer function for lateral direction ($\varphi = 90°$ $\theta = 0°$) in time domain - HRIR shows a time delay (ITD) and level offset (ILD) due to head shadowing. [21].

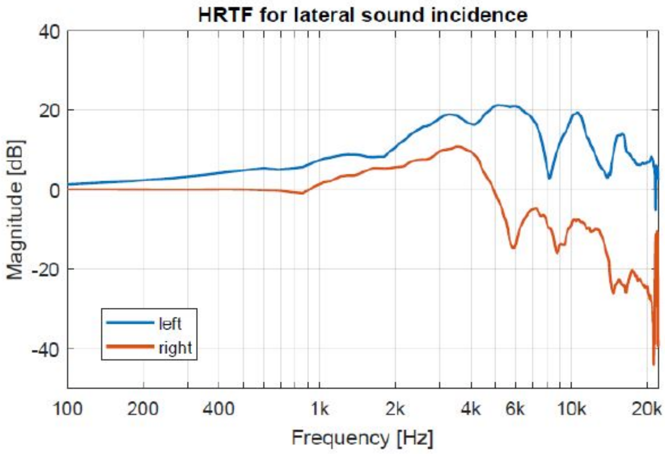

Figure 2.6. The magnitude spectrum of Head-related transfer function shows frequency dependent level offset (ILD) for higher frequencies as well as a direction- and subject-specific peaks and notches. [21].

2.2.1 Directional Transfer Function

The directional components of head-related transfer functions can be referred to as directional transfer functions (DTFs). The DTF can be calculated by dividing the complex transfer function measured at each location by the complex common component H_{com}. Therefore only information for specific sound-source location are provided, separating it from common component for all directions. [20].

$$DTF(f, \theta, \varphi) = \frac{H(f, \theta, \varphi)}{H_{com}(f)} \tag{2}$$

2.3 Mathematical Principles

Principal Component Analysis (PCA), Factor Analysis (FA) and Pearson's correlation coefficient will be introduced. PCA, a dimension reduction technique, can be in use of finding the necessary number of distance metrics given by the number of necessary dimensions. FA can be used for finding connections between different distance metrics. Pearson's correlation is usually used for determining a linear relationship between data. Spearman's correlation is used to determine a monotonic relationship between data.

Voronoi diagram, which can be used as a method for computing weights for unequally sampled points, will be also introduced.

2.3.1 Principal Component Analysis

The main principle of PCA is the projection of data into reduced space defined by principal components. Principal components are a linear combination of original data sorted by variance. The first few principal components contain most of the information from original data, which allows us to reduce the dimension of the dataset. PCA does so by solving the Eigenvector-Eigenvalue problem for the correlation matrix of the dataset $X \in R^{nxp}$ for p variables and n observations. [22–23]

PCs are found as defined ([23], nomenclature after [21]), see (3):

$$(\phi - \lambda I)V = 0 \tag{3}$$

where ϕ is pxp covariance matrix of X, the vector of eigenvalues $\lambda \in R^p andV \in R^{pxp}$ is the eigenvector containing PCs. V contains eigenvalues sorted from the highest to the lowest which signs the variance of data in directions defined by PCs.

In order to project X into new coordinate system, weighting is performed. W is nxp weight matrix:

$$V = XW \tag{4}$$

The data can be further converted to original space after preserving only important PCs (usually preserving 95 % of the variance of original data or keeping PCs with eigenvalues greater than one):

$$X = WV^{-1} = WV^T \tag{5}$$

eigenvector matrix V is orthogonal and therefore it's inverse and transpose are the same.

2.3.2 Factor Analysis

Factor analysis (FA) is a multivariate statistical technique for finding associations between observed variables. FA introduces latent variables, referred to as factors, that determine to some extent the values of the observed variables. Factor loading is a weight indicating the direct influence of a factor on a variable. [24]

FA is well-suited to psychology, or in our case, evaluating spatial audio quality. In psychology concepts such as "intelligence" (spectral coloration or distance when evaluating spatial audio quality) can be observed only indirectly [24]. Spearman noted in 1904 that if an individual performs well on intelligence-related task, he also inclines to do so on similar tasks. He introduced the general ability factor, which later developed into a general statistical procedure [25].

FA explains observed correlations in term of latent factors, PCA is a data reduction technique, where component scores represent a linear combination of the observed variables weighted by eigenvectors.

The factor model [24] declares that conditional expectation of variables X is a linear function of factors ξ:

$$E(X|\xi) = \Lambda\xi \tag{6}$$

where $X \in R^{px1}$ vector of random variables with zero means, $\xi \in R^{qx1}$ is a vector of random factors and $\Lambda \in Rpxq$ matrix of factor loadings.

The model assumptions are that variables are uncorrelated and the data are standardized with 0 mean and 1 variance. The covariance matrix Σ of the variables is defined as [24]:

$$\Sigma = \Lambda\phi\Lambda' + \Psi \tag{7}$$

where ϕ is correlation matrix, $\Psi = Var(X|\xi)$ is conditional covariance matrix. When holding model assumptions: $\Psi = Var(\varepsilon)$, where $\varepsilon = X - E(X|\xi)$ is a residual vector. Equation (7) shows, that each variable's variance is the sum of two sources. The first one is denoted as a variable's communality $h_2^i = Var(E(x_i|\xi))$ and is given by ith diagonal element of $\Lambda\phi\Lambda'$. The second source Ψ is referred to as a variable's uniqueness. A special case occurs if the factors are uncorrelated ($\phi = I$) and several equations simplify.

Maximum-Likelihood Factor Analysis

For maximum-likelihood (ML) estimation the assumption of multivariate normal distribution must be met. The estimation of loadings and uniqueness are obtained from minimizing following function [24]:

$$F(\Lambda) = log|\Sigma| + trace(S\Sigma^{-1}) \tag{8}$$

where S denoted sample covariance matrix and $\Sigma = \Lambda\Lambda' + \Psi$ (factors are assumed to be uncorrelated). Minimizing equation (8) w.r.t. λ and ψ requires iterative numerical methods. There are also other approaches for solving FA such as Principal-Axis FA. For determining the number of factors with the ML approach likelihood-ratio test can be used to test whether a specific number of factors can explain the sample covariance. The drawback is, that number of factors will be overestimated if the factor model is only approximately true (e.g. with a large sample size), also type-one error may highly increase due to multiple testing when the number of models is estimated varying the number of vectors. There are two other commonly used methods for determining the number of factors. Both rules apply when using the sample correlation matrix R rather than sample covariance matrix S to fit the model. The first rule chooses eigenvalues of R greater than 1, the second rule is a visual plotting procedure, where eigenvalues are plotted against their rank. Only the factors with the "largest" eigenvalues are kept. [24]

2.3.3 Pearson's Correlation Coefficient

The correlation coefficient between two random events is closely related to measures of correlation. Here Pearson's and Spearman's correlation coefficient will be described. Other correlation coefficients and measures of correlation can be found in [26]. The formula for Pearson's correlation coefficient ρ between two random variables X and Y can be written as:

$$\rho_{X,Y} = \frac{E(X,Y) - E(X)E(Y)}{\sqrt{E(X - E(X)^2)}\sqrt{E(Y - E(Y)^2)}} \tag{9}$$

2.3.4 Spearman's Correlation Coefficient

The formula for Spearman's correlation coefficient ρ is Pearson's correlation between the observation ranks $R(x_i)_1^n$, $R(y_i)_1^n$ and can be written as [26]:

$$\rho_{X,Y} = \frac{\sum_{k=1}^{n}[R(x_i) - \bar{R}_x][R(y_i) - \bar{R}_y]}{\sqrt{\sum_{k=1}^{n}[R(x_i) - \bar{R}_x]^2 \sum_{k=1}^{n}[R(y_i) - \bar{R}_y]^2}} \tag{10}$$

2.3.5 Voronoi Diagram

A Voronoi diagram separates a plane into regions (usually polygons) such that every point in the region is closer to a measurment point than to any other measurement point. These regions are also called Thiessen polygons. Thiessen polygons determine an area where prediction of characteristics at unsampled locations are provided by the nearest measurement point. If measurement points lie on a grid, then polygons are all equal. If the data are irregularly spaced then the result is an irregular grid of polygons. [27–28]

General construction of Voronoi Diagram:

- Every three sampling points form a triangle on which no other sampling point lies.
- These three points lie on a circle, the center of which is a node of the Voronoid diagram.
- The Voronoi points, which lie around a sampling point, span a polygon, the area of which corresponds to the weight of the sampling point.

The Delaunay triangluation is dual to the Voronoi diagram (in 2D space) in the graph theoretical sense [27], see Figure 2.7. In this project a Spherical Voronoi diagram is used to determine the weights for unequally sampled HRTFs.

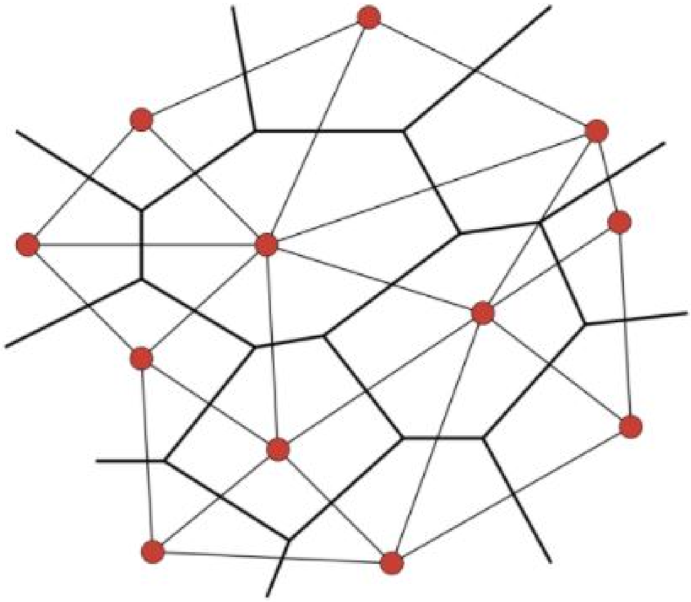

Figure 2.7. Example of Thiessen polygons (thick lines) and the equivalent Delaunay triangulation (thin lines). [28].

Chapter 3
State of the Art

This chapter discusses current research on binaural technology relevant to this thesis. A brief introduction to HRTF measurement is given as well as to a chosen HRTF individualization method, Multiple regression analysis (MRA) of principal components, which plays an important role in the present work. Objective and subjective measures to evaluate and compare HRTFs are presented. Finally, a short overview of common listening tests is given.

3.1 Measurement of HRTFs

Generic artificial head HRTF can be used for auralization of acoustic virtual scenes as the measurement of individual HRTFs can be exhausting and time-consuming. Nonetheless, this leads to localization errors such as front-back confusion and in-head localization. HRTF measurements summarize the direction-dependent acoustic filtering of a free-field point source due to the head, torso, and pinna [29]. An HRTF is ideally measured in an anechoic chamber over a discrete spatial grid. It can be also described as a free-field transfer function, which describes the sound pressure measured at the entrance of the aural canal in relation to the sound pressure, measured with the same sound source at the central point of a person's head (the person is absent during the measurement), see Figure 3.1. [17] The basic measurements method includes the movement of one loudspeaker around the subject to each desired direction. High resolution can be achieved, however, this method is highly time-consuming (up to 72 hours for 1° resolution) and undesirable for measurement with individual subjects. [17]

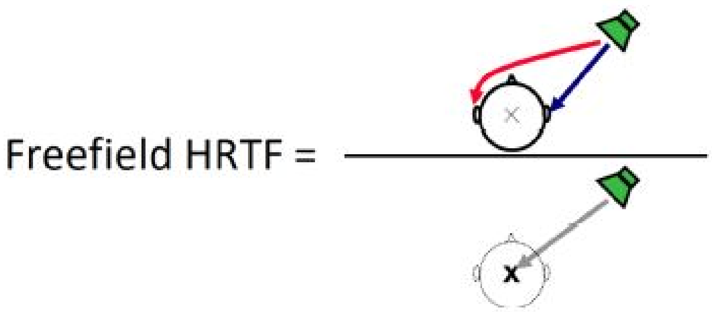

Figure 3.1. Depiction of free-field HRTF. [17].

Multiple fast HRTF measurement setups have been developed. One design uses an arc with loudspeakers as shown in Figure 3.2. Measurement can, therefore, be reduced below 10 minutes with a resolution of 2.5° with elevation 0° to 160°. [30] Further development of the arc at the Institute for Hearing Technology and Acoustics makes use of a continuous rotation to omit time periods exclusive for movement. The developed arc consists of 64 neighboring loudspeakers with a flat frequency responses between 450 Hz and 18 kHz. The speakers are integrated into the surface of the arc which eliminates the reflection from neighboring loudspeakers. Using this construction, placement of more loudspeakers using less space is made possible. The construction also allows rotating around the participant and it is possible to conduct stepwise and continuous measurements. The latter is proven to invoke fewer movements with participants due to shorter duration. [6]

Figure 3.2. The HRTF measurement arc of the Institute for Hearing Technology and Acoustics. The IHTA artificial head is used for the measurement. (Taken from [6].)

Spatial interpolation or weighting is of use for uneqally sampled HRTFs, as well as for reposition of sampling points due to unconscious movements. Another advantage of spatial interpolation is the ability to estimate HRTF values for directions that were not determined by measurement [31]. In [31] Spherical Harmonics Decomposition Method (SHD) is compared to other methods feasible for calculating HRTFs at arbitrary field points. SHD describes spherical objects as the weighted sum of spherical basis functions. It originates from solving Laplace's equation in the spherical domain.

Another method uses geometrical principles of the Voronoi diagram as mentioned in section 2.3.5. A Voronoi diagram is used to determine optimal triangulations. Converting this model to HRTF reconstruction, a frequency or time-dependent vector corresponds to each sampling point. The task is to identify the weight for a new point in an unsampled location, which is equivalent to inserting a new polygon as shown in Figure 3.3. This task is also called Sibson interpolation and is described in [32]. Interpolation on a surface of a sphere is introduced by Brown [33].

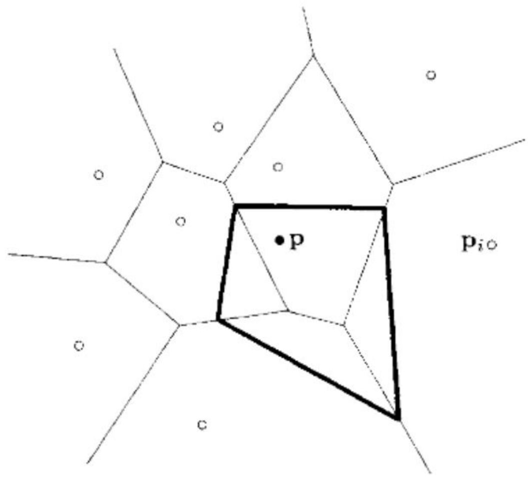

Figure 3.3. Insertion of a new point into Voronoi diagram, the polygon of new inserted point is highlighted (taken from [32]).

3.2 Individualization methods

Due to complex measuring conditions, it is not always possible to measure HRTF for every person individually, therefore many studies have been made in order to individualize it e.g. based on a person's preference or morphology such as the shape of the head and resonance properties of the pinna [4–5, 3]. There are methods that perform changes on existing non-individual HRTF and that split HRTF into components and combine them using weights exist. Anthropometric individualization of HRTFs using MRA of principal components offers a compromise between individual and non-individual HRTFs and offer a solution on how to quickly adapt HRTF to a certain listener.

3.2.1 MRA of Principal Components

Principal Component Analysis, as described in 2.3.1 is a dimension reduction technique. It allows in combination with multiple regression analysis to estimate individual HRTFs based on individual anthropometry. For the PCA model, every frequency bin (e. g. 129 bins) of the single-sided frequency spectrum can be understood as a variable. The accuracy of PCA increases with the size of input data. For that purpose, a database consisting of HRTFs of multiple individuals is used as an input to the analysis. Likewise, a mirrored set of HRTFs is provided in order to obtain twice as much data for the analysis. The left and right sides of the human head are not perfectly symmetrical, so HRTFs are not completely symmetrical around the median plane. Yet we can expect a symmetrical result when the average is built over a large group of individual HRTFs. Therefore it is feasible to use a mirrored set of HRTFs. [3]

A linear combination of principal components with direction and subject-specific weighing is used to reconstruct HRTFs. The PCs are sorted by the variance of the original data. The number of components required to reconstruct a HRTF is given by the threshold of 90 % cumulative variance of the original data [34]. PC1, which accounts for the highest variance in the input data, shows a maximum that corresponds to the first resonance frequency of the cavum concha at around 5 kHz. All other PCs show at least one minimum and one maximum. The number of PCs used for reconstruction determines the detail of the resulting HRTF. The required number of PCs is however dependent on the characteristic of input data. [3]

HRTF reconstruction in the frequency domain can be based on magnitude or complex spectra. When only magnitude spectra are used for the analysis, only real PCs are obtained and only absolute magnitude HRTF spectra can be reconstructed neglecting the phase information. An additional phase estimation is then needed. Hence, the minimum phase for given magnitude spectra is determined and a suitable ITD phase added, along with an (arbitrary) runtime phase to ensure causality of the filter. An overview of different methods for ITD estimation can be found in [35]. Complex spectra contain both magnitude and phase information. The drawback of using complex spectra is the higher number of PCs required to reconstruct HRTF accurately. [3]

3.2.2 PCA weight estimation using anthropometric measures

To determine HRTFs of subjects that are not included in the original database determination of weights is necessary. These weights are derived from anthropometric data, where a linear relationship between their geometric dimension and psychical effects at the ear is assumed. The anthropometric dimension α_i and each of the weights w (dependent on subject, direction and ear) of the matrix W can be used for regression analysis. Weights of a subject j can be expressed by a linear combination of anthropometric features $\alpha_{j,i}$ and regression coefficients β_i [3]:

$$w_j = \beta_0 + \sum_{i=1}^{n_{anthro}} \beta_i \alpha_{j,i} \tag{1}$$

The anthropometric features are expressed by a vector $\alpha_j = [1\ \alpha_{j,1}...\alpha_{j,anthro}]$ and the regression coefficients by a vector $\beta_j = [\beta_0...\beta_{anthro}]'$. The necessary amount of anthropometric features can be further reduced as some of the measures are of low importance. The determination of the suitable number of the features can be based on magnitudes of the regression coefficients β_i. Alternatively the determination can be based on collinearities between features where the features with the lowest pairwise correlation will be kept. [3]

3.3 HRTF Evaluation

3.3.1 Objective Measures

There are differences in individual HRTFs such as peaks and notches, which relate to an individual's anthropometry. To be able to compare them there is a need for objective metrics. These objective metrics are referred to as distance metrics, of which many have been introduced in different studies. Distance metrics can be of interest for many applications, for example, to compare HRTFs of a single listener in different setups or a comparison of individualized HRTFs. It also enables the comparison of intermodal HRTFs, such as the comparison of individual HRTF to individualized HRTF. Some examples of distance criteria can be [4–10]:

- Mean Squared Error (MSE Criterion),
- Critical Bands Mean Squared Error (CB-MSE),
- Inter-Subject Spectral Difference (ISSD),

- Correlation Distance (CD),
- Mel-Frequency Cepstral Distortion (MFCD),
- Spectral Difference (SD),
- The Loudness Level Spectrum Error (LLSE Criterion) ,
- Fahn Criterion,
- and many more.

They can be mainly focused on directional differences (MSE, CB-MSE, ISSD, CD and MFCD), frequency differences (SD) or they can take both into account (LLSE). Directional distance metrics quantify spectral changes for each direction individually whether the other metrics quantify differences for given frequency bins across all directions. The relationship between different directional distance metrics will be examined by different tools, such as correlation analysis, principal component or factor analysis in order to eliminate some metrics due to redundancy. A smaller set of distance metrics will be manageable regarding performing a subjective listening experiment with participants. We further hope to look for a correlation between distance metrics and subjective attributes of audible differences.

Mean Squared Error

One of the most obvious criteria that can be used for comparison of HRTFs is Mean Squared Error which is defined as [4]:

$$D_{MSE} = \frac{1}{N} \sum_{i=1}^{N} [H_1(i) - H_2(i)]^2 \tag{2}$$

where $H_1(i)$ is the magnitude spectrum of one HRTF and $H_2(i)$ that of another HRTF. N is the number of FFT points. The MSE is a symmetric criterion and it's value can be determined for every direction. A single-value distance metric is then achieved by averaging over all directions.

Critical Bands Mean Squared Error

The changing frequency resolution (logarithmic behaviour) of a human auditory system can be taken into account when computing MSE criterion. As stated in 2.1.1 the frequency resolution for low frequencies is higher than for high frequencies. It is therefore approached to lower the contribution of high frequencies part to the metric by frequency weighting. Frequency weights $\alpha(i)$ are computed as the inverse of the critical bandwidth [4]:

$$\alpha(i) = \frac{1}{a_0 \bigtriangleup (f_i)} \tag{3}$$

where a_0 is a normalization value [4]:

$$a_0(i) = \sum_{i=1}^{N} \frac{1}{\bigtriangleup (f_i)} \tag{4}$$

ensuring that [4]:

$$\sum_{i=1}^{N} \alpha(i) = 1 \tag{5}$$

.

The Critical Bands Mean Squared Error (CB-MSE) which includes frequency weighting is therefore computed as [4]:

$$D_{CB-MSE} = \frac{1}{N}\sum_{i=1}^{N}\{\alpha_i[H_1(i) - H_2(i)]\}^2 \tag{6}$$

By averaging CB-MSE over all directions a single value for comparison of two HRTFs is then determined.

Inter-Subject Spectral Difference

The directional components of HRTFs are referred to as DTFs as stated in 2.2.1. DTFs vary systematically between subjects in different frequency features for every direction. Since they subtract information common to all directions, they focus on differences between the fine structures of HRTFs. The inter-subject spectral difference was introduced by Middlebrooks [20] and is defined as [3]:

$$D_{ISSD} = \frac{1}{n_{dir}}\sum_{i=1}^{n_{dir}} var\left(20log10\frac{|HRTF_{1,i}f(j)|}{|HRTF_{2,i}f(j)|}\right) \tag{7}$$

The variance of the quotient ot absolute values corresponding to frequency bins $f(j)$ between 1 and 13 kHz is determined for each direction i. An averaging over all directions is then needed in order to obtain a single value metric to compare two HRTF sets.

Mel-frequency Cepstral Distortion

Mel-frequency Cepstral Coefficients (MFCC) are widely used in speech processing applications. They take into account human auditory perception by using the mel-scale. MFCCs are obtained using Cosine transform. The distance metric making use of MFCCs is Mel-frequency Cepstral Distortion and can be defined as [5]:

$$D_{MFCD} = \frac{1}{N_C}\sum_{i=1}^{N_C}(M_k - \hat{M}_k)^2 \tag{8}$$

where M_k and $\hat{M}_k$ are MFCCs determined for the HRTFs and N_C is the total number of the coefficients. This metric is computed for every direction, one value to compare two HRTFs can be achieved by averaging over all directions.

The Loudness Level Spectrum Error

The Loudness Level Spectrum Error is a distance metric based on evaluation of virtual source quality using binaural auditory model as proposed in [10]. The basis of the model lies in the inter-aural cross-correlation and shall estimate perceived localization cues and coloration.

A schematic model can be found in Figure 3.4 and consists of the following steps [36]. As excitation, pink noise is filtered with both HRTFs. Pink noise is used because it yields average spectral properties. The filtered noise is passed through a grammatone filterbank (GTFB) with 42 bandpass ERB (equivalent rectangular bandwidth) channels. For each channel, the resulting loudness is calculated.

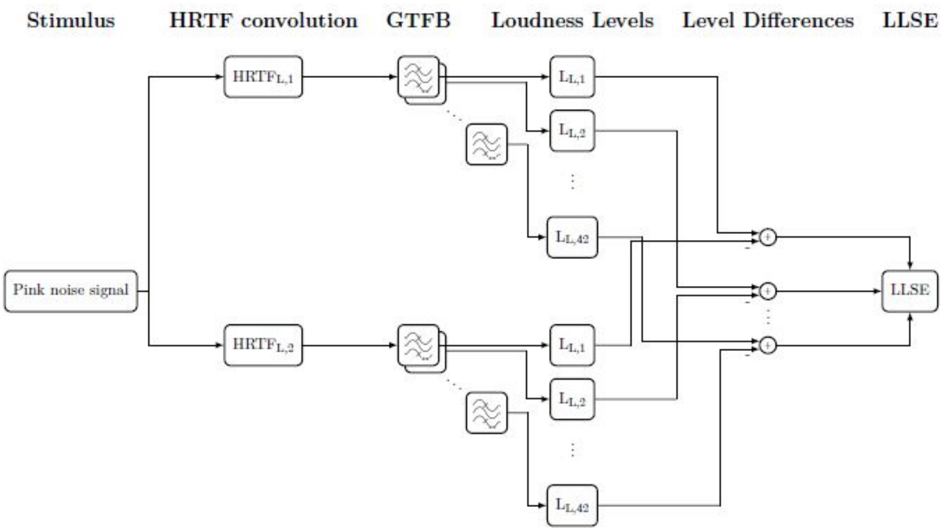

Figure 3.4. Binaural auditory model as described in [10], taken from [37].

In this thesis, the distance measure has been adjusted to indicate changes in the level difference as in [37]. This has been achieved by computing the variance in the frequency or in the spatial domain. The results are two distinct distance metrics. A single metric can be then obtained by averaging over all frequencies or directions, respectively.

3.3.2 Subjective Evaluation

The aim to create realistic virtual audio scenes is dependent on correctly reproduced binaural cues which are embodied in HRTFs. The quality of the reproduced sound can be at best assessed by a listening test. With regard to more complex virtual scenes, the perceptual impact of HRTFs can go beyond simple localization. It is therefore still a matter of research to find a suitable set of perceptual attributes [38].

For the perceptual evaluation of virtual audio rendering quality Spatial Audio Quality Inventory (SAQI) has been developed “to overcome limitations with respect to the relevance and completeness of vocabularies” [39]. Auditive qualities mentioned in different studies are spectral coloration, spaciousness, localizability, steadiness of movements, source width, loudness, loudness balance, distance, internalization vs. externalization, impulse-like artifacts, and dynamic responsiveness. However, for a comprehensive perceptual evaluation of the virtual environment, the existing vocabulary did not appear sufficiently complete. The vocabulary which consists of 48 descriptors of auditive qualities, that can be roughly sorted into eight categories (timbre, tonalness, geometry, room, time behavior, dynamics, artifacts, and general impressions) has been proposed. Some attributes are closely related to spectral properties of an audio signal, other attributes reflect higher-order psychological constructs (e.g. clarity, naturalness, presence) [39]. An exemplary listening test procedure implementing SAQI vocabulary is depicted in Figure 3.5. The listening test starts with comparing selected stimuli to another stimulus or reference. The participant is first asked whether they perceived any difference at all. In case they did not perceive any difference the test stops here, otherwise selected auditive qualities can be rated. Auditive qualities can be selected with regard to the research interest or with respect to the used stimuli [39]. This thesis will interrelation of different distance metrics and human auditory perception.

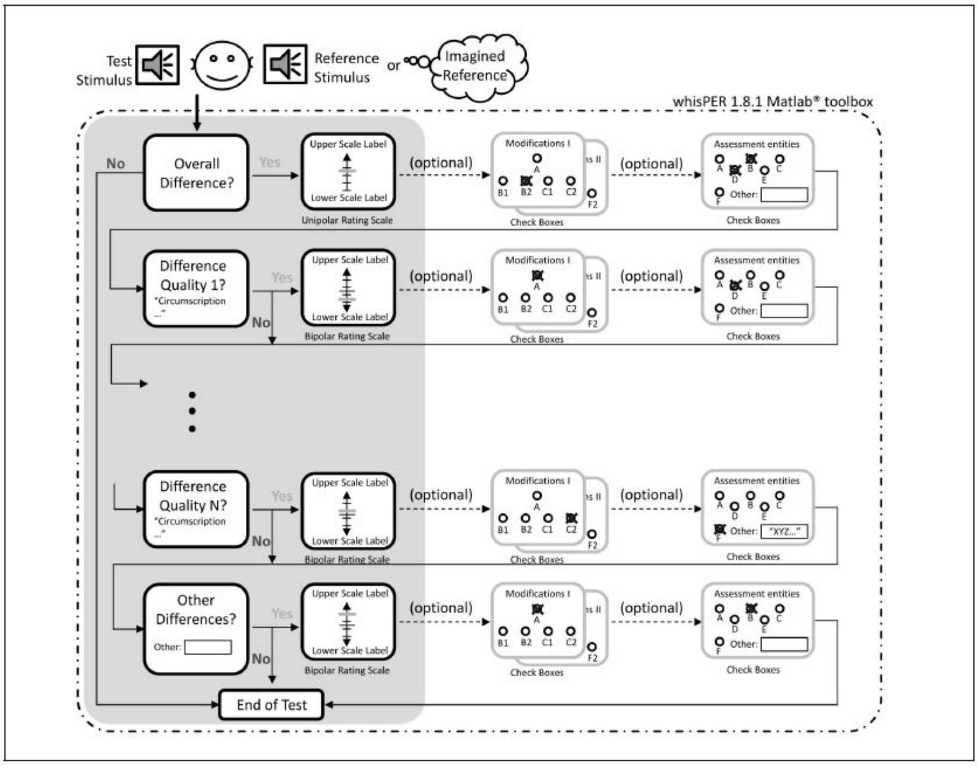

Figure 3.5. Exemplary listening test using SAQI vocabulary [39]. The procedure corresponds to the implementation in the free listening test software WhisPER [40].

3.3.3 Listening Tests

Listening experiments allow the transfer of subjective perception of audio scenes to quantitative measures. These measures establish a link between subjective evaluation and perceptual parameters (e. g. how different metrological values are perceived or how this connection influences the acoustic impression). [17] “A listening experiment is a scheduled, repeatable examination, conducted under variable conditions with test persons, who assess their acoustic perception according to given instructions” [17]. These assessments are usually acquired verbally or by motoric reactions.

One of the possible listening tests can be a just noticeable difference (JND) test. In other words, a listening test to determine threshold of audibility between two different stimuli. Another possible listening procedure can include scale assessment with different attributes in relation to stimuli. [17]

An inferential model that is often applied in detection tasks such as listening experiments is a psychometric function. It can model a probability of a “yes” answer in relationship with a selected feature of physical stimuli, in this case, acoustic stimuli. Therefore it can help to determine a threshold of examined acoustic measure. The result is linked to a linear combination of predictors, using a sigmoid link function. [41]

Depending of the study design 50 % point can correspond to the threshold of a likelihood of 50 % for audibility, while in some paradigms, e.g. the 72 % point must be applied to account for correct guesses. An example of the psychometric function is presented in Figure 3.6.

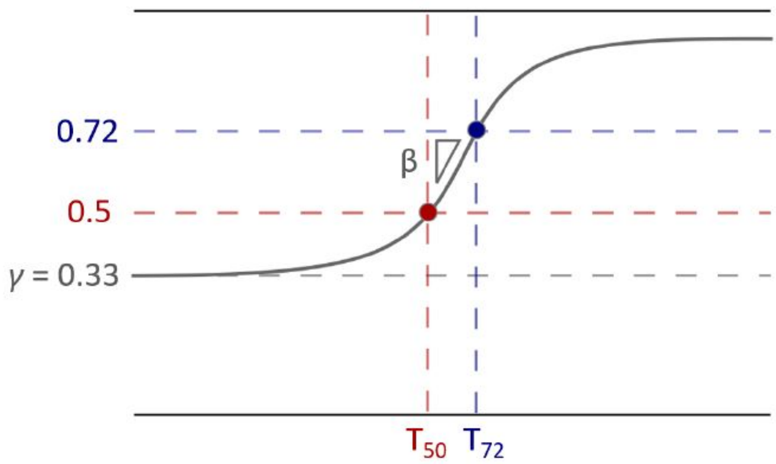

Figure 3.6. Psychometric function example for 3AFC test, taken from [21].

Various approaches to evaluate stimuli exist. The description of classical evaluation methods can be found in [42]. They include a method of limits, method of adjustment and method of constant stimuli. Another method allows evaluating more stimuli at the same time. It is referred to as n-alternative forced-choice (nAFC) where n is the number of stimuli being compared. The subject is presented with n stimuli and is forced to choose between them based on the selected criteria. For 3AFC experiment, the participant is presented with three stimuli and is asked to determine the odd one among the others. However, the participant is not provided with the option "I don't know". This guess rate needs to be taken into account when evaluating the experiment. [43, 17] For different test approaches or paradigms, the slope of psychometric function changes. E.g. for the yes-no paradigm, the slope is steeper which allows for more precise threshold detection, yet it is highly sensitive to "the position and stability of listeners' response criterion", an area in which nAFC test outperforms the yes-no paradigm, even though 3AFC experiment might be comparatively time-consuming. [43]

For the common listening test, several stimuli are pre-selected based on discrete levels of selected criteria and presented to the participant. Every presented stimulus is referred to as a trial. How many trials and how are they composed is part of the test design and strategy. The levels of the criteria shall cover the range of the slope of the psychometric function. For the 3AFC test, the T_{50} threshold is determined by the point of 2/3 of trials answered correctly. Adaptive methods may allow for more precise threshold determination as every next level of the presented trial is choosen based on the previous answer of the participant. In consequence to that, no prior knowledge regarding the threshold is needed. [43, 17]

By listening experiments, different scaling methods can be of use to help determine perceptual attributes of the stimuli [43]. Scaling methods can be divided into direct and indirect scaling, where indirect scaling is usually more demanding and should be preferrably used in laboratory experiments instead of field studies. [17] Multiple scales options come into consideration based on data's characteristics. The typical scales levels are nominal, ordinal, interval and ratio scales and they can differ in their properties.

One of the sub-classes of scaling methods are rating methods which present the test subject with questionnaire after playback of the acoustic stimulus. Different unipolar and bipolar scales can be utilized as seen in Figure 3.7. They can be continuous as well as discrete. For rating methods, the reference stimulus always needs to be taken into account in order to minimize bias. In order to align the reference for all participants, a common reference frame can be introduced. However, there exist a lot of possible biases that are discussed in [44–45].

1	2	3	4	5
very weak	weak	moderate loud	loud	very loud
not	little	medium	rather	very
not at all	slightly	moderate	highly	extemely

Unipolar five category scales.

– –	–	∘	+	++
-2	-1	0	+1	+2
very low	low	neutral	high	very high
very unpleasant	unpleasant	neutral	pleasant	very pleasant

Bipolar five category scales.

Figure 3.7. Example scales to assess stimuli (modified and taken from [17]).

Chapter 4
Materials and Methods

The practical part of the thesis consists of two main parts: In the first part, distance metric selection for further analysis is discussed. The approach for analysis of mutual information shared between mentioned distance metric is described as well as the selection of distance metrics to be examined in a listening experiment. The second part focuses on the design and implementation of the listening test.

4.1 Input Data

In this thesis, the analysis of HRTFs distance metrics' mutual information, as well as design of JND listening experiment was conducted on HRTFs from ITA HRTF-database [46]. HRTFs of 47 subjects were then utilized. The resolution of HRTFs is $5°$ x $5°$, with limited zenith angle from $0°$ to $160°$ which results in 2304 directions. Three datasets using ITA HRTF-database were considered for the analysis:

- Real measured HRTFs.
- Reconstructed HRTFs, using "ideal" weights PCA.
- Reconstructed HRTFs, using MRA of principal components with "reconstructed" weights.

The reconstruction methods are described in section 3.2.1 and in [3]. "Ideal" weights reconstructed HRTFs can be determined only for subjects whose individual HRTF is in the database. They usually serve as data compression approach. Individualized HRTF with "reconstructed" weights using individualization method MRA of principal components can be on the other hand computed for any arbitrary subject whose anthropometric features are known. All reconstructions were implemented in MATLAB with the use of signal processing tools in ITA-toolbox [47]. In this thesis, three HRTF datasets for the same subjects can be therefore compared. Each subject has real measured HRTF, and two reconstructed HRTFs, one with "ideal" weights, the other with "reconstructed" weights. The distances were computed between left ear data.

For the reconstruction with "ideal" weights 23 PCs were used for all 2304 directions. For the reconstruction with "reconstructed" weights anthropometric features are utilized to determine weights for the MRA of principal component procedure. A set of 6 features was chosen to estimate PC weights using the minimum correlation approach [3]. The reconstruction was performed on magnitude frequency spectra, therefore only real PCs were obtained and the phase information was neglected.

For above-mentioned datasets various distance metrics can be computed. The distance metrics can be determined for comparison of HRTFs of the same modality (e. g. reconstructed HRTF of one subject to reconstructed HRTF of another subject) and for comparison of HRTFs of different modalities (e. g. real measured HRTF to reconstructed HRTF). For purpose of this thesis the datasets will be referred to as: real measured HRTFs - Real HRTF, reconstructed HRTFs, using "ideal" weights PCA -

Ideal PCA HRTF; reconstructed HRTFs, using MRA of principal components with "reconstructed" weights - MRA PCA HRTF.

To analyze the interaction and mutual information of different distance metrics for HRTF comparison, the HRTFs of different modalities were used. The following intermodal comparisons were taken into account:

- Real HRTF to Ideal PCA HRTF for the same subject.
- Real HRTF to MRA PCA HRTF for the same subject.
- Ideal PCA HRTF to MRA PCA HRTF for the same subject.

For design of the JND listening experiment intramodal comparisons as well as intermodal comparisons of HRTF datasets were considered. A subset of comparisons was chosen, as discussed in section 4.3.2.

- Intermodal comparison:
  - Real HRTF to Ideal PCA HRTF for the same subject.
  - Real HRTF to MRA PCA HRTF for the same subject.
  - Ideal PCA HRTF to MRA PCA HRTF for the same subject.
- Intramodal comparison:
  - Real HRTF for different subjects.
  - MRA PCA HRTF for different subjects.
  - Ideal PCA HRTF for different subjects.

4.2 Analysis of Mutual Information of Distance Metrics

Many objective distance metrics have been introduced in an attempt for useful comparison of HRTFs. Comparison of HRTFs may be of interest to compare HRTF of single listener in different setups or for comparison of individualized HRTFs. Distance metrics can be focused on directional differences in frequency spectra magnitude and phase spectra or on overall frequency differences. Nine different distance metrics were examined in [37], where also more detailed descriptions of the measures can be found. The computation of selected distance metrics used for more detailed inspection in this thesis were described in section 3.3.1, that are MSE, CB-MSE, ISSD, MFCC and LLSE varied in spatial domain. That means that only distance metrics focused on directional differences were considered, i.e. quantifying the spectra error within discrete directions. Two of these metrics (MSE and ISSD) display mostly numerical errors between the pair of HRTFs, the other three distance metrics (MFCC, CB-MSE and LLSE) rely on psychoacoustic effects.

The most fundamental approach, the mean square error (MSE) introduced in [4], is rarely used for HRTF comparison. It provides wider range of values than any other considered metric for equal HRTF variations, however Bondu et al. [4] reports that MSE performs well for clustering purposes.

ISSD defined by Middlebrooks [20] focuses on the differences between the fine structure of HRTFs as the information common to all directions is omitted when subtracting DTF. According to [20], ISSD shall correspond well with virtual localization of participants. ISSD was originally computed up to 13 kHz, to enable comparison to other distance metrics it is determined for whole audible frequency range.

The intuitive approach to include psychoacoustic effects is the CB-MSE measure introduced in [4]. It includes frequency weighting resembling critical bands. Critical bands shall follow the changing frequency resolution of the auditory system.

The measure MFCD makes use of mel-frequency cepstral coefficients (MFCCs) which are widely used in automatic speech recognition. Hovewer, in [5] the measure was considered for HRTFs clustering and interpolation. It was shown to be suitable. MFCCs are useful in separating the periodic parts of the signal from additive noise and hence in representing the envelope of the signal. MFCD could be recognized as approach complementing ISSD which focuses more on differences in fine spectral structure.

Timbre is mostly defined as qualities of sound that help distinguish it “from other sounds of the same pitch and volume” [48]. It could be described as changes in volume distribution in auditory bandpass filters. The differences in timbre between stimuli are referred to as coloration. The measure investigating coloration changes in the present work shall be LLSE as defined in [36]. It computes disparities in loudness level for each direction and for each gammatone filter. The metric values strongly depend on the excitation signal, , e.g., pink noise can be used, as done in the original study. In this thesis it was generated using ITA-toolbox [47] function.

All the above-mentioned distance metrics provide a distance value for each direction. In order to obtain one distance metric value, the average over all directions needs to be computed. Weighing for unequally distributed spatial sampling of HRTFs needs to be considered when determining the mean. For that purpose spherical Voronoi was used to compute the weights for each direction. The distance metrics were computed for the left ear data.

The analysis of the interaction and mutual information between the mentioned distance metrics was performed on the set of directional distance metric values between HRTFs as well as on the set of mean distance measures, where there is only one distance metric between a pair of HRTF sets. That results into a 5 x 47 matrix. For the directional distance metric values, there are 2304 values between pair of HRTFs as HRTFs used for the analysis were measured for 2304 directions. That results into five 47 x 2304 matrices one for each of the five examined distance metrics, where 47 is the number of subjects. The analysis was then done on 5 x 2304 matrix for each person, i. e. five different distance metric values for each direction. For the analysis purposes, PCA, FA and correlation analysis were used. Mutual information and interaction of above-mentioned metrics was examined.

4.2.1 Choice of Distance Metrics for Listening Experiment

The next objective was to reduce the number of given metrics to a smaller set of measures suitable for a listening experiment. Based on findings from analysing mutual information metrics which cover the most information, regarding differences between different sets of HRTFs are inspected in the listening experiment. Selected metrics shall cover differences in the spatial and spectral domain.

Principal component analysis projects data into space defined by principal components. PCs are a linear combinations of original data sorted by variance, therefore the first few principal components contain the most information from the original data. It is possible to determine which variables (distance metrics) contribute the most to the first few PCs by noting which variables show the highest values (loadings) for the given PCs coefficients. Usually, the number of variables that explain at least 95 % variance of the original data shall be kept, see Section 2.3.1.

Factor analysis introduces latent variables (factors) that can explain associations between observed variables as is described in Section 2.3.2. The direct influence of factors on a variable is indicated by determining loadings of observed variables for the latent factors. Factor loading shows the variance explained by the variable on that particular factor and could be understood as the correlation coefficient for the variable and factor. PCA introduces new variables that are made of the linear combination of original data, factor analysis shall provide factors that describe variability among the observed and most probably correlated variables whilst extracting common variance. The highest specific variation from all observed variables can be also important in determining which variable provides the most information.

Individual correlation coefficients between all distance metrics can be also determined so that linear or monotonous relationships between these metrics can be observed, the calculations can be found in Sections 2.3.3 and 2.3.4.

4.3 Just Noticable Difference (JND) Experiment for Distance Metrics.

The second main objective of the thesis was the design and implementation of a listening test to determine JND for selected distance metrics. In other words to determine the audible threshold for selected distance measures between sets of HRTFs. Extension to the JND listening test shall also provide first insights into perceptual attributes of the selected distance metrics. Perceptual attributes cannot be exactly measured as they are subjective in their nature. They might include changes in sound coloration, timbre, spatial qualities or audible artifacts. There might be a possibility of prediction of the perceptual attributes by the respective distance metric.

The experiment is based on insights gained and discussed in chapter 5.1 regarding metric behaviour and correlation. The analysis of interactions of different distance metrics was conducted on datasets that included intermodal comparisons as such comparisons are more commonly needed, i. e. to compare HRTFs of one listener in different setups or to compare individualized HRTF to real measured HRTF. However, audible differences between several intermodal comparison datasets were too high to be considered for the experiment. For this reason, intermodal as well as intramodal comparisons of HRTF datasets were considered and examined when designing the listening test. Intramodal comparisons are also relevant when it comes to comparing different people's cues, or whether a person's HRTF is suitable for someone else.

4.3.1 Experiment concept

For the listening test several directions of compared HRTFs sets were chosen based on their distance metrics values. The purpose of the test is to address these issues:

- Identify distance metric values for which audible differences exists.
- The additional research question is concerned with prediction of perceptual attributes (coloration and localization) based on distance metric values.
- Identify, if there is a link between given perceptual attributes and any independent variable.

The distance metrics, localization and coloration were examined for following independent variables:

- Directions of HRTFs from ipsi-/contra-lateral side.
- Different comparisons (intermodal, intramodal) of HRTFs.
- Right and left ear playback (i. e. mirroring of the source stimuli as the analysis for choosing directions was performed on left ear data).

Ipsi- /Contra-lateral sides were considered because different distance metrics show very distinct values for these sides as can be seen in Figure 4.1.

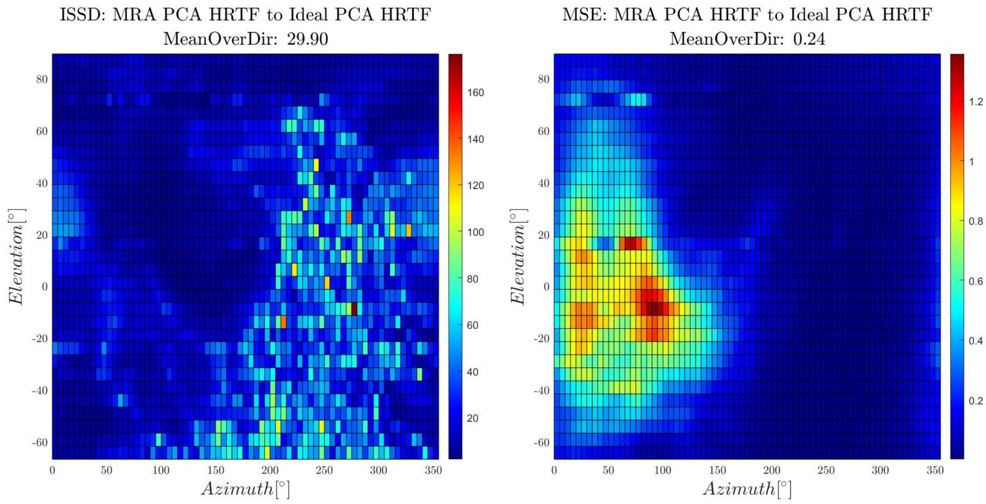

Figure 4.1. Comparison between ISSD and MSE metric for the same two HRTF sets. Some distance metrics show higher values for the ipsilateral side compared to the contralateral side and some show the opposite behaviour, therefore directions for the listening test were optimized also for ipsi-/contra-lateral side.

Mirroring of the sources takes into account the presumption of the right ear advantage, i. e. the fact that human auditory system is not symmetrical and right ear shall be in advantage in terms of binaural processing [49]. Hence listening to compared stimuli with the right ear might show different behaviour than to listening to compared stimuli with the left ear.

4.3.2 Test design

Test Subjects

A total number of 19 subjects participated in the experiment. All subject were IHTA researchers or employees, meaning that majority (15 out of 19) had experience with HRTF evaluation. There were 6 women among all subjects, all participants were aged between 23 to 35.

Intact hearing was screened for all the participants. For that purpose high frequency audiometry was conducted for each participant. All participants had intact hearing up to 8 kHz (considering the common 20 dB HL limit). For high frequencies up to 16 kHz alternative limit was considered as recruitment of subjects would be even harder. The limit for satisfactory hearing ability was defined as drawing a line from 20 dB HL at 8 kHz to 40 dB HL at 16 kHz. Trespassing the line at 16 kHz could be neglected as most monoaural cues of HRTFs lie below 13 kHz [20]. Subjects with such hearing ability did not present outliers in the data. Audiograms of all test subject are displayed in Figure 4.2.

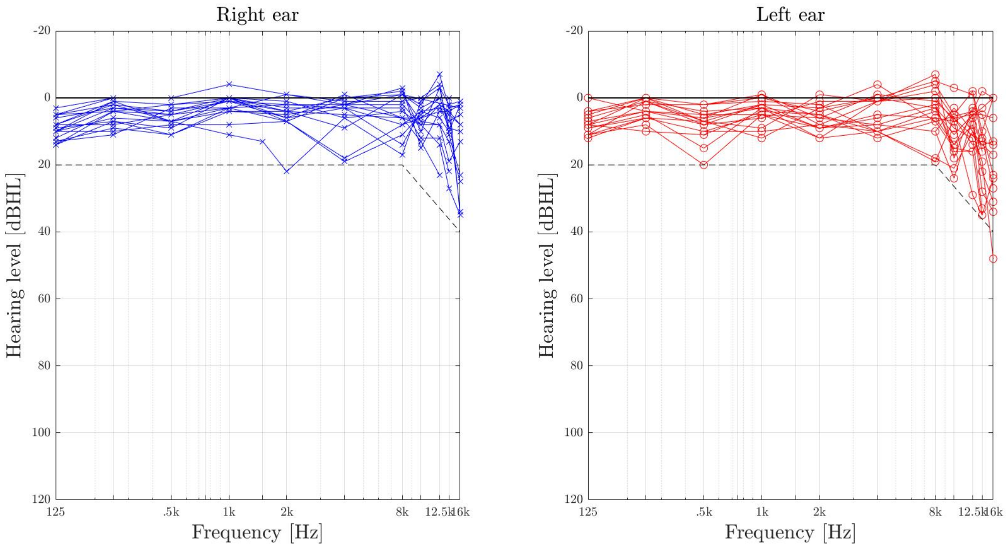

Figure 4.2. Audiograms of all test participants. The dashed line depicts the limits that were considered for intact hearing, just few subjects trespass the limit.

All participants were informed about the process of the experiment and signed the consent form. Pseudonym list for the participant's names and personal information was created and stored securely ensuring anonymity of the data.

Procedure

A 3AFC experiment was implemented for the JND test of distance metrics. Three stimuli (A, B, C) were presented and the participant was asked to determine the odd one, see Figure 4.3. There was one reference stimulus presented twice and a test stimulus once.

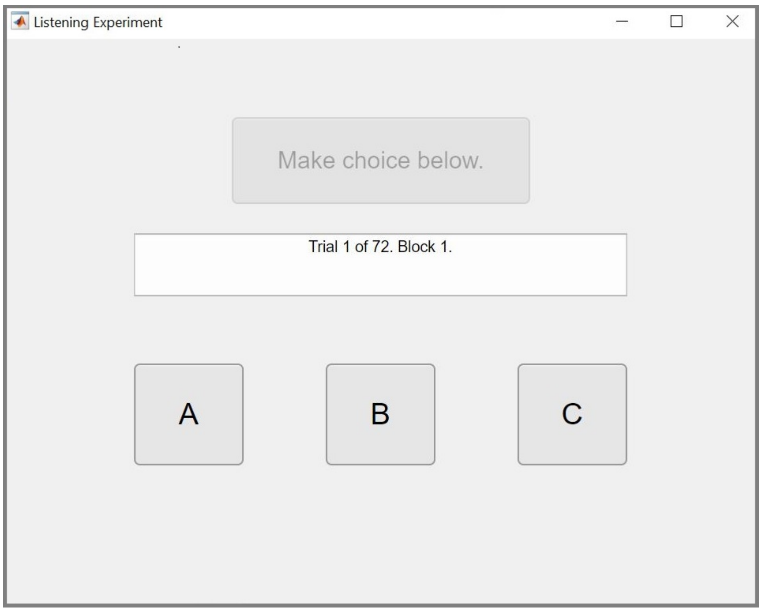

Figure 4.3. Graphical user interface for 3AFC test paradigm.

A problem is of course the order of stimuli within a trial. Direct comparison of A and C is difficult. Therefore, the choice and order of reference and test stimuli were randomized in each trial. To repeat the trial with different order of the stimuli, the

1	2	6	3	5	4
2	3	1	4	6	5
3	4	2	5	1	6
4	5	3	6	2	1
5	6	4	1	3	2
6	1	5	2	4	3

Table 4.1. Balanced latin square.

experiment consists of three blocks consisting of the same trials. The number of trials is 72 as will be discussed in the Stimuli selection part. The stimuli in one trial are always randomized and thus the order of stimuli (A, B, C) always differs.

The test design with repeated measures provides higher statistical power, however it may cause the subject's fatigue interfering with the results. Generally speaking, it is important to randomize the trials in order to prevent these undesirable effects. One common randomization measure is the Latin square which was utilized in this thesis. It presents the experimental design so that each trial/treatment occurs once in each row and in each column. Hence for 72 stimuli, 72 x 72 matrix was presented. For the listening test consisting of three blocks three rows of a given matrix were randomly selected and used as an order of the presented trials in each block. See an example of Latin square for 6 trials in Table 4.1.

The 3AFC experiment paradigm shall provide a more precise threshold determination, as mentioned before, with the drawback of taking more time.

In the second block of the experiment, additional questions about stimuli were asked. The questions were concerned with subjectively perceived sound quality. The procedure followed the proposed approach in Figure 3.5. The participant at first answered 3AFC paradigm question (which stimulus is different). Afterwards the participant was asked whether they perceived any overall difference between the stimuli (see Figure 4.4).

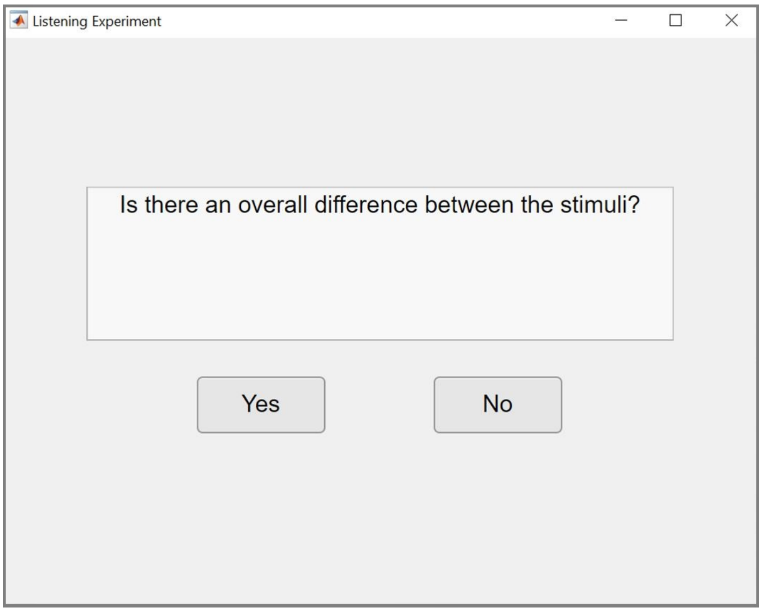

Figure 4.4. Graphical user interface to asses additional questions about subjective perception of presented stimuli - first the participant is asked whether they perceived an overall difference between stimuli.

If the answer was positive, two additional questions about perceived stimuli quality were asked. They are "How do the stimuli differ in terms of coloration?" and "How does the stimuli differ in the terms of source location?" The example stimuli regarding these qualities were presented together with their descriptions:

- Coloration describes sound characteristics other than loudness and virtual source location. As result sounds can become "hollow" or "metallic", but also "sharp" or "rough".
- The difference in localization between stimuli (whether the source changes location).

To assess these questions unipolar continuous scales were used. The range was set from 0 (the stimuli differ not at all.) to 100 (the stimuli differ extremely), see Figure 4.5. The listening test GUI was implemented in MATLAB App Designer.

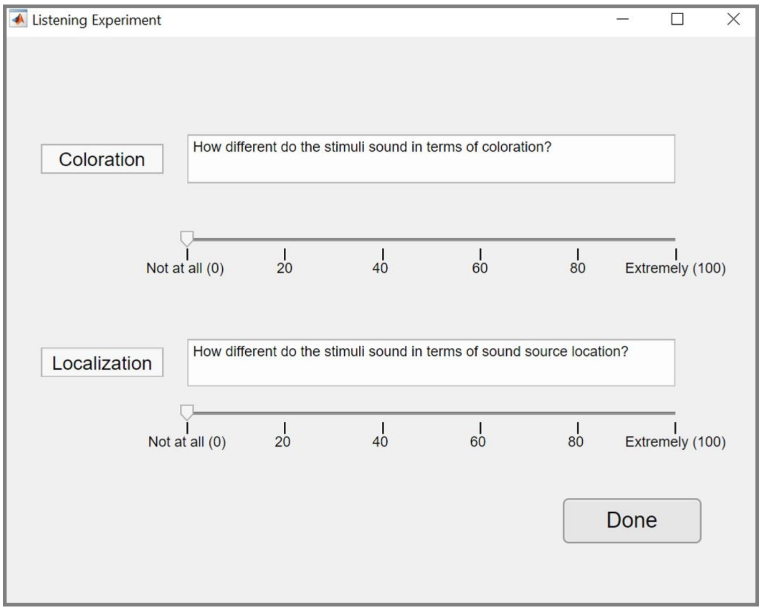

Figure 4.5. Graphical user interface to asses subjective perception of presented stimuli.

The overall experiment paradigm could be described by the following Figure 4.6. The expected time duration of the experiment procedure was determined as 45 minutes per participant for all three blocks, considering 15 minutes per block and therefore about 12,5 seconds per trial . Additional 10 minutes were considered for an introduction and an audiometry measurement. Summing up to the expected time of 55 minutes per participant.

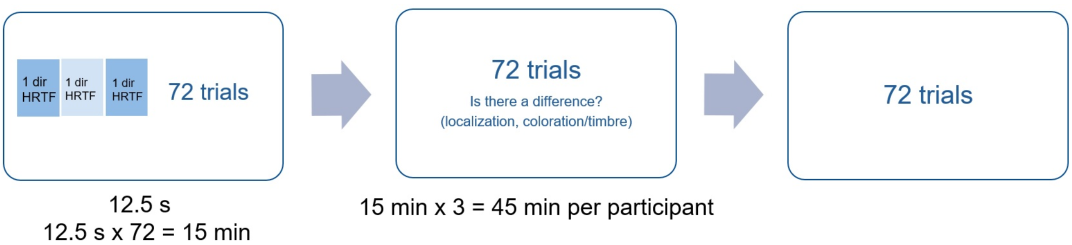

Figure 4.6. Depiction of the experiment blocks and expected time duration for each participant.

Stimuli Selection and Preparation

As will be seen in section 5.1, three metrics (MSE, ISSD and MFCD) cover the most information on the differences between HRTF sets and are applied in the experiment. The limits of the distance metric will be therefore selected in a non-adaptive manner as it would be difficult to optimize for all the three distance metrics at the same time from the reason that the distance metric values don't change linearly with each other.

For the selection of reproduced stimuli, HRTF directions from intermodal as well as from intramodal comparisons were considered. HRTF directions of different modalities were convoluted with pink noise generated using ITA-toolbox [47] function and replayed. The selection process was as follows for all independent variables (ipsi-/contra-lateral sides, different comparisons) except for the right and left ear playback as the selected source stimuli were only mirrored:

- Examination of the distance metric values of the three selected metrics (MSE, ISSD and MFCD) for different comparisons of HRTFs and ipsi-/contra-lateral sides. Pre-selection of the stimuli based on the values as can be seen in Figure 4.7.
- Listening to the compared pre-selected stimuli with different distance metric values.
- Selection of stimuli (directions) suitable for the experiment: directions of HRTFs between which no difference could be heard as well as where there was a slight audible difference.

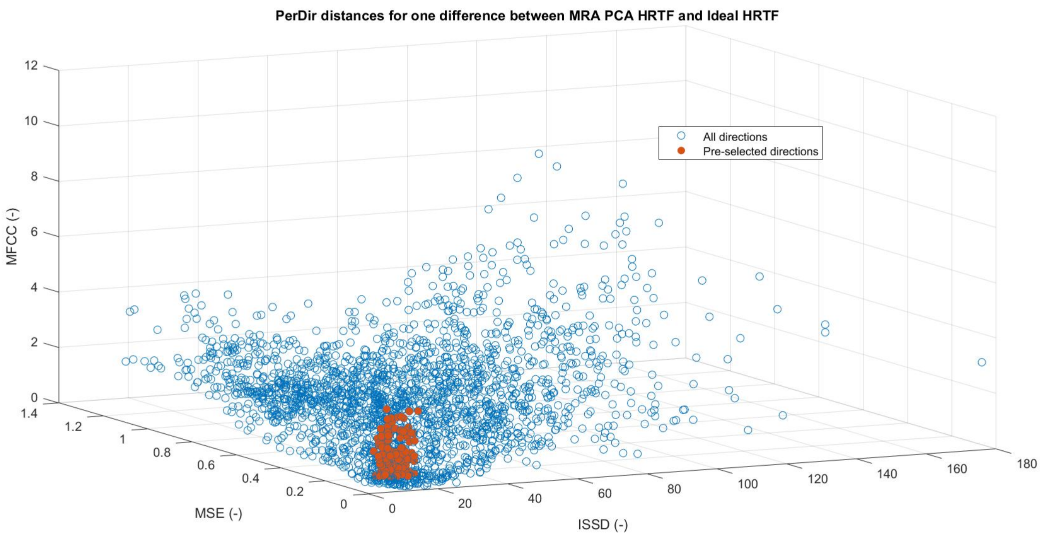

Figure 4.7. Example of pre-selection of stimuli directions for listening test based on three distance metric values between MRA PCA HRTF and Ideal PCA HRTF for the same subject.

During the selection process intermodal and intramodal comparisons with Real HRTFs were excluded from the experimental design as the audible differences and distance metrics values were too high compared to other cases. The following comparisons were used when designing the experiment:

- Intermodal comparison:
  - Ideal PCA HRTF vs. MRA PCA HRTF for the same subject.
- Intramodal comparison:
  - MRA PCA HRTF for different subjects.

- Ideal PCA HRTF for different subjects.

Six directions were selected for all independent variables combinations making it 72 stimuli in total (6 directions x 3 comparisons x 2 sides (ipsi- / contralateral) x 2 mirroring of the source stimuli for left and right ear playback). Six directions should be adequate to provide first outlook on audible threshold determination using fitting of psychometric function. For the intermodal comparison the HRTF directions were analysed for the first subject, for the intramodal comparison directions between the first two subjects were analysed.

As the reconstruction of HRTFs was performed on magnitude frequency spectra, the minimum phase of the reconstructed spectra was determined and the ITD phase component was added along with a time offset ensuring causality. ITD phase was determined according to the analytical ellipsoidal model [35]. All HRTF directions used for the test were calculated and saved locally prior to the test. That ensured avoiding the time-consuming calculation of the analytical ellipsoid model.

The distances were computed for left ear data, however, for binaural playback, data for both ears need to be provided. Thus, for the purpose of the experiment, it was decided to fixate right ear data for both stimuli in order to affect the participant's perception of the stimuli as little as possible. Transitions of interaural cross-correlation (IACC) between HRTFs' channels (ears) needed to be considered as we are able to notice changes in IACC, even if the magnitude spectrum is identical. Hence, if the binaural signal stimulus is less correlated than the stimulus replayed before we are more prone to notice change [50–51]. The right ear data from HRTF that showed smaller IACC transitions with left ear data of both HRTFs were used for the binaural playback as we were trying to select the stimuli below the JND for detecting IACC changes.

4.3.3 Test Execution

Prior the test, loudness calibration of the sound reproduction system was performed. It was calibrated to have sound pressure about 60 dB at the frontal direction and maximum of 66 dB at sides. The test was executed in hearing booth 2 at the Institute of Hearing Technology and Acoustics of RWTH Aachen to ensure a quiet environment during the test. Before the start of the experiment, a short introduction was given to the subjects, they were asked to sign the consent form and "Separate declaration on compliance with hygiene and infection control measures and Recording sheet for traceability according to § 2a CoronaSchVO".

High frequency audiometry up to 16 kHz was performed for every participant. For qualified subjects, a headphone transfer function was measured according to Masiero and Fels' approach [52]. Small microphones were inserted into the ears of participant and through the headphones, that were later used for reproduction of the stimuli, a frequency sweep was played. The used headphones were Sennheiser HD 650. The participant was then asked to take off the headphones and put them on again. This procedure was repeated 8 times. After that, the spectra of the left and right signals were averaged. Unsuccessful measurements (e.g. due to swallowing or low-frequency traffic noise) could be omitted before averaging. The result multiplied by the inverse sweep spectrum and the inverses of the known transfer function of the respective microphone, provide HpTF. It takes properties of the headphones and interaction of the signal with pinna geometry into account and serves for equalization.

During the test, the stimuli were loaded and convolved with pink noise train stimulus of the same sampling frequency of 44.1 kHz. The final stimuli spectra for playback were processed using the measured HpTF as:

$$S_i(f) = S_{noise}(f)\frac{HRTF_i(f)}{HpTF_i(f)}, \tag{1}$$

with $S_{noise}(f)$ being the pink noise pulse train, i signs indices 'L' and 'R' for the left and right ear.

The participant was then presented with the listening test. The test consisted of 3 test blocks with a possibility of a break between each block to prevent dizziness. The second block included the above-mentioned additional questions regarding the perceptual attributes of the stimuli and therefore took longer. The experiment duration was mostly between 20 to 30 minutes (excluding the introduction). All acquired subject data were saved in a pseudonymized way. Only a hard copy linking the data to the participant's name is kept.

Chapter 5

Results and Discussion

In this chapter results of both parts of the thesis are presented. At first, the results of an analysis of the interrelations and mutual information carried out on the distance metrics between different HRTF datasets are discussed. Based on these findings, it is proceeded to the selection of distance metrics suitable for the JND experiment. The findings from section 5.1 are used as basis for the listening test paradigm previously proposed in section 4.3. In the second part of this chapter, the results of the JND listening test are discussed.

5.1 Distance Metrics Selection for JND experiment

The range of values for each of the distance metrics for each examined dataset (as mentioned in section 4.1) was determined. Spectral components in between 20 Hz and 13 kHz were included in the computations.

Distance metric	Dataset	Minimum	Median	Maximum
D_{ISSD}	Real to MRA PCA HRTF	1.4	22.8	264.8
D_{ISSD}	Ideal to MRA PCA HRTF	0.6	16.3	221.3
D_{ISSD}	Real to Ideal PCA HRTF	0.4	8.8	180.8
D_{MSE}	Real to MRA PCA HRTF	0.001	0.101	16.204
D_{MSE}	Ideal to MRA PCA HRTF	0.0005	0.092	15.654
D_{MSE}	Real to Ideal PCA HRTF	0.0001	0.005	0.863
D_{CB}	Real to MRA PCA HRTF	$1.6 \cdot 10^{-7}$	$74.0 \cdot 10^{-7}$	$3406.5 \cdot 10^{-7}$
D_{CB}	Ideal to MRA PCA HRTF	$0.3 \cdot 10^{-7}$	$44.9 \cdot 10^{-7}$	$3393.4 \cdot 10^{-7}$
D_{CB}	Real to Ideal PCA HRTF	$7.1 \cdot 10^{-7}$	$23.4 \cdot 10^{-7}$	$511.8 \cdot 10^{-7}$
D_{MFCD}	Real to MRA PCA HRTF	0.06	1.83	36.26
D_{MFCD}	Ideal to MRA PCA HRTF	0.01	1.51	31.65
D_{MFCD}	Real to Ideal PCA HRTF	0.003	0.20	32.93
D_{LLSE}	Real to MRA PCA HRTF	0.04	0.20	0.94
D_{LLSE}	Ideal to MRA PCA HRTF	0.0002	0.004	0.10
D_{LLSE}	Real to Ideal PCA HRTF	0.04	0.20	0.78

Table 5.1. Range of values for all used distance metrics, computed for all HRTF comparison datasets on which analysis of interrelationships and mutual information was carried out.

The results can be found in Table 5.1. It is possible to observe that the highest distance metric values were determined for the Real HRTF to MRA PCA HRTF dataset. The second highest distance metric values were observable for Ideal PCA HRTF to MRA PCA HRTF and the lowest observable values were for Real HRTF to Ideal PCA HRTF. The exception can be found for the LLSE metric, where the Real HRTF to Ideal PCA HRTF dataset has higher values than Ideal PCA to MRA PCA HRTF.

However, the difference in distance metric values between the mentioned dataset is not as high as the difference between these two datasets and Real HRTF to MRA PCA HRTF dataset. That is understandable as MRA of PCA is an individualization method using anthropometric values to determine weights for MRA of PC reconstruction and therefore an approximation of PC reconstruction with ideal weights. Hence the biggest difference can be determined when comparing the twice approximated data (MRA PCA HRTF) to the measurement (Real HRTF).

To help to understand how the distance metric values are linked to the subjective perception of the difference between stimuli, a JND listening experiment is proposed in section 4.3. The difference between stimuli is not rated as greater or smaller, but rather as perceivable or not. Additionally, the test also includes questions about subjective perceptual attributes (coloration and localization) and thus tries to find a connection between distance metrics and these attributes.

5.1.1 Principal Component Analysis

To analyze the information provided by each distance metric, the PCA was used. At first, the analysis on the distance datasets with one weighted mean distance value (calculated over all spatial directions) between pairs of HRTF sets (resulting in 5 x 47 matrix for each dataset, respectively) will be discussed. The explained variation (i. e. information) by each PC for every inspected dataset can be found in Table 5.2. The distance metric values were normalized before performing the analysis. Usually, the number of PCs that explain at least 95 % variance of the original data is kept. Here it is visible that for every dataset the first two PCs are enough to keep the required variance/information of the original data. For Real HRTF to Ideal PCA HRTF dataset even the first PC would be enough to keep the required variance. In accordance with Table 5.1 the biggest variations (the lowest explained variation by PC 1) are found for dataset Ideal PCA HRTF to MRA PCA HRTF and the lowest variations in the data (the highest explained variation by PC 1) are found for dataset Real HRTF to Ideal PCA HRTF.

	Explained var. by PC (%)		
Dataset	Ideal to MRA PCA HRTF	Real to MRA PCA HRTF	Real to Ideal PCA HRTF
PC 1	93.7	94.2	96.0
PC 2	6.2	5.6	3.1
PC 3	0.1	0.1	1.0

Table 5.2. Principal component analysis performed on mean over all direction values for different distance datasets.

As can be seen in Figure 5.1, the highest loading (coefficient) for the first PC shows the metric MFCD. The highest loading for the second PC shows the metric MSE for two datasets out of three. For the Real HRTF to Ideal PCA HRTF dataset the metric LLSE shows the highest loading for the second PC. However, the second PC in this dataset keeps just 3.1 % of the explained variation and therefore MFCD and MSE can be considered as more important.

PCA on mean over all direction values dataset.

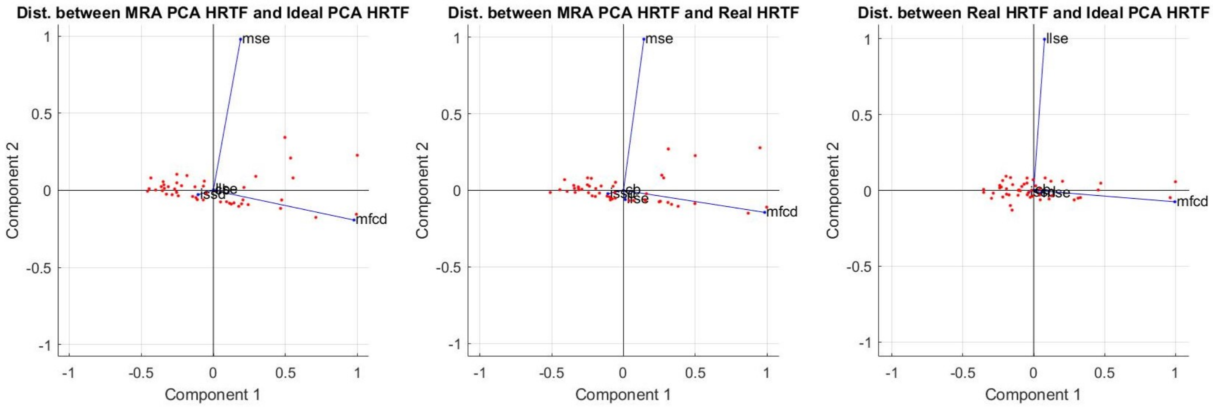

Figure 5.1. Biplot of first two principal components for each distance metric dataset with mean over all directions values.

Afterwards, the PCA was also performed on datasets that provide distance metric value for each direction between two HRTF sets (5 x 2304 matrix for each person, 5 distance metrics for 2304 directions, resulting in 47 matrices in total for each dataset). For every person for all three datasets, two PCs were enough to keep 95 % variance of the original data, hence the highest loadings for the first two PCs for each person were examined (which distance metric from the five examined metrics has the highest loading for PC 1 and which for PC 2). The results can be found in Table 5.3. It can be observed that the highest loading for the first PC also show metrics MFCD and MSE similarly to the mean over all directions datasets. The highest loadings for the second PC are more difficult to describe, yet LLSE and again MSE metrics seem to be important. LLSE metric again shows the highest loading values for the second PC mostly for the Real HRTF to Ideal PCA HRTF dataset and here additionally also quite often for the Ideal PCA HRTF to MRA PCA HRTF dataset.

Dataset	Distance metric	Highest loading PC 1	Highest loading PC 2
Ideal to MRA PCA HRTF	D_{ISSD}	0	11
	D_{MSE}	42	5
	D_{CB}	1	0
	D_{MFCD}	4	10
	D_{LSSE}	0	21
Real to MRA PCA HRTF	D_{ISSD}	0	2
	D_{MSE}	1	42
	D_{CB}	0	0
	D_{MFCD}	46	1
	D_{LSSE}	0	2
Real to Ideal PCA HRTF	D_{ISSD}	0	0
	D_{MSE}	0	1
	D_{CB}	0	0
	D_{MFCD}	47	0
	D_{LSSE}	0	46

Table 5.3. Principal component analysis performed on 5 x 2304 matrix for each person (47 in total), i. e. five different distance metric values for each direction. The highest loadings (coefficients) for first two PCs were recorded.

5.1.2 Factor Analysis

In the second part of the analysis of the interrelations and information provided by distance metrics between different HRTF datasets, factor analysis model was utilized. It attempts to describe the variables (distance metrics) in dependence on a smaller number of latent factors. They are also known as "common factors" as they may affect several variables in common. For each distance metric specific variance was observed, it describes an independent random variability that can't be described by the common factors. We are looking for metrics that show high specific variation as those metrics shall provide the most information outside of the factor analysis model and shall be less dependent on the other metrics. The specific variation for each distance metric (for mean over all directions values resulting into a 5 x 47 matrix for each dataset, respectively) for all three analysed datasets can be found in Table 5.4. The metric ISSD shows the highest specific variation for two datasets and the second highest for the third dataset and i. e. MFCD shows the second highest specific variation for the first two datasets and the highest specific variation for the third dataset. Therefore they can be considered as important and providing the most information outside of the model. LLSE metric shows the lowest specific variation for three datasets out of three.

	Specific variation (-)		
Dataset	Ideal to MRA PCA HRTF	Real to MRA PCA HRTF	Real to Ideal PCA HRTF
D_{ISSD}	0.81	0.68	0.33
D_{MSE}	0.01	0.13	0.29
D_{CB}	0.34	0.30	0.02
D_{MFCD}	0.42	0.48	0.78
D_{LSSE}	0.14	0.01	0.01

Table 5.4. Factor analysis performed on mean over all direction values for different distance datasets - specific variation not explained by common factor for each distance metric.

The loadings for the latent factors are displayed in Figure 5.2. CB and MSE metrics have always the highest loadings for the same latent factor and therefore cover almost the same information. This makes sense as CB is computed as MSE with frequency weighting, based on critical bandwidths (see section 3.3.1). ISSD and MSE (and hence CB) have always the highest loading for different latent factors and therefore probably cover different information. LLSE and ISSD have for two datasets out of the three examined datasets the highest loading for the same latent factor and also probably provide similar information. MFCD has similar loading for both underlying latent factors and thus can't be described by by a singe latent factor sufficiently.

The factor analysis was also performed on distance metrics datasets with values for each direction (5 x 2304 matrix for each person, 5 distance metrics for 2304 directions - 47 matrices in total for each dataset). In Table 5.5 it is recorded which metric has the highest specific variation as well as for which latent factor has the metric the highest loading (whether the distance metric for given person has the highest loading for F1 or F2, i. e. can be explained by the first or rather the second latent factor). Again, as for the mean over all directions distance values, we can observe that ISSD and MFCD have the highest specific variations.

FA on mean over all direction values dataset.

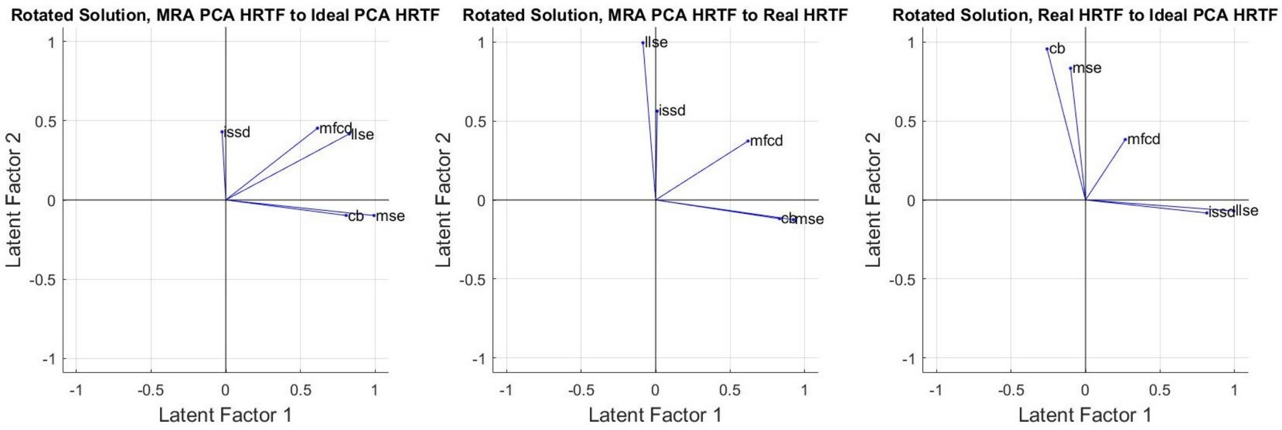

Figure 5.2. Biplot of loadings of different distance metrics for two underlying factors. Factor analysis was performed for each distance metric dataset.

For all datasets, we can also observe that ISSD and MSE have the highest loadings for different latent factors. MSE and CB, also as in the mean over all directions distance values, have mostly the highest loadings for the same latent factors and therefore cover the same information. For the Real HRTF to Ideal PCA HRTF dataset, LLSE shows the highest loading for the same latent factor as ISSD. Hence, LLSE might provide similar information as ISSD even for this dataset. This dataset also shows the smallest distance metric values as can be seen in Table 5.1 and LLSE shows the highest loading for the second PC - Table 5.3.

Dataset		Highest load. F1	Highest load. F2	Highest spec. var.
Ideal to MRA PCA HRTF	D_{ISSD}	7	40	33
	D_{MSE}	40	7	0
	D_{CB}	40	7	0
	D_{MFCD}	12	35	9
	D_{LSSE}	8	39	5
Real to MRA PCA HRTF	D_{ISSD}	7	40	30
	D_{MSE}	40	7	0
	D_{CB}	40	7	1
	D_{MFCD}	10	37	13
	D_{LSSE}	7	40	3
Real to Ideal PCA HRTF	D_{ISSD}	31	16	41
	D_{MSE}	16	31	3
	D_{CB}	14	37	0
	D_{MFCD}	10	37	3
	D_{LSSE}	31	16	0

Table 5.5. Factor analysis performed on 5 x 2304 matrix for each person (47 in total), i. e. five different distance metric values for each direction. Whether the metric had higher loading for underlying factor 1 or 2 was recorded as well as the highest specific variation from all distance metrics.

5.1.3 Correlation Analysis

Finally, correlation analysis was performed on all datasets. The correlation coefficients were computed for all three HRTF distance datasets both for mean over all directions distance values datasets and for datasets that provide distance metrics for each direction. Both Spearman's and Pearson's correlation coefficients were estimated. Here, mostly the findings using Spearman's correlation coefficient are presented and discussed as most of the metrics do not conform to the normal distribution. That can be seen from the histograms in the correlation plots (Figures 5.3, 5.5, 5.4) for the mean over all directions distance metric values. Correlation plots with Pearson's coefficient can be found in Appendix C.

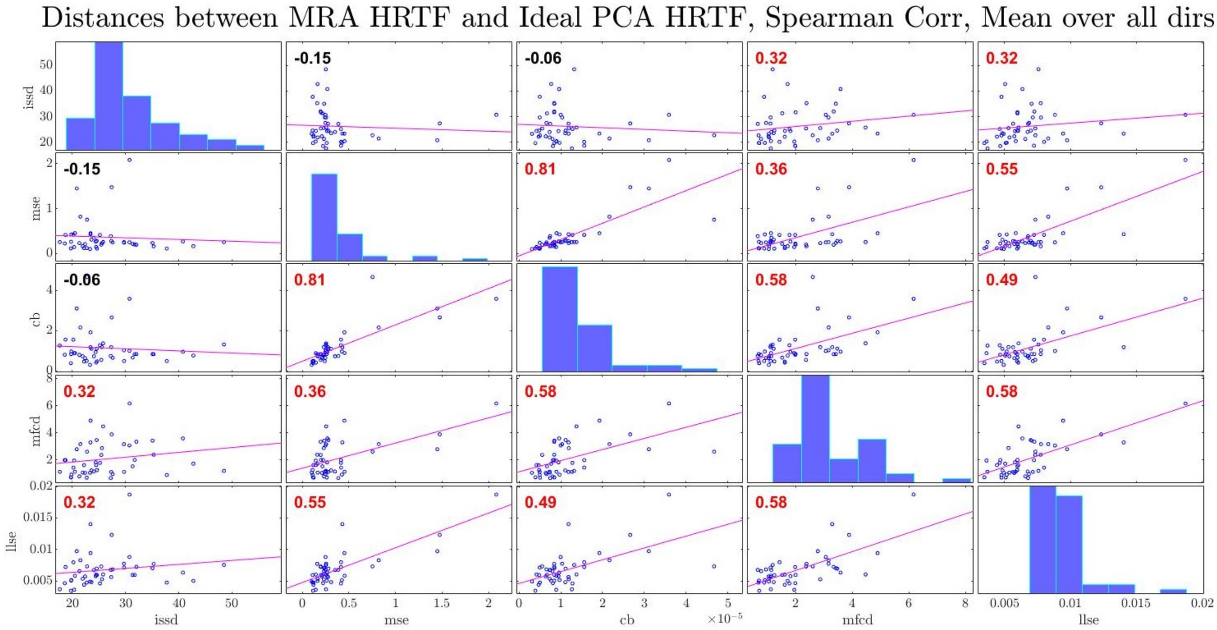

Figure 5.3. Correlation plot between all mean over all direction distance metric values - Spearman's coefficient, MRA PCA HRTF to Ideal PCA HRTF dataset. Significant correlation coefficients are displayed in red.

In all mentioned figures it can be shown that metric CB is highly correlated with MSE (the correlation coefficient is way higher than 0.5), which also corresponds with findings from the factor analysis. MFCD metric shows low correlation to MSE and ISSD metric (except for Real HRTF to Ideal PCA HRTF dataset) and seems to be correlated with CB and LLSE distance metrics. The correlation coefficient can for some cases be significant (displayed in red) however, a low correlation around 0.3 does not have such a high effect in predicting one distance metric based on the other. LLSE metric also seems to be correlated with most of the other metrics but lower than MFCD, which makes MFCD metric preferable for further considerations and analysis.

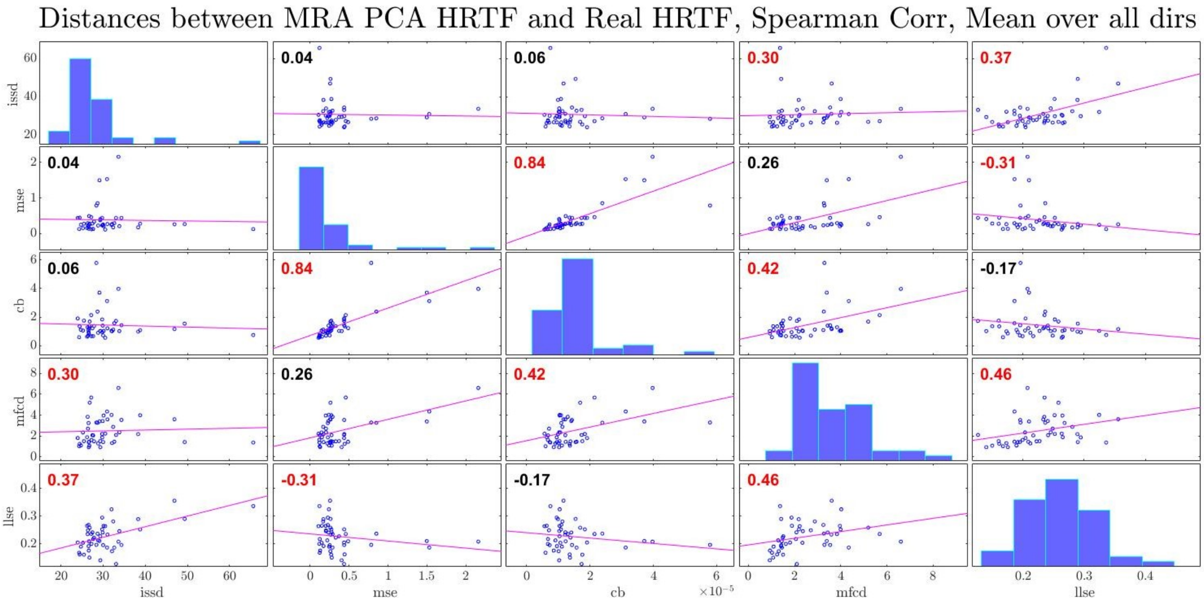

Figure 5.4. Correlation plot between all mean over all direction distance metric values - Spearman's coefficient, Real HRTF to MRA PCA HRTF dataset. Significant correlation coefficients are displayed in red.

In Figure 5.5 for the Real HRTF to Ideal PCA HRTF dataset it can be observed that LLSE metric is highly correlated with ISSD. That makes it also possible to leave out LLSE metric for further considerations as for this dataset LLSE showed the highest loading for the second PC and hence seemed more important.

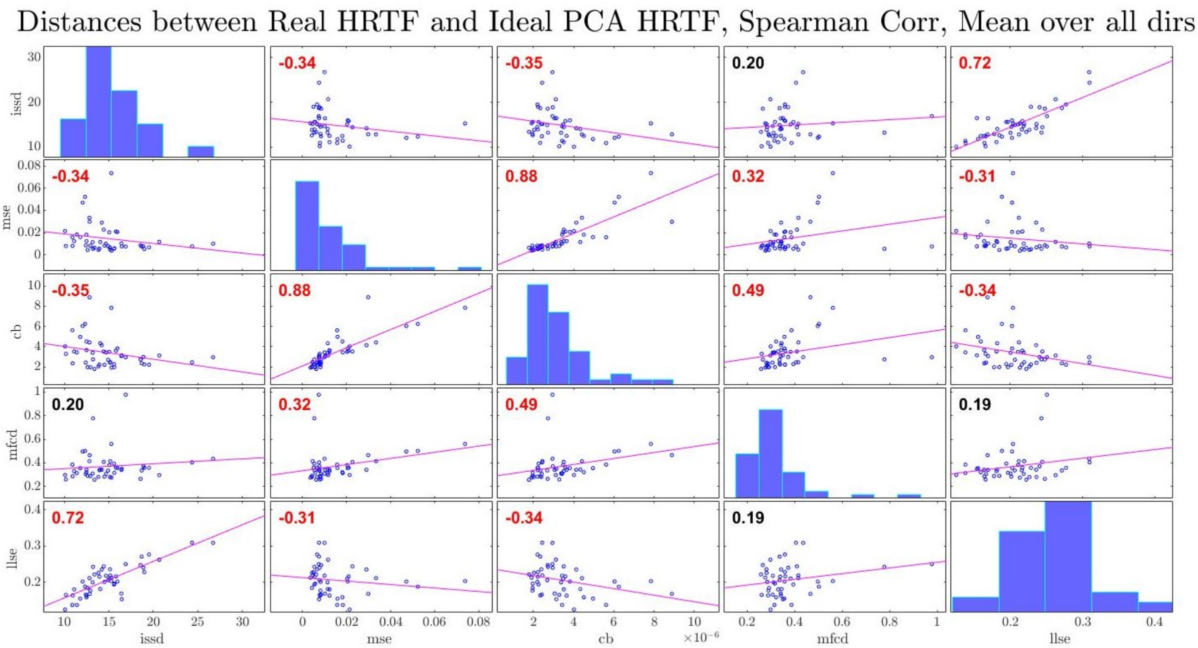

Figure 5.5. Correlation plot between all mean over all direction distance metric values - Spearman's coefficient, Real HRTF to Ideal PCA HRTF dataset. Significant correlation coefficients are displayed in red.

Further, the correlation analysis was performed on distance metrics datasets with values for each direction (5 x 2304 matrix for each person, 5 distance metrics for 2304 directions - 47 matrices in total for each dataset). Spearman's correlation coefficients were examined as the Spearman's coefficient is more robust to outliers and can determine the monotonous relationship between variables, while Pearson's coefficient tends to detect just linear relationships. It was decided to track significant correlation coefficients higher than 0.5 or lower than -0.5 because almost all computed correlation coefficients between two distance metrics were significant. The reason for that being a high number of input data (2304 values for each distance metric), therefore enough evidence existed even for a really low correlation between variables. Besides, low correlation does not provide enough information about the other variable and might not be that helpful in predicting one variable based on the other. In Tables 5.6, 5.7, 5.8 the significant Spearman's correlation coefficients higher than 0.5 or lower than -0.5 between different distance metrics are displayed.

	D_{ISSD}	D_{MSE}	D_{CB}	D_{MFCD}	D_{LSSE}
D_{ISSD}	-	6	1	8	9
D_{MSE}	6	-	47	0	0
D_{CB}	1	47	-	3	0
D_{MFCD}	8	0	3	-	20
D_{LSSE}	9	0	0	20	-

Table 5.6. Significant Spearman's correlation higher than 0.5 or lower than -0.5 between different distance metrics performed between two distance metrics on values for each direction (5 x 2304 matrix for each person, 5 distance metrics for 2304 directions - 47 matrices in total for each dataset), MRA PCA HRTF to Ideal PCA HRTF dataset.

	D_{ISSD}	D_{MSE}	D_{CB}	D_{MFCD}	D_{LSSE}
D_{ISSD}	-	3	0	8	10
D_{MSE}	3	-	47	0	36
D_{CB}	1	47	-	1	24
D_{MFCD}	8	0	1	-	12
D_{LSSE}	10	36	24	12	-

Table 5.7. Significant Spearman's correlation higher than 0.5 or lower than -0.5 between different distance metrics performed between two distance metrics on values for each direction (5 x 2304 matrix for each person, 5 distance metrics for 2304 directions - 47 matrices in total for each dataset), Real HRTF to MRA PCA HRTF dataset.

	D_{ISSD}	D_{MSE}	D_{CB}	D_{MFCD}	D_{LSSE}
D_{ISSD}	-	9	0	0	16
D_{MSE}	9	-	38	0	35
D_{CB}	0	38	-	35	1
D_{MFCD}	0	0	35	-	3
D_{LSSE}	16	35	1	3	-

Table 5.8. Significant Spearman's correlation higher than 0.5 or lower than -0.5 between different distance metrics performed between two distance metrics on values for each direction (5 x 2304 matrix for each person, 5 distance metrics for 2304 directions - 47 matrices in total for each dataset), Real HRTF to Ideal PCA HRTF dataset.

It can be noted that CB and MSE distance metrics are highly correlated for all datasets (some sparse exceptions can be found for Real HRTF to Ideal PCA HRTF dataset). LLSE distance metric seems to be correlated with almost all other metrics mainly for Real HRTF to MRA PCA HRTF and Real HRTF to Ideal PCA HRTF datasets and thus should not provide much new information.

ISSD metric shows overall low correlation to other distance metrics, the highest number of significant high correlations coefficients being for Real HRTF to Ideal PCA HRTF dataset for LSSE metric. That also corresponds to analysis for the same dataset for mean over all directions distance metric values. MFCD and MSE also present a low correlation between each other and ISSD.

5.1.4 Conclusion

On the basis of the analyses carried out ISSD, MSE and MFCD metrics shall provide the most diverse and varied information about differences between pairs of HRTFs. It shall hold true both for directional and mean over all directions dataset. These metrics shall be the least correlated in between. ISSD shall show the highest specific variation that could not be explained by "common factors". Also, ISSD and MSE should be correlated with different latent factors. MFCD and MSE showed also the highest loadings for the first PC in principal component analysis. These three distance metrics were used for stimuli selection as described in section 4.3.2 and later examined w. r. t. perceptual influence in section 5.2.

5.2 Listening Test Findings

The listening test was performed as described in section 4.3.3. The whole concept and design based on the findings from the previous sections are described in section 4.3. The results of the listening test are divided into two parts. The first one is concerned with findings of JND of the selected distance metrics. The second one discusses the additional questions, i. e. whether perceptual attributes can be predicted by distance metric values and if there is a link between perceptual attributes and any independent variable as introduced in section 4.3.1.

5.2.1 JND Test

The proposed listening test was mainly aimed as a JND test for the selected distance metrics (ISSD, MSE and MFCD). The test was meant as a pre-study to gain first insights into the audible threshold of these metrics and into their connection to perceptual attributes. The paradigm of the JND test was the 3AFC experiment, i. e. the participant was presented with three stimuli and they should determine the odd one. The stimuli were pre-selected during the design of the JND test so that they shall cover the range of the slope of the psychometric function, which is usually used to determine the audible threshold. The stimuli were preselected in such a way that the range of the slope of the psychometric function should have been covered for all three distance metrics. The audible threshold shall be at the point where for 2/3 of trials the odd stimulus was determined correctly. The distance metrics were computed for the datasets mentioned in section 4.1 - all datasets for intermodal as well as intramodal comparisons were included. Based on the pre-selection of the stimuli, only intramodal and intermodal datasets obtained for Ideal PCA HRTF and MRA PCA HRTF datasets were considered for the study. The choice of this subset is explained in section 4.3.2.

The JND for the distance metrics could be examined for each combination of the independent variables mentioned in Section 4.3.1. For each condition, 6 stimuli were presented and thus could be used for the threshold determination. The conditions are illustrated in Table 5.9.

Mirroring	Side	Comparison
Yes	Ipsi	Intermodal (Ideal to MRA)
No	Ipsi	Intermodal (Ideal to MRA)
Yes	Contra	Intermodal (Ideal to MRA)
No	Contra	Intermodal (Ideal to MRA)
Yes	Ipsi	Intramodal (Ideal)
No	Ipsi	Intramodal (Ideal)
Yes	Contra	Intramodal (Ideal)
No	Contra	Intramodal (Ideal)
Yes	Ipsi	Intramodal (MRA)
No	Ipsi	Intramodal (MRA)
Yes	Contra	Intramodal (MRA)
No	Contra	Intramodal (MRA)

Table 5.9. Combination of the independent variables for the listening test.

The mirrored and normal playbacks were displayed in the same plot as the distance metric values were the same. As can be seen in Figures 5.6, 5.7, C.4, C.5, C.6, C.7 the high values for one distance metric do not mean high values for other metrics as well (the stimuli/directions are distinguished by colors). Therefore, if we can see a trend resembling a psychometric function (e. g. in figure 5.6 for MSE metric) for one metric, it is very likely that we won't see such a trend for other metrics. The idea is that the audible threshold could be given by a linear or other combination of these metrics.

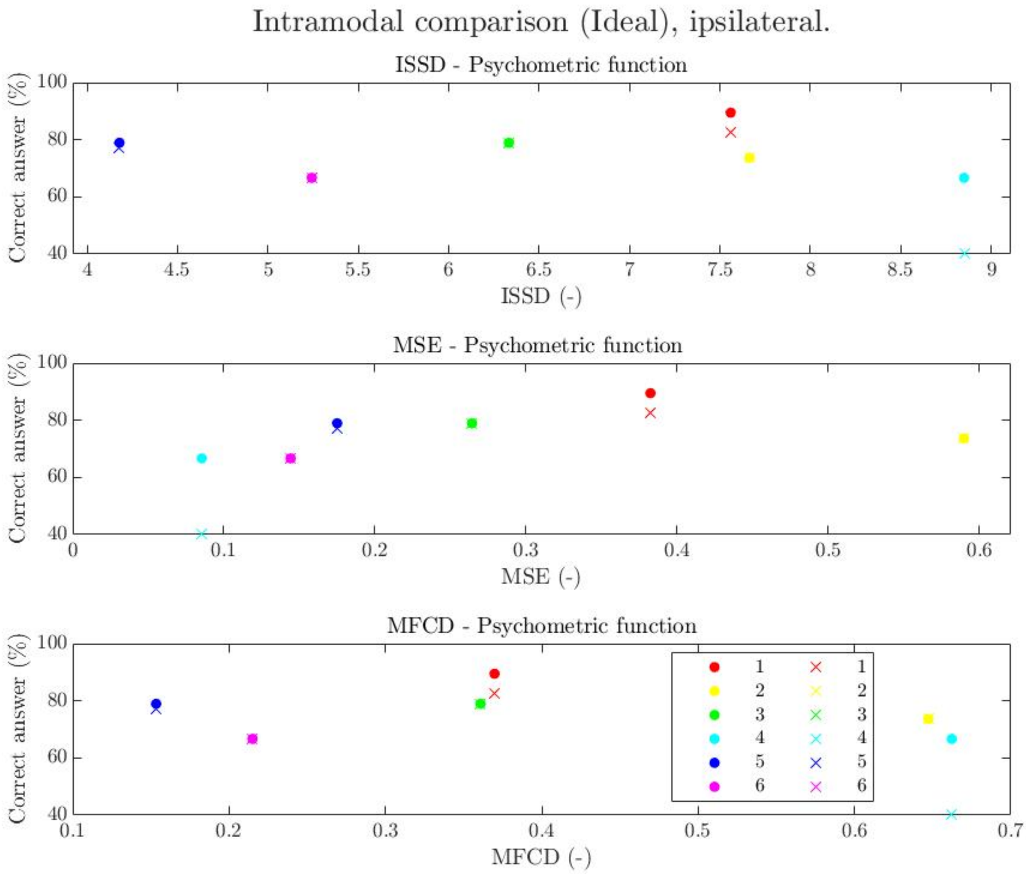

Figure 5.6. Probability of correct answer for three distance metrics. Intramodal comparison (Ideal PCA HRTF dataset) for normal (dots) and mirrored (crosses) playback, of stimuli optimized for ipsi-lateral directions.

It can be noticed that for most independent variable combinations there is a difference between correct answers for mirrored and normal playback. The biggest difference can be noted in Figure 5.7. However, there should be a right ear advantage, and thus more differences determined correctly should appear for the mirrored playback for the ipsi-lateral side (the source originally being on the left and after mirroring on the right side) and for normal playback for the contra-lateral side (the source being on the right side). That might be the case observing Figure C.5 (ipsi-, mirrored, more differences correctly determined) and e. g. Figure C.4 (contra-, normal, more differences correctly determined), where we can see the described tendency. This tendency may be true for some kind of differences (e. g. different spatial and spectral cues for intra-modal cases). That might imply that we react better to sources on the right side. Other independent variable combinations can be found in the Appendix, however, these findings can apply to all combinations of independent variables.

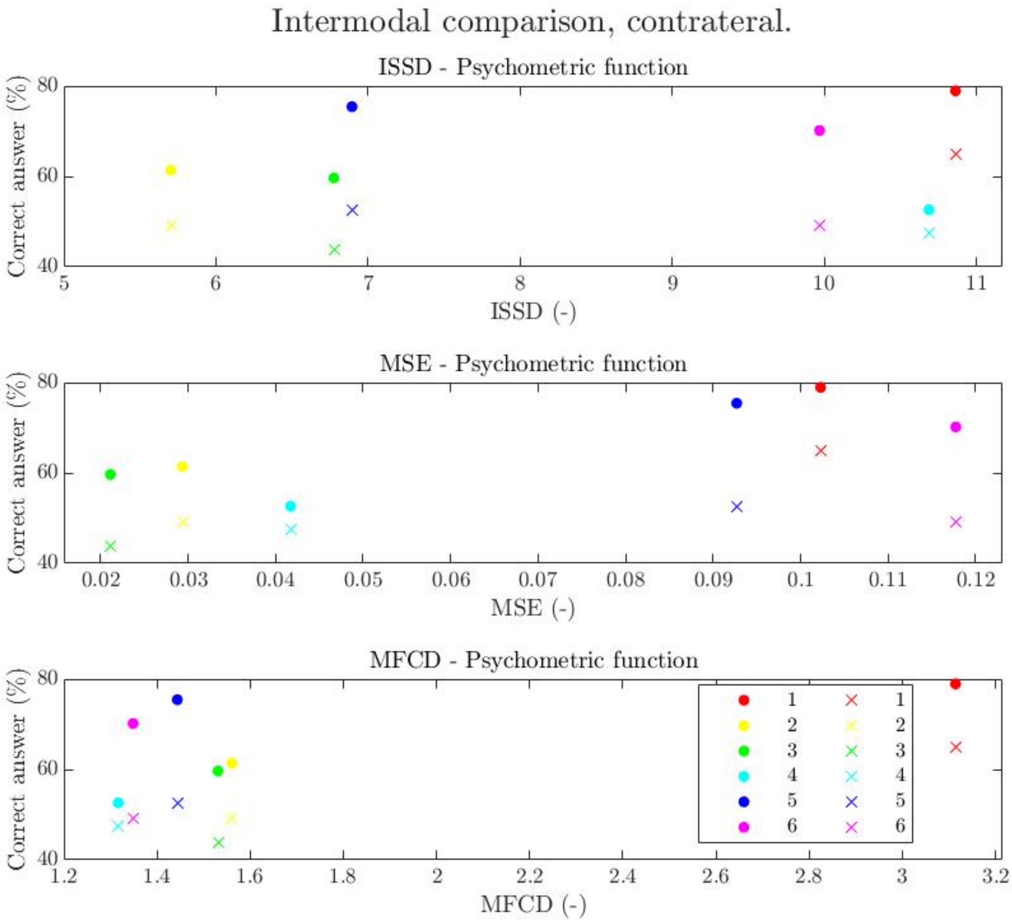

Figure 5.7. Probability of correct answer for three distance metrics. Intermodal comparison (Ideal PCA HRTF to MRA PCA HRTF dataset) for normal (dots) and mirrored (crosses) playback, of stimuli optimized for contra-lateral directions.

5.2.2 Additional questions related to perceptual attributes

The additional questions introduced in the second block of the experiment are concerned with the prediction of perceptual attributes (coloration and localization) based on distance metric values. The link between perceptual attributes and any independent variable is also examined.

In Figure 5.8 the dependence between each distance metric and coloration/ localization is displayed. The scales to assess these differences went from 0 to 100, where 0 meant the stimuli do not differ at all in terms of the stimuli source location/coloration and 100 meant the stimuli differ extremely in terms of the stimuli source location/coloration. The scales were continuous and the participant was provided with an option to express that there exists no perceivable difference. It can be noted that almost all differences were assessed below point 20 on the scale. That makes sense as it was a JND test and thus the differences between the stimuli were really low. Overall, the perceived differences in coloration were higher than for localization. These attributes were examined for all independent variables together as we are interested in the common link between the metric and the attribute. For most of the metrics, we can observe no obvious link. However, there might be a link between MSE and coloration/localization. The relationship between MSE and localization/coloration was evaluated using the Spearman's correlation coefficient. A weak link was found for MSE and coloration $r(70) = 0.234, p = 0.048$ as well as for MSE and localization $r(70) = 0.24, p = 0.04$. However these results could be strongly influenced by outliers. Also after interviewing some of the participants, the problem with assessing stimuli that differ in loudness arose. Some of the participants judged a difference in loudness as difference in coloration whilst others as difference in localization. For the future study a question concerning the loudness difference shall be considered. Also the stimuli chosen as example stimuli for differences in coloration might have been to extreme and hence providing more example stimuli with different kinds of coloration change might be advantageous.

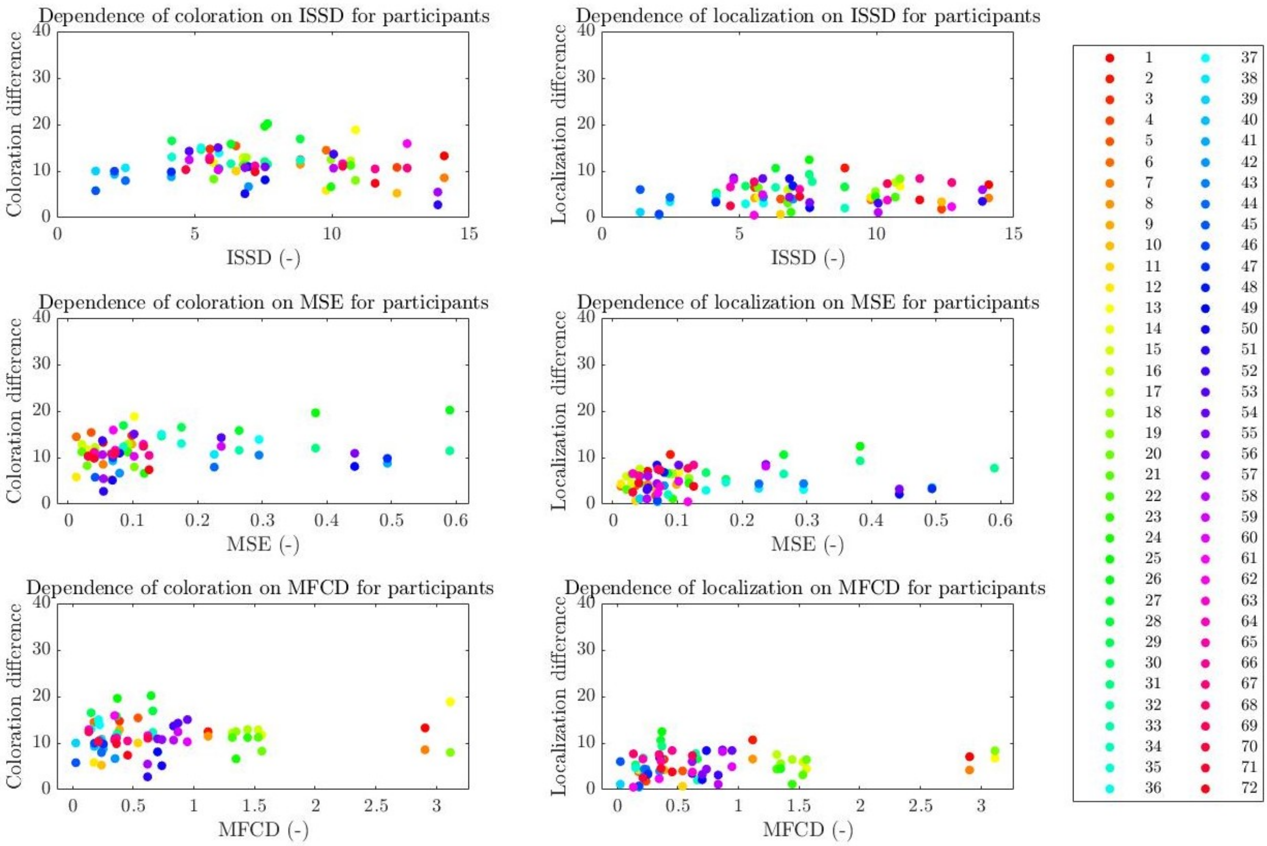

Figure 5.8. Dependence of coloration/localization on distance metrics (D_{ISSD}, D_{MSE}, D_{MFCD}) values. Different colors mark all 72 stimuli and allow for recognizing the same stimuli between subfigures.

Further, the relationship between independent variables and coloration/localization was analysed. In Figures 5.9, 5.10, 5.11 the boxplots for pairs of independent variables and localization/coloration differences are displayed. Again, it can be noted that the differences in coloration were overall higher than localization differences.

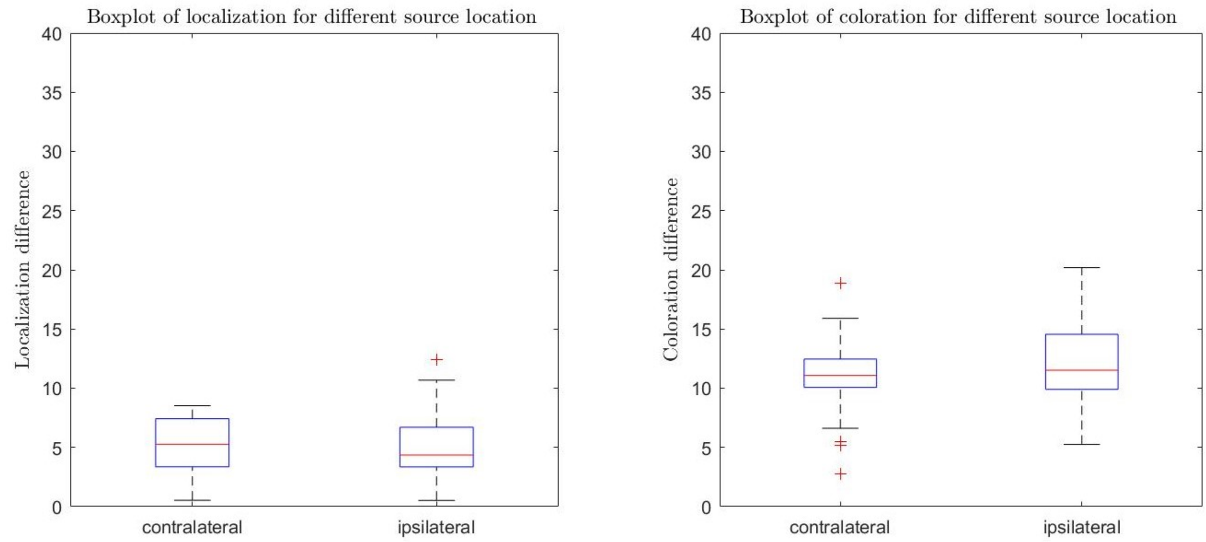

Figure 5.9. Dependence of coloration/localization on different stimuli source location.

The smallest difference in the perceivable attributes was probably for different stimuli source locations as can be seen in Figure 5.9. The broadest range of values can be observed for coloration assessment for stimuli from the ipsi-lateral side and surprisingly for mirroring of the stimuli source. The bigger changes in coloration for the ipsi-lateral side might be explained by the fact that there might be a more perceivable loudness level for the ipsi-lateral ear and the differences could be better assessed or the changes in loudness might be considered as changes in coloration.

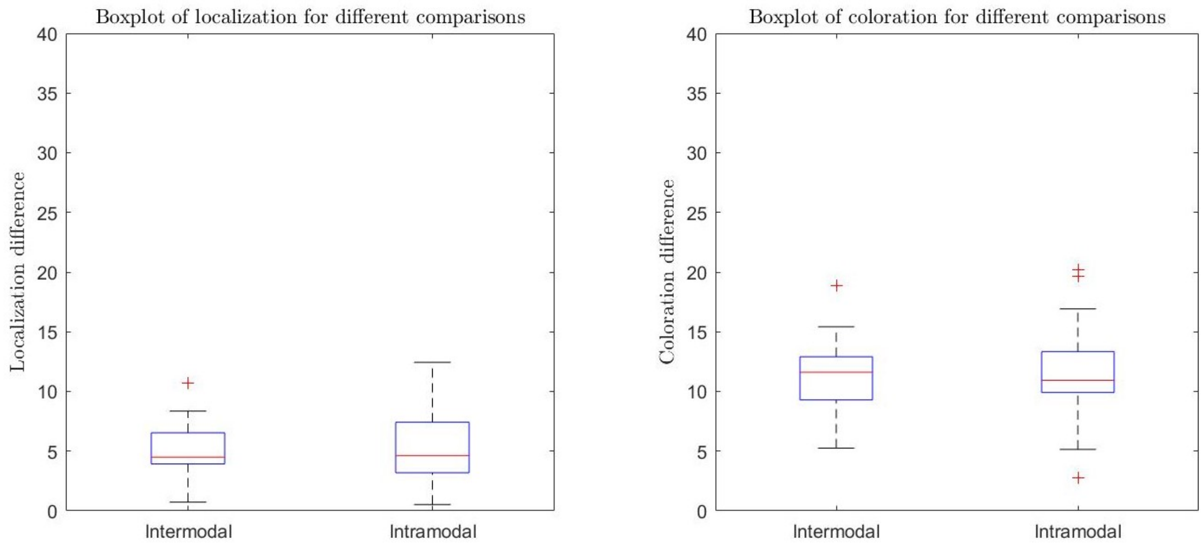

Figure 5.10. Dependence of coloration/localization on different comparisons.

The broadest range of values for localization assessment can be probably noted for intra-modal comparisons of HRTFs (Figure 5.10). That might suggest that there might be bigger localization differences when comparing HRTFs of the same modality for different people, rather than when comparing HRTFs of different modalities for the same person. However, further studies with bigger differences between stimuli shall be proposed in order to assess these differences.

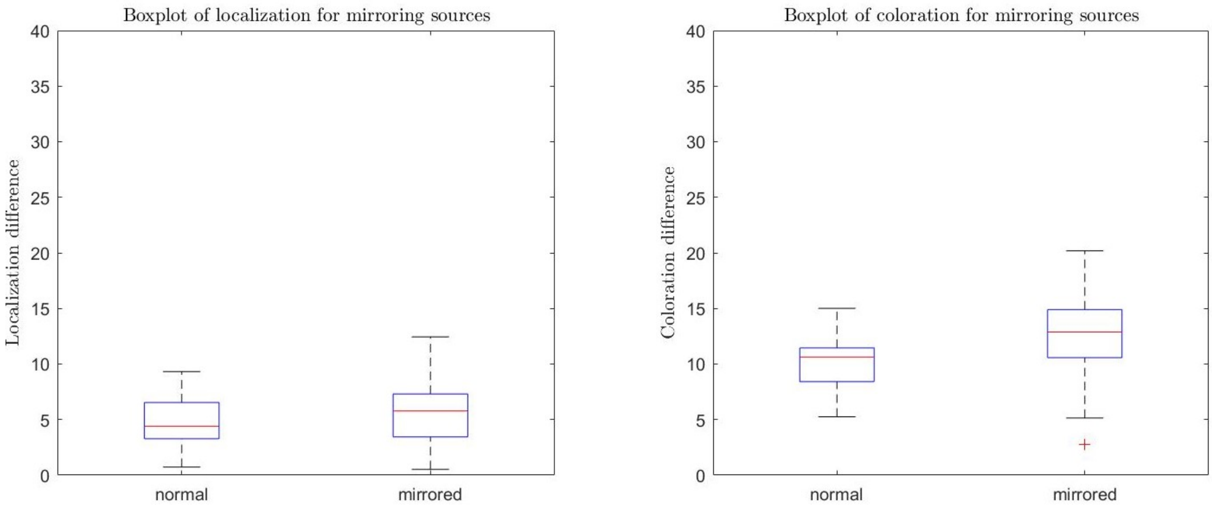

Figure 5.11. Dependence of coloration/localization on mirroring of the source stimuli.

The biggest difference in coloration and probably also in localization, although still not significant, might be observed for mirroring of the source stimuli in Figure 5.11. That might also correspond to findings from Section 5.2.1, where often more differences between stimuli were correctly determined for non-mirrored (normal) playback for the ipsi-lateral side and for mirrored playback for the contra-lateral side stimuli. Broader differences for coloration assessment can be observed (Figure 5.12) for mirroring sources for the stimuli directions form the ipsi-lateral side as for contra-lateral side (Figure 5.13) the differences were not so broad, confirming the theory that we can react better to sources on the right side. However, not so big differences could be observed for normal playback of the stimuli.

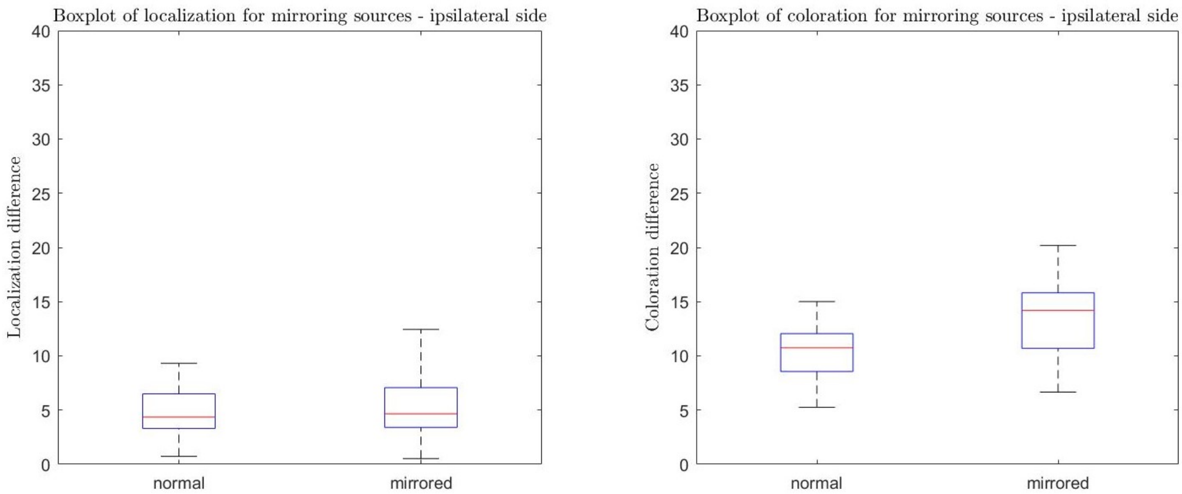

Figure 5.12. Dependence of coloration/localization on mirroring of the source stimuli and source location - ipsi-lateral side.

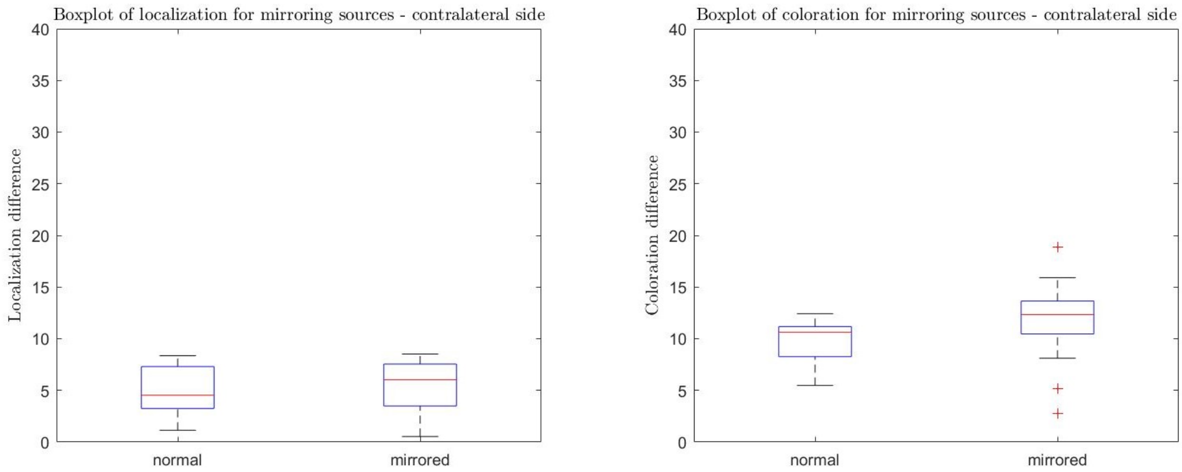

Figure 5.13. Dependence of coloration/localization on mirroring of the source stimuli and source location - contra-lateral side.

Chapter 6

Summary and Outlooks

The present work focused on the comparison of head-related transfer functions (HRTFs). Different levels of individualization of HRTFs exist and their comparison is enabled by various distance metrics. The thesis consists of two main parts. In the first part, the behaviour and interrelations of several objective distance metrics for HRTFs were inspected. In the second part, based on these findings a listening experiment was proposed to examine the audible threshold for given objective distance metrics. Furthermore, the test also assesses the additional questions whether perceptual attributes can be predicted by distance metric values and whether there exists a link between subjective perceptual attributes and any independent variable as mentioned in section 4.3.1 (i. e. different comparisons of HRTFs, mirroring of the source stimuli, different sides of the directions).

Five distinct distance metrics were considered for the analysis of the provided mutual information and their interrelations. Only metrics focused on directional differences were assessed, meaning that a direct comparison of spectra led to a set of distance values for each sound incidence direction, respectively. Three metrics (MFCC, CB-MSE and LLSE) rely on psychoacoustic effects, while the other two (MSE and ISSD) display mostly numerical errors. To analyze the interaction of different distance metrics for HRTF comparison, the distance metrics were computed between HRTFs of different modalities (Real measured HRTFs: “Real HRTF”, reconstructed HRTFs using “ideal” weights PCA: “Ideal PCA HRTF” and reconstructed HRTFs using MRA of principal components with “reconstructed” weights: “MRA PCA HRTF”, as described in section 4.1).

The range of values for the above mentioned distance metrics for different comparisons of HRTFs was inspected. The highest values could be observed for comparison of Real HRTF to MRA PCA HRTF dataset and the lowest values for comparison of Real HRTF to Ideal PCA HRTF. That corresponds to the idea that MRA of PC is an individualization method using anthropometric values and thus an approximation of PC reconstruction with ideal weights, which for its part is an approximation of the measured data. Hence the biggest differences were observed between the twice approximated data (MRA PCA HRTF) to the measurement (Real HRTF).

The analysis of the interaction and interrelations between the mentioned distance metrics was performed on the set of directional distance metrics between HRTFs as well as on the set of weighted mean distance measures, where there is one distance value between a pair of HRTF sets. On the basis of principal component analysis, factor analysis and correlation analysis ISSD, MSE and MFCD have shown to provide the most diverse and varied information. These metrics are the least correlated and carry the information of the other metrics analysed. MSE and MFCD showed the highest loadings for the first PC in principal component analysis. ISSD and MSE are correlated with different “common” factors and ISSD shall as well show the highest specific variation that could not be explained by them.

A listening test was then proposed and conducted to examine the audible threshold for ISSD, MSE and MFCD distance metrics. The test was implemented using MATLAB App GUI, the proposed paradigm was the 3AFC experiment. The audible threshold was examined for different conditions of independent variables as described in table 5.9 (i. e. combinations of different comparisons of HRTFs, mirroring of the source stimuli and different sides of the directions). It could be observed that high values for one distance metric do not mean high values for other metrics. In consequence, if we observe a trend resembling a psychometric function for one metric, it is very likely that we won't see such a trend for other metrics. Therefore the idea is that the audible threshold could be given by a linear or other combination of these metrics. In further work, the model for determining the audible threshold shall be examined. It might also be good to consider the values of other distance metrics. It could also be observed that there is a difference between the proportions of correct answers for mirroring of the source stimuli for different conditions. This difference could correspond to the right ear advantage presumption that listening with the right ear might show different behaviour than to listening compared stimuli with the left ear.

In the second part of the listening test, additional questions regarding subjective perceptual attributes (localization and coloration) were introduced. To assess these questions, unipolar continuous scales were used with a range from 0 to 100. The overall perceivable differences were low as the designed test was mainly a just noticeable difference test. However, the perceived differences were higher for coloration than for localization. There was a weak link observed between the MSE distance metric and coloration and localization, yet these results might be strongly influenced by outliers. Besides, by examining the relationships between independent variables and perceptual attributes, a broader range of coloration values for stimuli with ipsi-lateral sound incidence side could be observed, corresponding to the fact that for stimuli with ipsi-lateral sound incidence side might have higher loudness level and differences could be better assessed. Also a broader range of coloration values was observed for sources mirrored to the right side of the head and stimulus variation of the right ear signal, confirming the theory that we can react better to sources on the right side.

Based on feedback received from participants, a question concerning the difference in loudness shall be introduced in a follow-up experiment, as some participants judged a difference in loudness to contribute to the difference in coloration whilst others as a difference in localization. More example stimuli of different kinds of coloration changes should be presented before the start of the experiment. Stimuli covering a broader range of values could be introduced to assess a wider range of differences in perceptual attributes. Some conditions (combinations of independent variables as described in table 5.9) might be eliminated in the future as it might not be necessary to cover them all and in order to gain more values for the statistical evaluation. Further, for later examination of HRTFs it would be favourable to have one distance metric for binaural signal, as for now we have been using binaural signals but have been evaluating differences of one ear only. A weighting function to gain one distance metric for both ears might be proposed and verified in future listening experiments.

References

[1] Florian Pausch, and Lukas Aspöck. A hybrid real-time auralization system for binaural reproduction of virtual acoustic environments with simulated room acoustics adapted for hearing aid users. *The Journal of the Acoustical Society of America*. 2018, 143 DOI https://doi.org/10.1121/1.5035989.

[2] Janina Fels, and Iring Koch Josefa Oberem. Examining auditory selective attention in realistic, natural environments with an optimized paradigm. *The Journal of the Acoustical Society of America*. 2016, 28 DOI https://doi.org/10.1121/2.0000321.

[3] Ramona Bomhardt, Hark Braren, and Janina Fels. *Individualization of head-related transfer functions using principal component analysis and anthropometric dimensions.* In: Acoustical Society of America, 2016.

[4] Alexis Bondu, Rozenn Nicol, Vincent Lemaire, and Sylvain Busson. *Looking for a Relevant Similarity Criterion for HRTF Clustering: A Comparative Study.* In: 2006.

[5] Ki-Seung Lee, and Seok-Pil Lee. A Relevant Distance Criterion for Interpolation of Head-Related Transfer Functions. *Audio, Speech, and Language Processing, IEEE Transactions on*. 2011, 19 1780 - 1790. DOI 10.1109/TASL.2010.2101590.

[6] Jan-Gerrit Richter. *Fast measurement of individual head-related transfer functions*. Berlin: Logos Verlag Berlin GmbH, 2019. ISBN 9783832549060.

[7] J. Richter, G. Behler, and J. Fels. Evaluation of a Fast HRTF Measurement System. *Paper 9498*. 2016,

[8] S. Shimada, N. Hayashi, and S. Hayashi. A Clustering Method for Sound Localization Transfer Functions. *Audio Eng. Soc.*. 1994, 42 (7/8), 577-584.

[9] Xiao-Li Zhong, Feng-chun Zhang, and Bo-Sun Xie. On the spatial symmetry of head-related transfer functions. *Applied Acoustics*. 2013, 74 (6), 856–864. DOI 10.1016/j.apacoust.2013.01.004.

[10] Jyri Huopaniemi Ville Pulkki, Matti Karjalainen. Analyzing Virtual Sound Source Attributes Using a Binaural Auditory Model. *JAES*. 1999,

[11] Christopher A. Shera, John J. Guinan, and Andrew J. Oxenham. Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proceedings of the National Academy of Sciences*. 2002, 99 (5), 3318–3323. DOI 10.1073/pnas.032675099.

[12] Hugo Fastl Zwicker Eberhard. *Psychoacoustics - Facts and Models*. 2 edition. Springer, 1999. ISBN 978-3-662-09562-1.

[13] C.B. Pease. Combining the sone and phon scales. *Applied Acoustics*. 1974, 7 (3), 167-181. DOI https://doi.org/10.1016/0003-682X(74)90011-5.

[14] *Speech and Audio Processing in Adverse Environments*. 1 edition. Springer, 2008. ISBN 978-3-540-70602-1.

[15] Jaroslav Bouše. *Models and experiments of binaural interactions*. Ph.D. Thesis, Czech Technical University. 2020.

[16] Jens Blauert. *Spatial hearing: the psychophysics of human sound localization*. 1997.

[17] Janina Fels. *Lecture Notes on Psychoacoustics and Methods for Listening Experiments*. Master's Thesis, IHTA, RWTH Aachen University. 2021.

[18] Lord Rayleigh O.M. Pres. R.S. XII. On our perception of sound direction. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. 1907, 13 (74), 214-232. DOI 10.1080/14786440709463595.

[19] Jens Blauert. *Spatial hearing: the psychophysics of human sound localization.* In: MIT Press, 2001.

[20] John C. Middlebrooks. Individual differences in external-ear transfer functions reduced by scaling in frequency. *The Journal of the Acoustical Society of America.* 1999, 106 (3), 1480–1492. DOI 10.1121/1.427176.

[21] Shaima'a Doma. *Development of a demonstration and evaluation tool for individualized binaural technology*. Master's Thesis, Institute of Technical Acoustics, RWTH Aachen University. 2018.

[22] Davide Ballabio. A MATLAB toolbox for Principal Component Analysis and unsupervised exploration of data structure. *Chemometrics and Intelligent Laboratory Systems*. 2015, 149 Part B 1-9. DOI 10.1016/j.chemolab.2015.10.003.

[23] Hervé Abdi, and Lynne Williams. Principal Component Analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*. 2010, 2 433 - 459. DOI 10.1002/wics.101.

[24] Kimberly Kempf-Leonard. *Encyclopedia of social measurement: Factor Analysis*. Elsevier Inc, 2005. ISBN 978-0124438903.

[25] L.L. Thurstone. *Multiple factor analysis.* University of Chicago Press: Chicago., 1947.

[26] Hannu Oja Georgy L. Shevlyakov. *Robust Correlation: Theory and Applications*. WILEY, 2016. ISBN 978-1-118-49345-8.
`https://www.ebook.de/de/product/24526532/georgy_l_shevlyakov_hannu_oja_robust_correlation_theory_and_applications.html`.

[27] Zekai Sen. *2.8.1 Delaney, Varoni, and Thiessen Polygons.* In: Springer, 2016. Spatial Modeling Principles in Earth Sciences.

[28] Christopher D. Lloyd Peter A. Burrough, Rachael McDonnell. *Principles of Geographical Information Systems.* In: Oxford University Press, 2015. 8.11 Nearest neighbours: Thiessen (Dirichlet/Voroni) polygons.

[29] G. H. Wakefield. C. I. Cheng. Introduction to head-related transferfunctions (hrtfs): Representations of hrtfs in time, frequency, and space.. *In Audio Engineering Society Convention 107.* 1999,

[30] Martin Pollow, Bruno Masiero, Pascal Dietrich, Janina Fels, and Michael Vorlaender. *Fast measurement system for spatially continuous individual HRTFs.* In: 2012.

[31] Martin Pollow, Khoa-Van Nguyen, Olivier Warusfel, Thibaut Carpentier, Markus Müller-Trapet, Michael Vorländer, and Markus Noisternig. Calculation of Head-Related Transfer Functions for Arbitrary Field Points Using Spherical Harmonics Decomposition. *Acta Acustica united with Acustica.* 2012, 98 (1), 72–82. DOI 10.3813/aaa.918493.

[32] Gerald Farin. Surfaces over Dirichlet tessellations. *Computer Aided Geometric Design.* 1990, 7 (1-4), 281–292. DOI 10.1016/0167-8396(90)90036-q.

[33] J. L. Brown. Natural neighbor interpolation on the sphere. *An international conference on curves and surfaces on Wavelets, images, and surface fitting.* 1994,

[34] Oscar Alberto Ramos, and Fabián Carlos Tommasini. Magnitude Modelling of HRTF Using Principal Component Analysis Applied to Complex Values. *Archives of Acoustics.* 2015, 39 (4), 477–482. DOI 10.2478/aoa-2014-0051.

[35] Ramona Bomhardt, Marcia Lins, and Janina Fels. Analytical Ellipsoidal Model of Interaural Time Differences for the Individualization of Head-Related Impulse Responses. *Journal of the Audio Engineering Society.* 2016, 64 (11), 882–894. DOI 10.17743/jaes.2016.0041.

[36] Matti Karjalainen Jyri Huopaniemi, Nick Zacharov . Objective and Subjective Evaluation of Head-Related Transfer Function Filter Design. *AES.* 1999,

[37] Cosima Ermert. *Examining the influence of controlled HRTF-variation on variousdistance measures.* Master's Thesis, RWTH Aachen University. 2020.

[38] Laurent S. R. Simon, Nick Zacharov, and Brian F. G. Katz. Perceptual attributes for the comparison of head-related transfer functions. *The Journal of the Acoustical Society of America.* 2016,

[39] Alexander Lindau, Vera Erbes, Steffen Lepa, Hans-Joachim Maempel, Fabian Brinkmann, and Stefan Weinzierl. A spatial audio quality inventory (SAQI). *Acta Acustica united with Acustica.* 2014, 100 DOI 10.3813/AAA.918778.

[40] Simon Ciba, André Wlodarski, and Hans-Joachim Maempel. WhisPER – A New Tool for Performing Listening Tests. *126th Audio Engineering Society Convention 2009.* 2012, 1

[41] Wichmann, Felix A., and Frank Jäkel. *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience.* In: John Wiley and Sons, 2018. Methods in psychophysics. ISBN 9781119170174.

[42] G. A. Gescheider. *Psychophysics: the fundamentals.* Psychology Press, 2013.

[43] S. A. Gelfand. *Hearing: An introduction to psychological and physiological acoustics.* CRC Press, 2004.

[44] G. A. Gescheider. *Psychoacoustic measuring methods.* 2007.

[45] *Kompendium zur durchfuhrung von horversuchen in wissenschaft und industrieller praxis.* 2008.

[46] Janina Fels Ramona Bomhardt, Matias de la Fuente Klein. A high-resolution head-related transfer function and three-dimensional ear model database. *Proceedings of Meetings on Acoustics.* 2017, 29 DOI https://doi.org/10.1121/2.0000467.

[47] Pascal Dietrich et al. Matlab toolbox for the comprehension of acoustic measurement and signal processing. *Fortschritte der Akustik–DAGA.*

[48] Clipart.com Princeton University. *timbre.* 2003-2008. `https://www.thefreedictionary.com/timbre`.

[49] Jerger J. The remarkable history of right-ear advantage. *Hearing Review.* 2018, 25(1):12-16

[50] Kaigham J. Gabriel, and H. Steven Colburn. Interaural correlation discrimination: I. Bandwidth and level dependence. *The Journal of the Acoustical Society of America.* 1981, 69 DOI https://doi.org/10.1121/1.385821.

[51] Trittipoe W. J. Pollack I. Binaural listening and interaural noise cross correlation. *The Journal of the Acoustical Society of America.* 1959, 31 DOI https://doi.org/10.1121/1.1907852.

[52] Bruno Masiero, and Janina Fels. Perceptually robust headphone equalization for binaural reproduction. *Audio Engineering Society Convention.* 2011, 130

Appendix A
Thesis proposal

- Analyze the interaction and mutual information between different objective distance metrics for HRTF comparison, using tools like correlation analysis or factor analysis.
- Reduce the given metrics to a smaller set of measures suitable for a listening experiment and covering both spatial and spectral aspects.
- Design a listening experiment paradigm to examine just noticeable differences and give first insights into various perceptual attributes (e.g. changes in sound coloration, spatial aspects and audible artifacts), which might be predicted by the respective distance metrics.

Appendix B
Lists of Acronyms

HRTF	Head-related transfer function.
ITD	Interaural time difference.
ILD	Interaural level difference.
LTI system	Linear time independent system.
HRIR	Head-related impulse response.
DTF	Directional transfer function.
PCA	Principal component analysis.
FA	Factor analysis.
MRA	Multiple regression analysis.
SHD	Spherical Harmonics Decomposition.
MSE	Mean Square Error.
CB-MSE	Critical Bands Mean Squared Error.
IISD	Inter-subject Spectral Difference.
CD	Correlation Distance.
MFCD	Mel-frequency Cepstral Distortion.
MFCC	Mel-frequency Cepstral Coefficients
SD	Spectral Difference.
LLSE	The Loudness Level Spectrum Error.
GTFB	Grammatone Filterbank.
ERB	Equivalent Rectangular Bandwith.
JND	Just Noticable Difference.
IACC	Interaural Cross-Correlation.

Appendix C
Additional figures

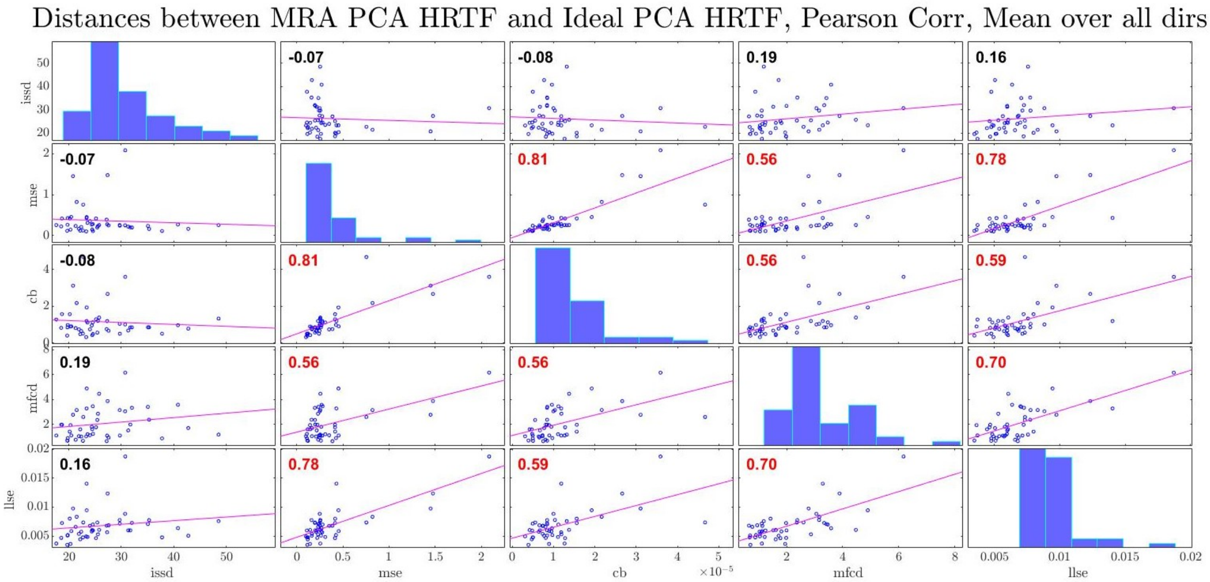

Figure C.1. Correlation plot between all mean over all direction distance metric values - Pearson's coefficient, MRA PCA HRTF to Ideal PCA HRTF dataset. Significant correlation coefficients are displayed in red.

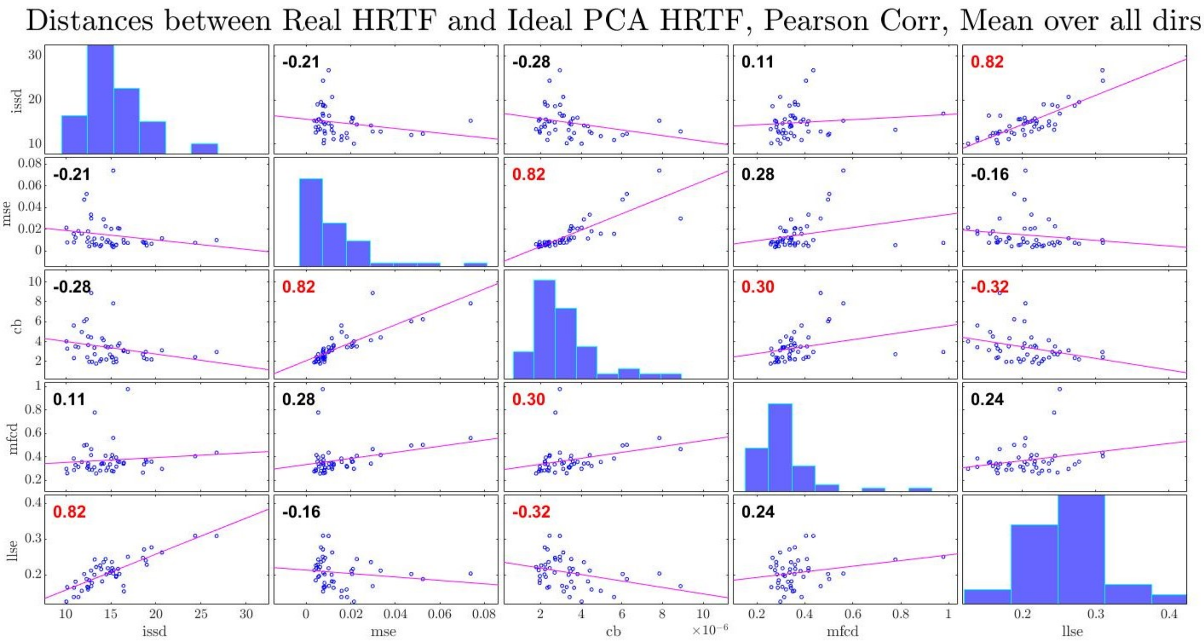

Figure C.2. Correlation plot between all mean over all direction distance metric values - Pearson's coefficient, Real HRTF to Ideal PCA HRTF dataset. Significant correlation coefficients are displayed in red.

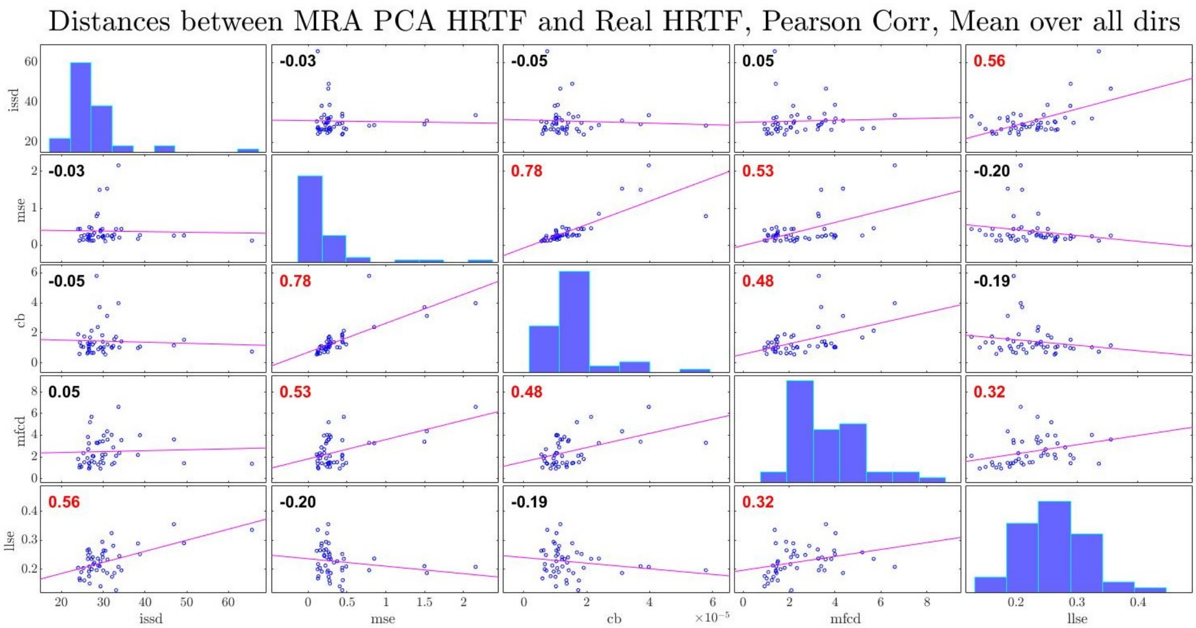

Figure C.3. Correlation plot between all mean over all direction distance metric values - Pearson's coefficient, Real HRTF to MRA PCA HRTF dataset. Significant correlation coefficients are displayed in red.

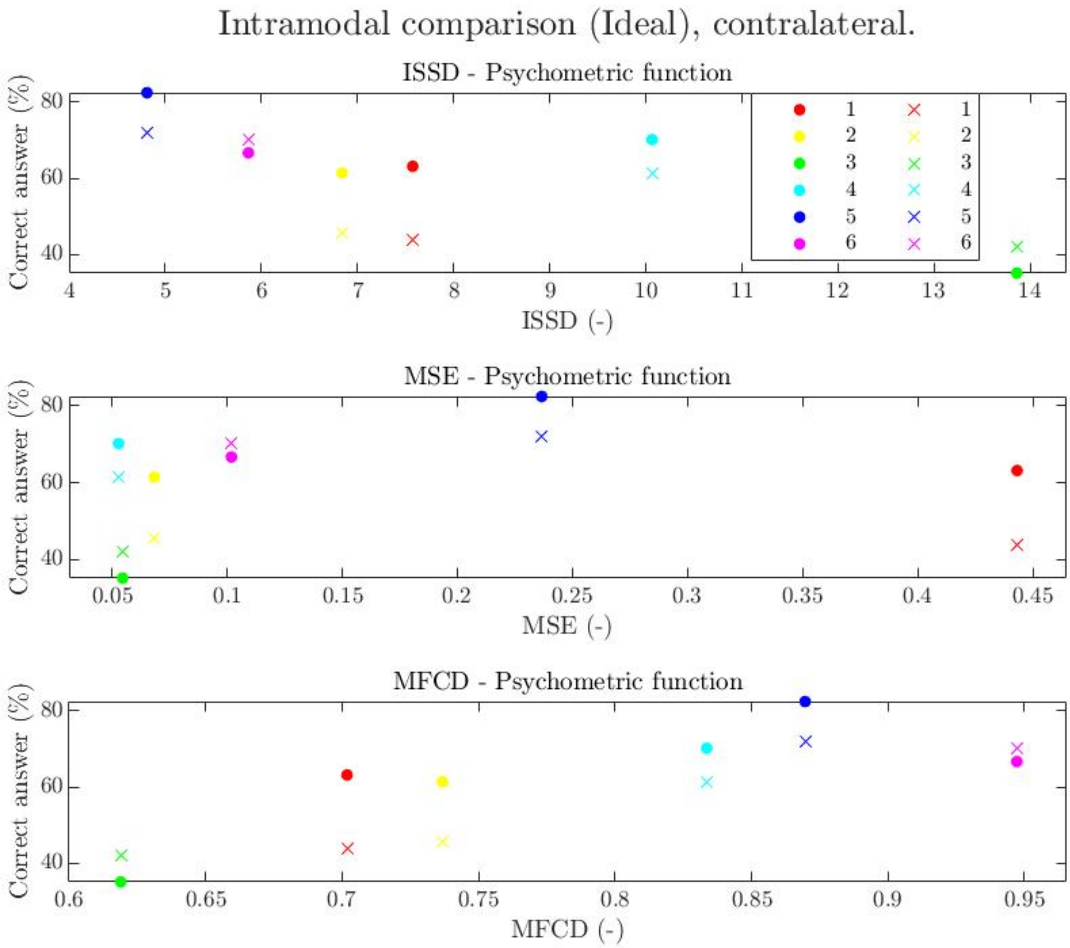

Figure C.4. Probability of correct answer for three distance metrics. Intramodal comparison (Ideal PCA HRTF dataset) for normal (dots) and mirrored (crosses) playback, of stimuli optimized for contra-lateral directions.

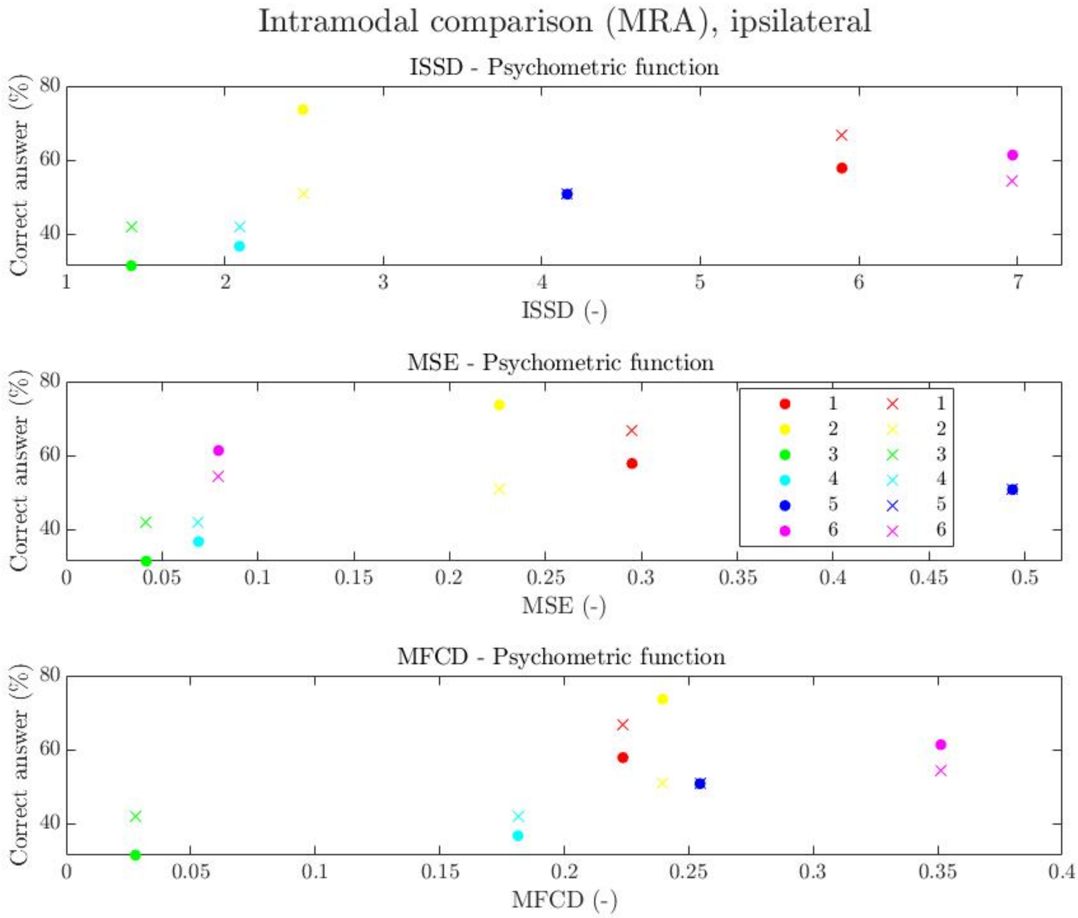

Figure C.5. Probability of correct answer for three distance metrics. Intramodal comparison (MRA PCA HRTF dataset) for normal (dots) and mirrored (crosses) playback, of stimuli optimized for ipsi-lateral directions.

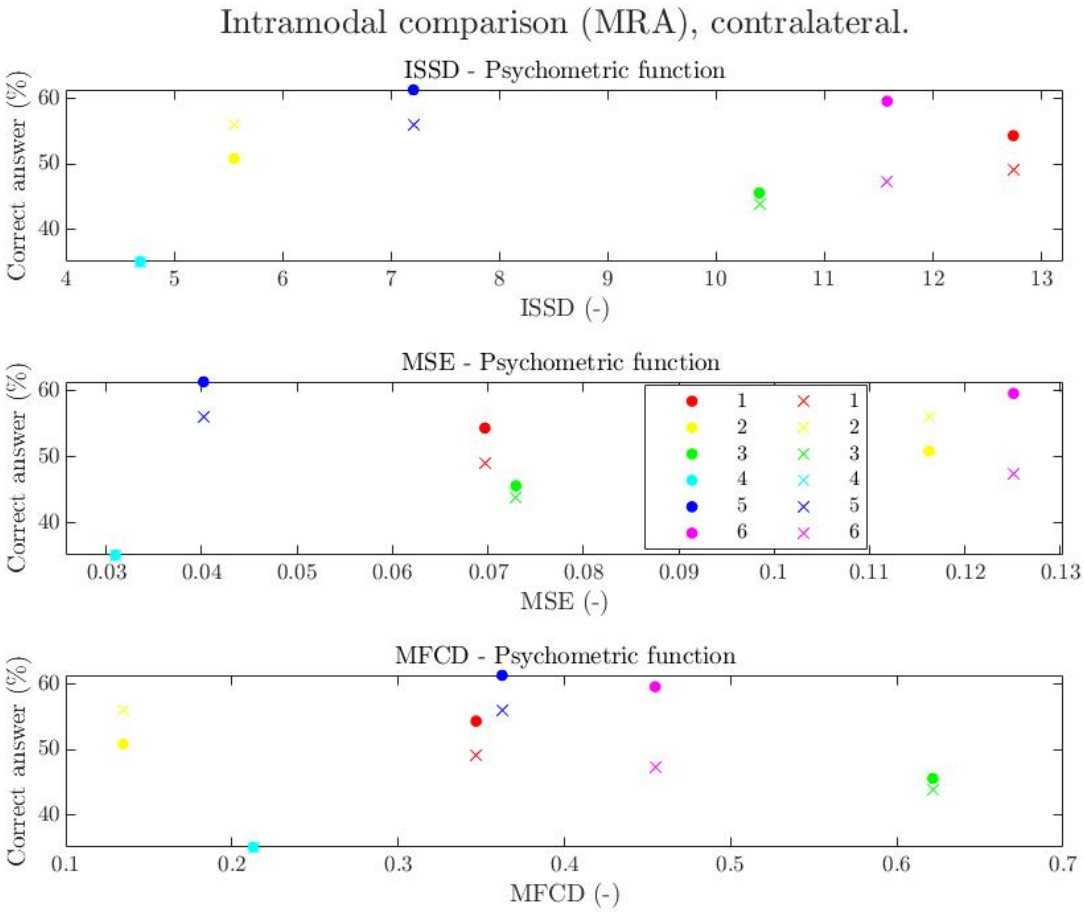

Figure C.6. Probability of correct answer for three distance metrics. Intramodal comparison (MRA PCA HRTF dataset) for normal (dots) and mirrored (crosses) playback, of stimuli optimized for contra-lateral directions.

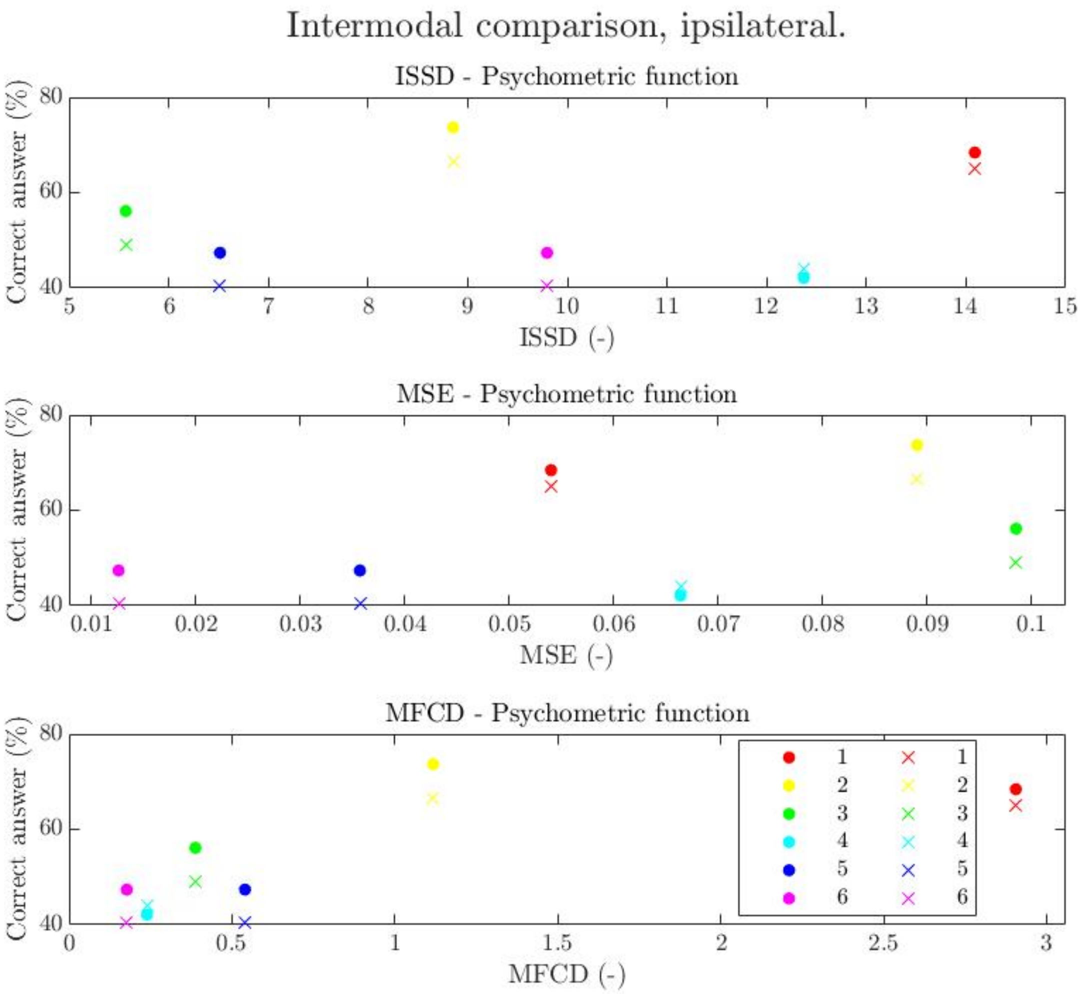

Figure C.7. Probability of correct answer for three distance metrics. Intermodal comparison (Ideal PCA HRTF to MRA PCA HRTF dataset) for normal (dots) and mirrored (crosses) playback, of stimuli optimized for ipsi-lateral directions.