A Global Model for HRTF Individualization by Adjustment of Principal Component Weights

Diploma Thesis by

Josef Hölzl

Graz, March 2014

Host Institution:
Institute of Electronic Music and Acoustics
University of Music and Performing Arts Graz

Graz University of Technology

Assessor: o.Univ.-Prof. Mag. art. DI Dr. techn. Robert Höldrich
Supervisor: Dr. Georgios Marentakis

institut für elektronische musik und akustik

Acknowledgments

First and foremost, I offer my sincerest gratitude to my advisor Georgios Marentakis. Thank you for your never ending stream of ideas and the resulting discussions throughout my work. You gave me a deeper understanding not only of the topic but also further afield. I would like to mention that Georgios was quite involved in developing the HRTF model.

I am very grateful for the opportunity to measure my own HRTFs at the Acoustics Research Institute in Vienna. It was a great experience and it took me another step towards my goals. The 11 test subjects who suffered through listening tests are sincerely thanked for their patience.

Special thanks go to you, my dear, for always having my back. I also want to thank my family for their support and caring throughout all these long years of study and work.

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Graz, _____ _____

Date Signature

Eidesstattliche Erklärung[1]

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommene Stellen als solche kenntlich gemacht habe.

Graz, am _____ _____

Datum Unterschrift

[1] Beschluss der Curricula-Kommission für Bachelor-, Master- und Diplomstudien vom 10.11.2008; Genehmigung des Senates am 1.12.2008

Abstract

Individual head-related transfer functions (HRTFs) can be used to generate virtual sound sources over headphones. According to the model of HRTF individualization using Principal Components (PCs), a Principal Component Weight (PCW) set is sought that when multiplied with a PC basis results in an HRTF set that yields good localization for a number of given directions of sound incidence. Although this is a promising model, the extent to which listeners can perform the individualization by hearing is debatable. The process requires adjustment for each location and PC of interest. In this work, the feasibility of a local and global method is numerically evaluated by estimating the accuracy with which a given basis component can model HRTFs regarding different kinds of input data. The number of required adjustments for a given direction set is then reduced by decomposing the PCW of individual users upon a Spherical Harmonics Basis. Optimal spherical model parameters are sought, depending on the order and reconstruction accuracy. In a listening test, subjects were asked to identify changes in localization when weights of individual directions are automatically modified. This allows a deeper inside into the usability of each technique.

Kurzfassung

Mit Hilfe von Außenohrübertragungsfunktionen (HRTFs) können bei binauraler Wiedergabe virtuelle Schallquellen im Raum generiert werden. HRTFs können durch Kombination von Hauptkomponenten (PCs) und deren Gewichte (PCWs) modelliert und adaptiert werden. Obwohl dieses Modell für manche Quellpositionen sehr gut funktioniert, ist die Genauigkeit und der Aufwand der Individualisierung noch nicht richtig erforscht. Die Gewichte müssen für jede einzelne Position und Hauptkomponente zeitaufwändig angepasst werden. In dieser Diplomarbeit werden zwei Methoden zur Anpassung der Basiskomponenten Gewichte diskutiert und numerisch für verschiedene Eingangsdaten ausgewertet. Um den Prozess der Individualisierung zu erleichtern, wird ein Kugelmodell basierend auf den Gewichten der Hauptkomponenten vorgestellt. Optimale Parameter für das Kugelmodell werden durch die Ordnung der Basiskomponenten und des resultierenden Rekonstruktionsfehlers berechnet. In einem Hörversuch bewerten die Probanden Unterschiede in Lokalisation wenn die Gewichte für individuelle Richtungen automatisch adaptiert werden und geben die wahrgenommene Quellposition an. Auf diese Weise wird ein Einblick in die praktische Handhabung beider Techniken gegeben.

Contents

Chapter 1

Introduction

1.1 Motivation

Individual head-related transfer functions (HRTFs) can be used to generate virtual sound sources over headphones. Describing the acoustic transmission path from a sound source to the ears, an audio stream can be reconstructed in a virtual audio scene by a simple filtering with the corresponding pairs of impulse responses. Nowadays, HRTFs are essential components of military environments, training simulations, room acoustic simulations, game consoles, augmented reality and many more consumer entertainment products.

Usually, a high spatial resolution for accurate source discrimination is required for 3D sound rendering. That is why recent research uses generalized HRTFs from an existing database to avoid expensive and long measurement procedures. Moreover, in most situations, a measurement is not acceptable. However, such averaged HRTFs suffer from perceptual problems, such as difficulty perceiving source elevation and front/back confusions. The same happens if an HRTF set is used by someone else. Since each person has different pinna, head and torso dimensions, the resulting transfer functions vary greatly between individuals. Therefore there is a strong need to adapt existing HRTFs to individual ones.

A large number of HRTF models have been proposed, above all using *Principal Component Analysis (PCA)* to decompose transfer functions into orthogonal components (PCs) and associated weights (PCWs). Other methods manipulate the frequency spectrum by scaling [Mid99b] or altering the energy in certain frequency bands that are crucial for spatial hearing [TG98, SL11]. Many approaches have the drawback only working in certain restricted areas, such as in the median plane or in the front region. To overcome this limitation, the thesis investigates in an applica-

ble HRTF customization tool that compromises the whole sphere, except the region directly below the head.

The main hypothesis can be stated as follows: Is it possible to adjust the localization of an existing arbitrary source position to another one and which parameter have to be changed? Secondly, can someone benefit from adjusting one specific source position to another one, thus is there a customization method to allow a global and consequently more efficient adjustment?

Another motivation is the lack of detailed comparisons of possible algorithms in HRTF models, such as PCA and *Spherical Harmonic Decomposition (SH)* and their combination with different input data. A detailed numerical evaluation of several model parameters is given. The applicability of the PCA model is verified trough a discrimination and localization test adapting the principal component weights.

Based on the findings and developments of the PCA model, a more effective way for adapting PCWs is inspected. A Spherical Model is introduced which projects the PCWs onto a sphere. By using the inherent properties of SH functions, the customization procedure can be enhanced through parameter reduction since not each single source positions has to be adapted.

Figure 1.1 indicates the main steps described in this work. Intentionally no detailed information on the type of transformation can be seen here because there are several options. While the analysis and transformation of the database can only be an offline process due to the large amount of data in an HRTF database, the processing of the modified parameters is implemented so that it can be used almost in real-time.

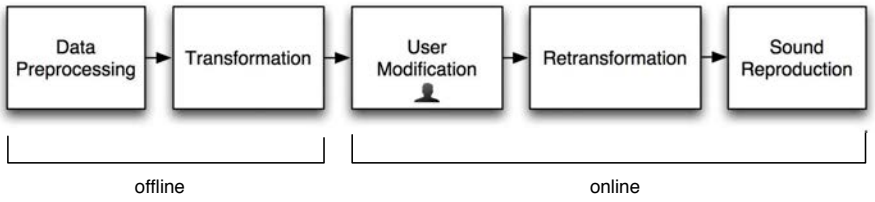


Figure 1.1: Overview of the key features in the HRTF model.

1.2 Thesis Layout

This document is organized as follows. **Chapter 2** goes through the fundamentals and properties of HRFTs and their importance in localization. **Chapter 3** gives

an overview of two basis function representations of HRTFs, namely Principal Component Analysis and Spherical Harmonic Decomposition, with underlying mathematical and statistical fundamentals. **Chapter 4** summarizes and discusses recent HRTF individualization techniques. Particular attention is paid to the commonly used minimum-phase and delay approximation in time domain. In **Chapter 5**, a numerical evaluation of the PCA Model indicates the reconstruction error and seals the model parameters. The focus is on processing and analysis of large data volumes and different kinds of input data. **Chapter 6** discusses a spherical method for global adaptation of the principal weights. An informal listening test for the PCA Model and evaluation of the experimental data is presented in **Chapter 7**. At least, **Chapter 8** summarizes the main hypothesis followed by concluding remarks and outlook on further work.

Appendix A demonstrates a graphical user interface which was initially created to test the HRTF model and analyze certain model parameters. In **Appendix B**, a chart overview of the implementation is given. **Appendix C** provides a graphical and tabular overview of different structures of the model input data matrix.

Chapter 2

Head-Related Transfer Functions

Head-related transfer functions are crucial for localization in azimuth and elevation over headphone. The encoded binaural cues give listeners the perception of a 3D sound display. While the HRTF $\underline{H}(s, \theta, f)$ refers to the frequency domain, the *Head-Related Impulse Response (HRIR)* $h(s, \theta, t)$ denotes the counterpart in the time domain. For each subject s and source position θ on a dense sampling grid, a pair of HRIRs is stored. Miniature microphones in the ear canals record the impulse response that are transmitted through a loudspeaker at a fixed distance of typically 1 meter to the head. While typically each acoustic transmission path must be processed individually and is therefore very time consuming, recent measurement environments are based on e.g. multiple exponential sweep method [MBL07] and automatic processing to decrease the total measurement time (see Figure 2.1). Such a method may speed up the whole process, however, the measurement still requires an anechoic chamber and expensive apparatus which both is not available for consumers. That is why several HRTF individualization techniques are proposed by researchers to circumvent the need of a measurement.

Typically, HRTFs are diffuse-field equalized prior further processing to exclude both ear canal resonance and measurement system response. This leads to the so called *Directional Transfer Function (DTF)*, which mainly includes the direction-dependent spectral parameters and is responsible for spatial hearing. Contrary, the *Common Transfer Function (CTF)* averages the spectrum across all directions N, which comprises subject dependent and position independent spectral information, such as the diffuse part and the ear canal resonance between 2 and 4 kHz. The latter is not desired in binaural processing because when using headphones for playback, the subject would listening in fact through two ear canals. Also through hearing tests, Middlebrooks *et al.* [MMG89] and Møller [Møl92] confirmed that the propagation in the ear canal is not dependent on directions.

The magnitude of the DTF, which is frequently used in this work, can be calculated in logarithmic frequency domain by subtracting the CTF from an HRTF set,

$$20 \log |\underline{D}(s, \theta, f)| = 20 \log |\underline{H}(s, \theta, f)| - \frac{20}{N} \sum_{\theta=1}^{N} \log |\underline{H}(s, \theta, f)| \ . \tag{2.1}$$

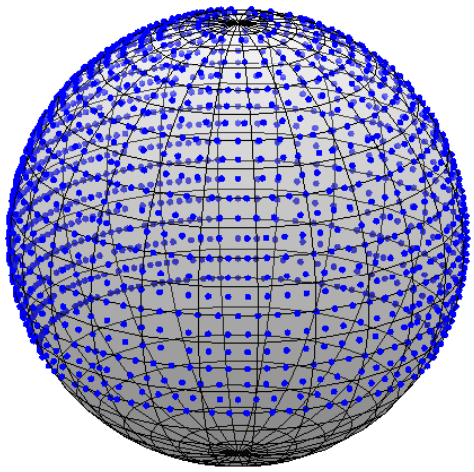
Whereas HRTFs incorporate the effects of the whole body, *Pinna-related Transfer Functions (PRTFs)* indicate only the contribution of the pinna and reduce the dependence with respect to azimuth. They can be calculated by applying a 1 ms Hann window at the beginning of the HRIR signal in order to eliminate reflections by torso and shoulders [SGA10] and then transformed into frequency domain. The use of these functions might be helpful when relating features in the magnitude spectrum to particular anthropometric dimensions.

Furthermore a critical band filter can be applied to smooth the spectrum of HRTFs and remove smaller features from the spectrum that humans are not sensitive to [HZK99].

2.1 Localization Cues

Binaural hearing is based on three basic localization cues. *Interaural Time Difference (ITD)* and *Interaural Level Difference (ILD)* are the primary cues for localization in azimuth. Their mechanism is formerly described in the *Duplex Theory* [MM02] by John William Strutt in 1907 that specified the relationship between sound propagation and the geometrical arrangement of the head. Later, Stevens and Newman [SN36] extended this theory and specified the effect of the physical properties within the frequency axis. Since the ITD can only be processed without ambiguity of the wavelength, time difference is limited to about 1.5 kHz when assuming a standard head diameter of 17 cm. Frequencies below 1.5 kHz are diffracted around the head whereas high frequencies above 1.5 kHz are shadowed, consequently the ILD becomes more dominating. That is why localization in the range of 1.5 and 2 kHz leads to the greatest inaccuracies because ITD as well as ILD information is not optimal. The binaural cues can be estimated by listening tests or computed from a pair of HRIRs. However, since the 70s, some studies have refuted that time differences are only localization cues for low frequencies. In fact, also the slow fluctuating frequencies of the sound envelope of higher-frequency stimuli can be evaluated by the binaural system [Gra95, Zim04]. Moreover, despite ITD is a frequency dependent parameter it is commonly simplified as a constant time delay [KW92, KC98].

(a) Anechoic chamber [Maj13].



(b) Spatial resolution (1550 directions).

Figure 2.1: Measurement setup at Acoustic Research Institute (ARI) in an anechoic chamber (a) and resulting spatial resolution (b).

According to the duplex theory, the interaural cues ITD and ILD can not provide vertical information. This is due the existence of the so called *Cone of Confusion* and *Torus of Confusion* which are solids centered on the sagittal plane with indefinite positions that have same ITD and ILD respectively. Consequently, they produce ambiguous perceptual coordinates that make it impossible to distinct between front/back or up/down. By slight movement of the head this confusion can be easily resolved in real environment because additional information about the source is obtained. In 3D simulation through headphone this is not always possible. Suitable time-variant filters and head movement detection must be applied. This additional judgment for localization, namely the energy division of the spectrum and spectral cues, is one of the most challenging issues in current HRTF models.

Monaural spectral features of the magnitude spectrum provide crucial localization cues for front/back and up/down discrimination [Bla70, HW74, LB02]. These are generally spectral peaks and notches between 4 and 16 kHz that are mainly effected by the shape of the outer ear. Spectral parameters below 3 kHz are mainly produced by head diffraction and torso reflections [Shi08]. Several studies [HW74, ST68] support the fact that a prominent 1-octave notch centered between 5 and 11 kHz in the frequency spectrum changes systematically with the vertical source location. However, the mechanism that filters the incoming sound and generates spectral features has not been fully understood, because the pinna shape is too complex and individual. Fact is, that pinna-based filtering only effects higher frequencies above 5 kHz because of

the small dimensions of the ears. In HRTF databases that also collect anthropometric dimension (e.g. CIPIC and ARI), currently up to 12 different pinna parameters are extracted.

Summing up, there are several interaural and spectral parameters that are crucial for localization of sound sources. Since the introduction of the duplex theory, major improvements in understanding binaural hearing has been established. However, as yet no comprehensive model for the spectral influence of the pinna was placed. Other promising studies based on analysis of the neural system will hopefully be of use here.

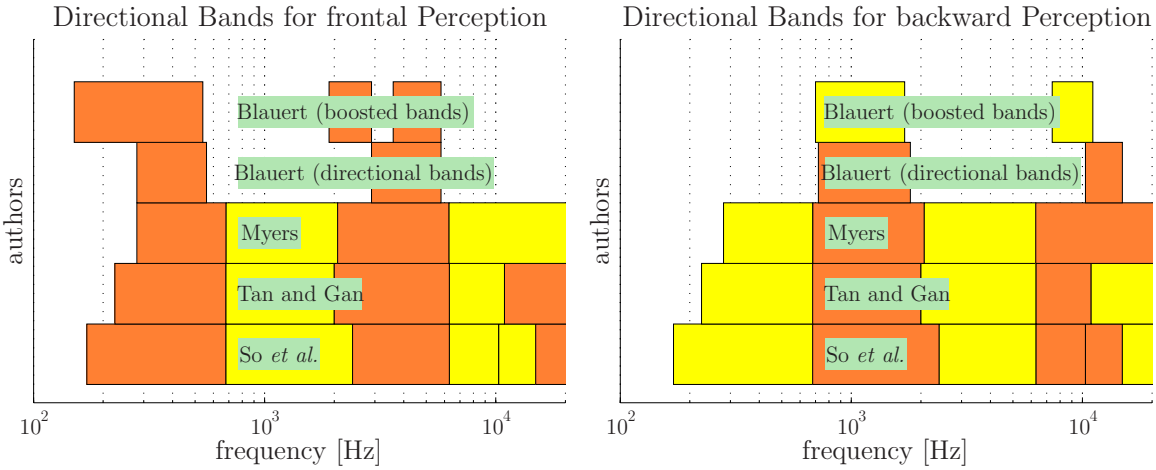
2.2 Directional Bands

Frequency bands describe certain parts of a spectrum. For the purpose of this work, an overview of frequency bands that are relevant for localization was prepared.

Among the first, Blauert [Bla70] discovered that the perceived source localization of 1/3 octave band noise was mainly affected by their center frequency. He described four different frequency bands that influence front/back discrimination and called them *Directional Bands*. By subtracting the HRTF spectrum of a rear sound from a frontal one, he analyzed the average differences and realized that a positive average difference indicates forward direction and a negative one perception from backwards. These bands are also known as *Positive* or *Negative Boosted Bands*. Hebrank and Wright [HW74] conducted three experiments including various filters to classify relevant frequency bands. For each of the directions of interest (frontal, behind or above) an equivalent peak or notch could be identified. Myers [Mye89] continued to investigate in this topic and classified four relevant frequency bands. Similarly, Tan and Gan [TG98] amplified and attenuated energy in five bands by ± 8 to ± 12 dB. Asano *et al.* [ASS90] also investigated the role of spectral cues and approximated the spectrum with an pole-zero model where poles represent resonances and zeros denote anti-resonances. By reducing the parameter of the model from 40 to 10 poles and zeros, relevant frequency regions were discovered in which the judgement errors dramatically increased. The authors concluded that spectral cues for front/rear discrimination are basically below 2 kHz, but also in some high frequency regions while vertical cues are located above 5 kHz.

Unfortunately, studies do not agree which frequencies are crucial for localization. Whereas Blauert reported that sounds between 280-560 Hz and 2.9-5.8 kHz are more likely to be perceived from front, Hebrank and Wright [HW74] claimed that spectral manipulation below 3.8 kHz does not affect front/back discrimination. Langendijk

and Bronkhorst [LB02] specified that important localization cues for front/back distinctions are mainly between 8-16 kHz. Asano [ASS90] *et al.* mentioned that spectral details above 2 kHz as well as unspecified high frequency components are important for this. Also Musicant and Butler [MB84] confirmed that frequencies below 1 kHz do not contribute to sound localization. The directional bands found in the literature are presented in Figure 2.2. It can be seen that the studies agree broadly, there are, however, several disagreements in the high frequency region. Moreover, except Blauert, the band modifications to make a sound event coming more from rear are exact the opposite to the front region.



(a) Front. (b) Back.

Figure 2.2: Comparison of different studies: Blauert [Bla70], Myers [Mye89], Tan and Gan [TG98], So *et al.* [SL11]. Manipulation of Directional Bands for frontal (a) and backward (b) perception. Orange and yellow colors indicate amplified and attenuated bands respectively.

2.3 Learning and Adaptation

Although in this work the focus is on adapting HRTFs *to* the individual subjects, it is worth noting that recent research shows that the human binaural auditory system can adapt to changes in HRTFs. An experiment with ear molds by Hofman *et al.* [HVRVO98] confirmed that within several weeks, a person can *learn* the HRTF set from another person resulting in an localization accuracy almost comparable to their original ones. Remarkably, the original inherent HRTF set is still usable without loss of accuracy, so the subjects can switch from one to a second, adopted HRTF.

Parseihian *et al.* [PK12] analyzed the process of adaptation by examining localization with virtual sound sources. They quantified the effect of training by assessing the improvement of localization accuracy over time. It was shown that a rapid adaption to a non-individual HRTF set through feedback is possible within three sessions of twelve minutes. These results can be applied to minimize front/back and up/down confusions.

Besides, when synthesizing HRTFs, filters and models should take into account the non-uniform resolution of the human auditory system [HZK99]. For example, a weighting function could be applied in spectral magnitude to incorporate the logarithmic amplitude resolution. From a psychoacoustical viewpoint, several distance measures using frequency smoothing or band-pass filters, such as Bark scale or Equivalent rectangular bandwidth (ERB) scale.

2.4 Spatial Coordinates

Several studies ([KP12], [PK12]) use the interaural polar coordinate system instead of the standard spherical coordinates. The first considers a lateral [-90°, 90°] and polar angle [-90°, 270°] which indicates the interaural axis while the latter denotes an azimuth [0°, 360°] and elevation angle [-90°, 90°]. Each point on a plane can be specified by an angle and radius r, but in HRTF processing typically the radius is neglected and set to one. The advantage of the interaural polar coordinate system is that the primary cues, which relates to ITD and ILD, can be expressed by the lateral angle while the monaural spectral cues are presented by the polar (or rising) angle. Therefore all front/back and up/down confusions are isolated in the polar angle.

In some cases, the cartesian coordinate system is applied [Zaa11]. However, in this work, the standard spherical coordinate system is used because the existing HRTF databases are specified therein. The cartesian coordinates are transformed into spherical coordinates [Wil99] by

$$\begin{aligned} r &= \sqrt{x^2 + y^2 + z^2}\ , \\ \theta &= \arctan\left(\frac{\sqrt{x^2 + y^2}}{z}\right)\ , \\ \varphi &= \arctan\left(\frac{y}{x}\right)\ , \end{aligned} \tag{2.2}$$

with radial distance r, azimuthal angle θ and polar or zenith angle φ. The transformation back to cartesian coordinates can be performed by

$$
\begin{aligned}
x &= r \cdot \sin(\theta) \cdot \cos(\varphi) \ , \\
y &= r \cdot \sin(\theta) \cdot \sin(\varphi) \ , \\
z &= r \cdot \cos(\theta) \ .
\end{aligned} \tag{2.3}
$$

In this work, the zenith angle φ is commonly replaced by the elevation angle which is set to zero at the horizontal plane. The measure of the azimuthal angle is counter-clockwise and starts in the frontal plane.

Chapter 3

Basis Function Representations of HRTFs

Orthogonal basis function representations of HRTFs have been widely used for understanding HRTFs and reducing their dimensionality, however, the application of such models in HRTF individualization has been limited due to the absent understanding of the perceptual nature. The most common basis functions used to decompose HRTFs are Principal Component Analysis and Spherical Harmonic Decomposition. Recent studies relating these techniques indicate that typically measured HRTFs contain a significant amount of data that is perceptually irrelevant. The aim of an efficient and compact HRTF model is to include only data that is relevant for localization in a certain way. Several researchers have attempted to model HRTFs using a small subset of orthogonal basis function decompositions and subsequently adapting HRIRs or HRTFs using the models. Most often, Principal Component Analysis [KW92, QE98, HP08] and Spherical Harmonic Transform [EAT98, ZKA09, ZAKD10] have been used for this purpose. Such decompositions often reduce the high-dimensionality of HRTF sets and also serve as a basis for the investigation of their numerical but also perceptual properties.

3.1 Principal Component Analysis

Principal Component Analysis is a robust statistical method for data representation. The technique projects an original dataset on an orthogonal subspace that is estimated by taking the covariance of the data into account. The technique can be used to unveil relationships between the independent variables in a dataset and in this way reduce a high-dimensional dataset into a more meaningful, low-dimensional space. It has been widely used in computer vision and pattern recognition to find

relevant structure in data and neglect redundant information. Usually the input data is pre-processed and aligned prior PCA to increase the performance. The resulting model parameters can be calculated directly from the input data through *Singular Value Decomposition (SVD)*. Through a linear combination of the new basis and their corresponding principal weights, the original dataset can be reconstructed with a controllable accuracy, because the orthogonal principal components are sorted according to their variance describing the original data.

3.1.1 Methodology

Before calculating the PCA, data need to be centered by subtracting their mean. This process is related in the case of an HRTF dataset to the calculation of the DTF, which also averages out specific singularities of test persons as well as measurement setup and recording artifacts. When each row of the input data represents a single DTF, actually the input data is already centered. However, as described in Chapter 5.2.3, there are several options for the structure of the input matrix, so it might be the case that DTFs need to be extracted from the HRTF set prior to subsequent analysis. Principal Component Analysis is normally applied onto a two-dimensional matrix, with columns defining the independent variables and rows containing observations.

Adjusted Dataset. The column mean $\bar{x}$ $(1 \times n)$ of the input matrix $\mathbf{X}$ $(m \times n)$, with m as the number of data points and n dimensions in the data set, can be calculated by

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \ . \tag{3.1}$$

A centered input matrix $\mathbf{Y}$ is computed from $\mathbf{X}$ by subtracting off column means,

$$\mathbf{Y} = \mathbf{X} - \bar{\mathbf{X}} \ , \tag{3.2}$$

with matrix $\bar{\mathbf{X}}$ $(m \times n)$ containing $\bar{x}$ as rows.

Eigenvectors and Eigenvalues. Mathematically, PCA is related to the well known singular value decomposition which splits the real-valued centered input matrix $\mathbf{Y}$ $(m \times n)$ into

$$\mathbf{Y} = \mathbf{U}\ \mathbf{S}\ \mathbf{V}^T \ , \tag{3.3}$$

with $\mathbf{U}$ $(m \times n)$ and $\mathbf{V^T}$ $(n \times n)$ as orthogonal matrices including *left* and *right eigenvectors* $\mathbf{u}_k$ and $\mathbf{v}_k$ respectively. $\mathbf{S}$ $(n \times n)$ is a diagonal matrix with nonzero elements only on the diagonal, so that $\mathbf{S} = diag(s_1, \ ... \ , s_n)$. These are nonnegative and real values, also known as *singular values* which are sorted in descending order of magnitude from top to bottom. Assuming that the matrix $\mathbf{Y}$ has a *rank* r leads to $s_k > 0$ for $1 \leq k \leq r$ and $s_k = 0$ for $(r + 1) \leq k \leq n$ [WRR03]. This means that if some eigenvalues are very close to zero, one can neglect those values and the corresponding eigenvectors to reduce the dimensionality of the new basis.

It has to be mentioned that SVD is not the only way to perform PCA. The singular values and associated eigenvectors also can be obtained directly through the covariance matrix $\mathbf{C_Y}$ that is formed as

$$\mathbf{C_Y} = \frac{1}{N-1} \ \mathbf{Y^T} \ \mathbf{Y} \ , \tag{3.4}$$

with $\mathbf{C_Y}$ as a symmetric, real-valued, square matrix. What the matrix actually says is that the diagonal elements are the variance and the off-diagonal elements are the covariance between the independent variables. For that reason, a large value in an off-diagonal element indicates a high redundancy whereas a large value in the diagonal potentially indicates an important pattern or reflects a significant dynamic in the data set. Consequently, an optimal and non-redundant structure of $\mathbf{C_Y}$ would minimize off diagonal elements. This can be achieved by eigendecomposition. Since the covariance matrix is symmetric, the matrix is diagonalizable, which follows

$$\mathbf{C_Y} \ \mathbf{V} = \mathbf{V} \ \mathbf{D} \ , \tag{3.5}$$

with a diagonal matrix $\mathbf{D}$ $(m \times m)$ containing the eigenvalues of $\mathbf{C_Y}$ and $\mathbf{V}$ as an orthonormal eigenvector matrix including the right eigenvectors as columns. The eigenvectors of the covariance matrix $\mathbf{C_Y}$ are termed as the principal components of $\mathbf{Y}$. When applying $\mathbf{Y} = \mathbf{U} \ \mathbf{S} \ \mathbf{V}^T$ to calculate the covariance matrix, and multiplying $\mathbf{Y^T}$ on the right side, it follows

$$\mathbf{Y} \ \mathbf{Y^T} = (\mathbf{U} \ \mathbf{S} \ \mathbf{V}^T) \ (\mathbf{U} \ \mathbf{S} \ \mathbf{V}^T)^T \ , \tag{3.6}$$

which leads to

$$\mathbf{Y}\ \mathbf{Y^T} = \mathbf{U}\ \mathbf{S}^2\ \mathbf{U}^T\ . \tag{3.7}$$

Consequently, the square root of the eigenvalues of $\mathbf{YY^T}$ are the singular values of $\mathbf{Y}$. In fact, using SVD is more efficient and robust because the formation of the covariance matrix is costly in terms of computational resources.

Transformation. The columns of $\mathbf{V}$ contain the principal components. So, the original centered data $\mathbf{Y}$ set can be transformed to the new basis by projecting it on the eigenvector basis,

$$\mathbf{W} = \mathbf{Y}\ \mathbf{V}\ , \tag{3.8}$$

where $\mathbf{W}$ represents the principal weight, or score matrix with the same dimensions as the input matrix, containing the loadings that when applied to the principal components would recreate the original data matrix.

Reconstruction. Finally, an approximation $\mathbf{X}^l$ to the input matrix $\mathbf{X}$ can be obtained through applying the PC weights on a basis of lower dimensionality $l < N$ and adding the subtracted mean $\bar{\mathbf{X}}$[1] again:

$$\mathbf{X}^l = \sum_{k=1}^{l} \mathbf{u}_k\ s_k\ \mathbf{v}_k^T\ + \bar{\mathbf{X}}\ . \tag{3.9}$$

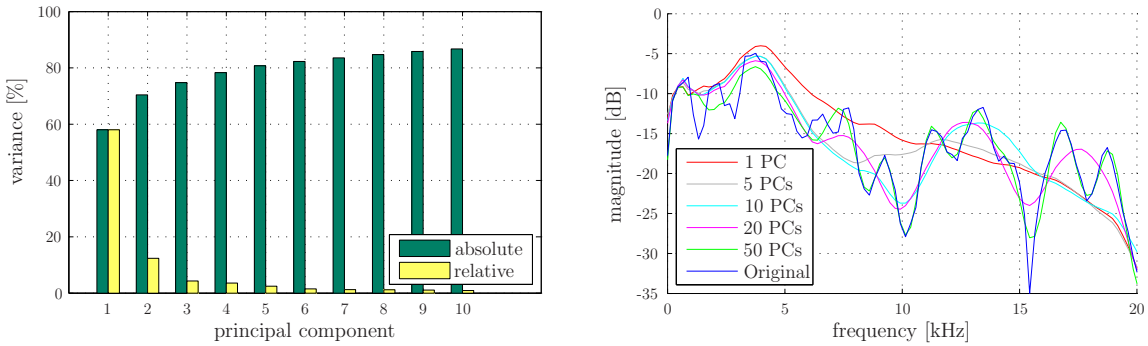
In fact, some information is lost in the reconstruction when using only l dimensions, however, through a good choice of l, a good reconstruction accuracy can be obtained. A reasonable value for l is usually obtained by calculating the percentage of the explained variance that can be explained using the lower dimensionality representation. This can be done by

$$var(l) = \frac{\sum_{k=1}^{l} s_k}{\sum_{k=1}^{N} s_k}\ \cdot 100\ [\%]\ , \tag{3.10}$$

where s_k is the k^{th} singular value, l is the number of a particular PC and N is the total number of components. The singular values correspond to the variance

[1] In MATLAB®, the function *princomp()* automatically centers the input data and returns PCs, PCWs and variance. It is highly recommended to calculate the subtracted mean before using this function, because otherwise one will loose this information.

explained by each component. Commonly, the number of principal components that are necessary to achieve 90 percent of the explained variance are used. Figure 3.1a illustrates the relative (yellow bars) and cumulative (green bars) explained variance for the first ten PCs and Figure 3.1b shows the resulting reconstruction accuracy for the corresponding PCs. It can be observed that the first PC already accounts for 80 percent of the total variance whereas the remaining ones only little by comparison. The variance explained by higher principal components in this case decreases rapidly and that they can be omitted without major loss of information. The difference in explained variance is not always as high as in this example, it depends on the type of the input data. More often than not, though, once components are sorted in terms of decreasing variance, the higher components contribute less to the explanatory power of the model.



(a) Total (green) and relative (yellow) variance of first 10 PCs.

(b) Reconstruction accuracy.

Figure 3.1: Reconstruction of one left ear DTF (340° azimuth, −39° elevation) in CIPIC database by including different numbers of principal components.

For the sake of completeness, it should be mentioned that a SVD of matrix $\mathbf{Y}$ is not unique, consequently, the principal components are not unique too. If one removes or adds records of the input matrix, it may happen that the components are mirrored because the dataset has some rotational symmetries. This can be easily discovered when comparing the same components in different HRTF databases.

3.1.2 Modelling HRTFs using Principal Components

A head-related transfer function set can be modeled by a set of common orthogonal basis functions and their associated principal component weights. This modeling approach can be used as an HRTF individualization method if PCWs are appropriately modified to yield the desired HRTF set. The number of the participating components

in the reduced model has however to be selected carefully, as this is directly linked to the success of the individualization process and affects its duration.

Most of the studies agree that using the first 4-5 components is enough [Mar87, KW92, Shi08], because they can explain about 90 percent of the variance in an HRTF set and capture the relevant properties for localization. However, the studies are not always comparable, as the HRTF sets used differ with respect to the number of subjects and measurement directions included (vary from 2 to 52 subjects, 1 to 1550 directions). They are usually preprocessed in different ways, while variations also exist in how the data matrix that is subjected to PCA is constructed. All these choices can affect the compression efficiency as well as the generalizability of the resulting model. A detailed overview of reasonable ways to construct this matrix can be found in Chapter 5.2.3.

Nevertheless, principal components obtained from variable HTRF datasets vary little as long as the number of measurement directions and subjects is reasonable. This invariance is more evident for components explaining a large amount of variance, as components of smaller variance reflect specificities that might not be shared across datasets. Middlebrooks and Green [MG92] were among the first who compared basis vectors calculated from their own measurement data (8 subjects, 360 positions) with an existing database by Kistler and Wightman [KW92] (10 subjects, 265 positions) and indeed confirmed a high correlation between the components, which however decreases with rising principal component order number. Similar results were observed with the databases used in this thesis.

A common difference is that some studies attempt to apply PCA directly on HRIRs [Shi08, HP08, HPP10] while others are focusing on HRTFs [KW92, MG92, CvVH93, QE98, GV07, XLS09, Xie12]. The HRIR approach has the advantage of including phase information in the model. Moreover, effects of pinna, head or shoulder can be better extracted in the time domain. In contrast, when PCA is applied in frequency domain, these effects are coupled. Beside the advantage that the logarithmic magnitude spectrum might be better related with the logarithmic sense of human hearing, minimum-phase reconstruction is most commonly used to transfer the reconstructed signal from frequency to time domain. Another difference is that sometimes PCA is applied simultaneously to all source positions in a database [KW92, Xie12], while in others to each direction separately [Shi08]. It is highly recommended to incorporate all directions for PCA because otherwise no relevant phenomena according direction dependency can be discovered. Table summarizes 3.1 individual differences in modeling HRTFs through PCA in the literature.

Author	Database	Structure	Data Format	PCs
Martens [Mar87]	own	2 subjects, 36 directions	24-point log DTF	4
Kistler [WK91, KW92]	own	10 subjects, 265 directions, $[5300 \times 150]$	150-point log DTF (0.2-15 kHz)	5
Middlebrooks [MG92]	own	8 subjects, 360 directions	133-point log DTF (3-16 kHz)	5
Chen [CvVH93]	KEMAR and own	-	magnitude and phase	-
Wu [Wu97]	own	2 subjects	HRIR	-
Qian [QE98]	Tucker-Davis	26 subjects, 360 directions	HRTF	6
Grantham [GWFA05]	Wightman	1 subject, 19 directions in azimuth	HRIR (256 samples)	90
Rodriguez [RR05a, RR05b]	CIPIC	5 subjects, 1250 directions	64-point PRTF	20
Grindlay [GV07]	CIPIC	45 subjects, 1250 directions, $[45 \times 1250 \times 181]$, (tensor SVD)	181-point HRTF (0.5-16 kHz)	10
Hwang [HP08, HPP08, HPP10]	CIPIC	45 subjects, 49 directions, $[67 \times 2205]$	median plane HRIR (first 1.5 ms after time delay)	12
Shin [Shi08]	CIPIC	45 subjects, 9 directions, $[10 \times 45]$ for each direction	left ear HRIR (first 10 samples after time delay)	4
Xu [XLS09]	CIPIC	45 subjects, 1250 directions	log HRTF	10
Rothbucher [RDS10]	CIPIC	30 subjects, 1250 directions (tensor SVD)	200-point HRTF	10
Xie [Xie12]	own	52 subjects, 493 directions	59-point HRTF	35
Fink [FR12]	CIPIC	34 subjects, 25 directions, $[850 \times 400]$	horizontal plane HRIR (200 samples)	25

Table 3.1: Overview of parameters for the PCA input matrix used in literature. Last column presents the proposed number of PCs for reconstruction.

Outliers or measurement errors in the dataset may have a serious effect on the PCA output, therefore a control mechanism is suggested. The sample mean and resulting covariance matrix is very sensitive to outliers. Several algorithms for outlier detection or missing data were proposed in literature [LC85, Che02, CMM09] to make PCA processing more robust and susceptible to outliers or missing data. A visual inspection of the input data can prevent further processing problems. However, this is practically impossible in HRTF databases because of the large amount of data.

In the preliminary work [Hol12], outliers were identified by checking if an individual weight is outside the normal range of the PCW distribution. Here, the intention was rather to find outliers in individuals as positions. PCA was applied on the dataset and an iterative implementation of the *Grubbs' Test* was applied separately on each of the first five principal weights. This procedure actually tests if the minimum and maximum values belong to the main population. According to this, the corresponding individuals have been removed from the dataset and the PCA was calculated again until no outlier was detected. Since the preprocessing of the HRTF model features is an off-line computation, this can be appropriate. In this work, for the proposed model, a visual inspection of the most relevant PCWs in ARI database did not reveal any outliers. However, subject ID 1034 in IRCAM database was detected as an outlier for PCW1 when using the dimension of **Struct1** [$subjects \times (signal * positions)$] with logarithmic frequency magnitude for the PCA input data (see Chapter 5.2.3, Page 40).

In literature, a more robust version of PCA to handle missing data and outliers was proposed by Roweis [Row98] and Chen [Che02]. The algorithm avoids the computationally intensive calculation of the covariance matrix and instead uses an expectation maximization (EM) algorithm. Similarly, Lee *et al.* [LYW13] proposed an *Online Oversampling Principal Component Analysis (osPCA)* algorithm to detect outliers. The idea is that the direction of the basis component are changing when an outlier is added to the covariance matrix. To enhance the computational performance during online detection not the entire covariance matrix is calculated. However, this is beyond the scope of this work. From experience in HRTF databases, as long enough measurements are presented, outliers can be detected on the weights, with not so much influence on the resulting PCs.

The requirements of the PCs may be different depending on the application. For example, one would obtain *statistically independent* components. *Independent Component Analysis (ICA)* is such an algorithm which also can be used for HRTF decomposition [HL09]. Similar to PCA, it is a projection technique but the major difference

is that ICA returns statistically independent components that are not orthogonal because ICA minimizes both second-order and higher-order dependencies [BDBS02]. In contrast, PCA only decorrelates second-order statistics [CPK06]. However, a drawback of ICA is a clearly higher computing power, which sometimes leads to problems with large datasets such as HRTF databases. PCA and ICA can also be combined to enhance performance, for example Berg *et al.* [BBL+05] proposed a blind source separation algorithm that first computes PCA to decorrelate the data and then performs ICA to separate the data.

PCA is not a continuous representation of HRTFs, because the principal weights for directions outside the dataset measurement do not exist. Obviously, locations that are not originally included in the dataset, can be estimated by involving PCWs of surrounding measured directions and apply one of many interpolation methods such as inverse-distance weighting or spherical splines [HBS99].

3.2 Spherical Harmonic Decomposition

Spherical Harmonic Decomposition, primary intended for the modeling and approximation of continuous functions on the sphere, has also been applied to model HRTFs. As HRTF measurements occur for positions distributed on a sphere, or spherical sections, such an approach is inherently appropriate. The dataset is projected onto spherical basis functions of a desired order, whose weighted combination can be used for modeling or approximation purposes. In contrast to PCA, where the basis functions are computed from the dataset, the spherical harmonic functions are fixed and defined hierarchically. On the basis of the Fourier Transform which decomposes a function $f(x)$ into an infinite sum of $\sin(nx)$ and $\cos(nx)$, the spherical harmonic decomposition expands a function $f(\theta, \phi)$ into an infinite sum of spherical harmonics. In this way, usually a better parametric description of a geometric body can be obtained [KSG99]. The spherical harmonics originate by solving the angular part of Laplace's equation in spherical coordinates.

3.2.1 Methodology

Spherical harmonic functions can be complex or real. In this work, the real valued spherical harmonics are used. Here, the notation is based on [Jar08]. The orthonormal real valued spherical harmonics $Y_l^m(\theta, \varphi)$ with polar angle θ $[0, \pi]$, azimuthal angle φ

$[0, 2\pi]$, order l and degree m $[-l, l]$ are defined as

$$Y_l^m(\theta, \phi) = \begin{cases} \sqrt{2} K_l^m \cos(m\varphi)\ P_l^m(\cos\theta) & m > 0\ , \\ K_l^0\ P_l^0(\cos\theta) & m = 0\ , \\ \sqrt{2} K_l^m \sin(-m\varphi)\ P_l^{-m}(\cos\theta) & m < 0\ , \end{cases} \tag{3.11}$$

with K_l^m as normalization constant,

$$K_l^m = \sqrt{\frac{2l+1}{4\pi} \frac{(l-|m|)!}{(l+|m|)!}}\ , \tag{3.12}$$

that ensures that the inner product of the basis functions with itself is one. P_l^m denotes the associated Legendre function which is described in detail in [PZ08]. Figure 3.2 depicts all real valued spherical harmonic basis functions up to order $l = 2$. The spatial complexity increases with the number of orders from top to bottom whereas the degree m is plotted horizontally. For each order l, $2l + 1$ basis functions exist, which leads in this case to a total of $(l + 1)^2 = 9$ different functions. Sometimes a single index i is used, such as $i = l(l + 1) + m$. Then, the orthogonality of the basis functions is demonstrated by

$$\int_{\Omega_{4\pi}} Y_i(\vec{w}) Y_j(\vec{w})\ d\vec{w} = \delta_{ij}\ , \tag{3.13}$$

with δ_{ij} as the Kronecker delta.

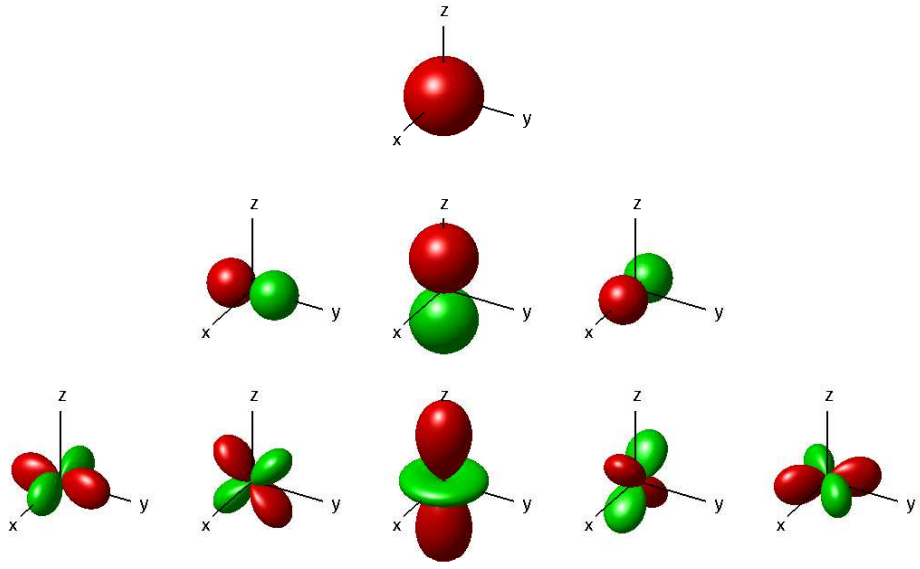


Figure 3.2: Real-valued spherical harmonics up to order $l = 2$. The color of the surface represents the value of the function. Each row illustrates a spherical order l whereas each column represents the corresponding degrees m $[-l, l]$.

A direction vector $\vec{w}$ using spherical parametrization can be defined as

$$\vec{w} = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta) \ . \tag{3.14}$$

Spherical Harmonic Transform (SHT). Given a function $f(\vec{w})$ on the sphere, f_l^m as the expansion coefficients can be obtained by the inner product of $f(\vec{w})$ with each of the basis functions,

$$f_l^m = \int_{\Omega_{4\pi}} Y_l^m(\vec{w}) \ f(\vec{w}) \ d\vec{w} \ . \tag{3.15}$$

The basis functions are orthonormal and ordered in spatial frequency [LH02]. Any real-valued function $f(\vec{w})$ on the sphere can be factored as a weighted linear combination of the basis functions by

$$f(\vec{w}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} Y_l^m(\vec{w}) \ f_l^m \ . \tag{3.16}$$

Theoretically an unlimited number of spherical harmonics is required, but in practice, the coefficients are truncated to order $l = N$, resulting in a lower angular representation, thus band-limiting.

Discrete Spherical Harmonic Transform (DSHT). In most cases a discrete-point presentation is required because of an infinite number of discrete sample points on the sphere. DHST allows to calculate a spherical wave spectrum computed by a set of spatially discrete points. Therefore, a set of discrete SH expansions has to be formulated and Equation 3.16 can be expressed as

$$\mathbf{f}_N = \mathbf{Y}_N \ \psi_N \ , \tag{3.17}$$

with $\mathbf{f}_N$ as a function at discrete points on the sphere, $\mathbf{Y}_N$ as a matrix containing $(N+1)^2$ sampled spherical harmonics and ψ_N as a position-independent vector with the corresponding SH coefficients [NZ11]. In order to calculate ψ_N, the spherical harmonic matrix $\mathbf{Y}_N$ has to be inverted. If the number of the discrete directions does not coincide with $(N+1)^2$, un under- or overdetermined system of equations has to be solved [Zaa11]. An optimal solution would be a regular sampled spherical surface. However, all HRTF databases are suffering from a measurement lack in the lower hemisphere. There even no exists a database with source positions exactly below the head. Consequently, the sampling points

are not homogeneous distributed resulting in biased models and mathematical regularization problems, like an ill-conditioned matrix $\mathbf{Y}_N$. A relevant property of $\mathbf{Y}_N$ is the condition number giving information about the ratio between the smallest and largest singular value and furthermore indicates the robustness and stability of its inverse.

The *Truncated Singular Value Decomposition (TSVD)* is a method for matrix regularization [Han87, HF09] yielding a reduced rank approximation. According to the standard SVD in Equation 3.3 (Page 12), neglecting the smallest singular values of the diagonal matrix $\mathbf{S}$ and using the truncated orthogonal matrices $\tilde{\mathbf{V}}$ and $\tilde{\mathbf{U}}^T$, the pseudoinverse of $\mathbf{Y}_N$ is regularized with $\mathbf{Y}_N^\dagger = \tilde{\mathbf{V}}\tilde{\mathbf{S}}^{-1}\tilde{\mathbf{U}}^T$. Finally the SH coefficients are obtained by

$$\psi_N = \mathbf{Y}_N^\dagger \, \mathbf{f}_N \ . \tag{3.18}$$

3.2.2 Modeling HRTFs using Spherical Harmonics

As with PCA, modeling using Spherical Harmonics has been attempted both in the time and the frequency domain. In such cases, the Spherical Harmonic Decomposition is applied either on each time sample of the HRIR, or each frequency bin of the HRTF. Again, the basis functions of the model are orthonormal and are usually ordered in spatial frequency. Similar to PCA, by truncating the SH series to certain degree, the accuracy of the model is affected. Each direction on the sphere is weighed differently by each spherical harmonic. The weight each spherical harmonic obtains, reveals to a certain extent the contribution of the associated dimensions on the original dataset. An advantage of the discrete spherical harmonic decomposition is that it approximates a continuous function on the sphere, it is therefore theoretically possible to obtain estimators for arbitrary points on the sphere, given a well-estimated model. Consequently, an HRTF representation in the spherical domain has been also used to estimate HRTFs in positions outside a given initial HRTF set.

Among others, Evans *et al.* [EAT98, EA98] described an HRTF representation using spherical harmonic expansion of HRIRs with and without onset time delays as well as magnitude and unwrapped phase in frequency domain. A dataset of 648 directions was analyzed whereas 32 additional directions that were not used for processing were measured in order to evaluate the accuracy of the model for interpolation. The authors suggested to use a 17th-order SH model (yielding 324 basis functions) to reproduce 90 percent of the HRTF energy and synthesize a pair of HRTFs at any

direction. The removal of ITD has improved the accuracy, and data representation in frequency domain showed a significant higher accuracy on reconstruction as in time domain. Most relevant spectral parameters for localization were included in the three first-order spherical harmonics. The authors concluded that highly efficient and compact representations of HRTFs are possible and the spherical harmonic functions would enhance our understanding of HRTFs. However, no psychoacoustic validation was carried out.

Romigh [Rom12] used the logarithmic magnitude instead of the linear one which has the side effect that the zero-order SH coefficients compose the diffuse-field part of the HRTF. In a localization experiment he confirmed that using an SH representation of order greater than four does not significantly affect localization accuracy. Romigh proposed a novel HRTF estimation method based on SH coefficients using only a small number of spatially distributed measurements. He claimed that only 12 measurement locations are necessary for adequate localization performance.

The success of spherical harmonic decomposition depends on the number and the distribution of the points on which the function has been approximated on the sphere as well as the spatial frequency with which changes in the function appear. Based on the spatial Nyquist criteria, at least N^2 measured source positions are required for an SH representation of order N, when assuming uniform distribution of measurement points on the sphere. Such a uniform distribution is not easy to achieve when considering HRTF measurement, as it is impossible to obtain measurements for points directly under the head for example. Finding an estimation technique that works well for an arbitrary measurement grid is still an open research question. To suppress problems caused by irregular sampled grids, one solution is to perform approximation by icosahedron subdivision [KSG99]. The use of a platonic solid that does not have a sampling point directly below can improve general model accuracy. Zhang *et al.* [ZK08] developed an iterative algorithm for extrapolating signals over the whole sphere when only a limited number of measurement points exist. Despite the lack of a quarter on the grid, the entire dataset could be successfully reconstructed with a 4th-order model. Zotkin *et al.* [ZDG09] addressed the limitation of under-sampled grids that are often used in HRTF databases. Several different grids (e.g. closed/open and high/low resolution grids) were tested and compared. They proposed a Least-Squares Fitting method that operates on any arbitrary grid. A Thikonof regularization was found to compensate in a satisfactory manner, when the problem of inverting arising in least-square estimation of SHWs is ill-posed.

Avni and Rafaely [AR10] studied the effect on the binaural cues ITD and ILD when using an incomplete representation with a finite order of HRTFs in the spherical harmonic domain. Original and several manipulated versions of the CIPIC HRTF database were used for analysis. On the basis of just-noticeable differences (JND) of ITD (10-20 μs) and ILD (1 dB), and r as the average human head radius, it was shown that the order $N \approx kr$ with $k = (2\pi f)/c$ as the wavenumber, is sufficient for reconstruction in order to preserve most of its spatial attributes. However, some directions in front or back of the head need more coefficients. Using this formula ($N \approx kr$) and assuming a human head with radius $r = 9$ cm and $f = 20$ kHz, a 33th-order model is required resulting in more than 1000 coefficients. Not every HRTF database has so many measured directions (see Table 5.1, Page 38).

In summary, SH expansion is a promising technique to produce continuous functions in spite of having only a small number of sampled positions. The difficulty is rather to bring the dataset into a suitable form that fits SH transformation. Almost all models are focusing on the frequency domain, since the use of HRIRs needs a higher spherical harmonic order. The dimensionality can be reduced while still maintaining relevant localization cues. This confirms the fact that a typically-measured HRTF includes a lot of perceptually-irrelevant data. Studies are in agreement that the first four SH coefficients are essential for understanding localization cues. They capture the information to distinguish between left/right, up/down and front/back. Also interesting is the fact that the expansion coefficients contain all HRTF information so a comparison between different databases and measurement methods is much easier than with other methods. However, currently no HRTF customization procedure where subjects adapt spherical harmonic weights was found.

Chapter 4

HRTF Individualization Methods

In last two decades, a variety of different methods for obtaining non-individualized HRTFs from a generalized HRTF set or an HRTF set of another individual have been proposed. The methodologies vary from short simple selection tasks to time-consuming tuning applications in which the subject has to adapt or compare several parameters. In almost all studies, before and after individualization, a short auditory localization test is carried out in order to validate the performance of the method.

4.1 Literature Review

4.1.1 Identifying a Near-Optimal Set

Qian *et al.* [QE98] proposed an HRTF individualization method that allows subjects to select the best-matching HRTF from an HRTF set by judging localization performance, coloration and externalization. Stimuli from an HRTF set that included 12 positions on a circle parallel to the horizontal plane, at three fixed elevations (-30°, 0°, 60°), were presented over headphones. Subjects evaluated externalization (yes/no), the form of the circle in azimuth and the vertical accuracy using a scale from 1 to 10. After 35 minutes, from the existing 26 HRTF sets, the six best-matching sets could be extracted. Then paired comparison were conducted on the best-matching ones, in which each position was presented by two different HRTF sets and the subject chose the best-matching one. Finally, through a cyclical presentation of the six HRTF sets, the test persons reevaluated the virtual sources only by a single criterion (scale from 1-10). By combining the results of these two examinations, a best-matching HRTF was identified. After the procedure, a subsequent hearing test confirmed some improvements in localization accuracy.

Seeber and Fastl [SF03] presented a fast method for selecting a best-matching HRTF set from 12 existing ones by evaluating localization, spatiality and externalization. In a preselection task, 5 of 12 sets could be extracted by neglecting HRTFs that suffer from in-head localization or front/back confusions. Within ten minutes, the subjects found an HRTF through several objective and subjective selection tasks. The authors also suggest to use direct comparison because the differences in HRTF sets are often very small to identify. Although subjective selection has proven useful in minimizing inside-the-head localization and front/back confusions, however, selecting from a larger database with more than 50 sets, results in increased test-time, which can yield listener fatigue. Furthermore an HRTF set from another person might only be valid on some local positions and this has not been examined by the authors. For this reason alone, such a method has limitations.

4.1.2 Anthropometry based Individualization

Other individualization methods operate by relating the physical anthropometric dimensions of the ear to specific HRTF parameters. Since measurement of anthropometric dimensions is not as expensive and tedious as measurement of HRTFs, this might be good alternative approach for automatic HRTF individualization. For example, the characteristic of a listener's outer ear can be used to predict certain spectral features. Positions in azimuth can be derived from the head circumference and consequently the resulting diffraction or shadowing effects. The physical parameters can be measured either manually by hand or automatically extracted from a pinna photography. Rodriguez and Ramirez [RR05a, RR05b] correlated PCWs and central frequencies of pinna notches (NCFs) with existing dimensions of the pinna. Instead of HRTFs, they used 64 pinna-related transfer functions (PRTFs) because the features of the pinna might be better related to them. The most important parameters turned out to be cavum concha height, pinna height and pinna rotation angle. To further improve the correlation factor, linear regression was applied. The goal was to estimate PCWs and NCFs from existing anthropometric parameters and model the resulting PRTFs. Similarly, Zotkin *et al.* [ZHDD03] proposed using a head-and-torso model for this task. This method could be useful for automatic extraction of pinna parameter from an image of the outer ears.

Xu *et al.* [XLS09] used the CIPIC database to extract local and global key anthropometric dimensions. The former describes parameters that vary for each position, while the latter provides general information about the subject. The synthesized HRTFs could reduce the reconstruction error by about nine percent compared to the

average HRTFs. Xie *et al.* [XZR07] and Watanabe *et al.* [WOI+07] put their focus on the estimation of the ITD. Xie claimed that the ITD varies through different ethnic groups. Watanabe used ITD estimation by physical dimensions to improve localization accuracy in the horizontal plane. Satarzadeh [ADS07] and Hu *et al.* [HCW08] continued this work and established pinna models as well as multiple regression models between the characteristic parameters of HRTFs and the anthropometric parameters. Also very common is the use of *Higher-order Singular Value Decomposition*, also known as *Tensor SVD* [GV07] [RDS10] to obtain a better determination between the eigenvectors and the three dimensions (subjects, directions, frequency) of a dataset. Results indicate a reduced reconstruction error than the models with PCA.

Boundary Element Method (BEM) has been used to calculate the HRTFs from a given 3D scan of an individual head. Through a finite number of small triangular elements, the physical model approximates the anthropometry. A fine grid and therefore high computational cost is necessary to simulate frequencies up to 20 kHz and all important individual features. Sottek [Sot99] compared measured and simulated HRTFs and concluded that the physical modeling of some pinna dimensions is still inaccurate. Similarly, Chen *et al.* [CK07] described that spectral features below 7 kHz can be estimated with good accuracy whereas peaks and notches in higher frequencies are still poorly represented.

On the whole, the introduction of the physical dimensions is a good approach, but the measurement is still flawed. Only a small measurement error of the anthropometric dimensions can lead to large discrepancies in the model, especially for high frequencies. Beside that, the correlation between anthropometric features is low, therefore it is not possible to predict a feature from knowledge of another one. Additionally, the simplified geometric models lead to inaccuracy, since the theoretical computation of more complex parameters is quite difficult and computationally expensive. For this reason a mixture of using anthropometric parameters as a starting point and additional subjective adaption would be a more robust procedure.

4.1.3 Frequency Band Adjustment

Several psychoacoustic studies investigated the role of spectral manipulation in certain frequency bands and their effect on sound localization. Silzle [Sil02] described an individualization process that considers smoothing and equalization of frequency magnitude as well as phase modification of two HRTF sets. A tuning expert adjusted each of the five directions in the horizontal plane (-135°, -90°, 0°, 90°, 135°) individually. Ten subjects participated the subsequent listening test with regard to in-head

localization, distance, coloration and localization accuracy. The results confirmed improved performance of the adapted transfer functions. However, the tuning process is very time consuming and the individual differences of the subjects could not be handled by the expert.

Middlebrooks *et al.* [Mid99b, MMO00] took a different approach to individualize HRTFs. Subjects scaled non-individual DTFs along the frequency axis to shift spectral peaks and notches and consequently minimize inter-subject variability. He even tried to estimate the optimal scaling factor through physical dimensions [Mid99a]. However, the number and location of spectral cues in frequency spectrum is often very different for individuals, so this method is not always successful. Anyhow, when a small subset of good-matching HRTFs is already chosen, this approach could be appropriate.

Tan and Gan [TG98] proposed a 2-step customization process. In the first part, the user chose a best-matching HRTF set, then fine tuning is carried out through ITD adjustment and spectral manipulation. A bank of four parallel bandpass filters and one high pass filter simulate the directional bands and the energy in each frequency band can be amplified or attenuated. In an localization test, ten subjects reported a reduced number of front/back confusions compared to non-individualized HRTFs.

So *et al.* [SL11] manipulated the energy in six frequency bands in the magnitude spectrum to enhance forward or backward perception. By altering the energy levels (± 0, ± 12, ± 18 dB) in these six bands, subjects had to judge several directions based on front/back discrimination. Using manipulated HRTFs, a sound could be coming more likely to be perceived from the front or back. Consequently, the localization error was reduced by 54 percent and the number of front/back confusions by 45 percent.

4.1.4 Principal Component Analysis

The number of studies investigating subjective adaptation of HRTFs using a PC model is small, the majority of them focusing on sounds on the median plane. PCA has been explored more as a modeling rather than a individualization process. Figure 4.1 indicates the main idea for using PCA as an adaptation tool. The individualization process assumes that by adjusting the weights of the principal components, one can eventually find a combination that results in the desired auditory impression.

Hwang *et al.* [HPP08] extracted 12 PCs of median plane HRIRs in the CIPIC database. Prior to PCA, the initial time delay was neglected and only the first 1.5 ms were included to extract effects of torso, shoulder, head and pinna. Principal

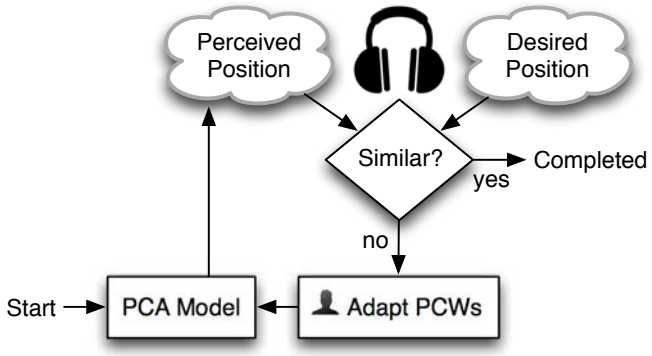


Figure 4.1: Overview of the adjustment process for the PCA model.

components were calculated for each position independently. Components were sorted in terms of the variability of the associated weights, which is directly related to the standard deviation of the PCWs. Participants were asked to adjust the weights of the first three components for nine source positions (-30° to 210°). The remaining PCWs (4-12) contributed to the reconstruction as a mean value over all individual weights. Three subjects participated the customization procedure by changing sliders using a graphical user interface (GUI). The range of the sliders was set to be mean ± 3 standard deviation. In a subjective listening test, the subjects reported enhanced front/back and up/down discrimination compared to the KEMAR HRTF set. Since also the original HRIRs were measured for the test subjects, no statistically significant difference in localization error was found between customized and measured impulse responses. Unfortunately, no information about the time needed for customization was given.

In [HPP10] they continued the work and proposed an individualization procedure for three source positions (0°, 70, 180°) in the median plane that are endpoints of two sectors. Participants adjusted three PCWs for each position, yielding a total of nine parameters. Weight values for the rest of the positions in the median plane were interpolated on the grounds of the adjusted PCWs. Nine subjects tuned their HRTFs taking on average 17 minutes, the results were consistent with the previous study.

Shin and Park [Shi08] applied PCA on the left ear HRIRs of the pinna response in the CIPIC database. On reconstruction, the right ear signal was estimated by the left ear and subjects could change the balance to center the sound. Nine positions in the median plane were adjusted by five PCWs respectively. A subsequent localization test of the four participants confirmed enhanced vertical and front/back discrimination

although the spectral features of individual and customized HRIRs were not matching. No indication about the duration of the test was made.

Fink and Ray [FR12] proposed a tuning model in the horizontal plane incorporating 34 left and right ear HRIRs of the CIPIC database. The data was arranged so that a principal component involves both ears. The five PCWs with the highest standard deviations were extracted (2, 4, 7, 8, 3) and tuned in three rounds. In the first experiment, the average HRTFs of five source positions in the frontal region were adjusted by the subject, after this the PCWs of the same HRTF have to modified so that the sound is perceived rotated by 180 degrees. In this way, front/back discrimination was analyzed. A GUI with six sliders indicates the five PCWs and the interaural time delay. Subsequent listening tests confirmed reduced front/back and average azimuth perception errors, however, only one subject participated in the tuning experiment. In addition, no further details were given about the slider range or test duration.

4.2 Technical Aspects

Beside the challenge of subjective adaption, several technical aspects have to be considered when synthesizing HRTFs or HRIRs in particular. Two important issues are discussed here in more detail. First the computation of the phase response and secondly the influence of the headphone transfer function to the frequency spectrum of HRTFs.

4.2.1 Phase Reconstruction

Almost all of the aforementioned techniques operate on the magnitude spectrum of HRTFs. During the individualization process, HRIRs are typically transformed into the frequency domain, manipulated and transformed back into time domain by Fourier Transform pairs. Since the spectrum magnitude has been changed in frequency domain, the original phase can no longer be used for reconstruction of the HRIR. Consequently, the phase information must be either estimated or approximated. Usually the HRTF phase is divided into a minimum-phase and an excess-phase component. Whereas the minimum-phase component contains all relevant spectral cues, the excess-phase part considers the time related cues for localization [Tol10], therefore this term is commonly estimated only by a constant time delay.

Each transfer function can be separated into a minimum-phase filter and a corresponding all-pass part. The former contains all poles and zeros that are within the unit circle on the z-plane which follows that also its inverse function is stable. All

zeros that lie outside the unit circle, however, are shifted to the all-pass system. In order to obtain a constant magnitude spectrum for the all-pass system, the zeros must be equalized by poles that are inside the unit circle, so the all-pass remains stable. Finally these extra poles have to be cancelled by adding zeros at the same position in the minimum-phase system to fulfill the original transfer function.

A complex spectrum $H(j\omega)$ at frequency ω can be separated into magnitude and phase response by

$$H(j\omega) = |H(j\omega)| \cdot e^{j\varphi(j\omega)} \ , \tag{4.1}$$

with the phase $\varphi(j\omega)$ calculated as

$$\varphi(j\omega) = \arctan\left[\frac{\Im\{H(j\omega)\}}{\Re\{H(j\omega)\}}\right] \ . \tag{4.2}$$

The original phase is usually divided into two parts,

$$\varphi(j\omega) = \varphi_{min}(j\omega) \cdot \varphi_{ex}(j\omega) \ , \tag{4.3}$$

with a minimum-phase and excess-phase term. The latter is described by

$$\varphi_{ex}(j\omega) = \varphi_{lp}(j\omega) \cdot \varphi_{ap}(j\omega) \ , \tag{4.4}$$

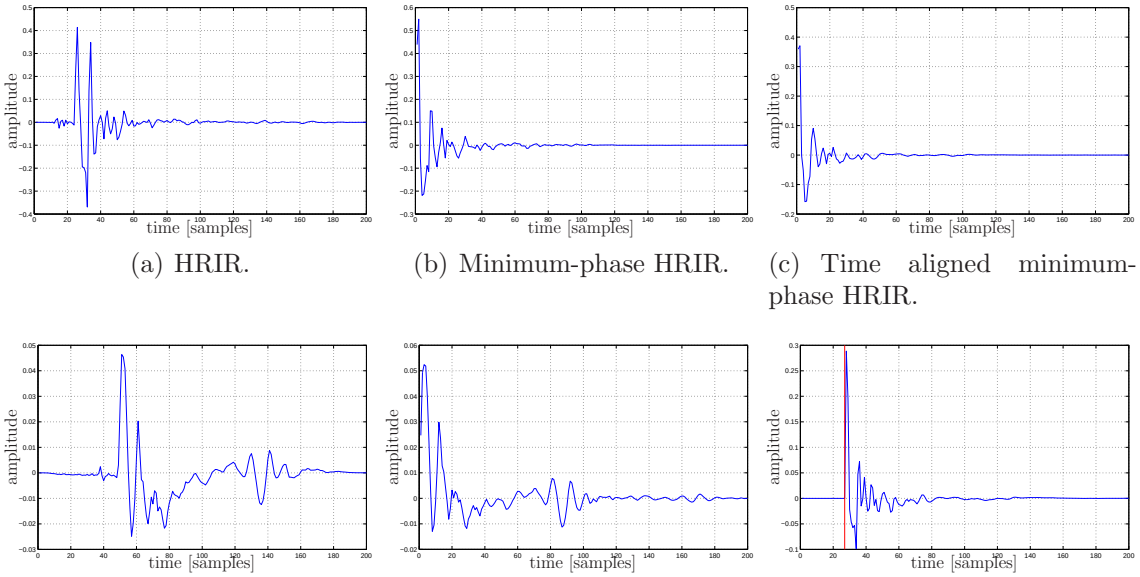
containing a linear-phase and an all-pass phase term. Commonly, the all-pass term is neglected because the auditory is sensitivity to absolute HRTF phase seems to be low [LEW10]. The remaining linear component is a simple frequency independent shift in time, which corresponds to the interaural delay. Consequently, an HRTF under minimum-phase assumption can be denoted by

$$H(j\omega) = H_{min}(j\omega) \cdot e^{-j\omega\tau} \ . \tag{4.5}$$

The phase response and logarithmic magnitude frequency response of a minimum-phase system are related through the *Hilbert Transform*. Thus, the desired minimum-phase $\varphi_{min}(jw)$ can be estimated through the logarithmic Fourier transform of the magnitude response,

$$\varphi_{min}(j\omega) = Im\left\{\mathcal{H}\left[-ln(|H(j\omega)|\right]\right\} \ . \tag{4.6}$$

Figure 4.2 shows a decomposition of left and right ear HRIRs into their minimum-phase versions and the subsequent time alignment. It must be pointed out that this relation can only be imposed when assuming causality of the minimum-phase system. Thus, the fact that HRTFs are almost minimum-phase can be an attractive quality.



(a) HRIR. (b) Minimum-phase HRIR. (c) Time aligned minimum-phase HRIR.

(d) HRIR. (e) Minimum-phase HRIR. (f) Time aligned minimum-phase HRIR.

Figure 4.2: Minimum-phase and delay decomposition for a pair of left (top) and right (bottom) ear HRIRs. Red line indicates the resulting interaural time difference calculated by the maximum of the cross-correlation between left and right ear HRIR. Sound source is given at azimuth 80° and elevation 0°.

Decomposition of HRTFs into a minimum-phase part and an excess-phase term is widely applied and verified through subjective hearing tests [KW92, KC98, KIC99, POM00, NKA08]. Kulkarni and Colburn [KC98] claimed that the remaining all-pass term can be neglected because the auditory sensitivity to the absolute phase spectrum is low, thus ITD can be modelled as a frequency independent time delay. Contrary, Mehrgardt and Mellert [MM77] emphasized that HRTFs are nearly minimum-phase only up to 10 kHz. Similarly, Nam *et al.* [NKA08] calculated the maximum cross-coherence between original and minimum-phase HRTFs and showed that HRTFs are essentially minimum-phase. A majority of the dataset had a cross-coherence greater than 0.9.

Since a minimum-phase HRIR is truncated in relation to the original one, the effect of truncation in relation to localization must also be examined. A welcome side effect is that the energy of the resulting impulse response is concentrated at the beginning, leading to shorter filter lengths and reduced computational complexity. Senova *et al.* [SMM02] investigated how truncated HRIRs effect localization accuracy. The spectral resolution of an HRTF magnitude is reduced when the corresponding HRIR is truncated in time domain. Several different durations ranging from

0.32 to 20.48 ms were tested and they found that the accuracy is affected even at modest smoothing. Localization performance was affected below a HRIR duration of 5.12 ms, but dramatically below 0.64 ms. This is in contrast to Kulkarni and Colburn [KC98] who claimed that extreme smoothing from 256 to only 16 coefficients does not significantly effect localization performance. Senova interpreted the different results so that Kulkarni has involved only four source positions in the interaural horizontal plane whereas 354 directions were tested by Senova. Consequently, some crucial spectral details that are necessary for localization judgments in 3D were not tested in Kulkarni's experiment. It has to be noted that in contrast to Senova, Kulkarni performed smoothing in frequency domain by truncating the Fourier series expansion of the logarithmic power spectrum (*power cepstrum*) which results in low-pass filtering. However, despite the different processing of the signals, the resulting psychophysical effect tend to be the same. Senova suggested to use at least a duration of 10.24 ms or even 1.28 ms when a slight decrease in accurateness is acceptable. Similarly, Zahorik *et al.* [ZWK95] investigated in differences between free-field and truncated virtual sound stimuli and concluded that from a spectral resolution of 195 Hz or higher the pairs are distinguishable.

Concluding, the majority of independent studies are in agreement that the described phase separation can be used without any significant impact on the localization accuracy. The HRTF model presented in this thesis is also based on this assumption.

4.2.2 Headphone Transfer Function

Many HRTF models and virtual acoustic simulation try to reproduce the spectral signal at the eardrum as coming from a natural sound source. Studies agree that headphone equalization is necessary, especially in synthesis of sound sources in vertical plane. Before each hearing test, first the transfer function should be measured and then compensated in the experiment. This is more true for closed cup headphones, open headphones are less susceptible to this.

Whereas the headphone transfer function (HPTF) does not include any directional or distance information, the function depends on how the apparat is placed on the head relative to the ear. Kulkarni and Colburn [KC00] pointed out that the HPTF also differs for each listener, because of the anatomical structures of the outer ear. In a study they described the large variability of the headphone transfer function using the KEMAR (Knowles Electronic Manikin for Acoustic Research) dataset and the widely-used headphone Sennheiser HD520. 20 different measurement were executed

and after each one the headphone was taken down and put back on again in order so simulate variability in headphone placement. They confirmed that using a mean HPTF to compensate differing headphone placements is not suitable. Moreover, the spectral characteristics of the headphone transfer function have almost the same properties as the directional features in HRTFs, that is why additional undesirable perceptual effect may be introduced.

However, no general solution for this issue has been proposed. For this work, the HPTF was measured once with a dummy head for the headphones AKG K271 Studio and AKG K272 HD.

4.3 Conclusion

HRTF individualization is still an open research but promising techniques have been proposed in recent years. The majority of the studies are based on selection of HRTF from an existing set, individualization using anthropometry and individualization using orthogonal basis function models.

Selecting an HRTF from a given set is interesting as it only takes little time to complete for example 10-20 min [SF03]. While this addresses the problem of finding an optimal HRTF within a dataset, it is not solving the problem of HRTF individualization. Such solutions could work for certain directions, however, the degree of individualization of HRTFs makes it difficult to assume that such a solution would work in general.

Anthropometric methods are interesting, however, a simple relation between the physical parameters and the HRTFs or the weights of basis function models has not been established yet. This makes it still a research method which is not ready to be taken to HRTF individualization experiments and needs to be examined further. Another problem with this method is that taking measurements of the ear is not simple and large problems with the accuracy of measurements exist that make this method difficult to apply especially in high frequencies. Related to this, through Boundary element method an HRTF set can be calculated from an 3D scan of the head. However, to obtain accurate results, this numerical approach still suffers from computational limits.

Individualization using adaptation of the energy in frequency bands has also been shown to be promising. However, given the lack of agreement and quite likely the individual nature in defining the frequency bands that determine the localization cues,

this method is also difficult to apply in practice and can be only considered to be partially successful.

Individualization using adjustment of the weights of orthogonal basis functions was found to provide interesting results in psychoacoustic experiments. Certain problems that emerge there is that the transformations that have been used do not yield weights that can be unambiguously associated with spatial perception. The obtained principal components have only been partially correlated to the variables in a cartesian or spherical coordinate system, such as the one that is use to describe spatial experiences. This is not always due to a small database or an incomplete measurement set, the difficulty is in extracting meaningful parameters that are in some way related to perceptual effects. Even if HRTF individualization could in principle be done using a basis function model, such a process could quickly become exhaustive, if all individual positions are to be adjusted. Given that the performance of the subjects may fall appreciably after hour of concentration, such a procedure would be infeasible. For this purpose, it also must be mentioned that in the literature individualization for only a small subset of directions s commonly investigated. Mostly the focus is on few positions in the median plane to start initial investigations and avoid estimation of the interaural time delay. The current state of the art is still far away from a global adjustment for all positions in a *tolerable* test time.

Currently, there is still no method being clearly superior to the others. Probably an efficient individualization technique has to combine different methods presented here. For this work, the focus is on Principal Component Models. This was done because they have been shown to be promising in the process of individualization of HRTFs and also because they have the inherent capability to describe in detail the individual nature of HRTFs. However, given the variability of the methods used in the literature, it was not possible to proceed further without examining in detail the impact of design and implementation aspects in estimating a PCA basis. In the next chapter, a first step into analyzing numerically the impact of design choices in the formation of PCA input matrix on the reconstruction accuracy and the feasibility of individualization is conducted.

Chapter 5

Numerical Evaluation of the PCA Model

In this section, a decision is made how to structure the principal component analysis model. To this end, design choices are evaluated, pertaining to how the HRTF dataset can be restructured and preprocessed to allow for PCA to be performed. In order to establish to what extent the results are reproducible across databases, three different HRTF databases were used that are available online for academic use. Due to the large number of parameters involved in constructing an appropriate PCA input matrix based on the HRTF data, a first screening among possibilities is done on the basis of the PCA compression efficiency (defined as the number of principal components required to maintain 90 percent of the variance in the data) and the applicability of the possible input structures for the purpose of individualization. Based on these results, an HRTF representation is selected which is then subjected to a more detailed investigation that considers the HRTF reconstruction error.

Numerical simulations were performed in MATLAB®. The software that was developed for this purpose offers certain flexibility in defining the parameters entering the simulations. It is possible to define: 1. a dataset containing some or all of the HRTF data, 2. the representation of the input data (time, frequency, logarithmic or linear), 3. smoothing and filtering preprocessing options, 4. the structure of the PCA input matrix and 5. the way HRTF data from the two ears are represented in the input matrix. The parameters are presented in detail in Section 5.2. Figure 5.1 summarizes the main processing steps.

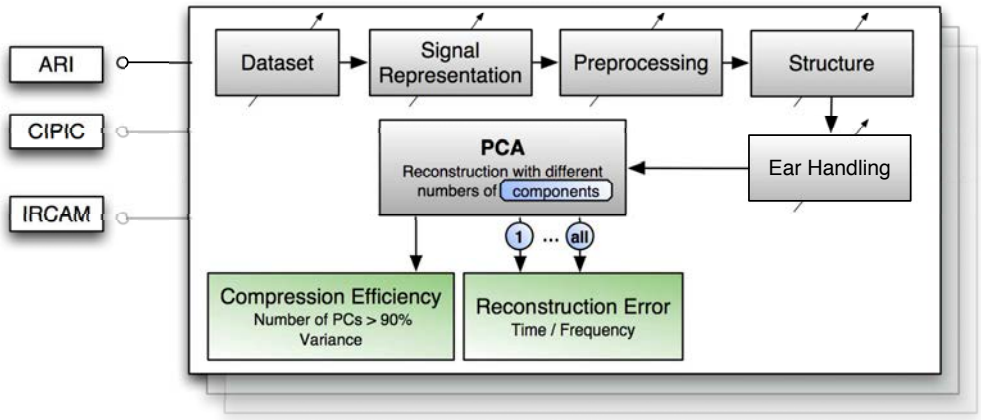


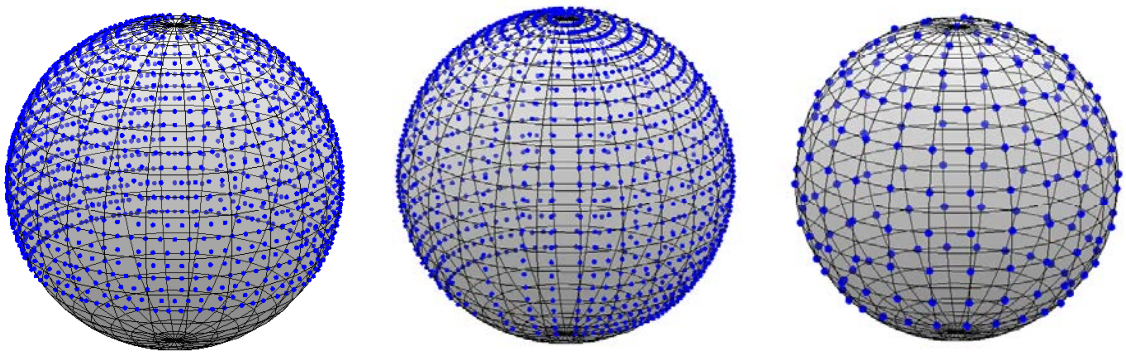
Figure 5.1: Numerical evaluation of input parameters.

5.1 HRTF Databases

Three open access HRTF databases from the Acoustics Research Institute (ARI) at Vienna, Institut de Recherche et Coordination Acoustique/Musique (IRCAM) at Paris and University of California at Davis (CIPIC) were used. Figure 5.2 and Table 5.1 display the number of subjects and the measurement positions in each database. ARI and CIPIC have the largest number of measurement positions (a total number of 1250 and 1550 positions respectively). ARI has the largest number of participants and the database grows continuously[1]. However, two subjects were excluded from calculations because impulse responses were not measured for all sound directions. In addition, subject ID 1034 in IRCAM database was detected as an outlier in several computations and therefore excluded from further processing.

In the following, most of the analysis results presented concern the ARI and CIPIC databases because: 1. they contain a large number of subjects and measurement directions, 2. to increase the consistency with other studies and 3. the ARI database was used in the listening experiment described in Chapter 7. It is worth mentioning that all existing HRTF databases contain few measurements in the lower hemisphere. The lowest elevation measurements in CIPIC and IRCAM are at -45 degrees.

[1]Majdak *et al.* [MIC+13] proposed a standardized but flexible format called Spatially Oriented Format for Acoustics (SOFA) for storing measured impulse responses which should overcome the incompatibility of current HRTF databases and enhance the exchangeability of measured data.



(a) ARI (b) CIPIC (c) IRCAM

Figure 5.2: Spatial resolution of HRTF databases. Each blue point indicates a measurement, thus a pair of HRIRs.

Identifier	Department	Subjects	Positions	Range
ARI	Acoustics Research Institute	85	1550	Az [0 360] El [-30 80]
CIPIC	University of California at Davis	45	1250	Az [-80 80] El [-45 230]
IRCAM	Institut de Recherche et Coordination Acoustique/Musique	50	187	Az [0 360] El [-45 90]

Table 5.1: Open access HRTF databases used for data analysis.

5.2 Simulation Parameters

Here, an overview of the various signal representations and their prefiltering used for PCA in literature is given. Particular attention is paid to the structure of the input matrix, since it mainly effects the application for HRTF individualization.

5.2.1 Variations in the Literature

In the literature, both time (HRIRs) [Wu97, GWFA05, FR12] and frequency [KW92, MG92, CvVH93, QE98, GV07, XLS09, Xie12] representations (HRTFs) have been subjected to PCA. Considering time representations, minimum-phase HRIRs have also been used [Shi08, HP08, HPP10]. In the case of HRTFs, PCA has been applied to both linear [QE98, RDS10] and logarithmic [Mar87, WK91, KW92, MG92, XLS09, Xie12] magnitude representations. Besides, PCA has not been always performed on the complete HRIR/HRTF dataset. Decomposition has been applied on the whole database [KW92, Xie12], smaller subsets, such as the median [HP08, HPP08, HPP10] or horizontal plane [FR12] and on single sound directions [Shi08].

The way the signals from the left and right ears enter the PCA input matrix has also been treated in different ways in literature. Sometimes only one ear is modelled and the second one is considered to be symmetric and therefore duplicated by the modelled one [Shi08, HPP08]. This solution is somewhat sub-optimal, especially given the lack of perceptual evaluation studies, however, there is currently no model that explains the differences of HRTFs across the two ears. Alternatively, it can be attempted to use PCA to explain the variability across the two ears. This can be done either by using the time/frequency signals from the second ear as independent variables in columns or expanded as observations in rows [WK91, KW92] in the PCA input matrix (Section 5.2.3).

Since databases of various sizes (2-85 subjects, 1-1550 directions) are used in the literature, it might be also relevant to investigate this. For that reason, a comparison in regards to compression efficiency between PCA of all sound directions and a smaller subset until only a single direction is done in the numerical evaluation. In addition, the impact of different numbers of subjects is inspected.

As there is a lack of comparison between the different studies, it is difficult to establish which representation is most useful for the purpose of individualization. To reach a conclusion, the impact of four different input data formats (raw HRIRs, minimum-phase HRIRs, DTFs with linear magnitude, DTFs with logarithmic magnitude), dataset structure as well as the way ears are inserted in the PCA input matrix on compression efficiency was investigated in the simulations that follow.

5.2.2 Spectral Smoothing

Kulkarni and Colburn [KC98] showed that smoothing the HRTF magnitude spectrum does not significantly affect localization performance. Smoothing was done by first taking the logarithm of the spectrum, performing FFT and then limiting the number of the Fourier coefficients that contributed in the spectrum reconstruction and transforming back to the linear domain. Even as few as 16 coefficients within a spectrum of 512 coefficients were found to yield satisfactory localization. This corresponds to a smoothing factor of 1/32 for effectively 512 frequency bins corresponding to an impulse response of 1024 samples. The impact of smoothing on PCA compression efficiency has not been evaluated per se. It can be expected, however, that as details of HRTF magnitude are smoothed out, compression efficiency will increase. For this reason, the smoothing factor was also included as a parameter in the simulations.

In order to achieve an equivalent smoothing for impulse responses of different length in the simulations, HRTF spectrum was smoothed by keeping $N/32$ spectral

coefficients upon reconstruction, where $2N$ is the impulse response length. For example, ARI database includes impulse responses of 256 time samples. Smoothed versions using 64, 32, 16, 8 and 4 coefficients upon spectrum reconstruction have been used. The resulting spectrum is shown in Figure 5.3.

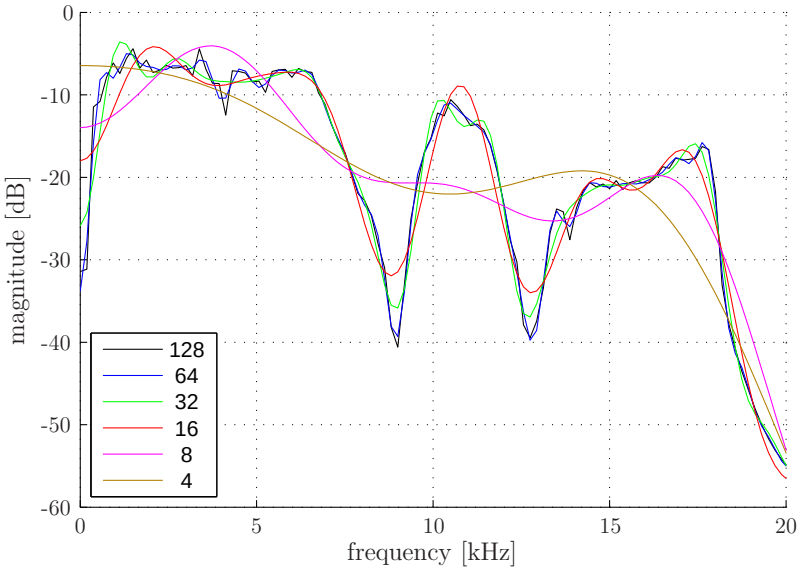


Figure 5.3: Spectral differences between unprocessed (128 coefficients) and smoothed DTF magnitude spectrum (with 64, 32, 16, 8 and 4 coefficients) in ARI database.

5.2.3 Structure of the PCA Input Matrix

Generally speaking, PCA works equally well on different realizations of a 2D input matrix with columns as variables and rows as observations. As the problem of HRTF modeling is multidimensional, because of the variability observed across subjects, time samples/frequency bins, sound directions and the two ears, the arrangement of the HRTF data in a two-dimensional matrix is open to different interpretations. Essentially, such an input matrix can be created based on, to a certain extent arbitrary, permutations of the four dimensions that appear typically in an HRTF dataset. The dimensions appear either in columns or rows of the input matrix. Considering the major trends in the literature as well as other possible arrangements, five structures which might be relevant for HRTF data decomposition were identified. These are called input data structures **Struct1** to **Struct5** in the following. These structures yield different principal components and corresponding weights, which need to be understood and made relevant to HRTF individualization.

In structure **Struct1** [$subjects \times (signal * sound\ directions)$], the complete dataset consisting of the aggregated data (either in frequency or time domain) in the different measurement positions are used as independent variables in columns, while the observations from different subjects are deployed in rows (Figure C.1, Page 102). PCA returns one weight that can be used to recreate the entire HRTF dataset for each person or two weights to recreate the HRTF dataset for each ear, depending on how ears are handled in the input matrix.

Structure **Struct2** [$(subjects * sound\ directions) \times signal$] is a pattern that has been used by Kistler and Wightman [WK91, KW92]. Here, signal bins (in frequency or time domain) are the independent variables in columns, while replications for the different subjects and measurement directions are used as observations in rows (Figure 5.4, Page 49). Four different realizations can be distinguished, depending on whether a single ear is used and on whether both ears are included as independent variables in columns or replicated as observations in rows. PCA returns PCs that weigh the contributions of each frequency bin differently and either PCWs for each direction and subject when data for the second ear are blocked in each column, or PCWs for each direction, subject and ear when ears are included as observations in rows.

In structure **Struct3** [$signal \times (subjects * sound\ directions)$] (Figure C.2, Page 103), the subjects and positions under consideration are used as independent variables in columns and the time or frequency domain signal bins are used as observations in rows. One obtains PCs that weigh the contribution of each subject and sound direction differently. Depending on how data from the second ear are included in the matrix, this can either be modelled using a single PCW (when ears are blocked as independent variables in columns) or one obtains one PCW for each ear. The resulting PCWs for each frequency bin or time sample can be used to recreate the variability of each frequency bin for a specific subject and sound direction.

Structure **Struct4** [$(signal * sound\ directions) \times subjects$] (Figure C.3, Page 103) deploys subjects as independent variables in columns whereas signal bins (in frequency or time domain) of all source positions are listed as observations in rows. PCA returns PCs that weigh the contribution of each subject or ear differently, depending on the handling of the ears. The PCWs can be used to recreate the contribution of a signal bin for a particular sound direction for each subject. When the data of the second ear are blocked as observations in rows, there are additional PCWs for each ear, otherwise ears are considered as independent variables and the PCWs take both ears into account.

Structure **Struct5** [$(subjects * signal) \times sound\ directions$] (Figure C.4, Page 104) was also used by Xie [Xie12] and lists each position on the columns as independent variables while frequency or time samples from all subjects are considered to be observations in rows. PCA of this representation is applied in the spatial domain, unlike previous structures in subject, frequency or time domain. One obtains principal components that weigh the contribution of each position and ear differently when ears are listed as independent variables in columns. The corresponding weights can be used to recreate each frequency or time bin for a given position for a specific subject.

5.3 Impact on Compression Efficiency

Table 5.2 presents the results of the manipulation of the main independent variables used in the simulations on the compression efficiency of a PCA model using the ARI database as input. Both ear signals are considered in the results. In Appendix C.2, a complete tabular analysis is also given for single ears and other databases.

		HRIR		Min HRIR		DTF lin		DTF log	
		E ↓	**E** →	**E** ↓	**E** →	**E** ↓	**E** →	**E** ↓	**E** →
Struct1	S/1	92	55	56	52	28	47	40	62
	S/2	92	55	56	52	26	46	35	61
	S/4	92	55	56	52	24	45	28	59
	S/8	92	55	56	52	18	39	19	55
	S/16	92	55	56	52	9	30	9	50
	S/32	92	55	56	52	6	24	5	46
Struct2	S/1	15	29	11	20	7	12	6	10
	S/2	15	29	11	20	6	12	5	8
	S/4	15	29	11	20	6	11	4	7
	S/8	15	29	11	20	5	9	3	6
	S/16	15	29	11	20	3	6	3	4
	S/32	15	29	11	20	2	4	2	2
Struct3	S/1	29	15	12	7	8	6	2	1
	S/2	29	15	12	7	8	5	2	1
	S/4	29	15	12	7	7	5	2	1

continued on next page

continued from previous page

		HRIR		Min HRIR		DTF lin		DTF log	
		E ↓	**E →**	**E ↓**	**E →**	**E ↓**	**E →**	**E ↓**	**E →**
	S/8	29	15	12	7	6	5	2	1
	S/16	29	15	12	7	3	3	2	1
	S/32	29	15	12	7	3	3	2	1
Struct4	S/1	48	85	13	19	6	8	1	2
	S/2	48	85	13	19	6	7	1	2
	S/4	48	85	13	19	5	7	1	2
	S/8	48	85	13	19	4	5	1	2
	S/16	48	85	13	19	2	3	1	2
	S/32	48	85	13	19	2	3	1	2
Struct5	S/1	151	174	9	8	5	6	2	1
	S/2	151	174	9	8	5	6	2	1
	S/4	151	174	9	8	5	5	2	1
	S/8	151	174	9	8	4	5	2	1
	S/16	151	174	9	8	3	3	2	1
	S/32	151	174	9	8	2	3	2	1

Table 5.2: Number of PCs required to yield 90 percent variance for different realizations of a PCA input matrix based on the ARI dataset. S/1 refers to no smoothing and S/2 ... S/N to different degrees of HRTF spectrum smoothing (see Section 5.2.2). Horizontally, the input signal representations are given and whether ears are blocked in rows (E↓) or columns (E→). Five different input structures and variations of spectral smoothing are listed vertically.

5.3.1 Impact of Input Structure

By observing Table 5.2, it can be seen that the structure of the input matrix has a great impact on compression efficiency. The effect of the input structure seems to be the same across different configurations of the other parameters. Using Struct1, 28 to 92 PCs are required to explain 90 percent of the variance on reconstruction, depending on the ear configuration and whether ear-transfer functions are in the time or the frequency domain. The number of components is less when using DFTs instead of HRIRs, but still too high. This is not surprising, since one PC weigh the data of all listening positions of a subject, consequently the variability of the corresponding component weights is very high. However, this greatly restricts the application of such a representation.

For Struct2, only 6 to 12 PCs are essential to express 90 percent variance when transfer functions are represented in the frequency domain. Only about half of the components (1-8 PCs) are necessary for Struct3. Interestingly, with logarithmic DTF

and ears blocked in columns, only one component is required. Apparently, the variation through subjects, positions and ears for each frequency bin can be well described by only one component.

Similarly, in Struct4, 1-8 components are necessary to achieve 90 percent variance in frequency domain whereas only 1-2 component are required with the logarithmic version. For Struct5, 90 percent variance is obtained in frequency domain using 5-6 components in the linear case and only 1-2 in the logarithmic version. This reveals that the variability observed in the different directions can be well presented by a smaller number of components, which, however, weigh the contribution of each direction in a way that optimally explains the variance in the dataset.

5.3.2 Impact of Ear Handling

Table 5.2 provides us with insight on the impact of the different ways to include ears in the database. In general, it can be seen that when ears are blocked in columns more components are required compared to the case in which ears are used in rows. This is to be expected as the dimensionality of the problem increases. Structures Struct1, Struct2 and Struct4 in both domains need more components when ears are listed as independent variables in columns. The increase of the required components is not surprising, since the signals from the two ears are stringed together. Consequently, a particular principal component must describe the double amount of data. Therefore the difference between the components is up to twice the number.

However, representation in time domain with input structure Struct1 indicates less components for ears blocked in columns whereas the opposite is the case in frequency domain. Similarly, structure Struct3 the number of components is decreased when ears are blocked in columns but this is consistent for all signal representations.

5.3.3 Impact of Dataset

Seven subsets of subjects, starting with only two up to the maximum number of individuals in the HRTF database, were used for PCA. Results for Struct2, Struct3 and Struct5 indicate that the usage between 2 and 10 individuals has an impact on compression efficiency, such as about 3-5 components can be saved with a smaller set of individuals. However, when using more than 10 subjects, there are hardly any differences. In contrast, when using Struct1 and Struct4, also great differences are evident for the entire range of subjects.

The number of sound directions was manipulated starting with only one direction and increased in six steps until the total amount of positions. For Struct2 and Struct3, a reduced number of sound directions leads to a worse compression efficiency whereas for the remaining structures the compression efficiency improves. In the same way as with subjects, differences are only visible on very small numbers, up to about ten positions. If one uses more than ten positions, there are no differences with respect to the compression efficiency for all structures.

5.3.4 Impact of Signal Representation

In general, time domain compared to frequency domain signals need more basis functions to explain the same variance due to the extensive nature of the data including both magnitude and phase. However, there are also great differences between the two representations in time domain, worth mentioning is the difference between HRIRs (174 PCs) and their minimum-phase versions (8 PCs) using Struct5. Above all, the removal of the onset delay before computing the PCA greatly reduces the amount of principal components.

Leung and Carlile [LC09] also investigated the PCA compression efficiency and came to the conclusion that the optimal format for PCA decomposition in terms of compression is the linear amplitude form in frequency domain. They used an HRTF dataset of 393 directions, but the number of subjects was not specified. Moreover, the structure of the PCA input matrix was not defined. From their results (5 PCs for 90 percent variance in linear magnitude), one can conclude that all described structures except Struct1 come into consideration.

Table 5.2 only indicates a better performance of the linear over logarithmic version for Struct1. A closer analysis of the remaining structures reveals that in general the logarithmic amplitude outperforms the linear version. Note that these results are not consistent over HRTF databases, because simulations with CIPIC and IRCAM produce fewer components with the linear representation. This can be seen in Figure 5.5 (Page 50), which indicates a better performance of ARI database (black line) with logarithmic spectrum than other databases. In addition, Table 5.2 shows that the more you smooth the frequency spectrum, the better results in the logarithmic spectrum. This is because of the nature of logarithmic compression. To this, Breebaart [Bre12] pointed out that despite a linear representation is more efficient in regards to the explained variance, using the logarithmic domain yields a smaller root mean square error (RMSE). Consequently, the total explained variance for a particular set of PCs does not always reveal the error in magnitude spectrum.

5.3.5 Impact of Frequency Smoothing

Spectrum was smoothed according to the rationale presented in Section 5.2.2. By observing Table 5.2, it can be seen that smoothing greatly reduces the number of components required for representing 90 percent of the variance for all manipulations of the other independent variables. In general, each time the Fourier coefficients are reduced by 50 percent, 1-2 components can be saved.

5.3.6 Conclusion

Concluding this investigation, the choice of a particular kind of HRTF representation in time or frequency domain indeed has impact on the performance of subsequent procedures. Assuming that minimum-phase HRTFs can provide sufficient auditory impressions, choosing a PCA model that operates on spectral data seems an obvious choice as one can explain more variance with fewer components.

The results were not clear enough to obtain insight on whether a linear or a logarithmic spectrum is more appropriate. For the majority of the input structures investigated and in the case of the ARI database, best compression efficiency was obtained for a PCA model based on logarithmic HRTF magnitude in contrast to other HRTF databases (see Table 5.2, Page 43), where linear magnitude offers some advantage. It is worth mentioning that using a linear magnitude can cause problems of HRTF magnitude undershoot (i.e. values smaller than 0) upon reconstruction with few components and/or interpolation between different principal component weights. This means that it is easier to obtain a stable way of interpolating between principal component weights using a logarithmic representation. In the case of ARI database, a compression efficiency advantage emerges as well, but for other databases the gain in explained variance when using linear magnitude is relative small.

Spectral smoothing seems also useful to use, up to the point where no important information is lost. Kulkarni and Colburn [KC98] proposed keeping 1/32 of the spectral coefficients, that is 8 coefficients in the case of the ARI database. This led to audible coloration in informal hearing tests that were performed, and for this reason, 32 of 128 coefficients were preserved, which led to no audible coloration, and a small increase in the explained variance due to a smoother spectrum.

The goal of this model is to capture and analyze individual differences, therefore all available subjects in the HRTF database were used. Similarly, all sound directions were included to obtain a spatial relationship between the principal component weights.

The appropriate PCA input structure is a choice that needs to balance two parameters. On one hand, it should lend itself easily to individualization and on the other, enough variance should be explained. Structure Struct1 refers to a global modification of all subject's positions and PCA returns PCWs for each subject (and ear when using ears as observations in columns blocked). That could be used to recreate the signal at all listening positions, thereby decreasing the overhead of an adjustment process. However, this does not seem to be very flexible because a single parameters modifies all directions *simultaneously*. More than 20 components are required to describe 90 percent variance of the dataset. An implementation of this was done and in fact, it appeared that such an adjustment is not useful. Most variance is explained with Structures 3/4/5. However, the weights one obtains from these structures cannot be easily used for individualization as they correspond to different frequency bins. Adjusting neighboring frequency bins is a task that is difficult to perform by hearing. In addition, the number of frequency bins that need to be adjusted even for a single position is quite high. For this reason, although the few components needed to explain variance for these structures is an interesting finding, the obtained model is difficult to be subjected to a perceptual adaptation task. An interesting alternative is structure Struct2, which requires relatively few components to explain a large proportion of the variance, in practice 10 PCs for 90 percent variance on reconstruction. At the same time, using this structures, one obtains a PCW set that can be easily used for individualization. As with this structure each PC is a function of frequency, by adjusting the weights one effectively interpolates between the PCW for each subject and sound direction. This is a task that can be easily translated to be used within a listening test.

Choosing an appropriate way to include both ears in the model is a difficult task. One obtains best results when only one ear is used or both ears are included as observations in rows, however, it was not possible to find neither subjective tests not a model that can sufficiently explain how data for the second ear can be obtained from data from the first ear. In addition, obtaining different weights for each ear leads to the situation where one needs to adjust HRTFs for both ears, which effectively doubles the number of adaptations one needs to perform. This naturally forms an interesting problem, which is, however, not investigated further in the thesis. In the following, the decision was made to combine left and right ear data by concatenating each pair of DTFs into a single vector, so that ears are blocked in columns. Consequently, both ears can be adjusted simultaneously using one PCW, albeit to an (unavoidable at

this point) sacrifice in the number of components required to explain 90 percent of the data variance.

5.4 Detailed Analysis of the selected Structure

Here the selected structure Struct2 is analyzed in a similar way as before, however, the analysis is replicated across the three databases and the reconstruction error for difference PC subspaces is also calculated. Figure 5.4 illustrates the selected structure and how information from the second ear could be handled. In the analysis and for the reasons explained before, both ears are used as independent variables in the calculations that are presented. In order to facilitate the comparison between databases, differences in the level of the measured HRIRs which affect the distribution of the PCWs, have been dealt with by applying global normalization of the raw data HRIRs in the databases to the amplitude of one. In this way, the distribution of the resulting PCWs becomes consistent across databases. This has no effect on the performance of the PCA, it facilitates though the comparison of the results between the databases.

5.4.1 Signal Representation

Figure 5.5 shows the explained variance as a function of the components used for four signal representations that are typically used in PCA for three HRTF databases. Consistent with previous analysis, PCA on raw HRIRs does not perform well in terms of compression efficiency, because of the direction-dependent onset delay (ITD) and the higher input variability of the time domain signal. Minimum-phase HRIRs need significantly fewer components to express 90 percent variance than their original versions. Best compression efficiency is achieved for frequency domain representations, where only the magnitude spectrum is taken into account. The choice of the database has an influence on the results. ARI with logarithmic magnitude yields the best compression efficiency, therefore this representation was chosen for the model.

In addition, spectral smoothing has a higher influence on DTFs with logarithmic than linear magnitude (Table 5.2, Page 43). When the Fourier coefficients are reduced to a quarter, the number of components for 90 percent variance can be reduced by three components in the logarithmic and only by one component in the linear case.

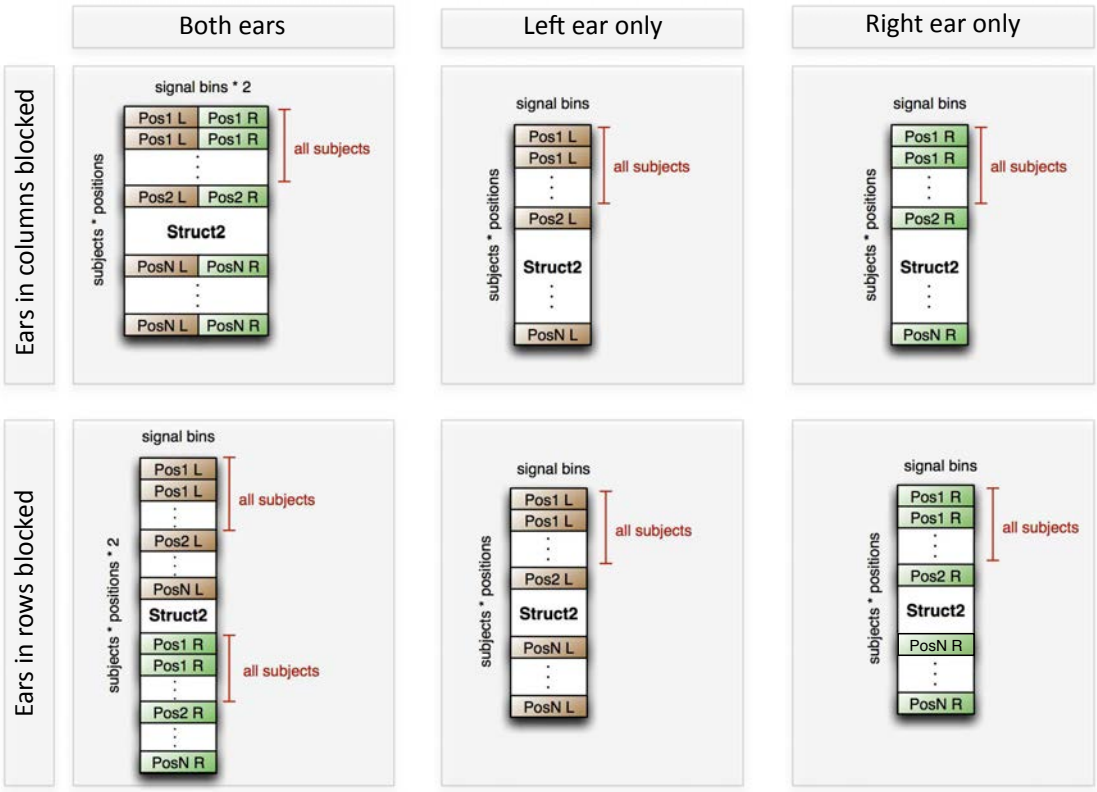


Figure 5.4: Dimension of input matrix **Struct2** $[(subjects * positions) \times signal]$ when choosing ears in columns (first line) or rows (second line) blocked.

5.4.2 Local or Global PCA

It also must be considered whether all positions are processed simultaneously [KW92, Xie12], or PCA is applied separately for each position [Shi08]. The common approach is the first one, because the latter yields different components for each sound direction and therefore the relationship between the directional weights in different positions is not possible to establish. This is an important limitation for the purpose of understanding the influence of PCW on perception, makes the development of PCW interpolation techniques difficult, but also complicates the construction of an individualization algorithm as PC sets for each direction of interest need to be available.

One could argue that as a particular position might share some spectral patterns, the decreased variability of the PCA input data might enhance the compression efficiency. However, as Figure 5.6a shows, this in not the case. For most practical applications, the number of principal components required is higher in the case of local compared to global PCA applied to logarithmic frequency magnitude of the structure under consideration.

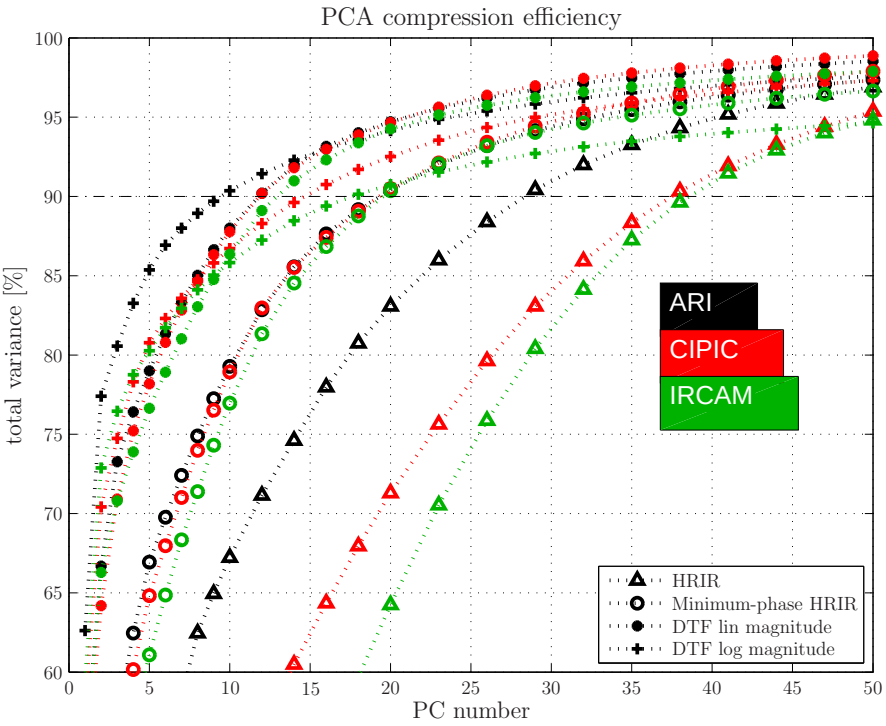
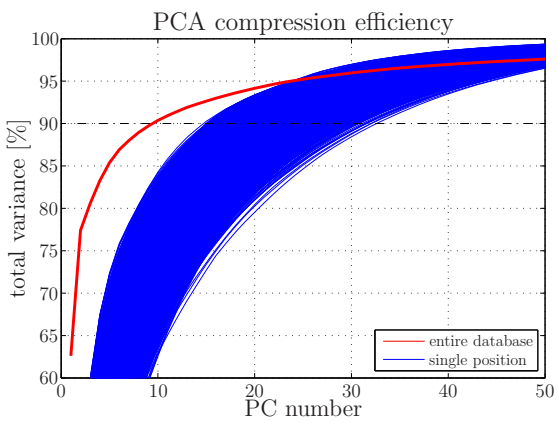
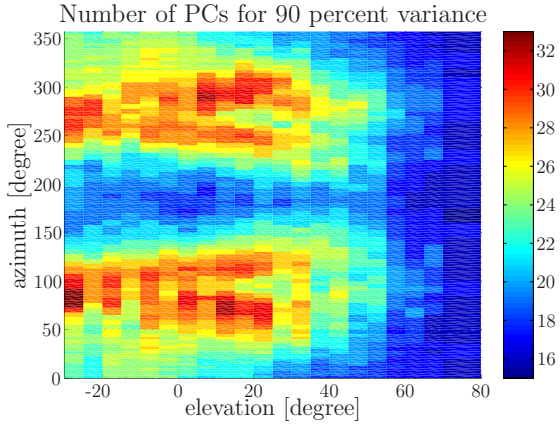


Figure 5.5: PCA compression efficiency of different input data formats and HRTF databases. Four markers indicate the representation of the input data: HRIR (triangle), minimum-phase HRIR (circle), DTF with linear spectrum (asterisk) and DTF with logarithmic spectrum (cross).



(a) Compression efficiency when computing PCA of entire database (red line) or on all single position separately (blue lines).



(b) Number of required principal components yielding 90 percent variance when computing PCA for each sound direction separately.

Figure 5.6: Differences in compression efficiency between local and global PCA with logarithmic DTF magnitude in ARI database.

It has to be noted that there are also great differences in compression efficiency comparing PCA of all single positions. Figure 5.6b indicates the required components to yield 90 percent variance for logarithmic magnitude. Up to 32 components are necessary for lateral directions below 40 degrees in elevation. In contrast, 20 PCs are required for sound directions above and behind the head.

5.4.3 Reconstruction Accuracy

In order to reduce the adaptation time, an PCA based HRTF model will necessary operate on a subspace of the principal component basis. Consequently, a reconstruction error is expected. To investigate this, the number of PCs for reconstruction was manipulated as shown in Figure 5.1 from only one to all PCs in five steps and the reconstruction error in time and frequency domain was estimated. In the simulation, the parameters in Table 5.3 (Page 58) reflecting the initial model choice were used to define the PCA input matrix.

5.4.3.1 Error Metrics

A measure for objective assessment of HRTF reconstruction in frequency domain is *Spectral Distortion (SD)*. The error metric for an arbitrary subject s and position θ between synthesized and real HRTFs is calculated by

$$SD(s,\theta) = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left[20\log_{10}\frac{|H(s,\theta,f_j)|}{|\hat{H}(s,\theta,f_j)|}\right]^2}\,, \tag{5.1}$$

where $H(s,\theta,f_j)$ and $\hat{H}(s,\theta,f_j)$ are measured and estimated HRTF logarithmic magnitudes respectively, f_j refers to the frequency index and N is the total number of frequency bins. The synthesized signal is more similar to the measured one when a small SD is obtained.

For a PCA HRTF model that operates on the magnitude spectrum, a relevant error measure in time domain is the reconstruction accuracy of the minimum-phase HRIR. When phase is neglected, the difference between original HRIR and reconstructed HRIR will be too large, which not necessarily reflects a high perceptual difference. Comparing to the minimum-phase functions reflects better the model reconstruction capacity as phase information is ignored. PCA is applied and reconstructed in frequency domain, afterwards HRIRs are synthesized using minimum-phase assumption

and compared to the minimum-phase functions of the measured HRIRs. *Signal-to-Distortion Ratio (SDR)* is defined as

$$SDR(s,\theta) = 10\log_{10}\left[\frac{\sum_{n=1}^{N} h^2(s,\theta,n)}{\sum_{n=1}^{N}\left[h(s,\theta,n) - \hat{h}(s,\theta,n)\right]^2}\right], \tag{5.2}$$

with $h(s,\theta,n)$ as original and $\hat{h}(s,\theta,n)$ as reconstructed impulse response. Index n indicates the time sample with a total number of N. Note that the higher the value, the better is the reconstruction. When the two signals are equal, the metric is infinite.

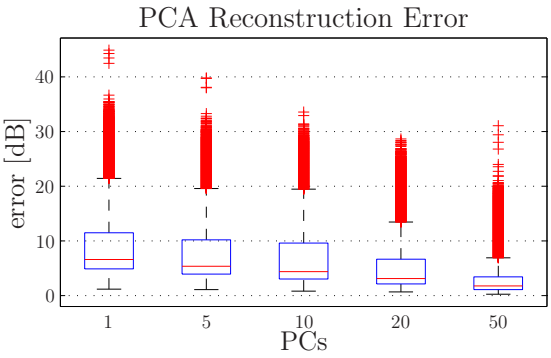
5.4.3.2 Results in Frequency Domain

The first row of Figure 5.7 shows the spectral distortion when using linear and logarithmic magnitude spectrum as input signal. Spectral distortion was calculated by comparing the original preprocessed HRTF dataset to the reconstructed one in which a different number of PCs has been used. For each of the PC subspace dimensionalities, a box plot based on the reconstruction error of all subjects, positions and ears is presented. The median value is indicated by the central mark, the edges of the boxes are the 25th and 75th percentiles and the whiskers extend to the most extreme data points. Outliers are plotted separately as red markers.

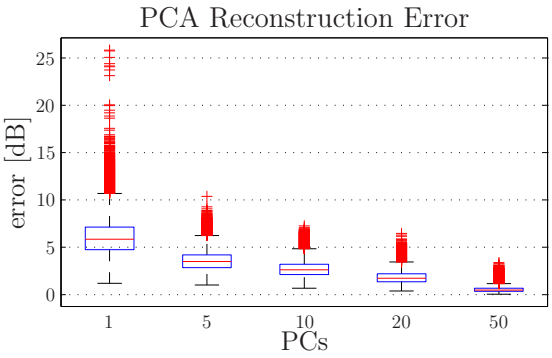
The results are consistent overall databases and indicate that at least five PCs are essential for an average error of about five decibel. As already mentioned by Breebaart [Bre12], the error is lower for processing in the logarithmic domain. To this, it should be noted that the difference between the average error values in linear and logarithmic magnitude is not great, but the range of the distribution and especially of outliers decreases in the logarithmic case.

The second row of Figure 5.7 shows the probability density function (pdf). It indicates the distribution of the error for linear and logarithmic magnitude when using ten components for reconstruction and generally indicates a skewed distribution in the case of linear amplitude and a normal distribution in the case of the logarithmic one.

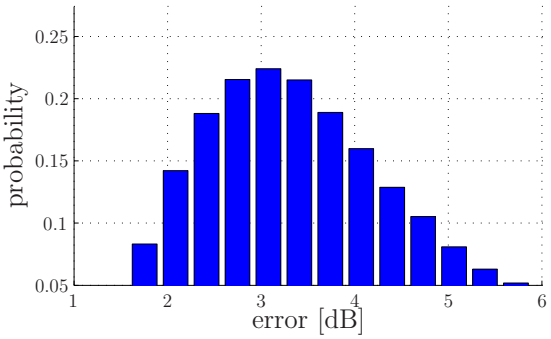
The bottom row of Figure 5.7 displays the mean error for each position when ten components are used for synthesis. The values were averaged across subjects and ears. Clearly, the error distribution follows a precise pattern. Both linear and logarithmic version indicate a higher error for the lateral region between -20° and 20° elevation.



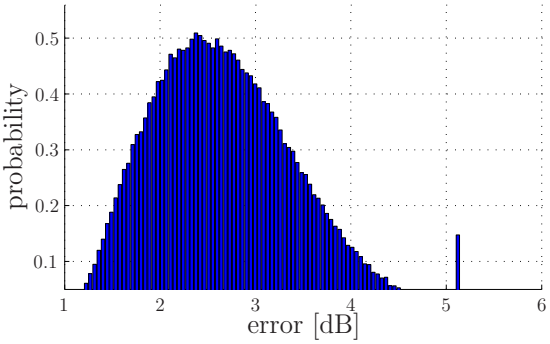
(a) SD using linear magnitude.



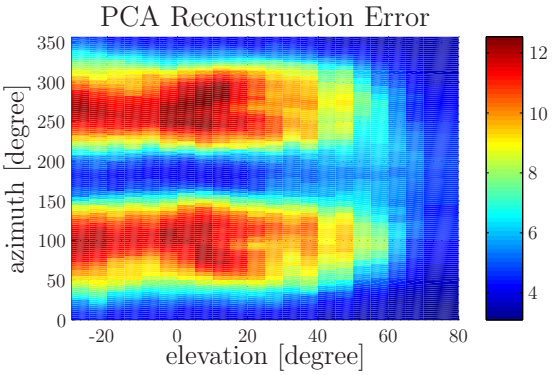
(b) SD using logarithmic magnitude.



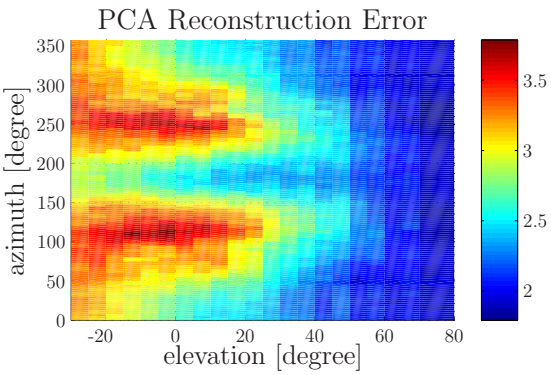
(c) SD distribution when using 10 PCs and linear magnitude.



(d) SD distribution when using 10 PCs and logarithmic magnitude



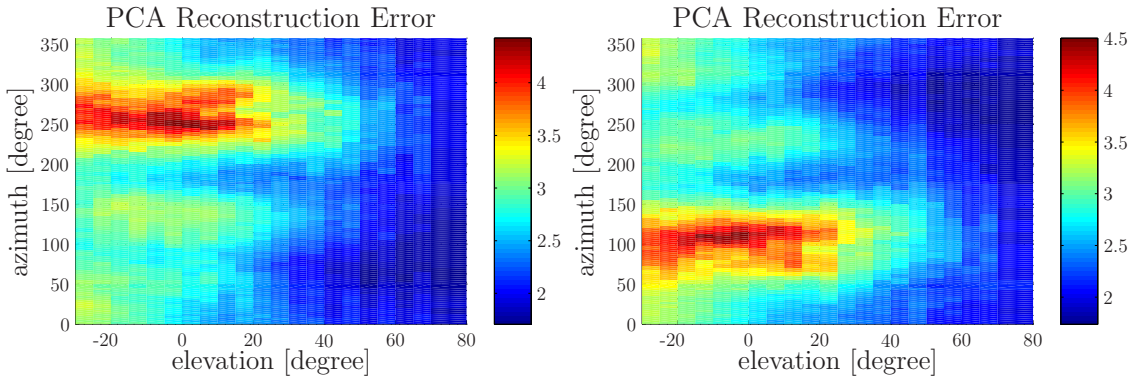
(e) Average SD across subjects and ears for each position when using 10 PCs and linear magnitude.



(f) Average SD across subjects and ears for each position when using 10 PCs and logarithmic magnitude.

Figure 5.7: Spectral distortion (SD) between PCA input matrix and reconstructed one when using linear (left column) and logarithmic (right column) frequency magnitude in ARI database.

Using a logarithmic amplitude again provides an advantage in reducing the spread error across the different directions. Closer analysis revealed, that the higher the number of components, the narrower is the shape of this pattern.



(a) Left ear.

(b) Right ear.

Figure 5.8: Average SD across subjects for each position and ear when using 10 PCs and logarithmic magnitude in ARI database.

Although the magnitude of the ears were simultaneously processed, the error can be plotted separately. Figure 5.8 shows the SD for left and right ear, averaged across subjects. Logarithmic magnitude for PCA and 10 PCs were used for reconstruction. For sound directions ipsilateral to the ear, SD is low whereas for sound directions contralateral to the ear SD is much higher up to four decibel. In general, due head shadow effect, the signal-to-noise ratio of the contralateral ear in the PCA input data is lower, therefore also modeling through PCA with a limited number of PCs leads to a higher reconstruction error. It has been shown, that the more components are used, the smaller is the difference between contra- and ipsilateral errors. However, for source elevations higher than 40 degrees, SD for the contralateral directions significantly improves. In practice, for lateral directions of more than 40 degrees, the contribution of the contralateral ear is far less important than that of the ipsilateral one [HB88]. Consequently, there should be only a small influence on localization performance.

Figure 5.9 indicates the error distribution over subjects and positions for each frequency bin in ARI and CIPIC database when 10 PCs are used for reconstruction. The median value is indicated by the white circle with a black point, the edges of the blue boxes are the 25th and 75th percentiles and the white whiskers extend to the most extreme data points. Outliers are plotted separately as blue points. An inspection of the figure reveals an increasing average error in higher frequencies. At first glance, it looks as if many outliers were included. However, one must bear in mind that for each frequency bin, more than 128 thousand (1550 positions $\times$ 83 subjects) points for ARI and 56 thousand (1250 positions $\times$ 45 subjects) points for CIPIC are represented. The results are almost the same with IRCAM database. The

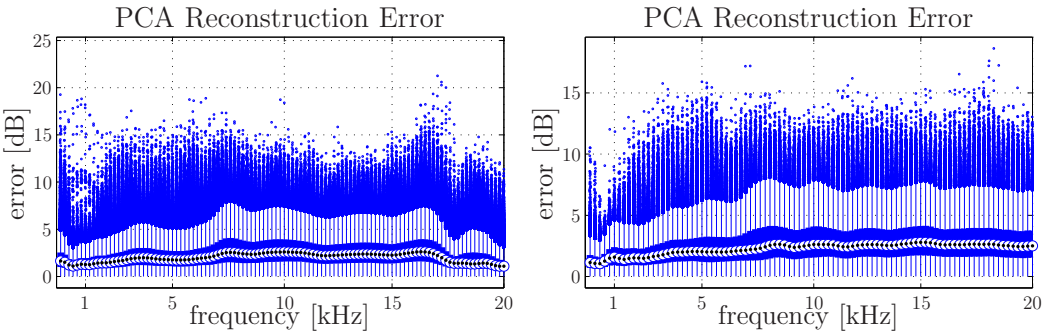
error increases for frequencies above 5 kHz, because the higher spectral variability in this range can not be modeled by the limited number of 10 principal components.

5.4.3.3 Results in Time Domain

The first row of Figure 5.10 shows the signal-to-distortion ratio. The average SDR is about 8 dB when using 10 PCs. Compared to Xie [Xie12], who used 35 basis functions and reached an average SDR across all positions of 21 dB, the SDR in this model is decreased to about 11 dB. However, as already mentioned, Xie used structure Struct5 for the PCA input matrix, this might also enhance the SDR. For example, Keyrouz and Diepold [KD08] proposed an HRTF interpolation method and indicated a SDR between 30 and 72 dB. In general, SNR of about 20 to 70 dB are specified in literature for acoustical measurements [WI03, NC10].

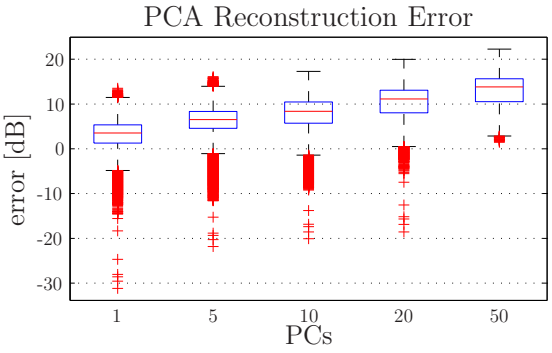
Unlike to the frequency domain, using linear data as input results in a small advantage in reconstruction error compared to using logarithmic magnitude. Similar findings were obtained when examining the envelope of the reconstructed and original minimum-phase HRIRs was compared.

The second row of Figure 5.10 shows the distribution of the error when using ten components for reconstruction, and similar to the frequency domain, a skewed distribution is obtained for linear and a normal one for logarithmic input. The bottom row of Figure 5.10 displays the SDR for each position when ten components are used for synthesis. The values were averaged across subjects and ears. As in the frequency domain, also in time domain a clear pattern for the error with respect to source positions is visible. Results are consistent with the frequency domain representations.

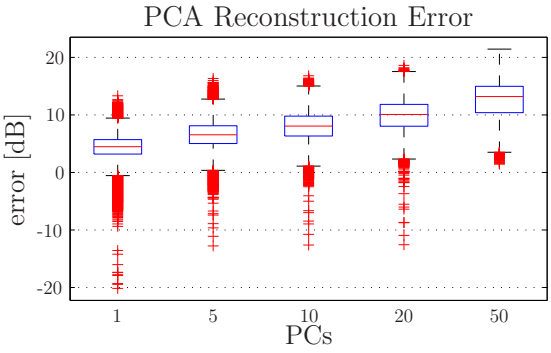


(a) ARI, logarithmic magnitude. (b) CIPIC, logarithmic magnitude.

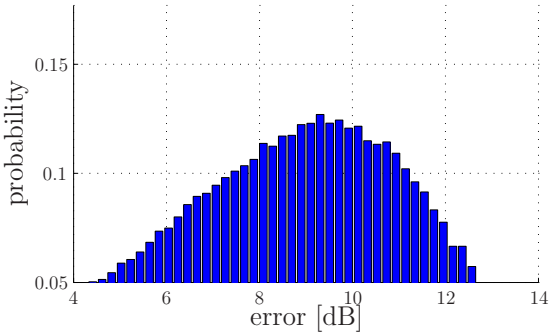
Figure 5.9: Distribution of spectral distortion (SD) over positions and subjects for each frequency bin when 10 PCs are used for reconstruction.



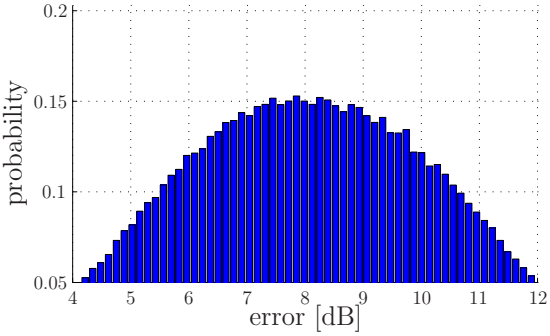
(a) SDR using linear magnitude.



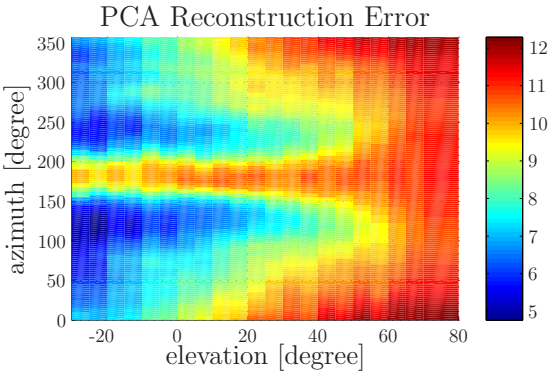
(b) SDR using logarithmic magnitude.



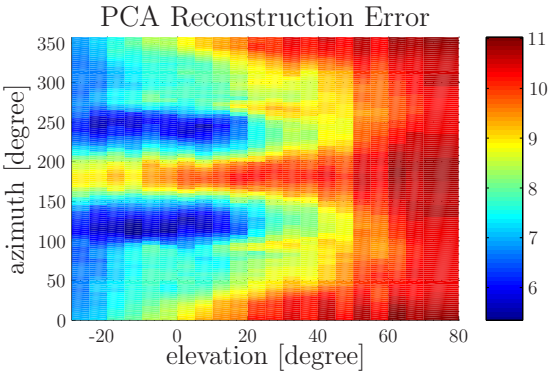
(c) SDR distribution when using 10 PCs and linear magnitude.



(d) SDR distribution when using 10 PCs and logarithmic magnitude.



(e) Average SDR for each position when using 10 PCs and linear magnitude.



(f) Average SDR for each position when using 10 PCs and logarithmic magnitude.

Figure 5.10: Signal-to-distortion ratio (SDR) between original and reconstructed minimum-phase HRIRs in ARI database.

5.5 Conclusion

In this chapter, different parameter sets that affect how the HRTF set can be transformed into a 2D matrix for PCA were described and evaluated through a numerical simulation. Based on the results of compression efficiency and also by considering

their suitability for HRTF individualization, the matrix structure Struct2 was chosen. In addition, it was decided to model ears independent variables in columns, as this yields a PCA model that allows the simultaneous adaptation of HRTFs for both ears, a very flexible approach for HRTF adjustment.

It was decided to construct a PCA model that operates on the logarithmic magnitude spectrum, because this yielded the maximum compression efficiency for a reasonable number of components in the case of the ARI database. Ten components are required for describing 90 percent variance of the dataset. Through smoothing of the frequency magnitude, the components could be reduced to seven. However, the results are not always consistent over HRTF databases. Therefore, for each database it must be considered separately whether linear or logarithmic signal representation leads to an adequate compression efficiency.

The reconstruction error in frequency and time domain was investigated for the selected structure. Using the proposed model configuration in Table 5.3, minimum seven components are required to yield 90 percent variance of the data, which leads to an average reconstruction error below five dB. In general, the error increases for higher frequencies since the spectral variability in this regions is increased.

In contrast to this, interpretation of the error in time domain is more complicated because some modifications of the PCA input matrix are not visible at the time domain error. Beside that, a large error in time domain is not directly linked to a degradation of localization performance [Rom12]. The difference between original and reconstructed HRIRs can not be used as an objective measure since the shape of the signal is completely altered by the minimum-phase approximation. Therefore, the reconstruction of the minimum-phase impulse responses was tested and indicate an average signal-to-distortion ratio of about eight decibel for reconstruction with 10 PCs.

Concluding this numerical evaluation, for the individualization process, the numerical results indicate that the modification for the first seven components is adequate since they describe about 90 percent variance of the dataset and yield an average reconstruction error of about 5 decibel.

The selected structure provides the possibility for adjustment of a single position by adjusting PCW for a PCA model using 7 components. Such a process can, however, be tedious as one needs to perform the individualization process separately for each sound direction of interest. The next chapter focuses on how the model could be extended to allow for a global HRTF individualization process. Finally, Table 5.3 summarizes the parameters of the PCA model.

Model Parameter	**Value**
Database	Acoustics Research Institute (ARI)
Dataset	All subjects and source positions
Matrix Structure	**Struct2** $[(subjects * positions) \times signal]$
Signal Representation	DTF with logarithmic magnitude
Ears	Both
Ear Handling	Ears added as independent variables in columns
Frequency Smoothing	Reducing the Fourier coefficients to a quarter (resulting in 32 coefficients)

Table 5.3: Input matrix properties of the PCA model.

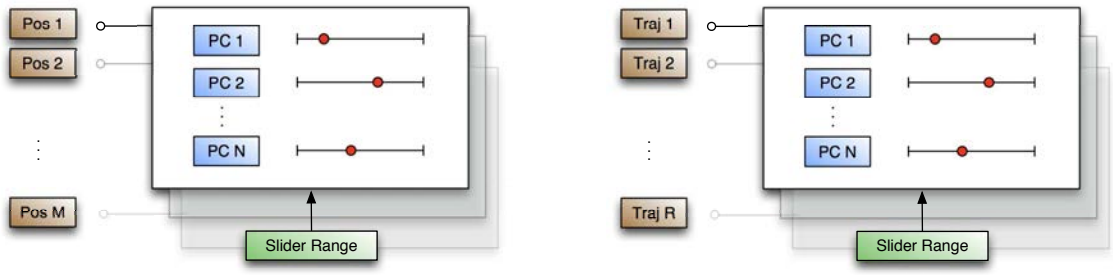
Chapter 6

Global Model of HRTF Individualization

Up to this point, existent HRTF models and their technical principles were discussed and a PCA analysis model was selected based on the results of a numerical evaluation. As mentioned earlier, these classes of models are most suitable for local HRTF individualization, in the sense that each position needs to be adjusted individually. This is a consequence of the fact that the algorithm returns a single PCW for each position and subject in the dataset. Such a practice has been used in the literature in the works of [HPP08, Shi08, HPP10, FR12] and will be referred as local adjustment in the following. After reviewing briefly the rationale of this methodology, this chapter attempts to investigate how such a model could be extended to allow for a global HRTF individualization, in which multiple sound directions are adapted simultaneously. To this end, a new HRTF individualization method is proposed which operates on a spherical harmonics model of the principal component weights. The model is evaluated numerically in order to understand how the order of the spherical harmonics representation affects the reconstruction accuracy of the HRTFs. The chapter concludes with a discussion on the feasibility of the technique and the possible use of regularization techniques.

6.1 Local Adjustment

In previous work [Hol12], a simple graphical interface for adapting PC directly for each source position or trajectory of interest was presented. Usually a trajectory includes a small subset of positions arranged parallel to the horizontal or median plane. Figure 6.1 depicts the method. The individualization process assumes that by adjusting the weights of the principal components, one can eventually find a combination that



(a) Each position.

(b) Each trajectory.

Figure 6.1: Local adjustment of source positions or trajectories.

results in the desired auditory impression. To this end, a simple model of the PCW distributions is used that employs the range of the weight distribution as estimated by the PCA. Consequently, the PCWs can be modified by

$$w_{pa} = \overline{w}_p \pm \sigma * v_a \ , \tag{6.1}$$

with

- w_{pa} as the adapted PCWs for one position,
- $\overline{w}_p$ as the mean value of w_p for a particular position over subjects,
- σ as the standard deviation of w_p over subjects and
- v_a as an *adaption vector* that essentially defines the allowable adaptation limits (typically between -3 and 3).

Note that using $v = 0$ leads to a generalized HRIR sample with average weight values. The unbiased sample standard deviation is given by

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{s=1}^{N}(w_p - \overline{w}_p)^2} \ , \tag{6.2}$$

with w_p as the sample value, $\overline{w}_p$ as the mean value and N as the total number of subjects. In order to enhance customization of particular components, the adaptation limits could be extended (for example $v = 10$).

Assuming a normal distribution of principal component weights, the range of $\overline{w_p} \ \pm \ 3\sigma$ contains 99.73 percent of all subjects weights. Hwang and Park [HPP08, HPP10] also used this range in their approach. Typically, the weights of the subjects should be very close to a normal distribution, but this is not always the case and

also depends on the signal representation used for the PCA input matrix. Without loss of generality, however, one could alternatively base the adaptation process on particular percentiles of the distribution (e.g. 1st, 12.5th 25th, 50th, 75th, 87.5th, 99th), where 25th and 75th percentiles are the first and third quartiles respectively, and the 50th percentile corresponds to the median value. The method essentially interpolates between the existing HRTFs in the dataset for a given position.

6.2 Global Adjustment

Although the approach of local adjustment is very flexible and powerful, an adaptation of each of the weights for every position of interest is time consuming. One way to overcome this problem is to provide a model of principal component weights. An important advantage of the selected structure for the model is that PCW are obtained for each subject and sound direction in the dataset. It is therefore straightforward to seek a spatial model of the principal component weights. Given the location of the sound directions on sphere, it would seem appropriate to use the spherical harmonic transform to create such a model. This transform has been used successfully to model and synthesize HRTFs [EAT98, EA98, ZDG09, AR10, Rom12]. The focus here is, however, not the same, rather it is on applying the transform to provide a model of the PCWs on the sphere. Thus, the spherical domain only serves to simplify the adaption of the PC weights.

6.2.1 Formulation

Assuming an HRTF dataset $\mathbf{D}_H$, comprising N_S subjects with HRTFs of signal length L, measured at N_D directions. The PCA input matrix $\mathbf{H}$ $[N_H \times M_H]$ is constructed according to Struct2 definition (Chapter 5.2.3, Page 40) as $N_H = N_D N_S$ and $M_H = 2L$.

In order to allow proper PCA processing, first, the column mean of $\mathbf{H}$ is subtracted resulting in a new input matrix $\mathbf{G}$ $[N_H \times M_H]$ and a matrix $\mathbf{M}$ $[1 \times M_H]$ including the mean data. PCA is applied on $\mathbf{G}$, resulting in a matrix $\mathbf{V}$ $[M_H \times M_H]$ containing the eigenvectors of the covariance matrix of $\mathbf{G}$, also called the principal components, and a corresponding principal component weight matrix $\mathbf{W}$ $[N_H \times M_H]$. The decomposition fulfills $\mathbf{W} = \mathbf{G}\ \mathbf{V}$.

Matrix $\mathbf{Y}_N$ $[N_D \times (N+1)^2]$ includes the spherical harmonics of order N, sampled at N_D positions on a sphere. The two-dimensional principal weight matrix $\mathbf{W}$ is reshaped to obtain three dimensions according to subjects N_S, positions N_D and

principal components M_H. The weights for each subject is presented by the matrix $\mathbf{W}_S$ $[N_D \times M_H]$. As described in Chapter 3.2, the spherical harmonic coefficients $\mathbf{\Psi}_{N,S}$ for each subject are given by

$$\mathbf{\Psi}_{N,S} = \mathbf{Y}_N^{\dagger}\ \mathbf{W}_S\ , \tag{6.3}$$

with $\mathbf{Y}_N^{\dagger}$ $[(N+1)^2 \times N_D]$ as the pseudo-inverse of $\mathbf{Y}_N$. Similar as for the principal component weights, the idea for adjusting the spherical harmonic weights is based on the assumption of a normal distribution of the individual weights for each basis function. Therefore, the range for the individualization process can be set to the mean value $\pm\ 3\sigma$ across all subjects.

Going back to the PCA domain, the principal weights for each subject that are modeled through spherical harmonics are reconstructed by

$$\hat{\mathbf{W}}_S = \mathbf{Y}_N\ \mathbf{\Psi}_{N,S}\ . \tag{6.4}$$

Afterwards, the individual principal weights are collected and formed again to the two-dimensional weight matrix $\hat{\mathbf{W}}$. Finally, the desired HRTF dataset $\hat{\mathbf{H}}$ is reconstructed from the PCA decomposition and the subtracted mean matrix $\mathbf{M}$ is added again by

$$\hat{\mathbf{H}} = \hat{\mathbf{W}}\ \mathbf{V}^T + \mathbf{M}\ . \tag{6.5}$$

The whole process is finished by reshaping $\hat{\mathbf{H}}$ to the original four-dimensional structure in order to evaluate the differences between original and reconstructed dataset. The accuracy of the model is based on the limitation of the PCs as well as the SH order.

6.2.2 Matrix Regularization

As already mentioned, the condition number of a matrix describes the ratio between smallest and largest singular value and is an indicator for the accuracy of its inverse. In general, it measures the sensitivity of a function in regards to how much error results in the output for only small changes in the input. A condition number of close to one means a matrix is well-conditioned and its inverse can be calculated with adequate accuracy. In contrast, a high value is referred to numerical errors computing its inverse and if the condition number is infinite, the matrix is singular and no inverse can be computed. For the numerical evaluation of this global HRTF model, the condition number of the spherical harmonic matrix, which contains the

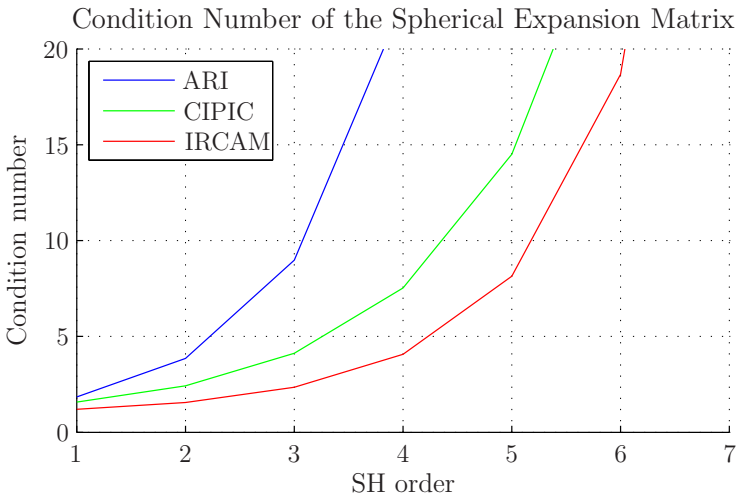


Figure 6.2: Condition number of the spherical harmonic expansion matrix in regards to the spherical harmonic orders in three different HRTF databases.

spherical harmonic expansion, was inspected for each SH order. Figure 6.2 indicates the rising condition number in regards to the SH orders for three databases with different sampling grids (cf. Figure 5.2, Page 38). For a better view, only 1st to 7th order are plotted, but the calculation was done until 20th order.

From the 3rd order, the number increases sharply in ARI whereas CIPIC has values above 7 until 4th order. In contrast to ARI database with -30°, the elevation values in CIPIC are sampled down to -45°, which results in a more stable inverse. Similarly, IRCAM database covers almost the same range of directions as CIPIC, but the number of 187 sound directions is significantly lower than in other databases with more than 1000 directions. Apparently, the arrangement of the sampling grid has a larger impact on the condition number compared to the total number of the positions.

The condition number increases rapidly for all databases after a certain SH order. The start of the rapid increase agrees roughly with the relationship N = 360/s with N as the SH order and s as the expansion in degrees of the open grid in the lower hemisphere of each HRTF database. In general, such an increase of the condition number is not avoidable due to the lack of measurements below the head. So to keep the condition number as low as possible, it must be considered whether a higher SH order for the modeling of PCWs is necessary at all. To this, an investigation in the reconstruction error of the PCWs and the resulting HRTF dataset must be carried out, which is done in the next section.

A better choice of the sound directions could decrease the condition number, but this was not further investigated here. Assuming an uniformly distributed density

grid of sampled sound directions, best results can be achieved when the number of sound directions equals the number of spherical basis functions. For other cases, an under- or overdetermined system of equations has to be solved. According to Zaar [Zaa11], a good sampling scheme for HRTF databases would be a high density in lateral direction and a lower density in polar plane direction. Actually, this is the case in ARI and IRCAM database, but in CIPIC the number of azimuthal sampled positions does not vary in respect to the vertical plane. In addition, the prominent spatial gap below the head in HRTF databases should be replaced with mirrored positions from above. This could avoid regularization problems and enhance the performance of spherical harmonic decomposition. However, no such replacement was carried out here.

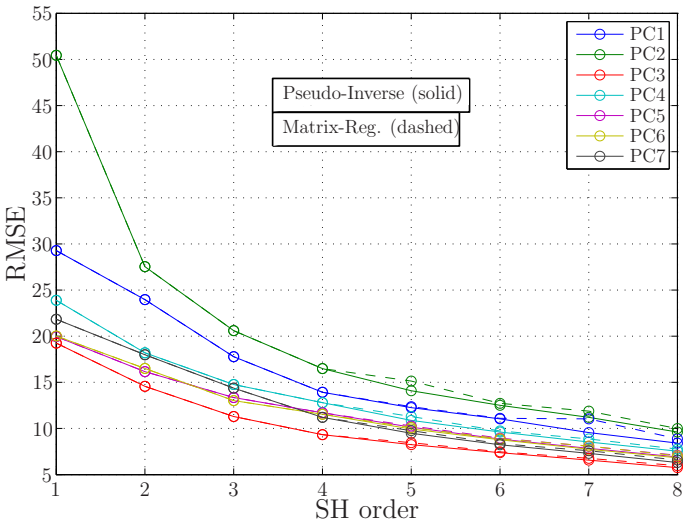


Figure 6.3: RMS error between original PCWs of a particular PC and PCWs when they are modeled with a different SH order in the IRCAM database. The inverse spherical harmonic expansion matrix is computed through the pseudo-inverse (solid lines) or through matrix regularization (dashed lines).

Compact Singular Value Decomposition (SVD) is commonly used for inverse problems. Only eigenvalues above a certain amplitude are involved for further processing. Chapter 3.2.1 (Page 19) deals with this topic briefly. In the weight model, the regularization process was implemented in such way that the condition number does not exceed the value of 5. This was done by neglecting all smallest eigenvalues that do not fulfill this requirement and calculating the inverse matrix by

$$\mathbf{Y}_N^{\dagger} = \tilde{\mathbf{V}}\tilde{\mathbf{S}}^{-1}\tilde{\mathbf{U}}^T \ . \tag{6.6}$$

Figure 6.3 indicates the RMS error between original PCWs and their modeled versions with and without regularization for different SH orders and for a single PC in the IRCAM database. The RMS error across positions was averaged over subjects. The solid lines indicate the usage of the pseudo-inverse whereas the dashed lines indicate the usage of the matrix regularization. PC2 has the highest error across all SH orders. The limited difference between the two types of matrix inversion can be explained that for points already existing in the dataset, regularization does not alter much. However, if interpolation is used in the spherical harmonic domain to estimate the weights of a missing position, then it is quite likely that without regularization problems appear (e.g. huge overshooting) when estimating the true weight of the missing position.

6.2.3 Numerical Evaluation of the Global Model

A numerical analysis is carried out, to evaluate the global HRTF model. Figure 6.4 depicts the main steps for processing. The principle of the calculation is similar to the PCA model, but the simulation parameters were fixed to the values in Table 5.3. After PCA, SH decomposition is applied to the principal weights and they are reconstructed with a limited SH order. Consequently, the error of the PCWs and the error of the reconstructed HRTF dataset is inspected and leads to the decision which SH order is necessary for an adequate reconstruction.

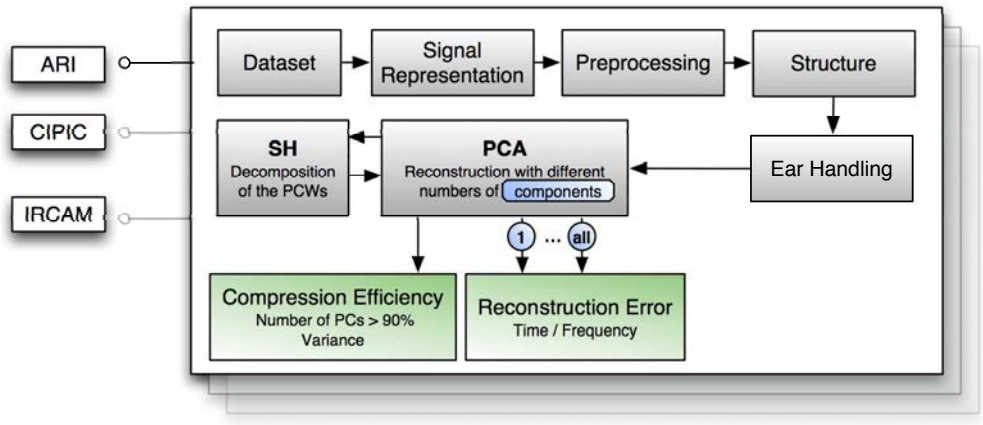


Figure 6.4: Numerical evaluation of PCA input parameters and SH order.

6.2.3.1 Reconstruction of the PCWs

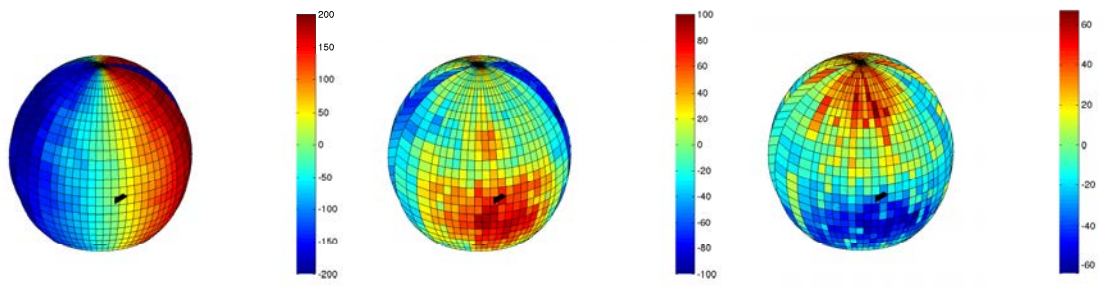
Figure 6.5 indicates original and modeled PCWs of PC1, PC2 and PC4 of a subject ID3 in CIPIC database for different SH orders. Obviously, at first sight, the usage of the spherical functions smooths the PCW distribution. Using the first order, the variation of the PCWs is oversimplified, but the main localization cues, that is left/right (PCW1), front/back (PCW2) and up/down (PCW3), can be modeled. From the 2nd order, the weights are modeled more or less sufficiently. Results from the other two databases show that an SH order of 2 or 3 is adequate to preserve the main distribution of the PCWs.

6.2.3.2 Dataset Reconstruction Error

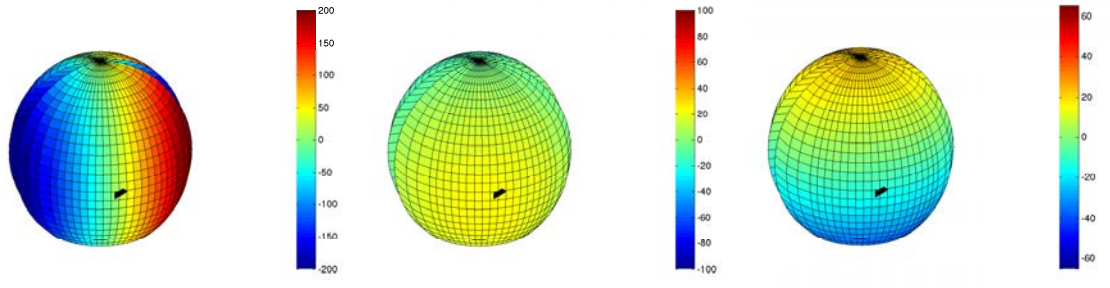
The first two rows of Figure 6.6 indicate the spectral distortion for different SH orders. As a reference, also the error without the SH model is given. The error measure was calculated by comparing the original preprocessed HRTF dataset to the reconstructed one in which a different number of PCs and SH orders have been used. For each of the PC subspace dimensionalities, a box plot based on the reconstruction error of all subjects, positions and ears is presented.

Naturally, the reconstruction error depends on both the number of principal components used in the model as well as the spherical harmonics order of the PCW model. Starting with only the first SH order, the error is on average 6 dB with 1 PC. This is about the same as the reference error. Probably the weights of the first PC can be modeled sufficiently by a 1st order SH model. For higher order PCs, the error drops to under 5 dB and remains on this value. This is in contrast to the reference, where an increase of the PCs significantly reduces the error. Similarly, by using SH order of 2 and 5, the error reduces to about 4 dB for 10 PCs. This is likely because the weights of higher PCs need a higher order SH model.

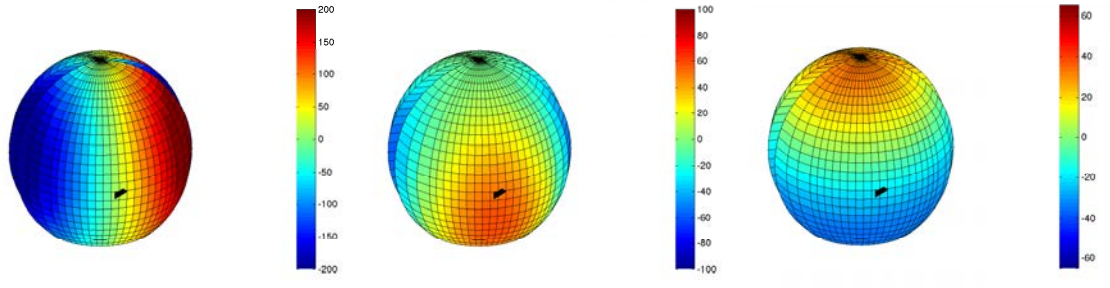
The bottom line of Figure 6.6 indicates the average spectral distortion across positions and subjects with respect to SH order of the PCW model and the number of PC used. Basically, when only a small SH orders are used, the error is only little dependent on the number of principal components. For example, when only the 1st SH order is used, it has no significant influence if 5 or more PCs are used. Similarly, when only 1 PC is used, the number of SH orders has no great influence on the error. This is related to a need for different SH orders depending on the PC number used, as weights become more complex for higher PCs.



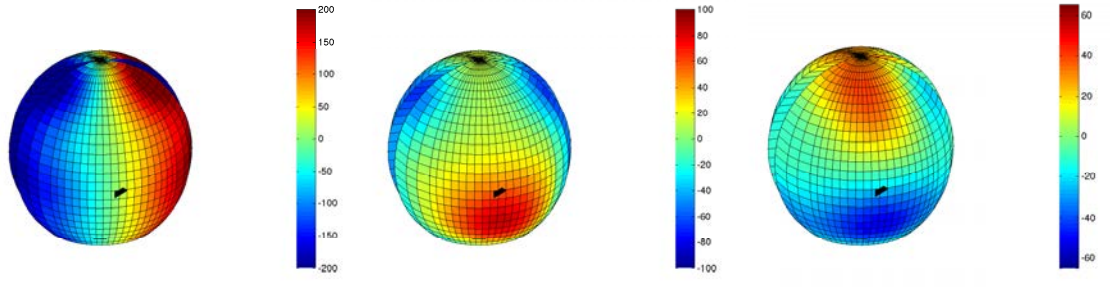
(a) PCW1, Original. (b) PCW2, Original. (c) PCW4, Original.



(d) PCW1, SH1. (e) PCW2, SH1. (f) PCW4, SH1.

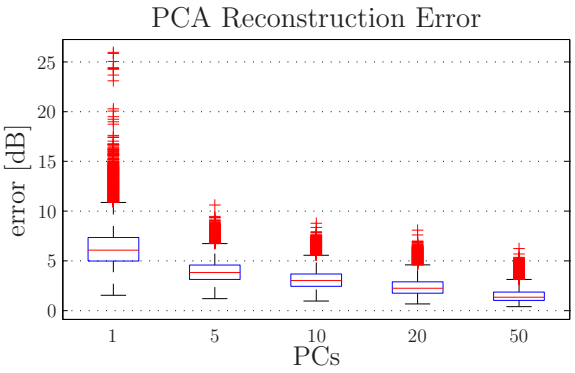


(g) PCW1, SH2. (h) PCW2, SH2. (i) PCW4, SH2.

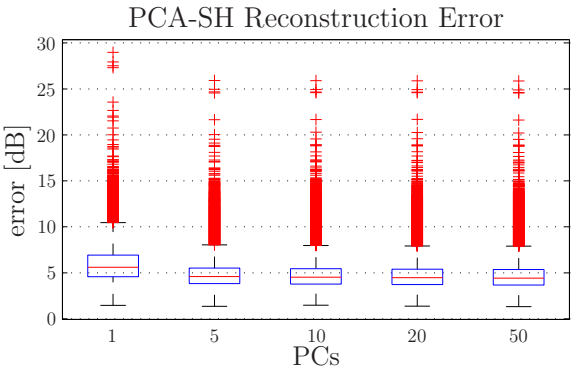


(j) PCW1, SH5. (k) PCW2, SH5. (l) PCW4, SH5.

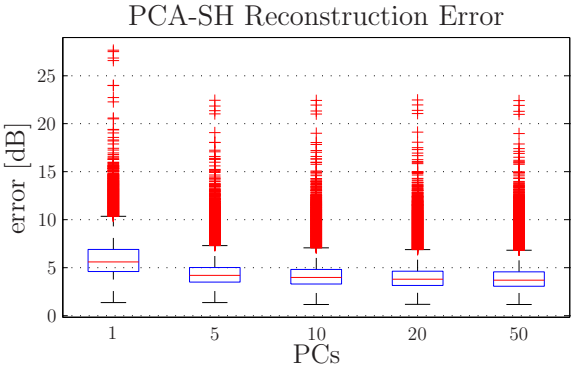
Figure 6.5: Original (first row) and reconstructed PCWs of the subject ID3 in CIPIC database when they are modeled through spherical harmonic functions with order $l = 1$ (second row), $l = 2$ (third row) and $l = 5$ (fourth row). The black arrow points at the frontal region.



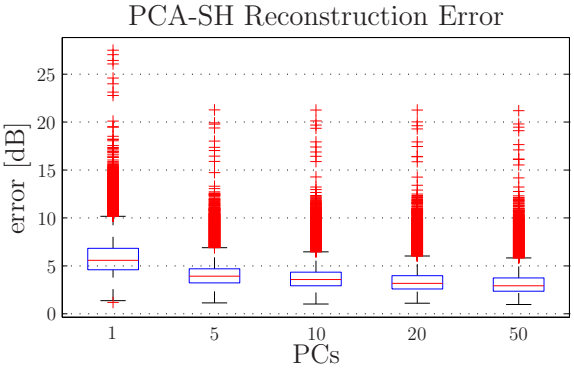
(a) Without SH Model



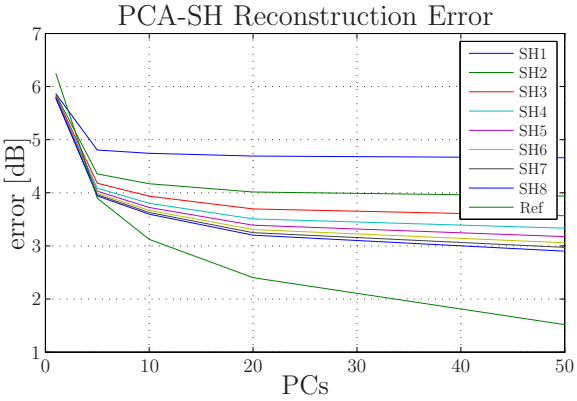
(b) SH Order 1.



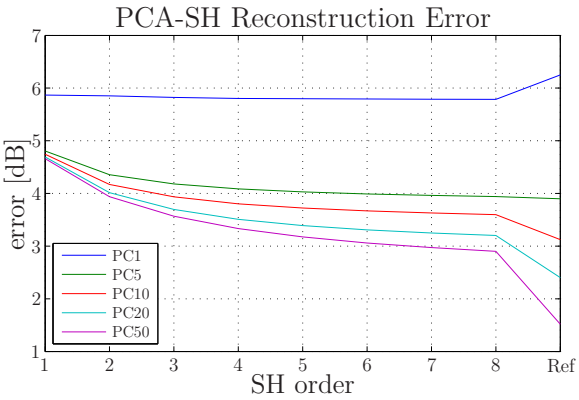
(c) SH Order 2.



(d) SH Order 5.



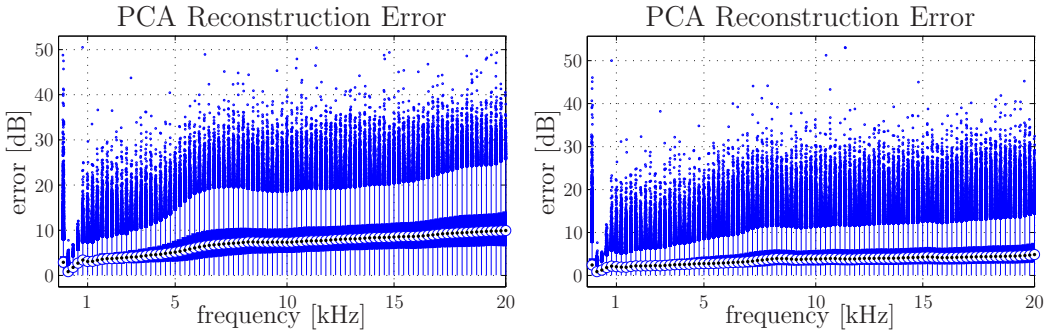
(e) Average error over positions and subjects.



(f) Average error over positions and subjects.

Figure 6.6: Spectral distortion (SD) between PCA input matrix and reconstructed one for different PC and SH orders in ARI database. First two rows: The median value is indicated by the central mark, the edges of the boxes are the 25th and 75th percentiles and the whiskers extend to the most extreme data points. Outliers are plotted separately as red markers. Third row: Mean SD over subjects and positions. The label *Ref* indicates the corresponding error of the PCA model.

Figure 6.7 indicates the variation of spectral distortion along the frequency axis. For each frequency bin, the distribution across subjects and positions is given. Using only the first SH order, frequencies up to 5 kHz are modeled with an average error of about 5 dB. For higher frequencies a greater number of PCs are required, therefore these regions also need higher SH order to model the higher variability of the PCWs.



(a) SH Order 1. (b) SH Order 5.

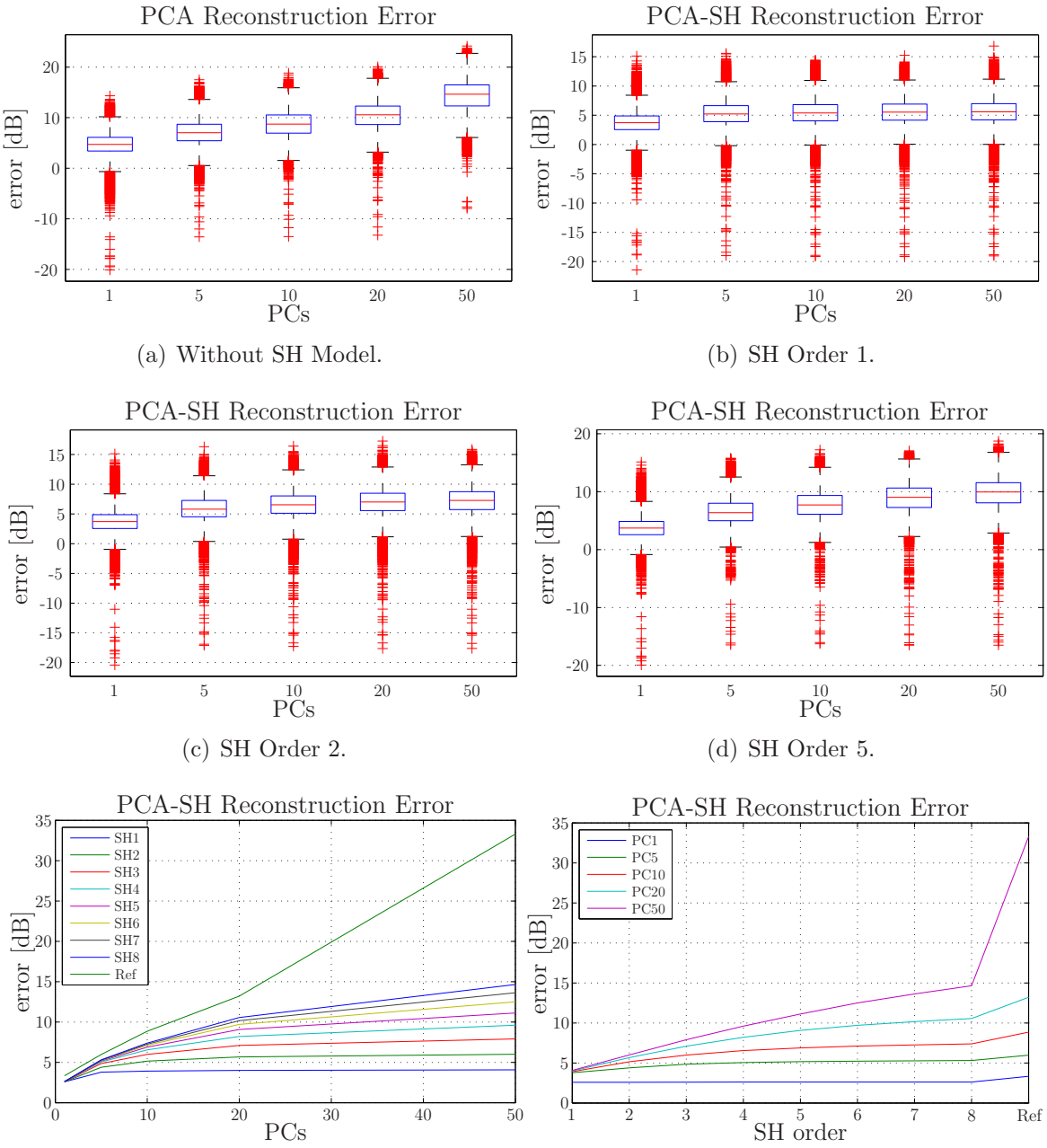
Figure 6.7: Distribution of spectral distortion (SD) over positions and subjects for each frequency bin when 10 PCs are used for reconstruction in CIPIC database. The median value is indicated by the white circle with a black point, the edges of the blue boxes are the 25th and 75th percentiles and the white whiskers extend to the most extreme data points. Outliers are plotted separately as blue points.

Figure 6.8 indicates the error in time domain. The first two rows indicate the SDR for different SH and PC orders. As a reference, also the SDR without the SH model is given. Using 10 PCs for reconstruction, the reference error is on average 3 dB. This can not be achieved with SH order 1 and 2, no matter how many components are used. Using the fourth SH order, the error is on average 4 dB for 10 PCs.

The bottom line of Figure 6.8 indicates the average SDR across positions and subjects and ears in reference to SH and PC order. In addition, the error of the PCA model is given as a reference. Also here it is evident that when modeling with only the first SH order, the choice of the PC number does not influence the resulting SDR.

6.2.4 Insight on the Operations of the Model

When operating on a spherical model of PCWs, one is essentially using PCA as a vehicle to reach a frequency independent representation of HRTFs. The dimensionality reduction achieved is depending on the number of positions to be adjusted. Instead of having to perform XP adjustments where X corresponds to the dimensionality of the PC subspace used and P to the number of positions of interest, one needs to adjust $(N_X + 1)^2 X$ adjustments, where N_X is the order of the spherical harmonics



(a) Without SH Model.

(b) SH Order 1.

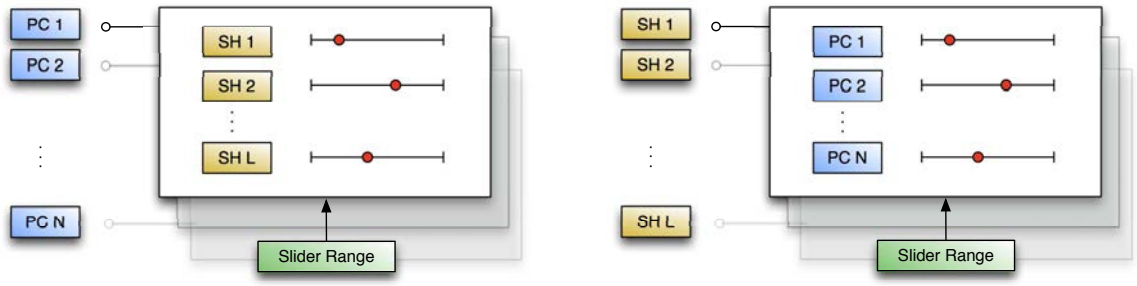
(c) SH Order 2.

(d) SH Order 5.

(e) Average error over positions and subjects.

(f) Average error over positions and subjects.

Figure 6.8: Signal-to-distortion ratio (SDR) between original and reconstructed minimum-phase HRIRs for different PC and SH orders in ARI database. First two rows: The median value is indicated by the central mark, the edges of the boxes are the 25th and 75th percentiles and the whiskers extend to the most extreme data points. Outliers are plotted separately as red markers. Third row: Mean SD over subjects and positions. The label *Ref* indicates the corresponding error of the PCA model.



(a) SHs for each PC.

(b) PCs for each SH.

Figure 6.9: Two different display options for customizing spherical weights using sliders.

weight model for PC X. As long as $(N_X + 1)^2 < P$, the total number of adjustments is smaller in the case of the spherical harmonics model. Thus, such a model could be useful in modifying large numbers of positions simultaneously, or even a complete HRTF set. Figure 6.9 illustrates two different options to arrange sliders in a GUI. Either all spherical harmonics are sorted by principal components or vice versa.

A second advantage of using spherical weights instead of directly adapting the principal weights is, that the balance of the weighting of principal components can be adjusted in different areas of the dataset, which leads to a faster and more effective adaptation procedure. The shape of the areas is determined by the spherical harmonic function whose weighting is adjusted at each instance. As can be seen from Figure 3.2 (Page 20), spherical harmonic functions for degree $m = 0$ (center column) are independent on lateral directions so they only represent vertical variation. In contrast, all basis functions which order equals the absolute value of degrees ($m = |l|$), only change in horizontal plane. For example, global scaling of the influence of each PC can be achieved by scaling the 0th spherical harmonic. First order spherical harmonics could be useful in adapting the weight of each component in front/back, left/right and up/down directions. By understanding the form of each spherical basis function, the weighting of each component in specific regions on the sphere can be increased or reduced.

However, the psychoacoustic impact of the adaptation is not yet understood, which means that a change that is applied for one position is not necessarily valid for other positions that are automatically adapted by the SH model. To this, it must be first considered which perceptual changes are caused by each principal component. An initial investigation into this with a listening experiment is given in the next chapter.

6.2.5 Conclusion

As a further development of the PCA model, a global model was presented. To overcome the position dependency of the PCWs, they are decomposed in spherical harmonic functions and corresponding weights, which can then be used to lead to adjustments over larger regions of the sphere. Depending on the number of positions being adjusted and the spherical harmonics order used for each PC, the proposed model could enable faster adjustment of the principal component weights, since a broader region is effected by the spherical harmonic weights. With only the first SH order, the localization cues for left/right, front/back and up/down can be modeled in general, but the individual variability of the PCWs can not be considered. Using the second SH order, the accuracy of the reproduction of the individual principal weights greatly improves.

Beside this advantage, the downside is calculation of the inverse of the spherical harmonic expansion matrix. Since in practice there are more or less spherical harmonic functions than sampled positions, this results in an under- or overdetermined system of equations. Consequently, a regularization method has to be applied to ensure a correct decomposition of the PCWs on the sphere. To this, the impact of the sampling grid on the reconstruction error has to be investigated further. As long as no interpolation needs to be done, the usage of the pseudo-inverse should be adequate.

Analysis of the Spectral Distortion and Signal-to-Distortion-Ratio, in comparison to a PCA model without SH modeling of PCWs, reveals that modeling the PCWs with an SH order of 2 or more leads to an average spectral distortion below 5 dB, which is promising result. This is an important finding because it greatly simplifies the complexity of the weight model. Consequently, assuming an SH order N, for each principal component $(N+1)^2$ spherical weights have to be adapted which is much less than adjusting PCWs for each position. For the sake of completeness it should be mentioned that also a single source position can be adjusted through this spherical model, however, in this case it might be more effective to adjust the corresponding principal weights directly.

Summing up, it has been shown that the model is promising but the process of adjustment has to be further inspected. Providing a thorough explanation of the way the model works perceptually, would require a series of experiments that were beyond the scope of this work. It is worth mentioning that the perceptual outcome of even simple adaptations of PCWs is not fully understood. For this reason and before proceeding further, a step inside this area is attempted by investigating how simple adaptations of the weights of a PC model affect the perceptual outcome. the

results of a listening test on four positions on the median plane is presented in the next chapter.

Chapter 7

Subjective Evaluation

In order to validate the PCA model described in Chapter 5 and to get a deeper insight into the process of HRTF individualization, a listening experiment was carried out which investigates the perceptual impact of the PCWs adaptation. This was done though an discrimination and localization test. Before presenting the methodology and results of the experiment, a short introduction is given about the variation of the PCWs in the database because this is the issue to be examined. Ideally, the results of the experiment and the findings of the variation of the PCWs in the database should coincide.

7.1 Variation of PCWs in the Database

Although the emerging PCs are mathematically orthogonal, they are not necessarily perceptually independent. Until further evidence is provided, one cannot assume that changes in the auditory percept as a result of changes in the weighting of the principal components reflect systematic changes within a spatial coordinate system. The emerging auditory percepts have been to a limited extend investigated analytically only for the first few components, there is, however, a lack of perceptual studies. In the first case, the variation of the PCWs across all angles can give further information on how particular principal components affect localization of the synthesized sound.

Hwang and Park [HP08] applied PCA on a dataset including 45 subjects and 49 directions in the median plane and interpreted the resulting weights of the 12 PCs. They found front/back and up/down cues in all of the first six PCWs whereas only inter-subject variations and no clear trend in respect to the median plane was found in the remaining 7-12 PCWs. In contrast, here all directions for data decomposition are included and a brief overview of the main findings for the first five PCs is presented.

Figure 7.1 shows the distributions separately for median and horizontal plane in the ARI database. Through this contrasting comparison, it should be easier to inspect the variation of the PCWs in reference to the two planes. The left column of Figure 7.1 indicates the first five left and right ear PCs that explain a total variance of 87.2% calculated using the input matrix Struct2 and ears in column blocked. The contribution of the remaining components is often too small to be noticed, or the resulting percept is too difficult to categorize. This happens when the inter-subject variability of the PCWs is greater than the variability with respect to the horizontal or median axis. In general, the higher the principal component number, the more difficult to define a direct relationship to spatial perception. Note that to provide a better illustration, the right ear basis function was plotted together with the left ear one, although the left and right ear basis functions are stringed together in the calculation. The middle and right columns of Figure 7.1 show the variation of the PCWs in the horizontal and median plane respectively.

The coordinate system that is used for the plots is described as follows. The azimuthal value is set to zero at the frontal side and increases counter-clockwise. Similarly, the elevation angle is set to zero at the frontal side, has its minimum at -90 degrees below the head and 90 degrees above the head. Finally it increases up to 270 degrees at the rear section below the head. Since the ARI database has only source positions until -30 degrees below the head, the elevation angle in the plots is limited from -30 to 210 degrees.

PC1. Clearly, the first component has no influence on elevation. PCW1 amplifies and reduces the corresponding component on the ipsilateral and contralateral side respectively, which basically controls the ILD. Due to the symmetry of the left and right ear basis function, one can well understand how a synthesized sound is affected differently for each ear through this component.

PC2. The weights have mainly positive values between azimuth of $\pm 45^\circ$ and negative values at other azimuths. In addition, along the median plane, positive weights are obtained for sounds in front (i.e. in $[-30^\circ, 90^\circ]$ elevation) and negative weights in $[90^\circ, 230^\circ]$ elevation). The fact that the left and right ear components are almost identical and their weights are symmetric with respect to the median plane supports the hypothesis that this component relates to cues supporting front/back discrimination and their variation in different azimuths and elevations.

PC3. The variation of the weights indicates a systematic variation with elevation. Roughly speaking, positive weights are obtained for sounds below and negative for sounds above the horizontal plane. In addition, the almost identical left and right ear basis functions and the relative invariance with azimuth indicate that this component has little impact on azimuth perception.

PC4. The weights change slightly in the median plane up to maximum value above the head. In the horizontal plane little systematic variation could be found. The basis function are not identical, although similar in shape, and the weights seem to affect the two ears in a slightly different weight. It is not clear what exactly is the influence of this component, although it appears to affect the energy balance between the ears as a function of azimuth and elevation.

PC5. In the horizontal plane, a slight variation of the weights is visible. Due the fact, that no variation in the median plane could be found, and the symmetric nature of the basis functions for the two ears, it appears as if this component might support PC1 for lateral discrimination, and its influence concerns more sounds in the very lateral directions i.e. in the areas between $[45°, 135°]$ and $[225°, 315°]$. The symmetric nature of the left and right ear principal component support the hypothesis of an influence to azimuthal perception through ILD manipulation.

More difficult is the description of the remaining components due to the large variability of the distribution. An additional post processing step after PCA, which rotates the PCs into an alternative linear coordinate system in such a way that interpretation might be easier, could help. For example, *Varimax Rotation* can be applied. Hence, a rotation matrix T to maximize the varimax criterion can be constructed to rotate the components and corresponding weights. It has to be noted, that the rotation does not change the explained variance proportion. For this analysis, such an additional rotation was applied on the PCA output with the MATLAB® function *rotatefactors()* but as at first inspection the interpretation was not simplified, therefore the original unrotated data was used here.

It is important to notice that the obtained PCs and PCWs are different when only a subset of certain positions such as the horizontal or median plane is used, since the PCA input matrix involves a different kind of dataset. Therefore, the prominent first principal component might not model left/right discrimination but up/down because no HRTFs with varying lateral information for left and right ears are included in the dataset. When a consistent model is used however, results are consistent as

calculations with the other databases IRCAM and CIPIC yielded reproducible results and confirm the fact that PCA is to a certain extent invariant with respect to the data source when the same type of measurements exist.

7.1.1 Research Questions shaping the Listening Test

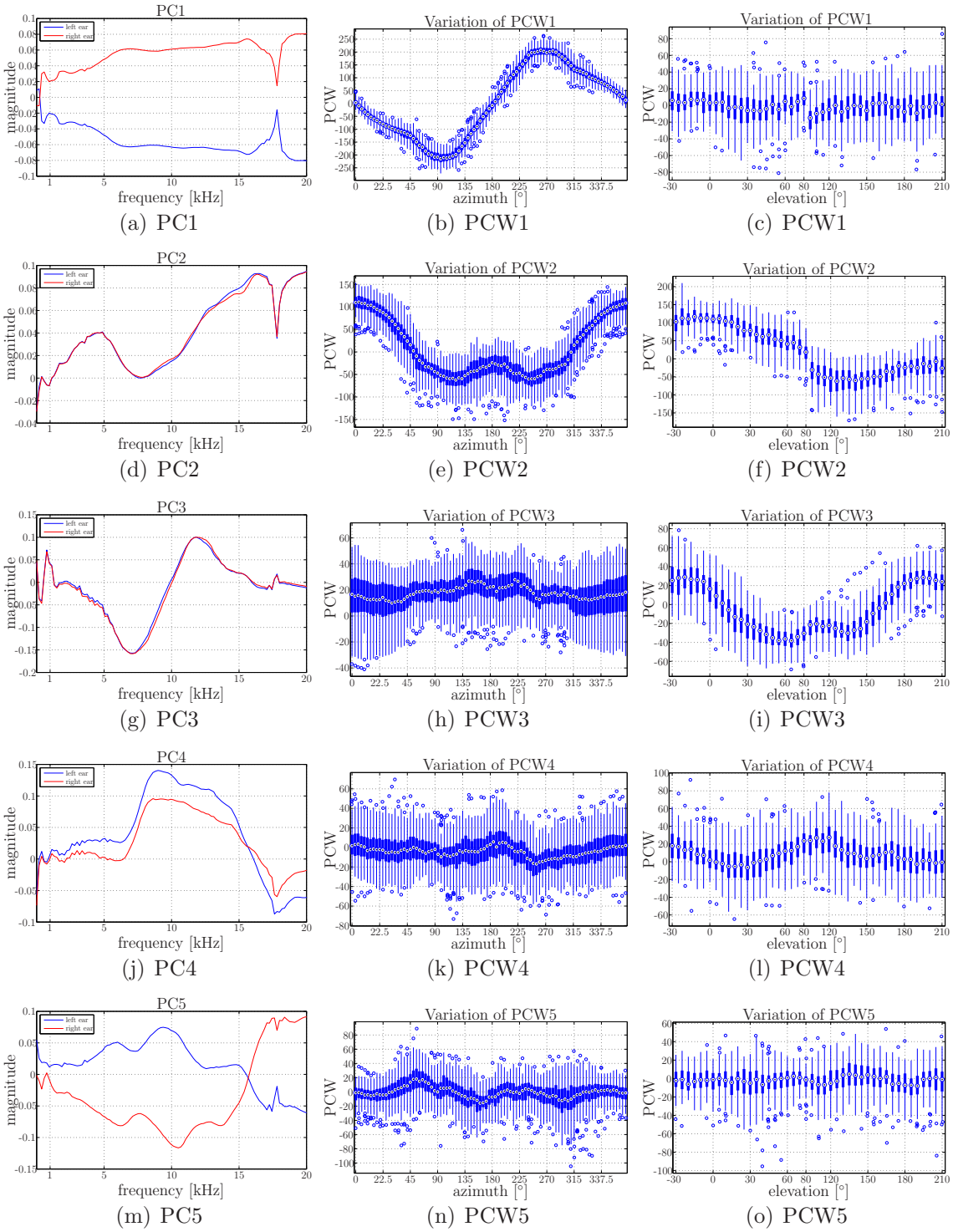
When considering HRTF individualization using PCWs, two questions arise from the analysis on the PCWs in the dataset, which are essential for the development individualization methods and are investigated in the experiment:

- To what extent are changes in the principal components weights audible, and what range of PCW adaptation is necessary to obtain consistent results across participants?
- What is the impact on localization induced by a change the principal weights for the first five components and is this consistent across participants?

Two related tasks were used in the experiment to answer these questions. The first task measured the sensitivity to changes in PCWs whereas the second one, measured the perceived position of the stimuli for similar changes to the PCWs. Based on the numerical evaluation and considerations in regards to HRTF adjustment described in previous chapters, the properties given in Table 5.3 (Page 58) were used for the PCA input matrix. PCA of the entire ARI database was computed, but only four positions in the median plane (elevation at -30°, 0°, 30°, 60°) were used in the experiment. Therefore no estimation of the interaural time difference was applied. Only the first five PCWs were adapted in the experiment, Table 7.1 lists the explained variance in the dataset of their components, which is in total 87 percent.

PC	**Relative Variance**	**Total Variance**
1	64.5 %	64.5 %
2	16.6 %	81.1 %
3	2.8 %	83.9 %
4	1.7 %	85.6 %
5	1.6 %	87.2 %

Table 7.1: Relative and total variance of the first five principal components in the ARI database.



(a) PC1 (b) PCW1 (c) PCW1

(d) PC2 (e) PCW2 (f) PCW2

(g) PC3 (h) PCW3 (i) PCW3

(j) PC4 (k) PCW4 (l) PCW4

(m) PC5 (n) PCW5 (o) PCW5

Figure 7.1: First five PCs (left column) and variation of the corresponding PCWs across all subjects in the **horizontal plane** (middle column) and **median plane** (right column) in ARI database. The median value is indicated by the central mark, the edges of the blue boxes are the 25th and 75th percentiles and the whiskers extend to the most extreme data points. Outliers are plotted separately as small circles.

7.2 Discrimination Test

7.2.1 Methodology

The goal of the discrimination test was to measure how much the principal component weights need to be adjusted to evoke a perceptible change in localization. Apart from providing calibration for the subsequent localization test, this experiment could also serve to inform local individualization algorithms. In each trial of the experiment, participants listened to a pair of sound samples: one was a reference sound, corresponding to one of the four tested locations, and the second an adapted version of it. The order with which the reference and test sounds were presented in each trial was randomized. The experiment included four reference positions in the median plane (elevation at -30°, 0°, 30°, 60°).

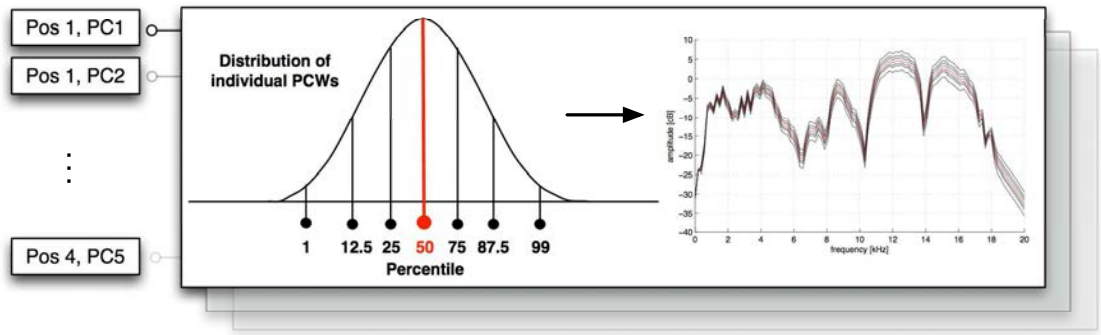


Figure 7.2: Theoretical Gaussian distribution of the individual principal component weights for each position and principal component. The adapted percentile values in percent for the discrimination test are marked in red.

The reference stimulus for each position and component was generated by adapting the PCWs of the first 5 PCs of a specific person in the database (ARI subject ID NH2) and setting them to the median value of the distribution of the corresponding PCWs for each PC and position. The weights for the remaining PCs (6-514) were not adapted to median values. The idea here was to start with principal weights of a measured subject, since the use of average weights lead to an unnatural sound. The test stimulus was generated by again adapting the weights of a single PC, out of the 5 principal components in the test, but this time in steps corresponding to seven different percentile values (1st, 12.5th, 25th, 50th, 75th, 87.5th, 99th) calculated from the distribution of PCWs for the particular reference position and PC. Figure 7.2 depicts the theoretical distribution of the individual weights and the choice of

the adaption values which results in a modified HRTF spectrum. The experiment included four positions, five principal components and seven degrees of adaption for each principal component weight yielding a total number of 140 sound pairs. As the 50th percentiles were used both to calculate the reference and one adaptation of the test stimulus, for each position and component a sound pair without any difference exists. Hence, a total of 20 catch trials were included in the experiment.

Although in another study the PCs to be adapted were determined by sorting them according their variance of the distribution of the corresponding weights [HPP08], here PCs were chosen by sorting them in terms of the total variance explained. To this, it must be mentioned that the standard deviation of the individual weights can not be a single indicator for the sorting process, without first establishing that adaptations of the weights of a particular PC lead to perceptible changes.

7.2.2 Procedure

Figure 7.3 shows the GUI in MATLAB®. Listeners were instructed to indicate whether they could detect a difference in the location of the two sounds in each trial pair by the answer *yes* or *no*. They were able to replay the stimuli as often as they wanted. The answer could be given either by clicking the buttons in the GUI or using the keyboard shortcuts which was much faster. The test stimuli was white noise amplitude with a duration 450 ms and a sampling rate of 44.1 kHz. 11 subjects participated in the task. In order to obtain more accurate results, the experiment was repeated once by each of the subjects. The test time for one repetition varied from 6 to 12 minutes, but two subjects required about 20 minutes. The subjects were instructed to set the playback level to a comfortable value, but more on the louder side. Diffuse-field equalized headphones (Sennheiser HD600 Avantgarde (8 subjects) and AKG K240 DF (3 subjects) were used and no headphone equalization was applied.

7.2.3 Results

Figure 7.4 depicts the percentage of trials, corresponding to a sampling of the associated psychometric function, in which participants detected a difference in the location of the stimuli, for each position, principal component and adaptation step. Depicted values were averaged across all test subjects. This function indicates how well a difference between the two stimuli was recognized with respect to different adaptation values.

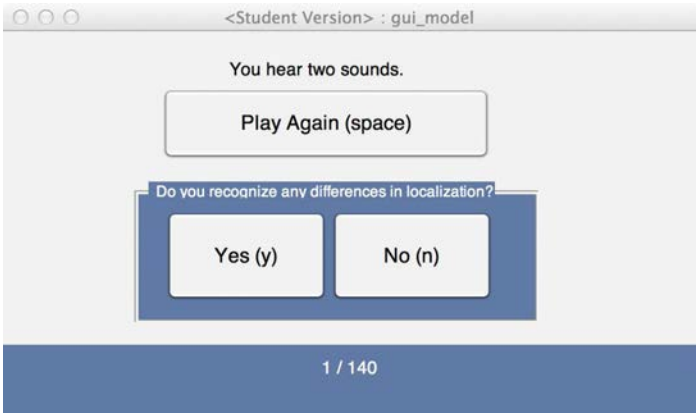


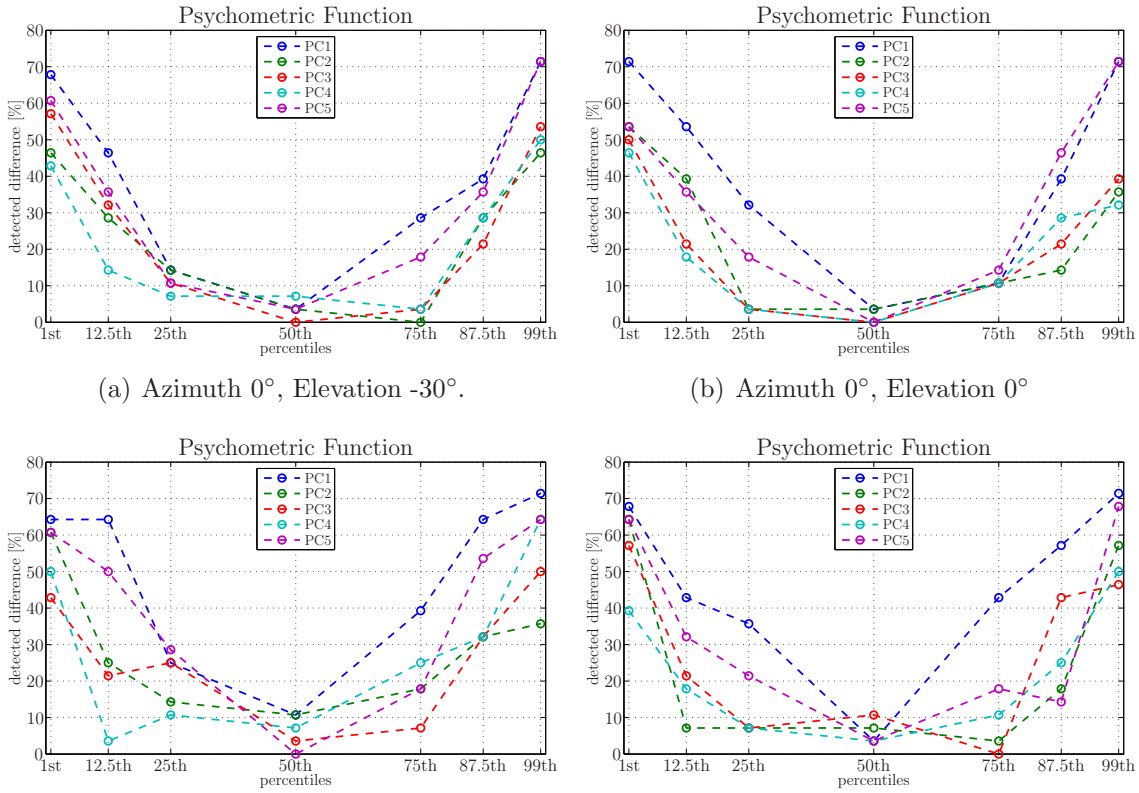
Figure 7.3: Graphical user interface for the discrimination task.

Obviously, in all four positions most differences are detected at PC1 and PC5. At the limits of adjustment (1th and 99th percentiles) about 70% of the stimuli are detected with PC1 and about 60% with PC5. Especially in the area between 25th and 75th percentile, the detection of differences with PC2, PC3 and PC4 is worse in contrast to PC1 and PC5. Especially striking is that PC5 has better results than PC2, PC3 and PC4, although they describe more variance in the dataset, as listed in Table 7.1. With few executions, there is a monotonic change with respect to the adaption strength. However, PC3 at the source position of 60° elevation decreases to 0% at 75th percentile. Also PC2 decreases to approximately 6% at 12.5th percentile.

A closer look at the median values reveals that in spite of the fact that the signal pairs at 50% percentile were the same, subjects answered up to 11% that a difference was recognized. This bias is different for each position and was at 8% for the lower position (-30°) and up to 11% for the two upper positions. However, only 4% bias was detected at the source position 0°.

In general, for PC1, modification of the weights at least to the 10th respectively 90th percentiles are necessary to report a change in sound location at 50% of the psychometric function. For the remaining components, a greater modification up to 1st and 99th percentile is required or even in some cases the 50% mark is not reached.

The psychometric function only shows the percentage of detected differences but it gives no information about the correctness of the answers. One can distinguish between a liberal response bias where the participant more likely responds *yes* regardless of the stimulus. In contrast, a more conservative response criterion biasses the data towards the response *no*. The presence of bias is predicted in Signal Detection Theory (SDT), therefore a sensitivity measure has to be calculated to exclude bias from data. Two major measures in SDT include the response bias, namely the



(a) Azimuth 0°, Elevation -30°.

(b) Azimuth 0°, Elevation 0°

(c) Azimuth 0°, Elevation 30°.

(d) Azimuth 0°, Elevation 60°.

Figure 7.4: Psychometric functions of the discrimination test for each position and principal component.

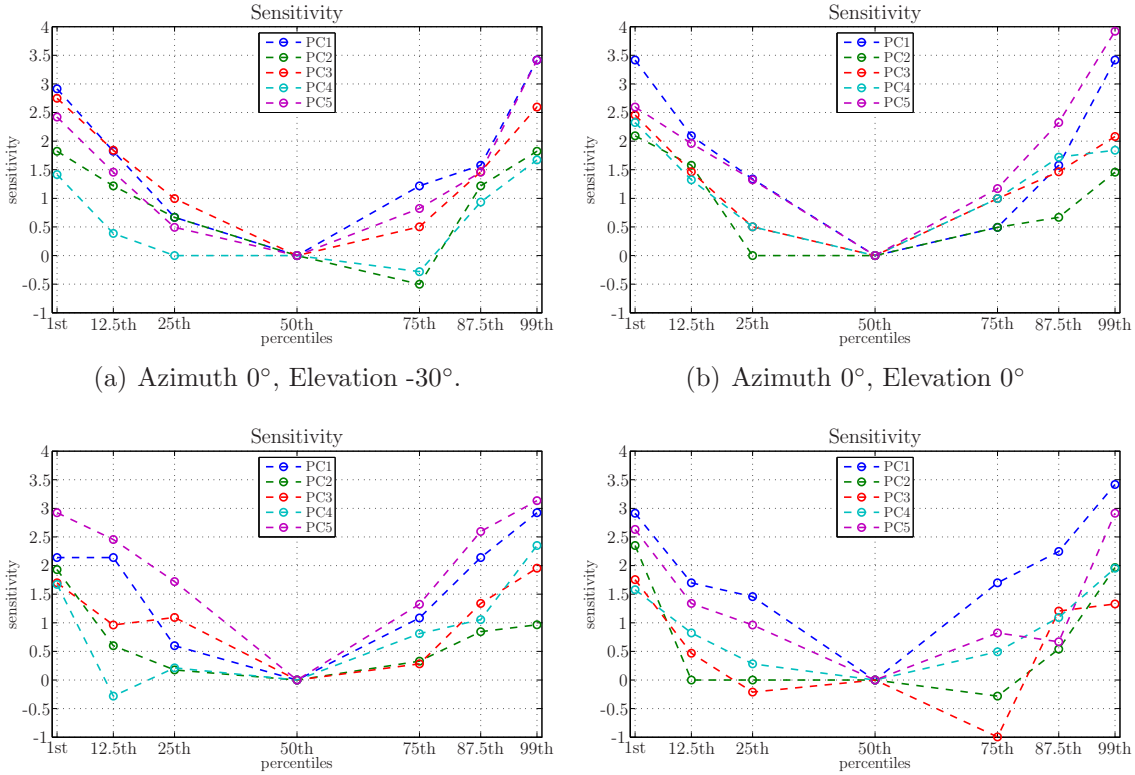
hit rate that considers adjustment of PCWs and subject detected it and false alarm that covers no adjustment of PCWs but the subject detected a difference.

According to this, the sensitivity index d' is an performance indicator for the difference between the signal and noise distribution and actually measures the proportion of the adapted sound pairs that are correctly identified as such. d' is measured in standard deviation units and can be estimated by the difference between the z-transform of signal and noise distribution,

$$d' = Z(hit\ rate) - Z(false\ alarm)\ , \tag{7.1}$$

with $Z(p)$ as the normal inverse cumulative distribution function of probability p. If the test subject is more sensitive to the adaptation, the difference between the two distributions is greater. However, a hit rate or false alarm of extreme values of 0 or 1 leads to a problem for the calculation of d', because the z-transform corresponds to negative or positive infinity. Therefore, for hit rate and false alarm, the minimum

and maximum values are set to be $1/(2N)$ and $(N - 0.5)/N$ respectively, with N as the total number of signal and noise trials [MK85].



(a) Azimuth 0°, Elevation -30°.

(b) Azimuth 0°, Elevation 0°

(c) Azimuth 0°, Elevation 30°.

(d) Azimuth 0°, Elevation 60°.

Figure 7.5: Resulting sensitivity measures of the discrimination test pooled over subjects for each position and principal component.

Figure 7.5 presents pooled sensitivity measures across subjects in reference to each position and PC. Similar as in the psychometric function, a clear trend in terms of adaptation steps is visible. The approximate values required to reach $d = 1$ depend heavily on the components but only little dependent on the positions. In general, PC1 and PC5 have the highest sensitivity in all positions. For 1st and 99th percentiles at the lower source elevations, a sensitivity value of 3 can be reached whereas for the two positions at a higher elevation the values are between 2 and 3 for the lower limit of modification. In contrast, PC2 has a lower sensitivity, most notably at source elevation -30° at 75th percentile, at source elevation 0° between 25th and 87.5th percentiles and in source elevation 60° between 12.5th and 75th percentiles. Also PC3 has a low sensitivity of -0.5 in this sound direction when the PCWs are adapted through the 75th percentile. In addition, PC4 has a negative sensitivity value of about -0.3 for source elevation 30° at 12.5th percentile.

According to Stanislaw and Todorov [ST99], also negative values for d' can occur through *"response confusion"*, which means that a participant responds *yes* when he actually wants to response *no* and the other way round. Obviously, this might happened between 12.5th and 87.5th percentiles or more likely when the subjects detected a signal by mistake.

Apart from that, for all positions the values are above -0.5 for each adaption step, expect PC3 at source position 60°, and increase at least up to 3 and sometimes 4 for the extreme adaptations. Apparently, the results are not always symmetric in respect to the 50th percentile. This is likely the case for higher order components. Good sensitivity values of more than 2 are reached in almost all positions by PC1, PC3 and PC5. PC2 could reach the value of 2 at the source elevations 0°, 30° and 60° for the 1st percentile. The values for PC4 were below 2 for almost all positions.

7.2.4 Statistical Analysis

Three-way *Analysis of Variance (ANOVA)* was applied which considered the three factors (position, component and adaptation) with different number of levels. The test investigated the within-subjects contrast. For the evaluation, a significance level of $p = 5\%$ was considered.

Table 7.2 summarizes the main results of the analysis of variance. The factors PC and adaptation were found to be significant whereas no significant effects could be found for the positions. This is an expected result and shows that the model works in the same way for different positions. Therefore, the null hypothesis which states that there is no difference among the levels of adaptions and components, can be rejected.

In addition, the interaction between PC and adaptation was found to be significant. This is obvious since each adaption of a particular PC produces different localization effects, hence the influence on the adaption levels is different for the PCs. In addition, also the interaction between position, PC and adaption is found to be significant. Considering that the interaction between PC and position is not significant, this can be explained that for *some* PCs the effect of adaption is also affected by the sound position.

Also pairwise comparisons of the different levels of the two factors that significantly influenced the results were analyzed. For the positions, no difference was found. Sensitivity for PC1 was significantly higher than all other PCs. Then the sensitivity in regards to PCs can be sorted by PC5, PC3, PC2 and PC4, the last three not being significantly different to each other. Analysis of the adaption levels reveals that the pairs 1st-99th, 12.5th-87.5th and 25th-75th have no significant difference within their

Source	Values
Position	F(3,27) = 0.436, p=0.729
PC	F(4,36) = 15.86, p=0.000
Adaptation	F(6,54) = 85.703, p=0.000
Position * PC	F(12,108)=0.926, p=0.524
Position *Adaptation	F(18,162) = 1,613, p=0.062
PC * Adaptation	F(24,216) = 2.571, p=0.000
Position * PC * Adaptation	F(72,648) = 2.152, p=0.000

Table 7.2: ANOVA output for tests of within-subjects effects.

members, but each pair is significantly different to the others, since these pairs might generate the same intensity in localization effect, but only in another direction. The 50th percentile is different to all another adaption levels.

7.2.5 Conclusion

Concluding the analysis of this discrimination task, modifications of the PCWs of the first five PCs causes a change in perceived location of sounds. This is an important result and crucial for the development of an HRTF adaption model based on PCA. Therefore, also the effect of higher order components up to the tenth component should be further investigated.

The greater the magnitude of the adaptation between the principal weights of the stimuli pairs, the more likely test subjects are to detect changes in localization. This is valid for all positions and generally for all components. The sensitivity for PC1 is greatest at source position 0°. The component with the second highest sensitivity for all positions is PC5, although PC2 to PC4 describe a higher variance in the dataset. Apparently, the modification of the weights of PC2-PC4 must be greater to cause a clear localization effect, or individual differences due to the individualized nature of HRTFs, reduce the perceptual impact of these components.

Analysis of variance revealed a significant effect of the factors PCs and adaptation. Since no effect could be found on the positions, this indicates a stable model implementation which is reproducible at each position that was tested. A closer look at pairwise comparison of the different levels of a factor revealed that PC1 has a significant difference to all other PCs, hence the sensitivity of PC1 was significantly higher than all other PCs. Moreover, the adaption levels do have a significant difference between themselves.

However, from this *yes/no* task, the exact localization effect can not be determined. Therefore, in the following, a localization test is carried out.

7.3 Localization Test

7.3.1 Methodology

In this test, subjects were asked to indicate the perceived sound directions in response to adaptions to the PCWs. As depicted in Figure 7.6, the modification of the weights up to 25th and 75th percentiles in the previous experiment were excluded, and instead the range of the distribution was enlarged. This was done by multiplying the absolute weight values of 1st and 99th percentiles with 1.5 (or divide by 1.5 if the value for the 1st percentile is positive) and adding them as additional adaption levels. The idea behind was that a trend of a localization effect may be easier to assess with extreme displacements, as sensitivity was relatively low for certain principal components in the previous experiment. For that reason, the adapted weights closest to the median value were neglected in order to keep a manageable number of test parameters and test time.

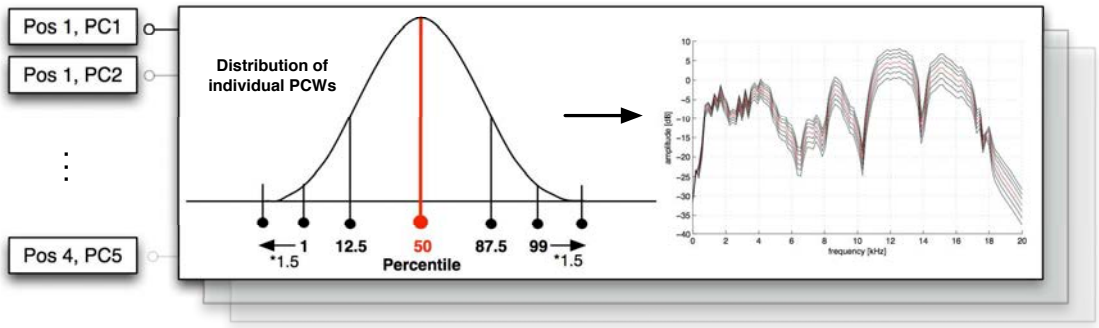


Figure 7.6: Theoretical Gaussian distribution of the individual principal component weights for each position and principal component. The adapted percentile values in percent for the localization test are marked in red.

7.3.2 Procedure

The 140 adapted samples, including the 20 catch trials (adapted with median value), were played. Listeners were instructed to specify the perceived sound location in a GUI in MATLAB® (Figure 7.7) and they were able to replay the stimuli as often as they wanted. Through sliders, the judgment in azimuth and elevation could be indicated.

The same white noise signal with a duration of 450 ms as well as headphones from the first test were used. 8 of 11 subjects from the previous test participated in this

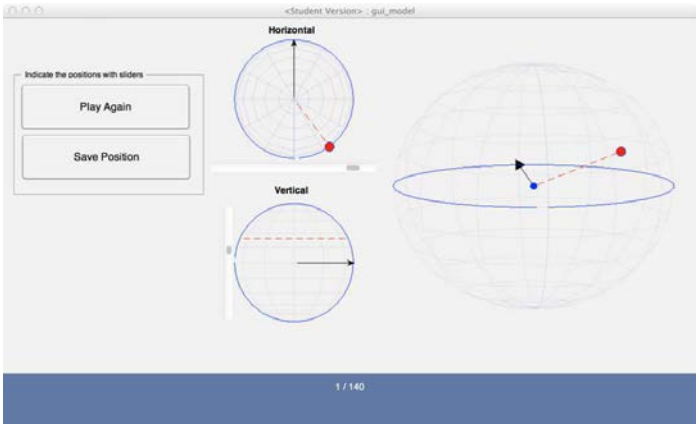


Figure 7.7: Graphical user interface for the localization task.

task. No repetition was carried out. Test duration varied from 8 to 40 minutes, on average 15 minutes. As in the previous test, the subjects were instructed to set the playback level to a comfortable value, but rather on the louder side.

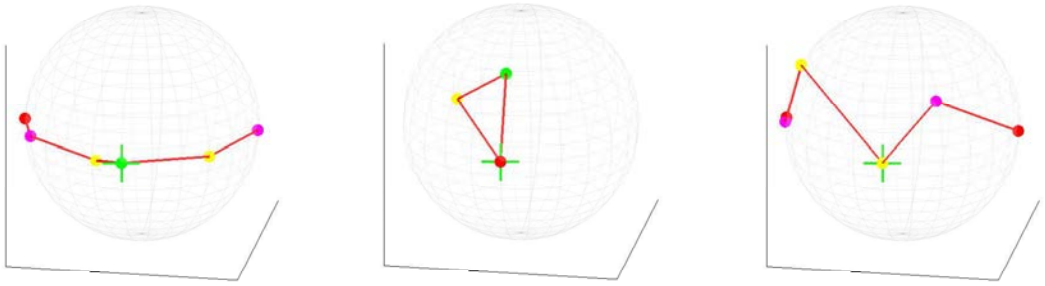
7.3.3 Experimental Results

As a start, the results of each subject, position and component was manually inspected and the resulting localization effects were interpreted. For example, Figure 7.8 shows the judgments for PC1, PC3 and PC5 at sound direction 0° azimuth and 0° elevation of subject ID 4. From this, one can conclude that PC1 describes lateral variation whereas PC3 mainly effects up/down and front/back cues. PC5 also supports PC1 for lateral discrimination. However, not all judgements are as clear as in the figure, therefore through numerical analysis, it was investigated which components are more sensitive to lateral, vertical or front/back discrimination.

Figure 7.9 provides an overview how the indicated source positions vary in respect to the adapted values. For better interpretation, results are shown separately for azimuth and elevation plane. Thus, the influence of the components can be better determined in terms of these planes.

The left column of Figure 7.9 indicates the judgements averaged across all test subjects for each PC and position in the horizontal plane. The centroid of the distributions was calculated by taking the mean direction across all individual judgements. For better readability, the markers for each position were staggered a little bit along the adaptation levels.

The centroid S of the distributions was calculated according to the formulas proposed by Leong and Carlile [LC98] and the following notation is based on them.



(a) PC1. (b) PC3. (c) PC5.

Figure 7.8: Judgements of source positions (subject ID 4) in respect to the adapted principal weights for the reference position 0° elevation. The green marker shows the reference position in the median plane whereas the green point indicates the perceived position adapted by the median value. Red points show the minimum and maximum of the adapted weights, yellow corresponds to 12.5th and 87.5th and magenta to 1st 99th percentiles.

First, the azimuth and elevation angles, where a point directly in front of the head corresponds to zero azimuth and elevation, were converted to the polar coordinates ($\theta = 90 - el,\ \varphi = az$). From a set of n data points on the sphere, the direction cosines are given as

$$\begin{aligned} x_i &= \sin(\theta_i) \cdot \cos(\varphi_i) \ , \\ y_i &= \sin(\theta_i) \cdot \sin(\varphi_i) \ , \\ z_i &= \cos(\theta_i) \ . \end{aligned} \tag{7.2}$$

The centroid of the distributions is computed as the vector sum of the unit vectors,

$$S_x = \sum_{i=1}^{n} x_i, \quad S_y = \sum_{i=-1}^{n} y_i, \quad S_z = \sum_{i=1}^{n} z_i \ . \tag{7.3}$$

The mean directions can be regarded as a measure of expansion and is given by

$$R = \sqrt{S_x^2 + S_y^2 + S_z^2} \ , \tag{7.4}$$

with small values for a high expansion and large values for a low expansion. The mean direction cosines are given by

$$\bar{x} = \frac{S_x}{R}, \quad \bar{y} = \frac{S_y}{R}, \quad \bar{z} = \frac{S_z}{R} \ , \tag{7.5}$$

and to convert them into the polar coordinates:

$$\theta = \arccos(\bar{z}), \quad \varphi = \arctan(\frac{\bar{y}}{\bar{z}}) \ . \tag{7.6}$$

When a sound direction was perceived to the back although the reference position was in the front, these judgments were averaged across subjects and marked as a cross. The upper table in each diagram indicates the number of times in percent when a sound was judged to be at the back for each position an adaption level. It has to be noted that there are no real termed *"confusions"* like in other localization tests, because in this test one actually does not know the *correct* perceived sound direction. Therefore, the judgments in the rear section that are shifted to the front are referred to as *corrected* judgments. The error bar along the curve represents the standard error across all individual corrected judgments.

For the first component, a clear trend in changing azimuth perception is visible, also when the perceived sound directions were in the back. Beginning from the minimum to the maximum adaptation level, the judgments go from -90 to 90 degrees. To this, also PC5 has similar distinct tendency to affect azimuth perception, but in the opposite direction of PC1. For the remaining principal components, the distance is in general between -30 and 30 degrees and does not reveal any systematic pattern in affecting azimuth perception.

A greater occurrence of corrected front/back judgments might lead to the interpretation that this component models front/back discrimination. Table 7.3 presents the percentage of sound directions that were perceived to be at the back although the reference position was in the front, averaged across adaption levels. PC3 and PC4 have the lowest percentage with an average of 16.5% across positions and PC1 and PC2 have the highest value with an average of 23.2% and 21.4% respectively. A closer look at the percentages across all components for each position reveals that the source position -30° has the highest value with 23.2%.

The right column of Figure 7.9 indicates the judgements averaged across all test subjects for each PC and position in the median plane. A trend for PC2, PC3 and PC4 is visible. Whereas in PC2, the indicated sound directions only change significantly for adaption levels above the 87.5th percentile to the upper direction, the moving along the adaptation steps makes elevation judgments decrease for PC3 and increase for PC4. As expected, PC1 and PC5 are did not influence vertical judgments.

Table 7.4 indicates the percentage of sound directions that were judged to be at the lower hemisphere although the reference position was in the upper hemisphere and the other way round, for different PCs and source positions in elevation, averaged

		Source Positions				
		-30°	**0°**	**30°**	**60°**	**Mean**
PCs	**1**	35.7%	25.0%	14.3%	17.9%	23.2%
	2	17.9%	21.4%	10.7%	28.6%	21.4%
	3	23.2%	12.5%	8.9%	17.9%	16.5%
	4	21.4%	19.6%	5.4%	14.3%	16.5%
	5	17.9%	23.2%	7.1%	32.1%	20.9%
	Mean	23.2%	20.4%	13.2%	22.1%	

Table 7.3: Percentage of sounds that were judged to be at the back although the reference stimulus was at the front for different PCs and source positions in elevation, averaged across adaptation levels.

across adaptation levels. Note, that the source position 0° was not listed here since no values can be evaluated for this position. The first component has a minimum of 4.1% averaged across the positions whereas the remaining components have almost the same values. In terms of positions, the source elevation -30° has significant higher percentage with 16.8% than the others.

		Source Positions			
		-30°	**30°**	**60°**	**Mean**
PCs	1	5.4%	3.6%	3.6%	4.1%
	2	14.3%	5.4%	8.9%	9.6%
	3	17.8%	5.4%	3.6%	8.9%
	4	25.0%	3.6%	0.0%	9.5%
	5	21.4%	1.8%	5.4%	9.5%
	Mean	16.8%	3.9%	4.3%	

Table 7.4: Percentage of sounds that were judged to be at the lower hemisphere although the reference stimulus was in the upper hemisphere and the other way round, for different PCs and source positions in elevation, averaged across adaptation levels.

7.3.4 Conclusion

Fortunately, a certain agreement between the interpretations of the principal weights in the database and the results of the experiment was found. Basically, all the important localization effects for lateral and up/down discrimination were recognized. Only front/back were difficult to inspect.

For two components, namely PC1 and PC5, the results show a clear trend to azimuth perception. It has to be noted again that no alignment of the interaural time delay was applied, since the four test positions were located on the median plane.

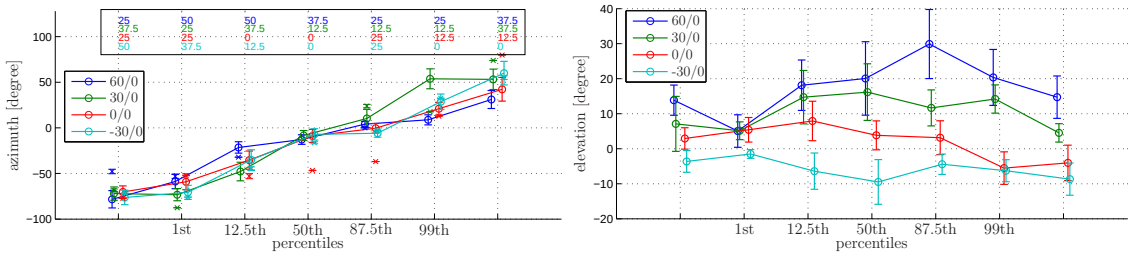
Nevertheless, the maximum indicated distance values between reference and adapted version were in the range of ± 90 degrees. Consequently, this lateral localization effect is only produced by the interaural level difference of the individual principal component weights.

PC3 and PC4 influence vertical discrimination and do not appear to systematically affect azimuth perception. Less easy is the interpretation of the effects for PC2. Analysis of the individual principal weights for this component in the database suggest a front/back cue. However, the number of positions that were perceived from the rear section is on average 21.4% and in the same range of PC1 and PC5. According to Figures 7.1e and f, the distinction between front and rear section can be found in variation of the principal weights in the horizontal plane. A closer analysis reveals, that the PCWs are negative in the rear section and positive in front of the head. However, the PCWs in the experiment were only varied according the four test position in the median plane and their PCW distribution is mainly positive. Therefore, no clear effect in regards to front/back discrimination was found in the evaluation. Apparently, the influence of this component should be further investigated with another listening task.

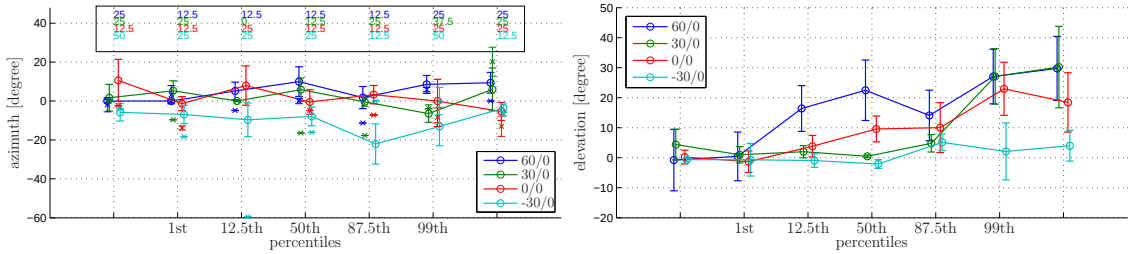
This raises the question of whether the range between 1st and 99th percentile is enough to cause a clear localization effect or the range should be enlarged which was done in this experiment. This might be different for each component that was examined. For PC1 and PC5 which produce lateral movement of the source position, the range between 1st and 99th percentile produced minimum and maximum judgments of about ± 50-80°. This is totally sufficient since only the individual weights of the source positions in the median plane were used. The additional adjustment levels enhanced the localization effect up to ± 90°. For PC3 and PC4 that modify the source position in the elevation, the judgments for 1st and 99th percentile were between -20° and 50°, depending on the source position. Here, the additional adjustment levels enhanced the localization effect especially in the upper hemisphere to additional 10° for almost each source position. This is a clear improvement and could be strengthened. Generally an increase of the adjustment range automatically means the inclusion of the weights of the neighboring positions. However, it should be taken into account that the PCA model only considers local adjustment, therefore a range beyond the distribution of the individual weights for each position stored in the database is not necessary.

To conclude, with only eight test subjects, the effect of lateral and up/down discrimination was easy to reproduce. Only the effects for PC2 cannot be clearly

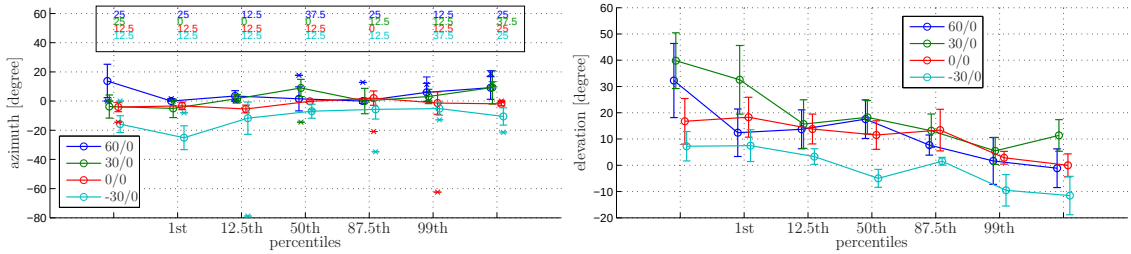
classified into one of the three main localization effects of HRTFs. Consequently, a different experiment in which subjects adapt the PCWs on their own to produce a source position at the rear would be beneficial.



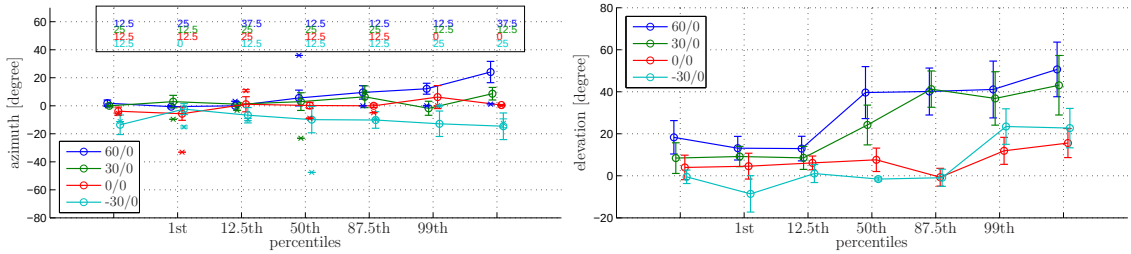
(a) Lateral judgments for PC1. (b) Vertical judgments for PC1.



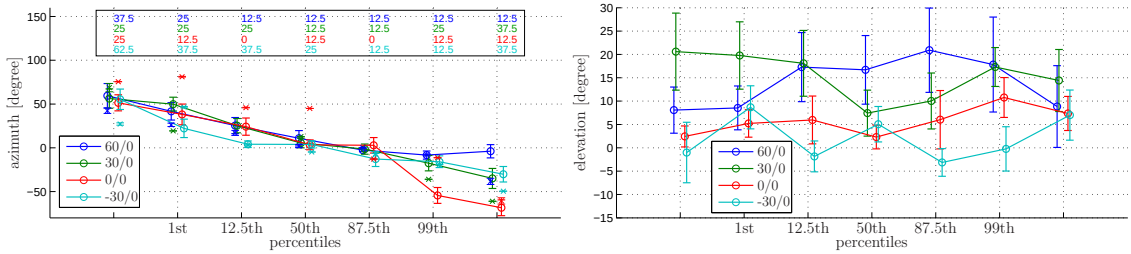
(c) Lateral judgments for PC2. (d) Vertical judgments for PC2.



(e) Lateral judgments for PC3. (f) Vertical judgments for PC3.



(g) Lateral judgments for PC4. (h) Vertical judgments for PC4.



(i) Lateral judgments for PC5. (j) Vertical judgments for PC5.

Figure 7.9: Averaged judgements across subjects in respect to the horizontal (left column) and vertical (right column) plane for each position and principal component.

Chapter 8

Conclusion

8.1 Summary of the Results

The work investigated in individualization of HRTF since direct measurement is time-demanding and requires special equipment and knowledge. This is a major research area and different approaches are currently under investigation. To this, a summary of current individualization methods, their advantages, disadvantages and technical aspects was given. Two data decomposition methods, namely Principal Component Analysis and Spherical Harmonic Decomposition were described and their potential with respect to HRTF adjustment was discussed.

PCA is very suitable for the decomposition of HRTFs because the resulting PCs and PCWs are reproducible for different databases with different sizes and measurement conditions. One can specify the total variance describing the original dataset which is directly related to the number of PCs used for reconstruction. The literature review has shown that PCA is used very differently for the modeling of HRTFs. Since an HRTF database is multidimensional, there are not only differences in the selection of signal representation, but also different realizations of a two-dimensional input matrix.

Adequate parameters for the PCA model were extracted through a numerical evaluation based on compression efficiency and suitability for HRTF individualization. A meaningful selection of the structure for the PCA input matrix returns principal component weights for each subject and position. In this way, adjustment of the weights is applied by navigating through the distribution of the individual weights for each position. Logarithmic frequency magnitude was used as signal representation because it was found as an optimum for the PCA compression efficiency. In addition, through spectral smoothing of the input data before PCA processing, only 7 PCs are required to describe 90 percent of the variance in the dataset. Analysis of the

reconstruction error in frequency domain reveals that at least 10 PCs are necessary for an average spectral distortion of about 5 decibel. In time domain, the minimum-phase HRIRs were used as a measure and reveals an average signal-to-noise ratio of about 8 decibel with 10 PCs.

A listening experiment investigated the variation of the PCWs that is required to produce a change in localization and provides reproducible results with respect to the analysis of individual weights in the database. Listeners indicated lateral localization effects when the weights of PC1 and PC5 were modified. PC3 and PC4 are important cues for up/down discrimination and PC2 is important for front/back discrimination. The resulting minimum and maximum judgements of the source positions in the localization test confirmed that an adjustment range between the 1st and 99th percentile across all individual weights of a source position is adequate to cause a clear localization effect. This is based on the fact that the individual weights are sufficiently different for each position that were examined.

Spherical Harmonics are widely used for decomposition of HRTFs resulting in weights for each frequency bin. However, in this work, the transform was used to model the PCWs on the sphere. The SH model attempts to overcome the directional limitation and enables global adjustment of all positions simultaneously. This is done by decomposition of the PCWs that are structured for each position and subject in spherical harmonic weights. Consequently, each PCW is modeled through a limited number of spherical harmonic weights. As long as $(N+1)^2 < P$, with an SH order N and a total number of P source positions, the approach is more effective than adjusting PCWs for each position, assuming that the same SH order is used for all positions. In this way, adapting a spherical weight effects PCWs of particular regions since each spherical basis function has different directional dependency. Analysis of the reconstruction error in frequency domain indicates that at least an SH order of 2 is necessary. When doing this, for each principal component, 9 basis function weights have to be adjusted which is in great contrast to adjusting the PCWs for each source position separately. It has to be mentioned that this proposed model has not been validated, perceptual studies are required to find an optimal model configuration.

Figure 8.1 depicts the main parts of the model implementation that was built in MATLAB®. The core functions can be used for parameter testing or an individualization process. To this, a GUI for adapting PCWs locally and globally was built and is described in Appendix A. Since virtual auditory display is a significant application in consumer products, this implementation is immediately interesting.

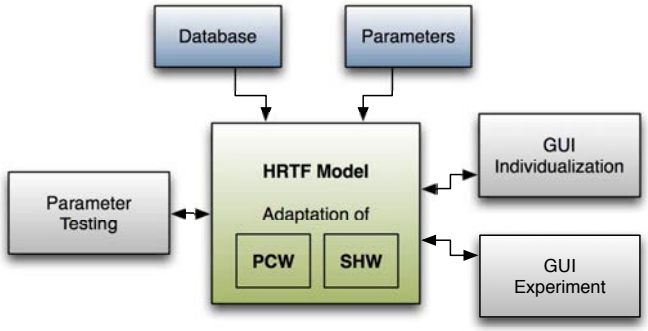


Figure 8.1: Overview of the model implementation.

8.2 Outlook

Based on the fact that the SH model was not perceptually tested, a more detailed investigation into this would be of interest. For this, a listening test to evaluate the performance of the adaption in the spherical domain is necessary. Particular attention must be paid to the usage and modification of the sampling grid in HRTF databases, To this, matrix regularization for the sampled source position grid has to be further investigated.

Two data decomposition methods beside PCA, namely *Independent Component Analysis (ICA)* and *Non-negative Matrix Factorization (NMF)* which are not treated in this work, can also be used to produce basis functions and corresponding weights. In the model implementation these methods were integrated, but not tested. Through a numerical evaluation, the resulting localization effects of the basis functions could be further investigated. To this, also the relation between directional bands and the modified HRTF spectrum that is adapted through the weights could be a fine step forward to better understand up/down and front/back discrimination.

A future work could be a mobile application and a standardized data format in which the individual adapted weights are stored and used in other application. As already a standardized format for HRTF databases was initially created by Majdak *et al.* [MIC+13], this could be a great symbiosis. The reduction in dimensionality enables HRTF customization on mobile phones, because only few parameters that are perceptual relevant have to be stored. Hence, using a clever test procedure to gather relevant individual localization parameters could avoid measurement of HRTFs.

Appendix A

HRTF Exploration Tool

For a better overview and understanding of the matter, a tool in MATLAB® was initially created. It focuses on synthesizing HRIRs with different methods that are described in this work and even more. All parameters of the input matrix for structure Struct2 that are described in Chapter 5 can be selected. Spherical Harmonic Decomposition can be patched to convert the adjustment process into the spherical domain. All directions and possible trajectories in azimuth and elevation can be adapted locally or globally. Several adjustments for the stimulus can be made, such as equalizing with a prior measured headphone transfer function and including several room impulse responses. Adjusted weights can be stored in filesystem so that they can be used or compared again later.

The variation of the spherical weights are categorized into three main groups: shifting, rotation and diminution/widening. Prior knowledge of these transformation effects can be used to overlay additional information in the GUI for each slider, e.g. that the sliders shifts trajectory left/right. In the GUI, different background colors were selected to distinguish between the transformation effects. It has to be noted that no subjective evaluation to verify these effects on spherical weights has been carried out. These overlay informations are only based on the subjective judgment of the author, hence, these effects might may be different for other people.

The computation time for the reconstruction of the time signal when a slider is changed is highly dependent on the number of the selected listening positions. For the PCA model, 10 ms are required for listening a single position and increases up to over 20 ms when a trajectory of source positions is selected. For the PCA-SH model, the reconstruction process takes at least 30 ms for a single position and increases up to 40 ms for several positions.

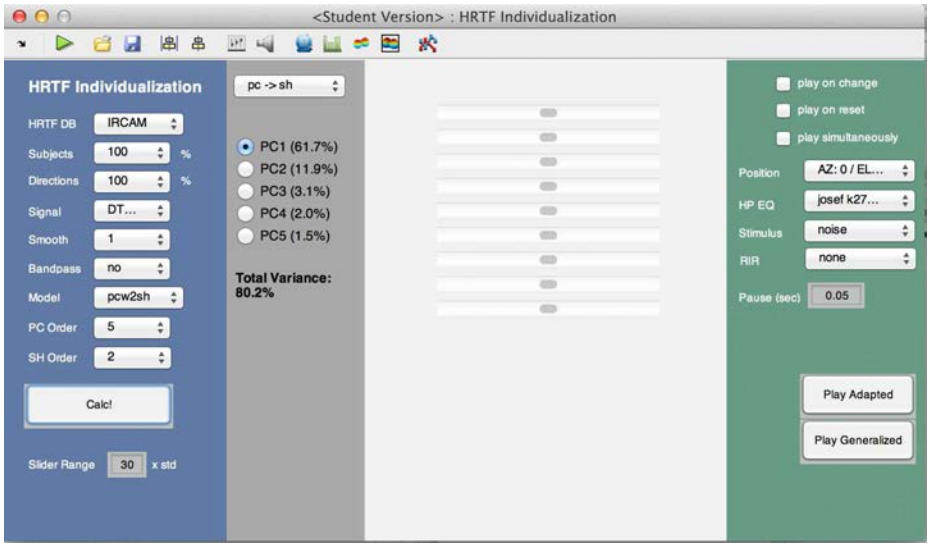


Figure A.1: Graphical user interface for testing input parameters for the PCA and PCA-SH model.

Appendix B

Overview of the Model Implementation

The implementation of the proposed model was done in MATLAB® and provides a flexible way to analyze and synthesize HRTFs. The core of the model are several functions (Figure B.1) that can be used to perform an error analysis of the entire dataset without adjustment of PCWs or modify the PCWs through sliders in a GUI. Table B.1 explains the possible model parameters. After calculation, the model data can be stored automatically in the filesystem for further use. If precalculated model data are available in filesystem, the file is imported automatically instead of computing the model again. This can save up to 10 minutes of computing time, depending on the model parameters.

The author would like to thank his advisor Georgios Marentakis who was quite involved in developing the model.

Parameter	Type	Values	Comment
Database	string	ari, cipic, ircam	
Dataset			
Subject Density	float	1-100	in percent
Direction Density	float	1-100	in percent
Ears	integer	1, 2, [1,2]	only left and right or both ears
Frequency Smoothing Ratio	integer	1, 2, 3 ...	until max. number of frequency bins
Bandpass	boolean	0, 1	
Model			
Model Type	string	pca, ica, nmf	
Model Order	integer	1, 2, 3 ...	number of basis functions
Structure	integer	1-5	dimension of the input matrix
Signal Representation	integer	1-4	HRIR, Minimum-phase HRIR, DTF linear, DTF logarithmic
Ear Handling	integer	1, 2	ears in rows or columns blocked
Weight Model			
Weight Model Type	string	local, global	
Weight Model Order	integer	1, 2, 3 ...	number of spherical basis functions
Matrix Regularization	boolean	0,1	
Reconstructed Set			
Order	integer	1, 2, 3 ...	until max. number of time/frequency bins
Weight Order	integer	1, 2, 3 ...	number of spherical basis functions
Direction IDs	integer	[1, 2, 3 ...]	vector of direction ids
Error Computation	boolean	0, 1	
Sound Reproduction			
Stimulus	integer	1, 2, 3, ...	id of predefined stimuli
Headphone EQ	integer	1, 2, 3 ...	id of predefined inverse headphone transfer functions
Room Impulse Response	integer	1, 2, 3, ...	id number of predefined room impulse responses

Table B.1: Explanation of the model parameters.

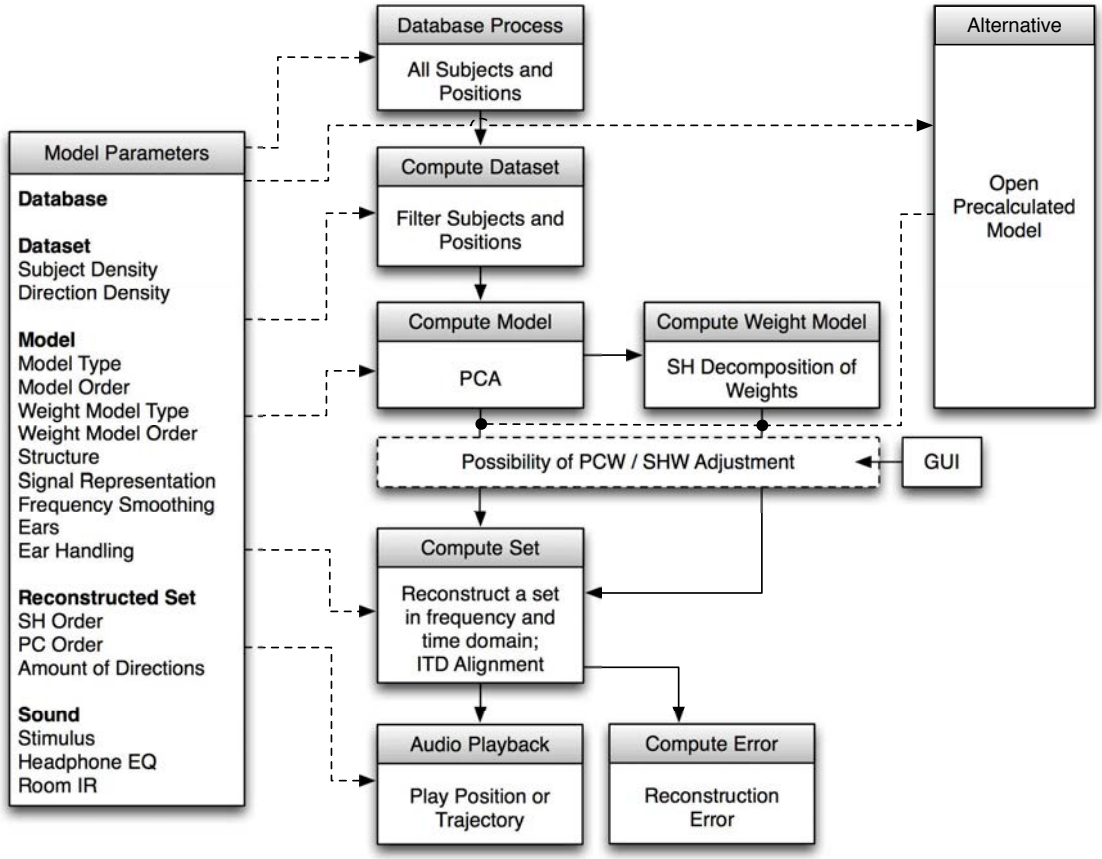


Figure B.1: Overview of the HRTF model implementation in MATLAB®.

Appendix C

Input Matrix Structures

C.1 Graphical Representation

A graphical representation of the five described structures allows a better overview of the multidimensional HRTF data and the resulting PCA output. Matrix dimensions for Struct2 can be found in Figure 5.4 on Page 49.

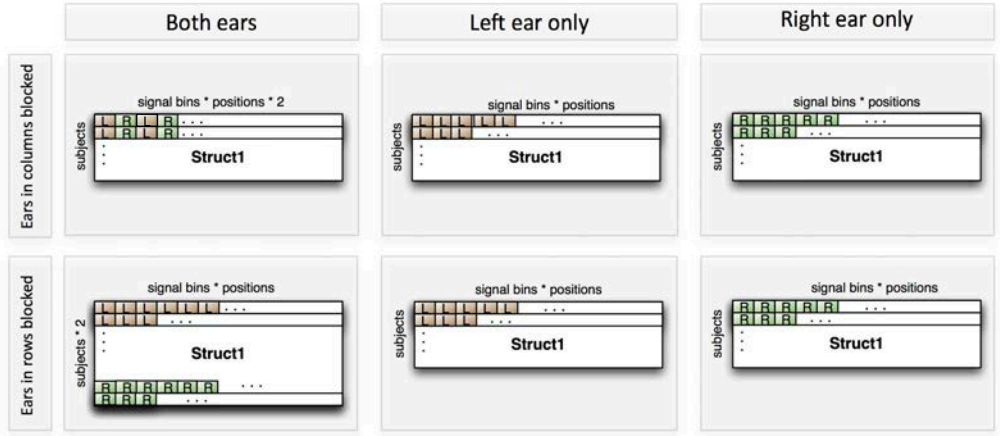


Figure C.1: Dimension of input matrix **Struct1** $[subjects \times (signal * positions)]$ when choosing ears in columns (first line) or rows (second line) blocked.

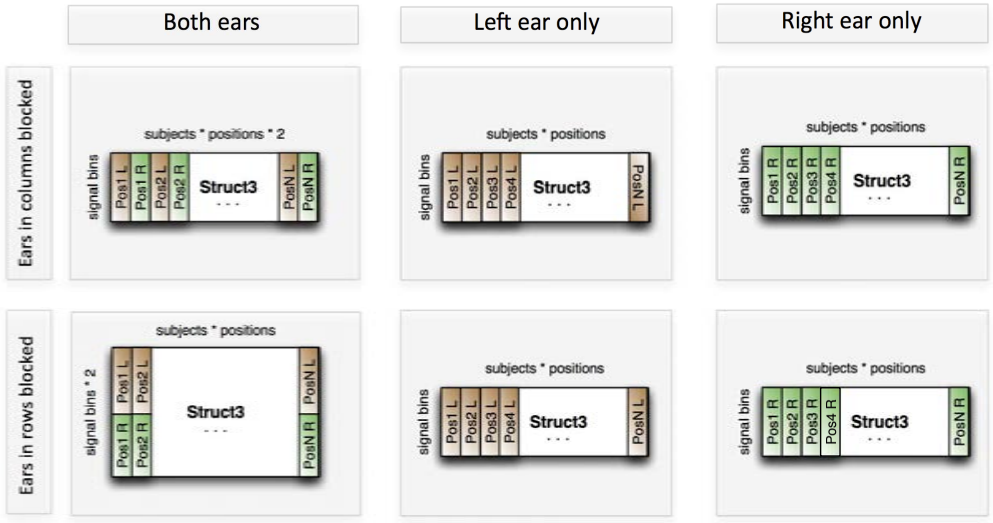


Figure C.2: Dimension of input matrix **Struct3** $[signal \times (subjects * positions)]$ when choosing ears in columns (first line) or rows (second line) blocked.

Figure C.3: Dimension of input matrix **Struct4** $[(signal * positions) \times subjects]$ when choosing ears in columns (first line) or rows (second line) blocked.

Figure C.4: Dimension of input matrix **Struct5** $[(subjects * signal) \times (positions)]$ when choosing ears in columns (first line) or rows (second line) blocked.

C.2 Variance Tables

Variance tables for three open access HRTF databases provide a full overview of the required PCs to obtain 90 percent variance for different properties of input data matrix.

			HRIR		Min HRIR		DTF lin		DTF log	
			E ↓	**E** →	**E** ↓	**E** →	**E** ↓	**E** →	**E** ↓	**E** →
Struct1	S/1	left ear	49	49	46	46	42	42	59	59
		right ear	50	50	46	46	41	41	60	60
		both ears	92	55	56	52	28	47	40	62
	S/2	left ear	49	49	46	46	40	40	58	58
		right ear	50	50	46	46	39	39	58	58
		both ears	92	55	56	52	26	46	35	61
	S/4	left ear	49	49	46	46	38	38	55	55
		right ear	50	50	46	46	38	38	56	56
		both ears	92	55	56	52	24	45	28	59
	S/8	left ear	49	49	46	46	32	32	51	51
		right ear	50	50	46	46	31	31	51	51
		both ears	92	55	56	52	18	39	19	55
	S/16	left ear	49	49	46	46	23	23	45	45
		right ear	50	50	46	46	22	22	45	45
		both ears	92	55	56	52	9	30	9	50
	S/32	left ear	49	49	46	46	17	17	40	40
		right ear	50	50	46	46	16	16	40	40
		both ears	92	55	56	52	6	24	5	46
Struct2	S/1	left ear	15	15	10	10	7	7	6	6
		right ear	15	15	11	11	7	7	6	6
		both ears	15	29	11	20	7	12	6	10
	S/2	left ear	15	15	10	10	6	6	5	5
		right ear	15	15	11	11	6	6	5	5
		both ears	15	29	11	20	6	12	5	8
		left ear	15	15	10	10	6	6	4	4

continued on next page

continued from previous page

			HRIR		Min HRIR		DTF lin		DTF log	
			E ↓	**E** →	**E** ↓	**E** →	**E** ↓	**E** →	**E** ↓	**E** →
Struct2	S/4	right ear	15	15	11	11	6	6	4	4
		both ears	15	29	11	20	6	11	4	7
	S/8	left ear	15	15	10	10	5	5	3	3
		right ear	15	15	11	11	5	5	3	3
		both ears	15	29	11	20	5	9	3	6
	S/16	left ear	15	15	10	10	3	3	3	3
		right ear	15	15	11	11	3	3	2	2
		both ears	15	29	11	20	3	6	3	4
	S/32	left ear	15	15	10	10	2	2	2	2
		right ear	15	15	11	11	2	2	2	2
		both ears	15	29	11	20	2	4	2	2
Struct3	S/1	left ear	16	16	7	7	6	6	1	1
		right ear	15	15	7	7	6	6	1	1
		both ears	29	15	12	7	8	6	2	1
	S/2	left ear	16	16	7	7	5	5	1	1
		right ear	15	15	7	7	5	5	1	1
		both ears	29	15	12	7	8	5	2	1
	S/4	left ear	16	16	7	7	5	5	1	1
		right ear	15	15	7	7	5	5	1	1
		both ears	29	15	12	7	7	5	2	1
	S/8	left ear	16	16	7	7	4	4	1	1
		right ear	15	15	7	7	5	5	1	1
		both ears	29	15	12	7	6	5	2	1
	S/16	left ear	16	16	7	7	2	2	1	1
		right ear	15	15	7	7	3	3	1	1
		both ears	29	15	12	7	3	3	2	1
	S/32	left ear	16	16	7	7	3	3	1	1
		right ear	15	15	7	7	3	3	1	1
		both ears	29	15	12	7	3	3	2	1
	S/1	left ear	41	41	10	10	4	4	1	1
		right ear	45	45	11	11	5	5	1	1
		both ears	48	85	13	19	6	8	1	2

continued on next page

continued from previous page

			HRIR		Min HRIR		DTF lin		DTF log	
			E ↓	**E** →	**E** ↓	**E** →	**E** ↓	**E** →	**E** ↓	**E** →
Struct4	S/2	left ear	41	41	10	10	4	4	1	1
		right ear	45	45	11	11	5	5	1	1
		both ears	48	85	13	19	6	7	1	2
	S/4	left ear	41	41	10	10	4	4	1	1
		right ear	45	45	11	11	4	4	1	1
		both ears	48	85	13	19	5	7	1	2
	S/8	left ear	41	41	10	10	3	3	1	1
		right ear	45	45	11	11	3	3	1	1
		both ears	48	85	13	19	4	5	1	2
	S/16	left ear	41	41	10	10	2	2	1	1
		right ear	45	45	11	11	2	2	1	1
		both ears	48	85	13	19	2	3	1	2
	S/32	left ear	41	41	10	10	1	1	1	1
		right ear	45	45	11	11	2	2	1	1
		both ears	48	85	13	19	2	3	1	2
Struct5	S/1	left ear	99	99	5	5	4	4	1	1
		right ear	98	98	6	6	5	5	1	1
		both ears	151	174	9	8	5	6	2	1
	S/2	left ear	99	99	5	5	4	4	1	1
		right ear	98	98	6	6	4	4	1	1
		both ears	151	174	9	8	5	6	2	1
	S/4	left ear	99	99	5	5	4	4	1	1
		right ear	98	98	6	6	4	4	1	1
		both ears	151	174	9	8	5	5	2	1
	S/8	left ear	99	99	5	5	3	3	1	1
		right ear	98	98	6	6	4	4	1	1
		both ears	151	174	9	8	4	5	2	1
	S/16	left ear	99	99	5	5	2	2	1	1
		right ear	98	98	6	6	2	2	1	1
		both ears	151	174	9	8	3	3	2	1
	S/32	left ear	99	99	5	5	3	3	1	1
		right ear	98	98	6	6	3	3	1	1

continued on next page

continued from previous page

			HRIR		Min HRIR		DTF lin		DTF log	
			E ↓	**E** →	**E** ↓	**E** →	**E** ↓	**E** →	**E** ↓	**E** →
		both ears	151	174	9	8	2	3	2	1

Table C.1: Number of PCs required to yield 90 percent variance for different realizations of a PCA input matrix based on the ARI database. S/1 refers to no smoothing and S/2 ... S/N to different degrees of HRTF spectrum smoothing (see Chapter 5.2.2). Horizontally, the input signal representations are given and whether ears are blocked in rows (E↓) or columns (E→). Five different input structures and variations of spectral smoothing are listed vertically.

			HRIR		Min HRIR		DTF lin		DTF log	
			E ↓	**E** →	**E** ↓	**E** →	**E** ↓	**E** →	**E** ↓	**E** →
Struct1	S/1	left ear	30	30	29	29	27	27	36	36
		right ear	29	29	28	28	26	26	36	36
		both ears	58	32	39	31	25	29	40	36
	S/2	left ear	30	30	29	29	27	27	35	35
		right ear	29	29	28	28	26	26	35	35
		both ears	58	32	39	31	24	29	37	36
	S/4	left ear	30	30	29	29	26	26	34	34
		right ear	29	29	28	28	25	25	34	34
		both ears	58	32	39	31	23	28	33	35
	S/8	left ear	30	30	29	29	24	24	33	33
		right ear	29	29	28	28	23	23	33	33
		both ears	58	32	39	31	19	26	25	34
	S/16	left ear	30	30	29	29	20	20	30	30
		right ear	29	29	28	28	19	19	30	30
		both ears	58	32	39	31	11	23	11	31
	S/32	left ear	30	30	29	29	17	17	26	26
		right ear	29	29	28	28	16	16	26	26
		both ears	58	32	39	31	6	20	4	28
	S/1	left ear	19	19	10	10	6	6	9	9
		right ear	20	20	11	11	7	7	8	8
		both ears	20	38	11	20	7	12	8	15

continued on next page

continued from previous page

			HRIR		Min HRIR		DTF lin		DTF log	
			E ↓	**E** →	**E** ↓	**E** →	**E** ↓	**E** →	**E** ↓	**E** →
Struct2	S/2	left ear	19	19	10	10	6	6	8	8
		right ear	20	20	11	11	7	7	7	7
		both ears	20	38	11	20	6	12	7	13
	S/4	left ear	19	19	10	10	6	6	6	6
		right ear	20	20	11	11	6	6	6	6
		both ears	20	38	11	20	6	12	6	11
	S/8	left ear	19	19	10	10	5	5	5	5
		right ear	20	20	11	11	5	5	4	4
		both ears	20	38	11	20	5	10	5	8
	S/16	left ear	19	19	10	10	3	3	3	3
		right ear	20	20	11	11	3	3	3	3
		both ears	20	38	11	20	3	5	3	4
	S/32	left ear	19	19	10	10	2	2	2	2
		right ear	20	20	11	11	2	2	1	1
		both ears	20	38	11	20	2	3	1	2
Struct3	S/1	left ear	20	20	7	7	7	7	7	7
		right ear	20	20	7	7	8	8	7	7
		both ears	39	20	13	7	10	8	7	7
	S/2	left ear	20	20	7	7	7	7	6	6
		right ear	20	20	7	7	8	8	7	7
		both ears	39	20	13	7	10	8	6	6
	S/4	left ear	20	20	7	7	7	7	5	5
		right ear	20	20	7	7	7	7	6	6
		both ears	39	20	13	7	10	7	5	5
	S/8	left ear	20	20	7	7	6	6	4	4
		right ear	20	20	7	7	6	6	4	4
		both ears	39	20	13	7	8	6	3	4
	S/16	left ear	20	20	7	7	4	4	2	2
		right ear	20	20	7	7	3	3	2	2
		both ears	39	20	13	7	4	4	2	2
	S/32	left ear	20	20	7	7	2	2	1	1
		right ear	20	20	7	7	2	2	1	1

continued on next page

continued from previous page

			HRIR		Min HRIR		DTF lin		DTF log	
			E ↓	**E** →	**E** ↓	**E** →	**E** ↓	**E** →	**E** ↓	**E** →
		both ears	39	20	13	7	2	2	2	1
Struct4	S/1	left ear	29	29	9	9	7	7	9	9
		right ear	28	28	10	10	8	8	9	9
		both ears	31	56	11	18	9	14	10	14
	S/2	left ear	29	29	9	9	7	7	7	7
		right ear	28	28	10	10	8	8	8	8
		both ears	31	56	11	18	9	13	8	11
	S/4	left ear	29	29	9	9	7	7	6	6
		right ear	28	28	10	10	7	7	6	6
		both ears	31	56	11	18	9	13	7	8
	S/8	left ear	29	29	9	9	6	6	3	3
		right ear	28	28	10	10	6	6	3	3
		both ears	31	56	11	18	7	10	4	5
	S/16	left ear	29	29	9	9	3	3	1	1
		right ear	28	28	10	10	3	3	1	1
		both ears	31	56	11	18	4	5	1	2
	S/32	left ear	29	29	9	9	2	2	1	1
		right ear	28	28	10	10	2	2	1	1
		both ears	31	56	11	18	3	3	1	2
Struct5	S/1	left ear	50	50	7	7	9	9	25	25
		right ear	53	53	7	7	8	8	31	31
		both ears	73	93	11	10	9	13	12	41
	S/2	left ear	50	50	7	7	8	8	16	16
		right ear	53	53	7	7	8	8	20	20
		both ears	73	93	11	10	9	13	8	28
	S/4	left ear	50	50	7	7	8	8	12	12
		right ear	53	53	7	7	8	8	13	13
		both ears	73	93	11	10	9	12	6	18
	S/8	left ear	50	50	7	7	7	7	7	7
		right ear	53	53	7	7	7	7	8	8
		both ears	73	93	11	10	8	10	4	11
		left ear	50	50	7	7	5	5	4	4

continued on next page

continued from previous page

			HRIR		Min HRIR		DTF lin		DTF log	
			E ↓	**E →**	**E ↓**	**E →**	**E ↓**	**E →**	**E ↓**	**E →**
Struct5	S/16	right ear	53	53	7	7	5	5	4	4
		both ears	73	93	11	10	5	6	3	5
	S/32	left ear	50	50	7	7	4	4	2	2
		right ear	53	53	7	7	4	4	2	2
		both ears	73	93	11	10	3	5	2	3

Table C.2: Number of PCs required to yield 90 percent variance for different realizations of a PCA input matrix based on the CIPIC database. S/1 refers to no smoothing and S/2 ... S/N to different degrees of HRTF spectrum smoothing (see Chapter 5.2.2). Horizontally, the input signal representations are given and whether ears are blocked in rows (E↓) or columns (E→). Five different input structures and variations of spectral smoothing are listed vertically.

			HRIR		Min HRIR		DTF lin		DTF log	
			E ↓	**E →**	**E ↓**	**E →**	**E ↓**	**E →**	**E ↓**	**E →**
Struct1	S/1	left ear	36	36	34	34	33	33	38	38
		right ear	37	37	34	34	34	34	38	38
		both ears	72	39	49	36	32	35	38	39
	S/2	left ear	36	36	34	34	32	32	37	37
		right ear	37	37	34	34	33	33	37	37
		both ears	72	39	49	36	31	35	32	38
	S/4	left ear	36	36	34	34	32	32	36	36
		right ear	37	37	34	34	33	33	36	36
		both ears	72	39	49	36	30	35	28	37
	S/8	left ear	36	36	34	34	32	32	35	35
		right ear	37	37	34	34	33	33	35	35
		both ears	72	39	49	36	29	35	24	36
	S/16	left ear	36	36	34	34	30	30	32	32
		right ear	37	37	34	34	31	31	33	33
		both ears	72	39	49	36	28	34	18	35
	S/32	left ear	36	36	34	34	28	28	28	28
		right ear	37	37	34	34	29	29	29	29
		both ears	72	39	49	36	20	32	9	31

continued on next page

continued from previous page

			HRIR		Min HRIR		DTF lin		DTF log	
			E ↓	**E** →	**E** ↓	**E** →	**E** ↓	**E** →	**E** ↓	**E** →
Struct2	S/1	left ear	21	21	11	11	7	7	10	10
		right ear	20	20	11	11	7	7	10	10
		both ears	21	39	11	20	7	13	10	18
	S/2	left ear	21	21	11	11	7	7	7	7
		right ear	20	20	11	11	7	7	7	7
		both ears	21	39	11	20	7	13	7	13
	S/4	left ear	21	21	11	11	7	7	6	6
		right ear	20	20	11	11	7	7	7	7
		both ears	21	39	11	20	7	12	7	12
	S/8	left ear	21	21	11	11	6	6	6	6
		right ear	20	20	11	11	7	7	6	6
		both ears	21	39	11	20	7	12	6	10
	S/16	left ear	21	21	11	11	6	6	5	5
		right ear	20	20	11	11	6	6	5	5
		both ears	21	39	11	20	6	12	5	8
	S/32	left ear	21	21	11	11	4	4	3	3
		right ear	20	20	11	11	5	5	3	3
		both ears	21	39	11	20	5	8	3	4
Struct3	S/1	left ear	21	21	7	7	14	14	41	41
		right ear	20	20	8	8	14	14	42	42
		both ears	39	21	13	8	14	14	19	42
	S/2	left ear	21	21	7	7	12	12	19	19
		right ear	20	20	8	8	13	13	20	20
		both ears	39	21	13	8	13	12	13	20
	S/4	left ear	21	21	7	7	12	12	14	14
		right ear	20	20	8	8	12	12	15	15
		both ears	39	21	13	8	13	12	11	14
	S/8	left ear	21	21	7	7	11	11	11	11
		right ear	20	20	8	8	11	11	12	12
		both ears	39	21	13	8	13	11	10	12
	S/16	left ear	21	21	7	7	9	9	8	8
		right ear	20	20	8	8	10	10	8	8

continued on next page

continued from previous page

			HRIR		Min HRIR		DTF lin		DTF log	
			E ↓	**E** →	**E** ↓	**E** →	**E** ↓	**E** →	**E** ↓	**E** →
		both ears	39	21	13	8	12	10	8	8
	S/32	left ear	21	21	7	7	6	6	5	5
		right ear	20	20	8	8	6	6	5	5
		both ears	39	21	13	8	8	6	5	5
Struct4	S/1	left ear	36	36	12	12	14	14	17	17
		right ear	37	37	12	12	14	14	16	16
		both ears	39	72	14	23	16	26	18	30
	S/2	left ear	36	36	12	12	13	13	13	13
		right ear	37	37	12	12	14	14	13	13
		both ears	39	72	14	23	15	25	15	24
	S/4	left ear	36	36	12	12	13	13	12	12
		right ear	37	37	12	12	13	13	12	12
		both ears	39	72	14	23	15	24	13	20
	S/8	left ear	36	36	12	12	13	13	10	10
		right ear	37	37	12	12	13	13	10	10
		both ears	39	72	14	23	15	24	11	17
	S/16	left ear	36	36	12	12	12	12	8	8
		right ear	37	37	12	12	12	12	7	7
		both ears	39	72	14	23	14	23	8	12
	S/32	left ear	36	36	12	12	9	9	5	5
		right ear	37	37	12	12	9	9	4	4
		both ears	39	72	14	23	10	16	5	6
Struct5	S/1	left ear	67	67	11	11	29	29	68	68
		right ear	66	66	10	10	30	30	71	71
		both ears	103	124	17	15	21	51	30	129
	S/2	left ear	67	67	11	11	26	26	57	57
		right ear	66	66	10	10	27	27	60	60
		both ears	103	124	17	15	19	45	21	105
	S/4	left ear	67	67	11	11	24	24	49	49
		right ear	66	66	10	10	25	25	51	51
		both ears	103	124	17	15	19	42	17	87
		left ear	67	67	11	11	23	23	40	40

continued on next page

continued from previous page

			HRIR		Min HRIR		DTF lin		DTF log	
			E ↓	**E →**	**E ↓**	**E →**	**E ↓**	**E →**	**E ↓**	**E →**
Struct5	S/8	right ear	66	66	10	10	23	23	42	42
		both ears	103	124	17	15	18	40	14	69
	S/16	left ear	67	67	11	11	20	20	28	28
		right ear	66	66	10	10	20	20	30	30
		both ears	103	124	17	15	17	34	11	47
	S/32	left ear	67	67	11	11	18	18	18	18
		right ear	66	66	10	10	18	18	19	19
		both ears	103	124	17	15	14	29	7	28

Table C.3: Number of PCs required to yield 90 percent variance for different realizations of a PCA input matrix based on the IRCAM database. S/1 refers to no smoothing and S/2 ... S/N to different degrees of HRTF spectrum smoothing (see Chapter 5.2.2). Horizontally, the input signal representations are given and whether ears are blocked in rows (E↓) or columns (E→). Five different input structures and variations of spectral smoothing are listed vertically.

List of Abbreviations

ANOVA Analysis of Variance

ARI Acoustics Research Institute

CIPIC Center for Image Processing and Integrated Computing

CTF Common Transfer Function

DRIR Directional Room Impulse Response

DHST Discrete Spherical Harmonic Transform

DTF Directional Transfer Function

FT Fourier Transform

GUI Graphical User Interface

HPTF Headphone Transfer Function

HRIR Head-related Impulse Response

HRTF Head-related Transfer Function

ICA Independent Component Analysis

IFT Inverse Fourier Transform

ILD Interaural Level Difference

JND Just Noticeable Difference

IRCAM Institut de Recherche et Coordination Acoustique/Musique

ITD Interaural Time Difference

KEMAR Knowles Electronic Manikin for Acoustic Research

NCF Central Frequencies of Pinna Notches

PCA Principal Component Analysis

PCW Principal Component Weight

PC Principal Component

PRTF Pinna-related Transfer Function

RMSE Root Mean Square Error

SDR Signal-to-distortion Ratio

SDT Signal Detection Theory

SD Spectral Distortion

SHW Spherical Harmonic Weight

SH Spherical Harmonic

SOFA Spatially Oriented Format for Acoustics

SVD Singular Value Decomposition

TSVD Truncated Singular Value Decomposition

Bibliography

[ADS07] V. R. Algazi, R. O. Duda, and P. Satarzadeh, "Physical and Filter Pinna Models Based on Anthropometry," 2007.

[AR10] A. Avni and B. Rafaely, "Sound localization in a sound field represented by spherical harmonics," *... Symposium on Ambisonics and Spherical ...*, 2010.

[ASS90] F. Asano, Y. Suzuki, and T. Sone, "Role of spectral cues in median plane localization," *The Journal of the Acoustical Society of America*, 1990.

[BBL+05] M. Berg, E. Bondesson, S. Y. Low, S. Nordholm, and I. Claesson, "A Combined On-Line PCA-ICA Algorithm for Blind Source Separation," in *Communications, 2005 Asia-Pacific Conference on*, 2005, pp. 969–972.

[BDBS02] K. Baek, B. A. Draper, J. R. Beveridge, and K. She, "PCA vs. ICA: A comparison on the FERET data set," *Joint Conference on ...*, 2002.

[Bla70] J. Blauert, "Sound localization in the median plane(Frequency function of sound localization in median plane measured psychoacoustically at both ears with narrow band signals)," *Acustica*, 1970.

[Bre12] J. Breebaart, "Effect of perceptually irrelevant variance in head-related transfer functions on principal component analysis." *Journal of the Acoustical Society of America*, vol. 133, no. 1, pp. EL1–EL6, Dec. 2012.

[Che02] H. Chen, "Principal component analysis with missing data and outliers," *Electrical and Computer Engineering Department ...*, 2002.

[CK07] Z. Chen and W. Kreuzer, "A Fast Multipole Boundary Element Method for Calculating HRTFs," 2007.

[CMM09] T. Chen, E. Martin, and G. Montague, "Robust probabilistic PCA with missing data and contribution analysis for outlier detection," *Computational Statistics & Data Analysis*, 2009.

[CPK06] P. S. Chanda, S. Park, and T. I. Kang, "A Binaural Synthesis with Multiple Sound Sources Based on Spatial Features of Head-related Transfer Functions," in *Neural Networks, 2006. IJCNN '06. International Joint Conference on*, 2006, pp. 1726–1730.

[CvVH93] J. Chen, B. D. van Veen, and K. E. Hecox, "A spatial feature extraction and regularization model for the head-related transfer function," *The Journal of the Acoustical Society of America*, 1993.

[EA98] M. J. Evans and J. A. Angus, "Spherical harmonic spectra of head-related transfer functions," *PREPRINTS-AUDIO ENGINEERING SOCIETY*, 1998.

[EAT98] M. J. Evans, J. A. Angus, and A. I. Tew, *Analyzing head-related transfer function measurements using surface spherical harmonics.* JOURNAL-..., 1998.

[FR12] K. J. Fink and L. E. Ray, "Tuning principal component weights to individualize HRTFs." *ICASSP*, pp. 389–392, 2012.

[Gra95] D. W. Grantham, "Hearing," in *Spatial Haring and Related Phenomena*, 1995.

[GV07] G. Grindlay and M. Vasilescu, "A Multilinear (Tensor) Framework for HRTF Analysis and Synthesis," *IEEE International Conference on Acoustics, Speech and Signal Processing. Proceedings*, vol. 1, pp. I–164, Apr. 2007.

[GWFA05] D. W. Grantham, J. A. Willhite, K. D. Frampton, and D. D. Ashmead, "Reduced order modeling of head related impulse responses for virtual acoustic displays," *The Journal of the Acoustical Society of America*, 2005.

[Han87] P. C. Hansen, "The truncated SVD as a method for regularization," *BIT Numerical Mathematics*, vol. 27, no. 4, pp. 534–553, 1987.

[HB88] R. A. Humanski and R. A. Butler, "The contribution of the near and far ear toward localization of sound in the sagittal plane." *Journal of the Acoustical Society of America*, vol. 83, no. 6, pp. 2300–2310, May 1988.

[HBS99] K. Hartung, J. Braasch, and S. J. Sterbing, "Comparison of different methods for the interpolation of head-related transfer functions," *Audio Engineering Society Conference: ...*, 1999.

[HCW08] H. Hu, L. Chen, and Z. Wu, "The estimation of personalized HRTFs in individual VAS," *Natural Computation*, 2008.

[HF09] Q. Huang and Y. Fang, "Interpolation of head-related transfer functions using spherical fourier expansion," *Journal of Electronics (China)*, vol. 26, no. 4, pp. 571–576, 2009.

[HL09] Q. H. Q. Huang and K. L. K. Liu, "A reduced order model of head-related impulse responses based on independent spatial feature extraction," *IEEE International Conference on Acoustics, Speech and Signal Processing. Proceedings*, pp. 281–284, Apr. 2009.

[Hol12] "An initial Investigation into HRTF Adaptation using PCA," *iem.kug.ac.at*, Jul. 2012.

[HP08] S. Hwang and Y. Park, "Interpretations on principal components analysis of head-related impulse responses in the median plane," *The Journal of the Acoustical Society of America*, vol. 123, no. 4, pp. EL65–EL71, 2008.

[HPP08] S. Hwang, Y. Park, and Y.-s. Park, "Modeling and Customization of Head-Related Impulse Responses Based on General Basis Functions in Time Domain," *Acta Acustica united with Acustica*, vol. 94, no. 6, pp. 965–980, Nov. 2008.

[HPP10] ——, "Customization of Spatially Continuous Head-Related Impulse Responses in the Median Plane," *Acta Acustica united with Acustica*, vol. 96, no. 2, pp. 351–363, Mar. 2010.

[HVRVO98] P. M. Hofman, J. G. A. Van Riswick, and A. J. Van Opstal, "Relearning sound localization with new ears," *Nature neuroscience*, vol. 1, no. 5, pp. 417–421, Sep. 1998.

[HW74] J. Hebrank and D. Wright, "Spectral cues used in the localization of sound sources on the median plane," *The Journal of the Acoustical Society of America*, vol. 56, no. 6, pp. 1829–1834, 1974.

[HZK99] J. Huopaniemi, N. Zacharov, and M. Karjalainen, "Objective and subjective evaluation of head-related transfer function filter design," *Journal of the Audio ...*, 1999.

[Jar08] W. Jarosz, *Efficient Monte Carlo Methods for Light Transport in Scattering Media*. ProQuest, 2008.

[KC98] A. Kulkarni and H. S. Colburn, "Role of spectral detail in sound-source localization," *Nature*, 1998.

[KC00] ——, "Variability in the characterization of the headphone transfer-function," *The Journal of the Acoustical Society of America*, vol. 107, no. 2, pp. 1071–1074, 2000.

[KD08] F. Keyrouz and K. Diepold, "A New HRTF Interpolation Approach for Fast Synthesis of Dynamic Environmental Interaction," *Journal of the Audio Engineering Society*, vol. 56, no. 1/2, pp. 28–35, 2008.

[KIC99] A. Kulkarni, S. K. Isabelle, and H. S. Colburn, "Sensitivity of human subjects to head-related transfer-function phase spectra," *The Journal of the Acoustical Society of America*, vol. 105, no. 5, pp. 2821–2840, 1999.

[KP12] B. F. G. Katz and G. Parseihian, "Perceptually based head-related transfer function database optimization," *The Journal of the Acoustical Society of America*, 2012.

[KSG99] A. Kelemen, G. Székely, and G. Gerig, "Elastic model-based segmentation of 3-D neuroradiological data sets," *Medical Imaging*, 1999.

[KW92] D. J. Kistler and F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *The Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1637–1647, 1992.

[LB02] E. H. A. Langendijk and A. W. Bronkhorst, "Contribution of spectral cues to human sound localization," *The Journal of the Acoustical Society of America*, vol. 112, no. 4, pp. 1583–1596, 2002.

[LC85] G. Li and Z. Chen, "Projection-pursuit approach to robust dispersion matrices and principal components: primary theory and Monte Carlo," *Journal of the American Statistical Association*, 1985.

[LC98] P. P. Leong and S. S. Carlile, "Methods for spherical data analysis and visualization." *Journal of Neuroscience Methods*, vol. 80, no. 2, pp. 191–200, Apr. 1998.

[LC09] J. Leung and S. Carlile, "PCA Compression of HRTF and localization performance," in *International Workshop on the Principies and Applications of Spatial Hearing*, 2009.

[LEW10] A. Lindau, J. Estrella, and S. Weinzierl, "Individualization of dynamic binaural synthesis by real time manipulation of the ITD," *Proc of the 128th AES Convention ...*, 2010.

[LH02] J. Li and A. O. Hero, "A spectral approach to statistical polar shape modeling," in *Image Processing. 2002. Proceedings. 2002 International Conference on*, Dec. 2002.

[LYW13] Y.-J. Lee, Y.-R. Yeh, and Y.-C. F. Wang, "Anomaly Detection via Online Oversampling Principal Component Analysis," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 25, no. 7, pp. 1460–1470, 2013.

[Maj13] P. Majdak. (2013, Dec.) ARI HRTF Database. [Online]. Available: http://www.kfs.oeaw.ac.at/content/view/608/570/lang,8859-1/

[Mar87] W. Martens, *Principal components analysis and resynthesis of spectral cues to perceived direction.* Proceedings of the 1987 International Computer Music ..., 1987.

[MB84] A. D. Musicant and R. A. Butler, "The influence of pinnae-based spectral cues on sound localization," *The Journal of the Acoustical Society of America*, 1984.

[MBL07] P. Majdak, P. Balazs, and B. Laback, "Multiple exponential sweep method for fast measurement of head-related transfer functions," *Journal of the Audio Engineering Society*, 2007.

[MG92] J. C. Middlebrooks and D. M. Green, "Observations on a principal components analysis of head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 92, no. 1, pp. 597–599, 1992.

[MIC+13] P. Majdak, Y. Iwaya, T. Carpentier, R. Nicol, M. Parmentier, A. Roginska, Y. Suzuki, K. Watanabe, H. Wierstorf, H. Ziegelwanger, and M. Noisternig, "Spatially Oriented Format for Acoustics: A Data Exchange Format Representing Head-Related Transfer Functions," 2013.

[Mid99a] J. C. Middlebrooks, "Individual differences in external-ear transfer functions reduced by scaling in frequency," *The Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1480–1492, 1999.

[Mid99b] ——, "Virtual Localization Improved by Scaling Nonindividualized External-Ear Transfer Functions in Frequency," *The Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1493–1510, 1999.

[MK85] N. A. Macmillan and H. L. Kaplan, "Detection theory analysis of group data: estimating sensitivity from average hit and false-alarm rates." *Psychological Bulletin*, vol. 98, no. 1, pp. 185–199, Jun. 1985.

[MM77] S. Mehrgardt and V. Mellert, *Transformation characteristics of the external human ear.* The Journal of . . . , 1977.

[MM02] E. A. Macpherson and J. C. Middlebrooks, "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," *The Journal of the Acoustical Society of America*, vol. 111, no. 5, pp. 2219–2236, 2002.

[MMG89] J. C. Middlebrooks, J. C. Makous, and D. M. Green, "Directional sensitivity of sound-pressure levels in the human ear canal," *The Journal of the Acoustical Society of America*, vol. 86, no. 1, pp. 89–108, 1989.

[MMO00] J. C. Middlebrooks, E. A. Macpherson, and Z. A. Onsan, "Psychophysical customization of directional transfer functions for virtual sound localization," *The Journal of the Acoustical Society of America*, vol. 108, no. 6, pp. 3088–3091, 2000.

[Møl92] H. Møller, "Fundamentals of binaural technology," *Applied Acoustics*, vol. 36, no. 3-4, pp. 171–218, Jan. 1992.

[Mye89] P. H. Myers, "Three-dimensional auditory display apparatus and method utilizing enhanced bionic emulation of human binaural sound localization," Patent, 1989.

[NC10] K. V. Nguyen and T. Carpentier, "Calculation of head related transfer functions in the proximity region using spherical harmonics decomposition: Comparaison with the measurements and ...," *... and Spherical ...*, 2010.

[NKA08] J. Nam, M. Kolar, and J. Abel, "On the minimum-phase nature of head-related transfer functions," *Proceedings of AES 125th convention*, vol. 7546, 2008.

[NZ11] M. Noisternig and F. Zotter, "On the decomposition of acoustic source radiation patterns measured with surrounding spherical microphone arrays," *old.iem.at*, Mar. 2011.

[PK12] G. Parseihian and B. F. G. Katz, "Rapid head-related transfer function adaptation using a virtual auditory environment." *The Journal of the Acoustical Society of America*, vol. 131, no. 4, pp. 2948–2957, Mar. 2012.

[POM00] J. Plogsties, S. K. Olesen, and P. Minnaar, "Audibility of all-pass components in head-related transfer functions," *PREPRINTS-AUDIO ENGINEERING SOCIETY*, 2000.

[PZ08] H. Pomberger and D. Zotter, "Angular and radial directivity control for spherical loudspeaker arrays," *Thesis*, 2008.

[QE98] J. Qian and D. A. Eddins, "The role of spectral modulation cues in virtual sound localization," *The Journal of the Acoustical Society of America*, vol. 123, no. 1, p. 302, 1998.

[RDS10] M. Rothbucher, M. Durkovic, and H. Shen, "HRTF customization using multiway array analysis," *Proc Europ Signal ...*, 2010.

[Rom12] G. D. Romigh, "Individualized Head-Related Transfer Functions: Efficient Modeling and Estimation from Small Sets of Spatial Samples," 2012.

[Row98] S. Roweis, "EM algorithms for PCA and SPCA," *Advances in neural information processing systems*, 1998.

[RR05a] M. A. Ramirez and S. G. Rodriguez, "HRTF Individualization by Solving the Least Squares Problem," 2005.

[RR05b] S. G. Rodríguez and M. A. Ramirez, "Linear Relationships Between Spectral Characteristics and Anthropometry of the External Ear," *parameters*, 2005.

[SF03] B. U. Seeber and H. Fastl, "Subjective selection of non-individual head-related transfer functions," in *Proceedings of the 2003 Internation Conference on Auditory Display*, 2003, pp. 259–262.

[SGA10] S. Spagnol, M. Geronazzo, and F. Avanzini, "Fitting pinna-related transfer functions to anthropometry for binaural sound rendering," *Multimedia Signal Processing (MMSP), 2010 IEEE International Workshop on*, pp. 194–199, 2010.

[Shi08] K. H. Shin, "Enhanced vertical perception through head-related impulse response customization based on pinna response tuning in the median plane," *IEICE Transactions on Fundamentals of Electronics*, 2008.

[Sil02] A. Silzle, "Selection and tuning of HRTFs," in *Audio Engineering Society Convention Paper 5595*, 2002.

[SL11] R. H. Y. So and N. M. Leung, "Effects of Spectral Manipulation on Nonindividualized Head-Related Transfer Functions (HRTFs)," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 53, no. 3, pp. 271–283, Jun. 2011.

[SMM02] M. A. Senova, K. I. McAnally, and R. L. Martin, "Localization of virtual sound as a function of head-related impulse response duration," *Journal of the Audio Engineering ...*, 2002.

[SN36] S. S. Stevens and E. B. Newman, "The Localization of Actual Sources of Sound," *The American Journal of Psychology*, vol. 48, no. 2, p. 297, Apr. 1936.

[Sot99] R. Sottek, "Physical modeling of individual head-related transfer functions (HRTFs)," *The Journal of the Acoustical Society of America*, 1999.

[ST68] E. Shaw and R. Teranishi, "Sound Pressure Generated in an External-Ear Replica and Real Human Ears by a Nearby Point Source," *The Journal of the Acoustical Society of America*, 1968.

[ST99] H. Stanislaw and N. Todorov, "Calculation of signal detection theory measures," *Behavior Research Methods, Instruments, & Computers*, vol. 31, no. 1, pp. 137–149, 1999.

[TG98] C.-J. Tan and W.-S. Gan, "User-defined spectral manipulation of HRTF for improved localisation in 3D sound systems," *Electronics Letters*, vol. 34, no. 25, pp. 2387–2389, 1998.

[Tol10] D. Toledo, "Technical and Perceptual Issues on Head-related Transfer Functions Sets for Use in Binaural Synthesis," Nov. 2010.

[WI03] G. Wersényi and A. Illényi, "Test Signal Generation and Accuracy of Turntable Control in a Dummy-Head Measurement System," *Journal of the Audio Engineering Society*, vol. 51, no. 3, pp. 150–155, 2003.

[Wil99] E. G. Williams, *Fourier Acoustics*, ser. Sound Radiation and Nearfield Acoustical Holography. Academic Press, 1999.

[WK91] F. L. Wightman and D. J. Kistler, "Localization of virtual sound sources synthesized from model HRTFs," *... of Signal Processing to Audio and ...*, 1991.

[WOI+07] K. Watanabe, K. Ozawa, Y. Iwaya, Y. Suzuki, and K. Aso, "Estimation of interaural level difference based on anthropometry and its effect on sound localization," *The Journal of the Acoustical Society of America*, 2007.

[WRR03] M. Wall, A. Rechtsteiner, and L. Rocha, "Singular value decomposition and principal component analysis," *... approach to microarray data analysis*, 2003.

[Wu97] Z. Wu, "A time domain binaural model based on spatial feature extraction for the head-related transfer function," *The Journal of the Acoustical Society of America*, vol. 102, no. 4, pp. 2211–2218, Oct. 1997.

[Xie12] B.-S. B. Xie, "Recovery of individual head-related transfer functions from a small set of measurements." *Journal of the Acoustical Society of America*, vol. 132, no. 1, pp. 282–294, Jun. 2012.

[XLS09] S. Xu, Z. Li, and G. Salvendy, "Identification of Anthropometric Measurements for Individualization of Head-Related Transfer Functions," *Acta Acustica united with Acustica*, vol. 95, no. 1, pp. 168–177, Jan. 2009.

[XZR07] B.-s. Xie, X.-L. Zhong, and D. Rao, "Head-related transfer function database and its analyses," *Science in China Series G ...*, 2007.

[Zaa11] J. Zaar, "Phase Unwrapping for Spherical Interpolation of Head-Related Transfer Functions," Ph.D. dissertation, iemkug.ac.at, Dec. 2011.

[ZAKD10] W. Zhang, T. D. Abhayapala, R. A. Kennedy, and R. Duraiswami, "Insights into head-related transfer function: Spatial dimensionality and continuous representation," *The Journal of the Acoustical Society of America*, vol. 127, no. 4, pp. 2347–2357, 2010.

[ZDG09] D. N. Zotkin, R. Duraiswami, and N. A. Gumerov, "Regularized HRTF fitting using spherical harmonics," *Applications of Signal Processing to Audio and Acoustics, 2009. WASPAA '09. IEEE Workshop on*, pp. 257–260, 2009.

[ZHDD03] D. N. Zotkin, J. Hwang, R. Duraiswaini, and L. S. Davis, "HRTF personalization using anthropometric measurements," *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on*, pp. 157–160, 2003.

[Zim04] A. Zimmermann, "Eigenschaften des Richtungshörens beim Menschen ," Universität Ulm, Fakultät für Informatik, Abteilung Neuroinformatik, Tech. Rep., Jul. 2004.

[ZK08] W. Zhang and R. A. Kennedy, "Iterative extrapolation algorithm for data reconstruction over sphere," *Acoustics*, 2008.

[ZKA09] W. Z. W. Zhang, R. A. Kennedy, and T. D. Abhayapala, "Efficient Continuous HRTF Model Using Data Independent Basis Functions: Experimentally Guided Approach," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 17, no. 4, pp. 819–829, Apr. 2009.

[ZWK95] P. Zahorik, F. Wightman, and D. Kistler, "On the discriminability of virtual and real sound sources," *Applications of Signal Processing to Audio and Acoustics, 1995., IEEE ASSP Workshop on*, pp. 76–79, 1995.