

# An initial Investigation into HRTF Adaptation using PCA

IEM Project Thesis

Josef Hölzl

Supervisor: Dr. Georgios Marentakis

Graz, July 2012



institut für elektronische musik und akustik



### **Abstract**

Recent research indicates that it is necessary to provide ways of adapting HRTFs to individuals since generalized HRTF models result in perceptual problems and individualized HRTFs are costly to measure. The project investigates HRTF individualization using Principal Component Analysis (PCA) and the related literature on HRTFs is discussed. The process of HRTF individualization using PCA is formalized and a Least Squares feasibility analysis is presented. To this end, existing HRTF databases are analyzed, with a focus on the impact of estimation methodology and database configuration on the coherence of the produced Principal Components (PCs) and Principal Component Weights (PCWs). Some initial thoughts on the development of a system for HRTF individualization are presented together with a user interface that allows the adjustment of the PCWs using the different methods discussed for a number of existing HRTF databases.

## Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
1.1	Psychoacoustic Importance of HRTFs . . . . .	5
1.2	HRTF Properties . . . . .	7
1.2.1	Sensitivity to Phase . . . . .	8
<b>2</b>	<b>HRTF Estimation and Modelling</b>	<b>10</b>
2.1	Why is Individualization necessary . . . . .	10
2.2	Measurement . . . . .	12
<b>3</b>	<b>Modelling</b>	<b>12</b>
3.1	Orthogonal Basis Functions Models . . . . .	15
3.1.1	Principal Component Modelling of HRTFs . . . . .	16
3.1.2	Principal Component Modelling for Anthropometry . . . . .	18
3.2	Conclusion . . . . .	23
<b>4</b>	<b>HRTF Individualization Techniques</b>	<b>25</b>
4.1	Methodologies for Subjective Adaption . . . . .	25
<b>5</b>	<b>HRTF Model</b>	<b>28</b>
5.1	Database Analysis . . . . .	28
5.1.1	PCA Compression Efficiency . . . . .	30
5.1.2	Correlation of PCs . . . . .	32
5.1.3	Variation of PCWs . . . . .	34
5.1.4	Least Squares Reconstruction of HRTFs . . . . .	35
5.2	Evaluation of Input Matrices . . . . .	46
5.3	Methodology . . . . .	47
5.4	Self Tuning of PCWs . . . . .	51
5.5	HRTF Tuning Tool . . . . .	53

<i>J. Hözl: HRTF Adaptation</i>	4
<b>6 HRTF Analysis Tool</b>	<b>55</b>
6.1 Basic Operations . . . . .	55
6.2 PCA Operations . . . . .	57
6.3 Visualizing Correlations . . . . .	58
<b>7 Conclusion</b>	<b>58</b>
<b>A Principal Component Analysis (PCA)</b>	<b>60</b>
A.1 Covariance Matrix . . . . .	61
A.2 Singular Value Decomposition (SVD) . . . . .	62

# 1 Introduction

*Head-related transfer functions (HRTFs)* describe the acoustic transmission path from a sound source to the left and right ear. The HRTF is defined as the ratio of sound pressure in the ear of a subject in relation to the sound pressure level if the subject is not there [RD05]. They are useful both for understanding the perceptual mechanisms of hearing and reproduction of virtual spatial sound scenes. These functions are different for each person. Individualized HRTFs provide an accurate representation of the sound pressure due to a three-dimensional sound field at the eardrum using headphones.

One way to obtain individual HRTFs is to measure it. However, this conventional method is cost-intensive and time consuming, because expensive equipment and expert knowledge in acoustical measurement is necessary. Typically, a measured HRTF set includes a large number of impulse responses, measured for each specific direction of interest. Due to the high spatial resolution that is necessary for good localization accuracy, this process is normally time-consuming.

Because of this reason, there has been an effort to develop generalized HRTFs. This however did not yield the expected results. Although in several experiments ([Shi08],[HPP10]) it was confirmed that individualized HRTFs result in localization performance comparable to free-field localization, when generalized HRTFs were used, poor performance emerged, expressed primarily by poor localization and a large number of *front-back* and *up-down confusions* [KW92].

Recently, therefore there is a strong research trend for a convenient customization technique based on existing HRTF databases. Nowadays, a common method to improve localization with non-individual HRTFs is to individualize them by subjective selection, scaling or grouping. Many studies investigated in the approach of spectral manipulation, first of all *Middlebrooks* [Mid99b] [Mid99a], *Wightman and Kistler* [KW92] [SL11] and *Hwang and Park* [HPP10] [HPP08]. An alternative research direction sought to individualize HRTFs by modelling their behavior in relation to anthropometric parameters, such as *head width*, *shoulder width*, *pinna offset* and so on. Some studies ([XLS09], [Rod05]) have found linear relationships of anthropometric dimensions with *spectral cues*, such as pinna spectral peaks or notches that influence spatial hearing.

In this spirit, this project intends the research and develop a manageable HRTF customization method based on subjective tuning of PCA basis functions. To this end, related HRTF literature is reviewed and the particular method is investigated in order to understand its capabilities and limitations. The results of this project could be applied to the processing of tunable non-individual HRTFs for 3D audio rendering.

## 1.1 Psychoacoustic Importance of HRTFs

HRTFs provide localization cues for spatial hearing in virtual auditory display. While localization in azimuth plane can be simply modeled by ILD and ITD, localization in elevation plane and discriminating between front-back sounds is more complex and subject-dependent. Obviously, ITD and ILD are also subject-dependent, but these cues can be

modelled easily using physical dimensions, such as head size or radius.

The perception of the location of sound source can be evaluated in terms of the accuracy of the perception of its direction, the degree of its externalization and the focus. Externalization means that a sound event is localized outside the head, like in real world situations. The term focus is associated with the correct reproduction of the spatial extent of a sound stimuli.

Different times of arrival on both ears occur when a sound source is not located in the median plane. Low frequencies bend around the head, a phenomenon that is called diffraction, whereas high frequencies are reflected. According to Rayleigh's duplex theory of localization [Che99], this results in *interaural time difference (ITD)* and *interaural level difference (ILD)* between the two ears, which are used by the brain to estimate the azimuth of a sound source. ILD is defined as the difference in the energy between each ear, and is dominating at frequencies higher than 1.5 kHz (Figure 1a) because in this frequency range, the contralateral ear is shadowed by the head and less sound energy reaches this ear. According to Blauert [Bla83], the smallest perceptible threshold (*just-noticeable difference, JND*) is about 1 dB. Thus, the ILD is a key parameter for horizontal localization above 1.5 kHz. Similarly, interaural time difference describes the time difference between the incoming sound wave on the right and left ear (Figure 1b). In addition, important is the time difference between the signal envelopes in high frequencies. The ITD can be modelled as

$$\Delta(T) = \frac{r}{c_0}(\theta + \sin \theta) , \quad (1)$$

with head radius  $r$ , angle of incidence  $\theta$  of the sound and  $c_0$  as the speed of sound [Kuh77]. This approximation holds for low frequencies only. If the ITD is longer than a wavelength, it can not be assigned to a unique angle. Based on the absolute refractory period of neurons, this mechanism is limited to 1.5 kHz. According to Kyriakakis [Kyr98], human can discriminate time differences in the order of  $7\mu s$ .

The duplex theory fails to explain localization in an open field where sound sources can be emitted from every imaginable source position (azimuth and elevation). These positions form the *cone of confusion* (same ITD) or *torus of confusion* (same ILD) [Mid99b]. Based on the duplex theory, elevation perception and front-back discrimination cannot be achieved. However, incoming sound is spectrally colored by the physical structure of a listener, such as head, pinna, shoulder and torso [Che99]. Thus, localization in elevation strongly depends on pinna (above 3 kHz) and head/torso response (below 3 kHz). Unlike the ITD and ILD, these spectral cues are effective for monaural as well as binaural directional sounds [SNH<sup>+</sup>10].

In real environments, subjects normally resolve front-back and up-down ambiguities by changing the head position [Bla83]. New localization cues about the sound stimuli is obtained and the spectral details, such as peaks or notches give information about the source position.

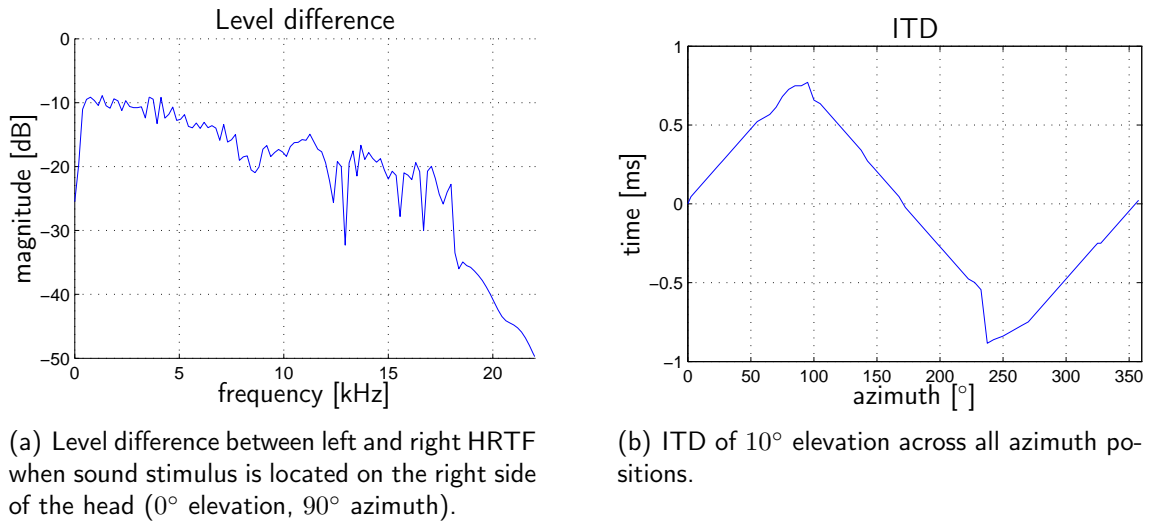


Figure 1: Level difference and interaural time difference (ITD) between left and right ear of subject ID 103 in ARI database.

## 1.2 HRTF Properties

The impact of the presence of the body, head and outer ear is to cause peaks and notches in the HRTF frequency response. These spectral cues are mostly contained in the first 0.7ms of the HRIR [BD98]. The impact of the different components can be extracted from the HRIR, if the reflections caused by shoulders, torso and pinna are separated. Spectral cues occurring at middle and low frequency bands ( $< 3\text{kHz}$ ) are mostly due to head diffraction and torso reflections [Shi08]. Pinna generally influences frequencies in the range of 2-14 kHz [Gie92] and becomes dominant above 5 kHz [RD05].

In addition, interference of direct and reflected sound waves in rooms result in sharp notches in the spectrum (*pinna spectral notches*) with a periodicity that is inversely proportional to the time delay [SGA10].

Psychoacoustic experiments have proved that the location of the spectral peaks and notches in frequency are closely associated with vertical localization and front-back discrimination. Blauert [Bla70] found out, that energy in specific frequency bands (280-560 Hz, 2.9-5.8 kHz) result in frontal perception whereas others (720-1800 Hz, 10.3-14.9 kHz) tend to be perceived as coming more from the back. These are called directional bands. Myers [Mye89] focused on amplifying and attenuating energy in four frequency bands within a range between 6 and 12 kHz (boosting bands with center frequencies 392 Hz and 3605 Hz, attenuating bands with center frequencies 1188 Hz and 10938 Hz) and filed a US patent. Tan and Gan [Tan98] continued Myers work and implemented a set of parallel filters based on five frequency regions ( $f_1$ : 225-680 Hz,  $f_2$ : 680-2000 Hz,  $f_3$ : 2-6.3 kHz,  $f_4$ : 6.3-10.9 kHz,  $f_5$ : 10.9-22 kHz). For frontal localization  $f_1$ ,  $f_3$  and  $f_5$  were amplified whereas  $f_2$  and  $f_4$  were attenuated for backward perception. However, they did not describe the full experiment and statistical analysis.

On the contrary, Hebrank and Wright [Heb74] reported that removing frequencies below

3.8 kHz did not affect forward nor backward perception. Musicant and Butler [Mus84] confirmed this thesis. Bronkhorst [Bro95] reported that localization is not affected by sound energy at frequencies above 9 kHz. Unlike to Blauert and Hebrank, he used headphone instead of loudspeakers for his experiment.

The exact determination of the frequency bands that determine elevation perception and front-back discrimination has not been established. This might also be due to the observation by Langendijk and Bronkhorst [LB02], that cues from different bands interact with each other. Through systematic manipulation of the spectrum, Langendijk and Bronkhorst investigated the contribution of spectral cues to human sound localization. They found out that the up-down cues are essentially in 1-octave band from 6 to 12 kHz and the front-back cues are coded mainly in the high 1-octave band from 8 to 16 kHz [QE98]. However, studies are still in disagreement which exact frequencies have an effect on the localization and it is not yet explored, which peaks are really important [Bla83].

Kulkarni and Colburn mentioned that a smoothing of the HRTFs by reducing the Fourier coefficients in the reconstruction of the magnitude spectrum did not significantly affect localization performance, even when the coefficients were reduced from 256 to 16 [KC98]. Furthermore, in a test focusing on azimuth location, all subjects reported complete externalization of the sound stimuli.

Middlebrooks [MMO00] introduced a new domain called *Spectral modulation frequency domain (SMF)* with units of cycles per octave. It describes the Fourier transform of the frequency spectrum. Thus, smoothing of the HRTF in frequency domain is similar to a low-pass filtering in modulation frequency domain. It was shown that low-pass filtering at 2 cycles/octave did not significantly influence localization accuracy. From this, Middlebrooks concluded that the major cues for sound localization are in the SMF region below 2 cycles/octave.

### 1.2.1 Sensitivity to Phase

In order to study HRTF phase it is usually decomposed into three components: a minimum-phase and an excess-phase component represented by a constant time delay:

$$H(e^{jw}) = H_{min}(e^{jw}) \cdot H_{ep}(e^{jw}), \quad (2)$$

with the excess-phase function modeled as a linear phase term and an all-pass:

$$H_{ep}(e^{jw}) = H_{lp}(e^{jw}) \cdot H_{ap}(e^{jw}). \quad (3)$$

Commonly, the excess-phase function is modelled as a linear phase term only:

$$H(e^{jw}) = H_{min}(e^{jw}) \cdot H e^{-jw\tau_{ap}}. \quad (4)$$

The minimum-phase assumption allows to specify the phase of a HRTF by its magnitude



response only. Through Hilbert transform, the logarithmic magnitude frequency response and phase response of a minimum-phase causal system are connected. In general, a minimum-phase system can be constructed by finding the roots of a polynomial that are outside the unit circle in the  $z$ -plane and mirroring them on their reciprocal positions. Basically, the all-pass term with the same order as the original function is obtained by dividing the original function by its minimum-phase version in frequency domain. However, the aim is to get an all-pass term that contains only low-order filters, therefore each complex conjugate pair of poles and zeros correspond to a second order section and each real pair correspond to a first order section [POM00].

Several researchers investigated the minimum-phase nature of HRTFs. Mehrgardt and Mellert [MM77] claimed that HRTFs are nearly minimum-phase up to 10 kHz whereas Nam *et al.* [NKA08] pointed out that HRTFs are essentially minimum-phase. The maximum of cross-coherence between HRIRs and minimum-phase versions was used to evaluate the similarity between the signals. They found out that the values for coherence are above 0.9, which means that the biggest part of HRTF energy is minimum-phase. HRTFs in frontal and ipsilateral directions tend to have coherences below 0.9, so in these directions the HRTFs had non-minimum-phase zeros in high frequency regions but almost none of them below 8 kHz. Nam suggested to model HRTFs by pure delays followed by minimum-phase filters.

Kulkarni and Colburn [KC98] examined whether HRTF decomposition into a minimum phase and an all phase component results in HRTFs that can not be distinguished from measured ones. The remaining all-pass term is ignored because the auditory sensitivity to the absolute phase spectrum is low. Therefore the ITD can be assumed as frequency independent time delay and is calculated as the difference of the group delays of left and right ear and rounded to integer samples. The validity of the model was confirmed by subjective testing.

Kistler and Wightman [KW92] compared real HRIRs with ones that were synthesized using the assumption of minimum-phase. Results from the two conditions were almost similar. This means that the phase of synthesized HRTFs can be calculated by a combination of minimum-phase functions and a pure time delay without loss of spatial hearing.

Also Plogsties *et al.* [POM00] found out that the all-pass component of HRTFs can be removed without audible consequences. In an three-alternative forced choice (3-AFC) experiment the audibility of the all-pass component was tested. For some HRTFs, the absence of the all-pass term was detected. In this case, he suggested to replace the all-pass components by pure delays that are calculated as the group delays of the all-pass terms at 0 Hz. The interaural time difference was modelled as the interaural group difference evaluated at 0 Hz calculated from the excess phase.

Lindeau [LEW10] realized that using Hilbert transform for calculating the phase term could alter the impression of sound source distance. Another common ITD extraction method is onset or leading edge detection by finding the sample where the impulse response exceeds 5% of the maximum value. However, it must be ensured that cutting into the rising edge of impulse responses including a larger SNR, such as the ipsilateral impulse response, is avoided.

Typically HRTFs are implemented as a cascade of a pure delay term and a minimum-phase filter. This model has some advantages. Firstly, the length of the FIR filter can be shortened, because the main energy occurs at the beginning of the impulse response. Secondly, smooth interpolation of HRIRs when simulating moving sounds can be better implemented with minimum-phase filters. Several subjective hearing tests confirmed that the all-pass term can be neglected in almost all cases, thus minimum-phase HRTFs with a pure delay as ITD are now widely applied.

## 2 HRTF Estimation and Modelling

HRTFs are usually decomposed into a directional and a non-directional part, which are called *Directional transfer function (DTF)* and *Common transfer function (CTF)* or *Average transfer function (ATF)* [LB02], respectively. The latter is mainly influenced by the resonance of the ear canal and includes only the diffuse part. It can be calculated by averaging the relevant (logarithmic or linear depending on the representation used) spectrum of the HRTF  $\underline{H}_{1...N}(f)$  across all source positions:

$$C(f) = \frac{20}{N} \sum_{i=1}^N \log|\underline{H}_i(f)| . \quad (5)$$

Hence, the DTF can be obtained by subtracting the CTF from the HRTF. In literature, the calculation of the DTF is slightly different, for example Middlebrooks [Mid99b] took the root mean square of the HRTF to calculate the CTF. Nevertheless, the resulting directional function primarily consists of the direction-dependent spectral parameters of the HRTFs.

### 2.1 Why is Individualization necessary

When using generalized HRTFs, occurring ambiguities may not always be solved properly. Compared to binaural simulation, head movements can not affect directional sounds due to the fact that the listener is wearing a headphone. This results in localization error, especially in elevation plane.

As shown in Figure 2, HRTFs are unique for each individual. Several parameters determine this difference: the location of the ears, the differences between the two ears, the size and shape of the head and the size and shape of the torso. Each ear has its own transfer function, thus it is not symmetric even in the median plane (asymmetric, especially from 5 kHz) and unique for each source position. In the course of life, HRTFs change because of body changes. Nevertheless, there is a certain plasticity in the auditory system, that helps it to tune to these changes. This also explains the finding that we can be gradually accustomed over time to HRTFs others than our ones [Bla83]. Mendonca [CM10] dealt with the issue of learning and found out that there is a learning

effect after a period of time, although this a slow and sometimes inconvenient process that has not been thoroughly studied.

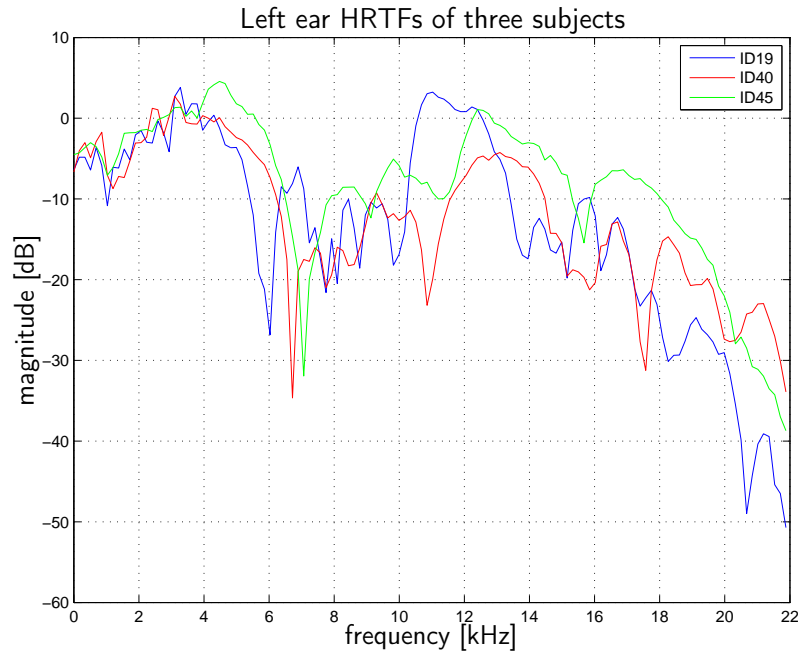


Figure 2: Left ear HRTFs of one source position ( $0^\circ$  elevation,  $25^\circ$  azimuth) of three different subjects (ID19, ID40 and ID45) in CIPIC database.

Because the spectral cues contained in HRTFs are significant for localization, localization errors occur if a non-individual HRTF does not match because of the existing diversities. Several studies investigated in this issue ([LB02], [SL11], [SF03]).

Since the effort required for the measurement of individual HRTFs is significant, there has been an effort to use generalized HRTFs for reproduction in binaural setups (e.g. headphones). This is usually an averaged HRTF, that is calculated from an HRTF database. The perceptual consequences are lack of presence, that sounds appear "in head" instead of being externalized, poor localization and a large number of front-back and up-down confusions. Wenzel examined how the perception of synthesized HRTFs changed over time [Wen88]. However, it is important to mention, that even in real environment some confusions (up-down or front-back) can occur. In an experiment ([WAK93]), the accuracy of localization in free-field was compared to the virtual stimuli (over headphone) using a generic HRTF set from a study by Wightman and Kistler [WK89]. 16 subjects evaluated 24 source positions. Even in the free-field condition, the mean error rate for front-back and up-down confusions was 19% and 6% respectively. When using synthesized HRTFs, the mean error rate significantly increased to 31% and 18%.

One of the main causes of degradation in localization in virtual sound reproduction is an unmatched head size [Xie02]. Confusions and degradation in elevation perception occur in the process of averaging. The spectrum is smeared, because the location of the peaks and notches of subjects occur at different magnitude and at different frequencies.

Hence, the spectral variance of human HRTFs is too large to enable a simple averaging model to work.

## 2.2 Measurement

In order to obtain HRTFs of a listener, HRIRs are measured and stored as a pair of head-related impulse responses. Typically HRIRs are measured at a certain fixed radius in relation to the head. To achieve high spatial resolution and high fidelity reproduction, a large number of impulse responses has to be measured. In existing databases, the number of measured source positions varies from 24 to more than 1500 for each subject. The measurement of HRIRs is very time consuming and costly, as each acoustic transmission path must be measured individually. An anechoic room and high quality equipment is required. In addition, specialized knowledge in the preparation of the microphones and speakers is needed. It is recommended to use a software tool that operates and controls the whole measurement process in order to minimize errors and speed up predefined tasks.

For the sake of completeness, it should be mentioned that a method for reciprocal measurement already exists ([ZDG06]). The sound coming from a small loudspeaker in the ear canal is measured by a surrounding microphone array. In [Zaa10], the measurement time of 64 source positions could be reduced to only one minute.

When measured, it is easy to play a sound event at a desired source position through headphones, by convolving the sound with the measured HRIRs of the left and right ear at the desired location:

$$\begin{aligned} s_{left}(t, \theta, \phi) &= s_{stimuli}(t) * hrir_{left}(t, \theta, \phi) , \\ s_{right}(t, \theta, \phi) &= s_{stimuli}(t) * hrir_{right}(t, \theta, \phi) , \end{aligned} \tag{6}$$

with  $hrir_{left/right}(t, \theta, \phi)$  as the left and right ear HRIR of a source position (azimuth  $\theta$  and elevation  $\phi$ ),  $s_{stimuli}(t)$  as the monaural stimuli and  $s_{left/right}(t, \theta, \phi)$  as the resulting signal at the source position.

As part of the process and in order to analyze and validate the tuning process that is described in Section 5, my own head-related transfer functions were measured at Acoustic Research Institute (ARI) in Vienna (Figure 3).

## 3 Modelling

A large number of methods to model HRTFs have been proposed. These can be categorized in signal models, anthropometric models and orthogonal basis function expansions. Signal models attempt to provide a simplified transfer function (plus time delay), characterized by complex poles and zeros. Most frequently, these model the diffraction effects

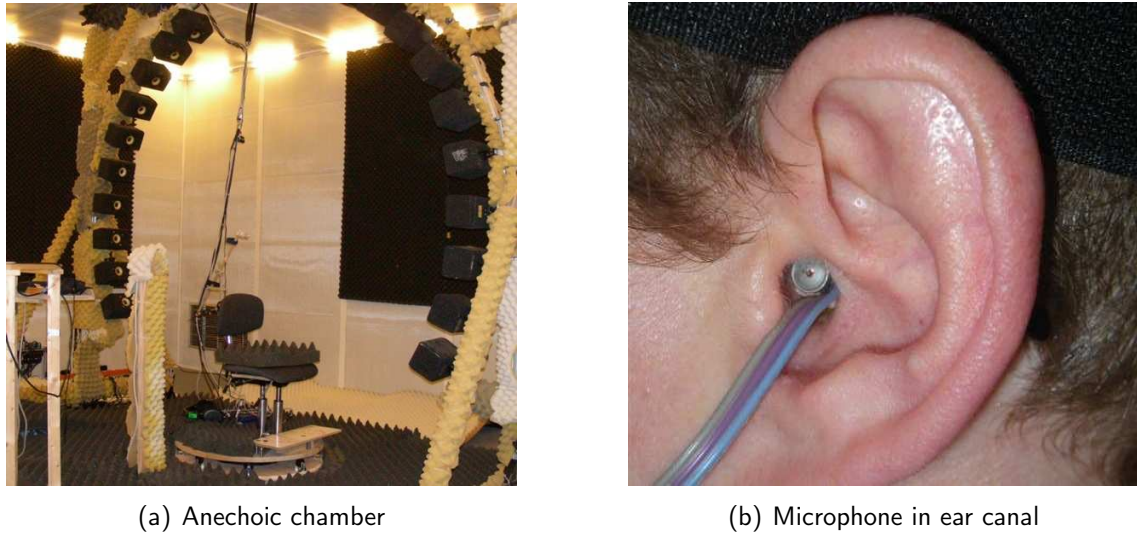


Figure 3: Measurement setup at Acoustic Research Institute (ARI) in an anechoic chamber. The chair rotates automatically during the measurement.

of the head (e.g. head-shadow filter). Thus, only azimuth effects can be modelled convincingly. To ensure localization in the elevation plane as well, a changing time delay in a monaural model can be added. This could produce a spectrum with a moving notch and consequently the effect of a vertical motion [BD98]. A low-order IIR representation has been proposed by Kulkarni and Colburn [KC04]. Asano *et al.* [AS90] introduced a model with 40 poles and zeros. Quite commonly, linear prediction theory (LPC) or weighted-least-squares (WLS) is applied to eliminate artifacts from windowing and extracting the poles respectively. The reduced representations ease the physical interpretation. The peaks and notches in the DTF of an external ear could be characterized by using a pole-zero model. Kulkarni used an all-pole model (position-dependent) and the coefficients were estimated using the autocorrelation method for linear prediction, where the magnitude in the shadowed ear (more than  $60^\circ$  from the midline) was smoothed.

The phase is usually not considered in detail and instead a minimum phase, plus time-delay plus all-pass time delay is used. The minimum-phase can be obtained by the Hilbert transform of the logarithmic magnitude spectrum. Secondly, the all-pass filter of HRTFs can be modelled as a simple time delay (at least up to 10 kHz) [HF98].

Anthropometric HRTF models try to identify and explore the relationship between the HRTF spectral properties and the shape of the ear. Several models deal with the problem of individualization by using anthropometric data ([Rod05], [Mar10]). Although there is evidence that anthropometric data can explain certain properties of the HRTF, a direct relationship has not been established yet. In order to model HRTFs based on anthropometric data, accurate measurements of the pinna, external ear, torso and head are necessary. Table 2 provides an overview of the physical dimensions used in the CIPIC database. The ARI (Acoustic Research Institute) database includes the same parameters and few more that have been considered to be important.

In anthropometric models, the *pinna-related transfer function (PRTF)* is of interest. PRTFs isolate the influence of the pinna ear and can be obtained by extracting the relevant part of the HRIR. This is not always easy. Raykar et al. [RDD03, RD05] proposed that reflections of shoulder and torso, can be eliminated by using a right half-sized Hanning window with a length of about 1 ms. Geronazzo and Spagnol [GS10] described an algorithm to separate reflections and direct sound from the PRTF. Thus for reconstruction, the influence of each physical phenomenon can be modeled separately. The original PRTF could be re-synthesized through a low-order filter model. Upon isolating the PRTF, the resonant and reflective influence of the pinna are identified by using residue computation and multi-notch filter parameter search. The PRTF magnitude spectrum is compensated iteratively until no significant notches are left. Then a low-order filter estimates and synthesizes the PRTF. In [SGA10], the authors continued the investigation in the extraction of the most important notches and their relation to anthropometry. Using the CIPIC database, three major resonances at 4, 7 and 12 kHz could be identified. A notch tracking algorithm was used to exploit the most distinct spectral notches. Then a structural model was formed with two resonances and three notches (resulting in an eight-order global filter) by two filter blocks respectively. Finally a bandpass suppressed undesired frequencies.

Wenzel *et al.* [Wen88] found a relationship between a listeners accuracy in vertical localization and the characteristics of a listeners external ears. Begault *et al.* [Beg94] pointed out that concha and fossa of the helix are major parameters for localization at high frequencies. According to Satarzadeh *et al.* [DS07], only depth and effective width of the ear are crucial for modelling the PRTF. On the contrary, Rodriguez *et al.* [Rod05] identified even more dimensions (cavum concha height, fossa height, pinna height, and pinna rotation angle) that are closely related to the PRTF. Whereas spectral features of the pinna are dominant above 5 kHz, head diffraction and torso reflections are providing additional cues in lower frequency regions. Some studies ([BD98], [XLZ07]) investigated the importance of various anatomical structures in relation to sound localization.

Based on the shape of the pinna, numerical methods have been used to model HRTFs [XLS07]. In such methods, the physics of wave propagation and diffraction are modelled [BD98]. Quite often the models do not include the influence of HRTFs and are rather simplified. Spherical head model, snowman model and ellipsoidal head model have been developed in this direction. The models are often accompanied by a simplified filter representation, where the filter parameters of the model could be related to certain anthropometric dimensions. Currently, there are still issues to measure the pinna accurately. Even small deviations in measurement can cause large errors in modelling. However, if there is a precise measurement, this approach can have good results [Sot99]. Despite high performance computing, there is still a gap between measured and synthesized HRTFs, so these methods need to be improved and compared with existing HRTF databases. The accuracy of the model always strongly depends on the measured data used by the model. In general, a large amount of samples for a specified position in a three-dimensional space is essential in order to synthesize this source position. This goes with large memory size and high-performance equipment. In order to save memory, it is a common way to model the HRTFs using only few distinct parameters that are relevant

for spatial hearing.

Raykar *et al.* [RDD03] decomposed the HRTF into different components and extracted features that could be perceptually relevant for sound source localization. By relating these features to the anthropometry, it could be possible to model HRTFs for any subject and source position. Linear prediction (LP) analysis was applied to extract the poles of the HRIR. For any given azimuth, a 2D image (HRIR or HRTF) of all elevations was calculated from the CIPIC database. Three distinct features could be marked by visual inspection: ITD with head diffraction with pinna effects, torso reflections and knee reflections since the subject were measured seated. In order to isolate various spectral peaks and notches, poles were extracted by *Linear Prediction Analysis (LP)*. These poles correspond to the resonances of the pinna. The results were in agreement of the thesis by Shaw [Sha97] describing six modes of resonance.

### 3.1 Orthogonal Basis Functions Models

Orthogonal basis modelling model HRTFs on a set of orthogonal basis functions. *Principal component analysis (PCA)*, *Spherical Basis Functions* and the *Karhunen-Löve expansion* has been used for this purpose. Because a PCA model is used in this work, it is explained further below.

Models using spherical harmonics try to express each bin of the HRTF magnitude spectra in the form of a weighted sum of surface spherical harmonics (SSHs). The HRTF set is decomposed into spherical basis functions in spatial domain. These are functions that are orthogonal upon a surface of a sphere. Since the resulting spherical harmonics representation is continuous, any location in space, not just the source positions where measurements were obtained, can be estimated. Consequently, this model is well suited for synthesizing smooth auditory motion. Evans *et al.* [EAT98] proposed a model that allows a direct, continuous, and accurate synthesis of a pair of HRTFs for any arbitrary direction. Similarly, Zotkin [ZDG09] presented a method that works well for an arbitrary grid.

Principal component analysis is a powerful statistical technique that is used reduce the dimensionality in a correlated dataset and simplifying it while keeping the important information. In this section, several studies based on PCA compression are discussed as these are extensively used to form the model we will work on.

Generally, PCA is an analysis of variance. Directional information is highlighted whereas redundant information is neglected. The resulting data is a set of orthogonal principal components that are sorted according to their variance in the original data. The mathematical approach can be found in the Appendix A.

When arbitrary HRIRs and HRTFs are modelled, PCA is used to estimate the principal components that are later used as a set of general basis functions. Synthesized HRTFs can be generated from individual components through a linear combination of basis functions and they can reconstruct any HRTF in the dataset. PCA has been done both on plain HRIRs and HRTFs. When applied to HRTFs, a minimum phase model is

required for reconstruction. Across all subjects, a tendency for variation of weights as a function of azimuth and elevation was found and most studies claimed that the first 3-5 basis functions provide information about front-back discrimination of a source position.

### 3.1.1 Principal Component Modelling of HRTFs

Martens [Mar87] presented a model that is based on PCA and minimum-phase reconstruction. The data set included critical-band-filtered HRTFs of 36 source positions in the horizontal plane of two subjects. He used only 4 of 24 components for reconstruction to explain 90% of the data set. Unfortunately no experiment for perceptual evaluation was carried out.

Kistler and Wightman [KW92] analyzed HRTFs of 10 subjects using PCA and showed that more than 90% of HRTF variance can be approximated with only 5 of 150 basis spectral functions. For PCA, a matrix  $\mathbf{X}_{\text{input}}$  ( $5300 \times 150$ ) was composed by including 10 subjects, 265 source positions and corresponding DTFs of both ears. Only the 150 log magnitude points in a specific frequency region (0.2 - 15 kHz) were used for analysis. It is worth mentioning that the resulting basis function are almost close to zero below 2-3 kHz. This points out that the DTFs have almost no direction-dependent variation in this frequency range. For reconstruction, HRIR were assumed to be minimum-phase functions and ITD was approximated by a constant time delay. This interaural delay was estimated by calculating the maximum of the cross-correlation function of measured left and right HRIRs. The model was validated, in an experiment with 5 listeners. Five different conditions were compared. In the first, "baseline" condition, the measured HRTFs were used. In the control condition, the original HRTF magnitude spectrum reconstructed using minimum-phase was used. In the other conditions the HRTF were reconstructed using one, three or five PCs, respectively. There were 10 runs, including 36 pseudorandomly selected positions, for each condition. In one run, each stimulus was repeated 8 times with 300ms pause between. After listening to a stimulus, the subject reported verbally the perceived azimuth and elevation and the experimenter recorded the values on keyboard. Before testing, the headphone transfer function was compensated. In contrast to other studies in which confusions were "resolved" or inspected separately ([WK89]), the raw data including confusions was analyzed. Results from control condition were almost similar to the baseline condition. This means that the phase of synthesized HRTFs can be calculated by a combination of minimum-phase functions and a pure time delay. Moreover judgments on the horizontal plane were accurate even when only the first PC was used for reconstruction. The first principal component therefore contains most of the interaural intensity information necessary for lateral discrimination. Front-back and up-down performance dramatically decreased when using less than five PCs. Basis functions 2-5 turned out to be crucial parameters for front-back and up-down discrimination. So when using less than five components, the fine spectral details are not accurately represented by the model, therefore subjects are not able to distinguish properly between front-back and up-down.

Middlebrooks et al. [MG92] also examined HRTF reconstruction using PCA with a minimum-phase plus time-delay model. First, he verified if PCA of different data sets



give almost the same results. Because in theory, differences in measurement setup and subjects should have only little effect on the resulting components, he compared own measured HRTFs (8 subjects, 360 source positions) with the database by Kistler and Wightman (10 subjects, 265 source positions) [KW92]. A high correlation between the two different sets of basis vectors was found (PC1: 96%, PC2: 88%, PC3: 70%). This reveals that PCA is relatively robust to different measurement techniques. In addition, Middlebrooks divided the data into two groups according to their physical size. He found out, that basis vectors of smaller subjects were shifted systematically to higher frequencies. However, no evaluation was carried out.

Qian and Eddins [QE98] investigated in the importance of the spectral modulation frequency (SMF), that is the Fourier transform of the frequency spectrum and can be considered as the rate of change in HRTFs. They were able to find SMFs that are critical to sound localization. First, PCs in SMF domain were analyzed, then a localization experiment by comparing original and manipulated HRTFs was conducted. PCs for each ear of each HRTF set (including 360 source positions) were derived. To obtain the spectral modulation characteristics, Fourier transform of all PCs was calculated. This results in FTPCs (FT of the PCs), presented in spectral modulation frequency domain. It was shown, that only seven FTPCs can present 99% of the total variance in the data set. The main energy of these components is located in the lower regions, below 2 cyc/oct. Moreover, the first FTPC contains a prominent peak at the first SMF across all HRTF sets. Prior to experiment, an HRTF customization procedure has been applied, because non-individualized HRTF sets were used. After preselecting 6 best matching HRTFs from 26, a single best matching HRTF and a proper scale factor was found. Both subjective and objective criteria were used. In the experiment, specific regions in the SMF domain were filtered (by applying notch and low pass filters) or enhanced. For reconstruction, the original phases of the database were used because the modifications were limited to the spectral cues. 10 normal hearing subjects passed seven different conditions (one baseline and six HRTF-modified conditions). In the test session, 10 judgments were made for each of the 72 directions. The subjects identified the direction of the stimuli by using a mouse in a graphical interface. Azimuth localization was quite accurate across all conditions when front-back confusions were resolved (average front-back confusion rate was 31%, standard deviation 7.9%). Up-down confusion rates were relatively low (10.8%, s.d. 4%), therefore these errors were not resolved. Localization performance in elevation was mainly effected by notch filters in the SMF domain between 0.1 and 0.4 cyc/oct and 0.35 and 0.65 cyc/oct. However, low pass filtering had little effect on elevation localization. In summary, low regions in spectral modulation frequency domain can be associated with sound localization at low elevations.

Instead of using HRTFs for PCA processing, Hwang and Park used HRIRs to model arbitrary impulse responses. In [HPP10], they performed PCA on median-plane HRIRs in CIPIC database. Prior to that, HRIRs were time-aligned by removing the interaural time delay (ITD) between two related impulse responses. The maximum of the cross-correlation function between two HRIRs indicates the time delay. Ear-symmetry was assumed, so only the left-ear HRTF was modelled and the right-ear channel was driven by the same signal. Due the fact that they focused in the inter-subject variations in

PCWs, the PCW can be modelled as a simple function of elevation, because a common elevation dependency of PCW across all subject exists. Static source positions as well as spatially continuous HRIRs in the median plane were used for customization. Subjective listening tests were carried out to evaluate the model. Nine subjects customized both stationary and moving sounds. For seven of nine subjects there was no statistically significant difference between using measured or customized HRIRs. Contrary, localization performance for all probands increased when using individual instead of KEMAR HRIRs. The results when presenting moving sounds were about the same.

In [HPP08], they took the approach from above to model HRIRs with 12 PCs resulting in modelling error bound of 5%. Again, only the median plane HRIRs in CIPIC database were involved. In addition, the individual HRIRs of six male subjects were measured using different measurement conditions and source positions. A short localization test at 9 positions in elevation plane was conducted to evaluate the performance of the model. Headphone transfer function was equalized and each source position was presented 10 times, including 4 different conditions (measured, 12, 8, 4 PCs). The subject located the position in a graphical interface by changing a slider bar with a resolution of 1 degree. There was no statistically difference in localization performance of the first and second condition (measured vs. reconstruction with 12 PCs). When using 12 PCs, the average error was about 15% but the spectral features in the measured set were substantially reproduced. This was concluded by visual inspection of 2D diagrams of all log-magnitude HRTFs. Again, the error increased dramatically as the number of PCs was reduced. Although using two different datasets for modelling and measurement, the empirical mean does not contribute to the localization performance, but it has an effect on the modelling error.

Chen *et al.* [CvVH93] proposed a model that consist of a set of eigenfunctions which are formed by using the *Karhunen-Löve expansion (KLE)* or HRTFs. This expansion is used to present data in a low-dimensional space. The model consists of weighted combinations of the basis vectors, but in contrast to PCA, the complex valued eigentransfer functions of KLE generate HRTF magnitude *and* phase. Focusing on the time domain, Wu *et al.* [Wu97] calculated the KLE of HRIRs and proposed a low-computational model. However, these two models have been tested only on an anesthetized live cat.

### 3.1.2 Principal Component Modelling for Anthropometry

The general methodology in these models is the connection of anthropometric parameters with principal component weights. The calculation of the correlations is performed either for each individual target position or for the entire database. Since some individual dimensions do not reveal any strong correlation, the result can be improved by using linear regression. The final goal of all these studies is the estimation of HRTFs using only physical parameters.

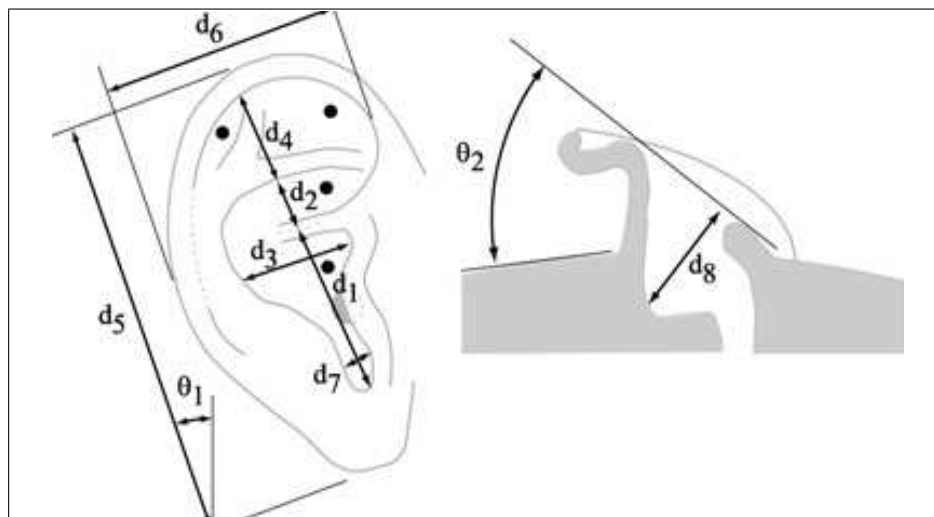
Rodriguez and Ramirez [Rod05] [Ram05] showed that the first five principal component weights (PCWs) of each source position were correlated with the anthropometric data in a way, but the resulting correlation coefficients were low, consequently they were not

Variable	Calculation
$d_{11}$	$d_1 + d_2$
$d_{12}$	$d_1 + d_2 + d_4$
$d_{13}$	$(d_1 + d_2) * d_3$
$d_{14}$	$(d_1 + d_2) * d_3 * d_4$
$d_{15}$	$(d_1 + d_2) * d_3 * \theta_2$
$d_{16}$	$(d_1 + d_2) * d_7 * d_8$
$d_{17}$	$(d_1 + d_2) * d_7 * \theta_2$
$d_{18}$	$d_1 * d_3$
$d_{19}$	$d_5 * d_6$
$d_{20}$	$d_5 * d_6 * d_8$
$d_{21}$	$d_5 * d_6 * \theta_2$
$d_{22}$	$d_4 * d_6$

Table 1: Multidimensional linear regression with existing anthropometric data [Rod05].

well correlated. Because pinna relevant dimensions are better related to the transfer function of the pinna, PCA was performed on 64 PRTF sets and the resulting weights were correlated with existing physical dimensions. The best predictors turned out to be fossa height, pinna flare angle and pinna width. Rodriguez suggested that combinations of anatomical features could provide a better match to HRTFs. Twelve extra parameters were derived from existing dimensions, listed in Table 1. Some of them have physical meaning while others are bidimensional or tridimensional. It turned out that the correlation between the new parameters and PCWs increased, thus they were better related with the concha, especially  $d_{13}$ . In addition correlation coefficients between pinna parameters and central frequencies of spectral notches (NCF) were calculated. The correlation between pinna parameters and NCFs were stronger than the correlation between pinna parameters and PCWs, but still poor for a linear regression. Finally, 20 PCWs and 2 NCFs could be estimated based on their anthropometric parameters by solving the least square problem. Notch positions were adjusted from the estimated NCFs by using an moving notch algorithm. A useful application of this method could be the extraction of parameters from a pinna photography. Image processing and automatic segmentation could lead to a good result.

Xu *et al.* [XLS09] proposed a weighted correlation method (WCM) to correlate PC weights with local and global key anthropometric measurements (KAMs). Local KAMs are dependent on the source position, whereas global KAMs are independent of positions and represent general dimensions for individualization at all positions. Both key measurements use measurements of pinna, head and torso. 10 typical source positions of 45 subjects from the CIPIC database were analyzed. In order to relate the PCA scores and the weighted correlation between listeners anthropometric measurements, a method for identifying local and global KAMs was introduced. Local KAMs were found by calculating the correlation of PCWs and anthropometric data for each position and sorting according to their importance. Spectral distortion (SD) defined by Equation 11 (Page 35) was adopted to evaluate estimated DTFs. When using local KAMs, SD was less than



Variable	Measurement	Variable	Measurement
$x_1$	head width	$x_{15}$	seated height
$x_2$	head height	$x_{16}$	head circumference
$x_3$	head depth	$x_{17}$	shoulder circumference
$x_4$	pinna offset down	$\theta_1$	pinna rotation angle *
$x_5$	pinna offset back	$\theta_2$	pinna flare angle *
$x_6$	neck width	$d_1$	cavum concha height *
$x_7$	neck height	$d_2$	cymba concha height *
$x_8$	neck depth	$d_3$	cavum concha width*
$x_9$	torso top width	$d_4$	fossa height *
$x_{10}$	torso top height	$d_5$	pinna height *
$x_{11}$	torso top depth	$d_6$	pinna width *
$x_{12}$	shoulder width	$d_7$	intertragal incisure width *
$x_{13}$	head offset forward	$d_8$	cavum concha depth *
$x_{14}$	height		

Table 2: Pinna measurements and anthropometric parameters stored in the CIPIC database [ADT01]. Marked dimensions are specified separately for both ears.

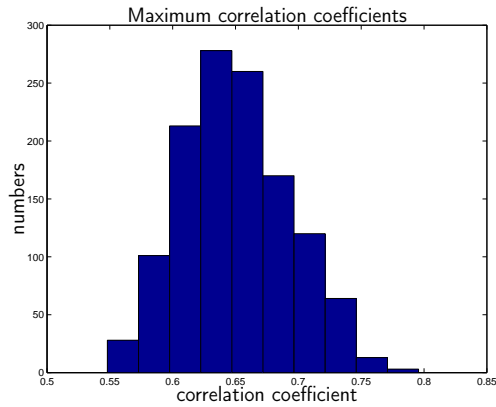
5.7 dB. Compared with averaged HRTFs, individualization with local and global KAMs could reduce SD by 9.4% and 9.7% respectively. Therefore, rough sound localization was satisfied when using local and global KAMs. However, no significant differences between the two could be found.

Xie *et al.* [XZR07] pointed out that the mean values of maximum ITD for male and female are significantly different. Because the most existing HRTF databases include measurements and anthropometric parameters from western people, they cannot represent special features of different individuals or ethnic groups, such as people in China. Therefore, the authors have begun to build a new HRTF database, which now includes 52 subjects (26 male and 26 female). The results were compared to the CIPIC database (16 male, 27 female). The mean ITD for Chinese people was significantly less than that of western subjects.

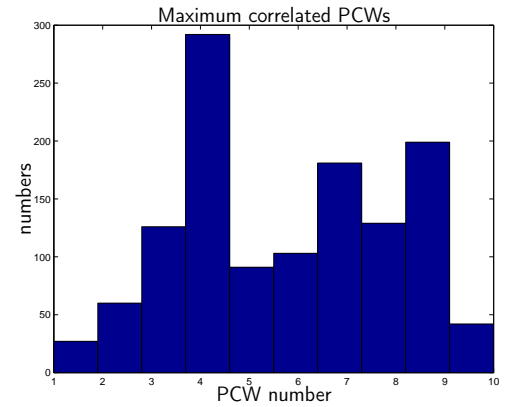
The models discussed above tend to solve the problem of HRTF individualization using anthropometric dimensions. Different approaches try to model PRTF or HRTF using correlation between local source positions or global parameters and PCWs or peak/notch positions. Especially cultural differences in body's physical structure could be better represented by such models. However, the process of connecting physical dimensions with related model parameters is a complex task and the identification of crucial anthropometric parameters is not always clear. Furthermore, there is currently no large HRTF database that can be used to present statistically significant results. This area still needs a lot more work to be invested. In addition, measurement of anthropometric dimensions is also time consuming and it is relatively difficult to measure accurately. Probably, automatic processing of 2D images of left and right pinna could replace manual measurement.

Figure 4a-c indicates the maximum correlation values between first 10 left PCWs and 43 anthropometric features across all subjects and source directions in the CIPIC database. In addition, the proposed dimension by Rodriguez and Ramirez [Rod05] were included (Table 1), resulting in 67 different dimensions. First PCA was applied on the entire database, then the resulting weights of the subjects were correlated with the corresponding anthropometric dimensions. In Figure 4a, a histogram about the maximum correlation values of each position is depicted. It turned out that the maximum correlation was not as high as expected. The mean correlation value is 65% and the maximum correlation in the entire database is only 79%. Figure 4b indicates how often the first 10 PCWs are maximum correlated with some physical parameters. No significant difference between the PCWs was found. Remarkably, the number of PCW1s and PCW2s is very low. In Figure 4c, the numbers of maximum correlation for each physical parameters are listed. The *head offset forward* showed the most correlations in all source positions, indicated as dimension number 13. Figure 4d shows the correlation between left PCW2 and fossa height left across all source positions.

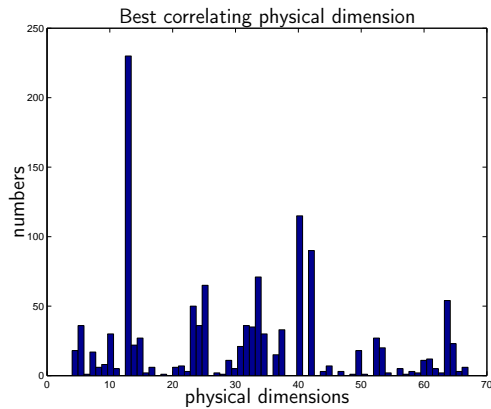
The results for Figure 4a-c were almost the same when using right ear PCWs. For this reason, in the model (Section 5) proposed in this work, anthropometric data was not used, because it is still not proved which physical parameters are dominant for the HRTF. In order to obtain global parameters that have impact on all source positions, further



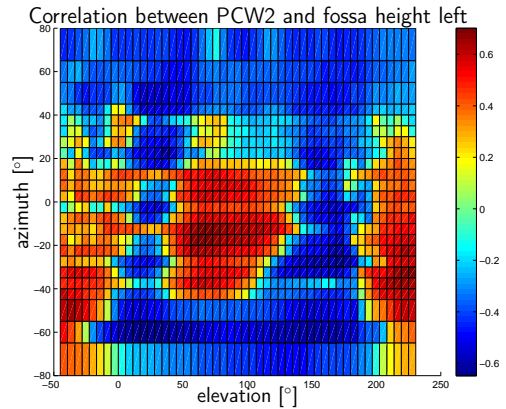
(a) Histogram of maximum correlation coefficients for each source position.



(b) Histogram about maximum correlated PCWs for each source position.



(c) Histogram about maximum correlated physical dimension for each source position.



(d) Correlation between left PCW2 and fossa height left across all source positions.

Figure 4: Correlation between physical dimensions and first 10 left PCWs of each source position in CIPIC database.

investigation regarding correlation of the principal weights in relation to different source positions would be appropriate.

### 3.2 Conclusion

Overall, PCA has been used with somewhat encouraging results. The method is not limited to a particular domain, so both HRTFs and HRIRs can be modeled. Due the fact that the resulting total variance of the reconstruction can be simply calculated and controlled by adding or neglecting individual PCs, a desired representation of the data set can be easily achieved.

Before computing PCA, important decisions have to be made. Firstly, there are several ways to analyze HRTFs according to its dimensions, e.g. analysis on frequency, source positions or inter-subject variations. A comparison between the different techniques is given later in Section 5.2. Secondly, the choice of the input data can have a strong influence on the results. In general, three different input data for PCA calculations are found in literature:

1. DTFs by subtracting global mean (database mean) from HRTF magnitude
2. DTFs by subtracting the mean of each subject from HRTF magnitude
3. HRIRs (original or minimum-phase functions)

Before computing the PCA in frequency domain, it is necessary to eliminate the differences in measurement conditions within a HRTF database [MG92]. Therefore the empirical mean must be subtracted from individual HRTFs. This is tolerable because the mean of an individual has no direction dependent information, thus it is irrelevant to PCA. Afterwards, the input matrix has to be centered by subtracting off columns means. This is a crucial operation in order to get relevant values after PCA processing.

As described in [HP08], there are some advantages and disadvantages using HRIRs or HRTFs for PCA. HRIRs can be grouped into temporal sound events, such as the initial time delay, pinna response, effects of head and shoulders or reflections from measurement device. Therefore PCA processing and interpretations can be more accurate. Importantly, no phase model for reconstruction is necessary. On the other hand, the log-magnitude spectrum of HRTFs is more equatable with the logarithmic sense of human hearing. However, the responses of certain body parts are coupled in frequency domain and can not be simply decomposed. In addition, the approach of minimum phase assumption results in HRIRs that are much shorter than the original ones, consequently pinna and torso contributions are merged.

In summary, the studies discussed above perform PCA decomposition to explain about 90% variance of the original data, while the numbers of selected PCs vary from 5-12. Radriguez and Ramirez [Ram05] took even 20 PCs describing 99% of the total variance to approach almost perfect reconstruction. Table 3 summarizes selected HRTF models discussed above. To provide an objective evaluation criterion for synthesized HRTFs, spectral distortion (SD) defined by Equation 11 (Page 35) can be calculated.

To use the powerful mechanism of component decomposition, several individualization techniques based on PCA were established. By subjective adjustment of the component weights, localization performance could be significantly increased. In Section 4, fundamental studies are presented.

	Kistler & Wightman [KW92]	Qian & Eddins [QE98]	Hwang & Park [HPP08]	Shin & Park [Shi08]	Martens [Mar87]
HRTF database	own measured	Tucker-Davis	CIPIC	CIPIC	own measured
PCA input data	DTF log magnitude in specific frequency range (0.2 - 15 kHz)	HRTF magnitude	first 1.5 ms of HRIR	first 10 samples of left-ear HRIR after direct impulse	DTF log magnitude, critical band filtered
PCA input matrix	$5300 \times 150$	-	$67 \times 2205$	$10 \times 45$ for each position	-
HRTF sets	10	26	45	45	2
Source positions	265	360	49	9	36
Azimuth range	entire plane (24 pos.)	entire plane (36 pos.)	$-80^\circ$ to $80^\circ$ , 25 pos.	$-80^\circ$ to $80^\circ$ (25 pos.)	entire plane (36 pos.)
Elevation range	$-48^\circ$ to $72^\circ$ (11 pos.)	$-30^\circ$ to $60^\circ$ (10 pos.)	$-35^\circ$ to $230^\circ$ (50 pos.)	$-35^\circ$ to $230^\circ$ (50 pos.)	-
Rec. phase	minimum phase	original	not required	not required	minimum phase
Suggested PCs	5	-	12	4-5	4
Test subjects	5	6	6	4	no test
User feedback	verbal response	graphical inter-face	GUI with sliders	GUI with sliders	-

Table 3: Various parameters used in studies for HRTF individualization based on PCA.



## 4 HRTF Individualization Techniques

In the last two decades, much research has been devoted to develop various models for HRTF individualization. In general four methods for adapting non-individual HRTFs exists. All of these approaches have advantages and disadvantages, but they all have been shown to improve localization relative using a generalized set of HRTFs. The first one intends scaling in frequency axis and was proposed by Middlebrooks. When scaling the dataset, peaks and notches are shifted. Secondly, *Directional band equalization* could have great improvement in spatial hearing, because it has been shown that interaction between bands can also affect localization.

In [Mid99a] [Mid99b], Middlebrooks used scaling of the spectrum along the frequency axis. Existing peaks and notches are shifted and so inter-subject variability in individual HRTFs is minimized. By scaling, the inter-subject differences can be reduced to 6.2 dB. To find a good starting value for the scaling factor is not an easy task. The optimal scale factor can be estimated from the physical dimensions of a subject. Based on pinna-gravity height and head width, a scaling factor can be estimated as a starting point for further adaptation. The number of spectral maxima and minima is often different. It is shown that certain spectral peaks and notches move as a function of azimuth and elevation.

Mehrgardt and Mellert have found an optimal scale factor as a function of incidence angle, while the study by Middlebrooks intends a global scale factor for all positions. The pinna of individuals are different in many more ways than just a simple scaling, therefore this approach has limited success.

While in the study by Middlebrooks the magnitude is kept constant and the center frequencies of spectral cues are shifted, So *et al.* proposed a method to manipulate the magnitude in certain regions of the spectrum. In [SL11] six ear-level directions were manipulated by changing the gains in six different frequency bands (170-680 Hz; 680-2400 Hz; 2.4-6.3 kHz; 6.3-10.3 kHz; 10.3-14.9 kHz; 14.9-22 kHz). For example, band 1 was amplified and band 2 was attenuated to make a sound event more likely to be perceived from front. Each of the six directions were manipulated in 18 different ways, resulting in 108 HRTFs. In a graphical interface the 42 subjects indicated the perceived sound directions by moving the mouse cursor. Sound stimuli were presented through headphones and each manipulated stimuli was scaled so that the overall sound pressure level stayed on the same level. Up to 66% less front-back errors occurred and localization error decreased up to 70%. The spectral manipulations can be used as a set of add-on filters to increase directional accuracy. Moreover, the authors suggested to combine the approach by Middlebrooks and his own to reach less localization error.

### 4.1 Methodologies for Subjective Adaption

In recent years, several methods for tuning or adaption of generalized HRTFs were proposed. Nearly all of the them use no accurate phase model, rather the minimum phase plus time-delay approach. Silze [Sil02] proposed a method in which the transfer

functions were tuned by a tuning expert. The aim of this study was the reproduction of multichannel signals over headphone. Since HRTFs for each source position were selected and changed on the authors experience, this process was very difficult and time consuming. Listening tests results fully confirmed the selection and tuning of the expert. However, the main focus of this report deals with the adaptation of the subject itself.

In [QE98], Qian and Eddins described a short procedure for HRTF customization including three steps. First, sound stimuli were presented over headphone forming a horizontal circle at a fixed elevation. Because of individual differences, subjects could evaluate the stimuli on criteria like externalization (yes / no), azimuth or elevation position (scale from 1-10). For three positions in elevation ( $69^\circ, 0^\circ, -30^\circ$ ) and each of the 26 HRTF sets, the circle presentation was repeated. After this the six best matching HRTF sets were identified by summing up the total rating score and excluding sets without externalization (about 35min).

In the second phase, a best-matching HRTF was found by paired comparison of the remaining sets. Each direction was presented by two different sets and the subject chose the one that sounds closer to the corresponding virtual position. Again, circle presentation for the six HRTF sets was performed and subjects had to reevaluate the stimuli with a single criterion based on general impression. Finally the best set was calculated by using the results of paired comparison and cycle presentation (about 18min).

To reduce differences of selected and subjects individual HRTF, the magnitude of the DTF can be scaled in frequency [Mid99a]. The proper scaling factor was obtained by paired comparison (about 20min). So, within about one hour, a best-matching HRTF set from 26 possible sets was found and in addition a scale factor to reduce inter-subject variation was estimated.

In [HPP08], a novel customization procedure of HRIRs based on self-tuning of PCWs in median plane was introduced. A HRIR model including 12 PCs was starting point for further processing. Three subjects participated in the experiment. The number of tuning parameters was reduced by sorting the PCWs for each position according to their standard deviation. Based on the assumption that PCWs with large standard deviation contribute significantly to the inter-subject variation, only the three largest PCWs according to their magnitude of standard deviation were used for customization. For each elevation, the subject tuned three PCWs by moving a slider bar. The minimum and maximum bounds are set to be mean  $\pm 3$  standard deviation. Thus the customized model was formed by a linear combination of three adjusted PCWs and nine remaining ones whose weights were the mean values of all subjects in the database. During the procedure, a subject could play the adapted HRIR and a reference stimulus.

In [HPP10], the procedure was repeated with nine subjects in the upper hemisphere. The sphere was divided into two sectors ( $0-70^\circ$  and  $70-180^\circ$ ) including 3 different source positions respectively. The median-plane PCWs were modeled using a linear interpolation of the inter-subject variation.  $\Delta$ PCWs 1-3 at each endpoints of the sectors were tuned, so the whole upper hemisphere could be adjusted by the tuning of nine parameters only. The customization process took about 17 minutes on average. In the end, a short localization test with 7 elevation angles was carried out with individual, customized

and KEMAR HRTFs. All subjects reported enhanced localization performance with customized ones.

Seeber and Fastl [SF03] described a fast method to enhance localization performance. 12 different HRTF sets from AUDIS-catalogue were used in this experiment. 17 subjects without training in acoustics evaluated non-individualized HRTFs based on various criteria, such as spatiality, localization and externalization, by rating with the numbers 0-9. In the preselection, 5 of 12 HRTFs with greatest spatial perception were extracted. In the final step, the remaining criteria were compared. Within 10 minutes, a best-matching HRTF from another person could be found. In general, subjects tend to choose a HRTF from a subject with a slightly larger head. The authors emphasize that direct comparison of different HRTFs is a crucial task, because mostly there are only minor differences between pre-selected sets. Due to direct access, the selection process can be performed more efficiently. The method could be applicable in teleconferencing systems or computer games without the need for special equipment.

Shin and Park [Shi08] isolated the pinna responses from the median HRIRs using 45 subjects from CIPIC database. Only the first 10 samples after the direct impulse were extracted to include pinna activity with largest interject variation. For each elevation, the model included 4-5 basis functions. In the experiment, 4 subjects tuned the pinna response of 9 elevation angles in the median plane by changing the weights of the corresponding PCs. Because only the left-ear HRIRs were included for processing, subjects could adjust the balance of left and right channels for reconstruction. Finally subjective listening tests with measured, customized and KEMAR HRTFs were carried out. Front-back confusions were reduced when using customized HRTFs.

Middlebrooks *et al.* [MMO00] introduced a customization procedure by scaling the transfer function in frequency. Various scale factors were estimated by transforming the DTF in time domain, interpolating with a factor of 32 and decimating in the time domain by a factor of an integer between 15 and 42. This resulted in an impulse response scaled in frequency by a factor of 0.47 to 1.31. 20 listeners evaluated the various DTFs in respect of vertical localization. The procedure took about 2-4 hours for each subject. It was shown that the preferred scale factor was highly correlated with the physical scale factor that could be estimated based on head width and some pinna dimensions. Middlebrooks recommend to use physical dimensions at first to define a narrow range and then perform a psychophysical procedure to find the preferred scale factor in that range.

So *et al.* [SNH<sup>+</sup>10] tried to reduce front-back ambiguity in non-individualized HRTFs (KEMAR) by providing several choices for the listener. Several spectral features of 196 HRTF sets were quantified based on previous studies and then clustered into six near-orthogonal groups for forward and backward directional sounds respectively. In the within-subject experiment, 15 listeners evaluated 7 different HRTF sets in 4 different source positions. Each stimulus was repeated six times over headphones. Subjects indicated the perceived incident angle using a hand-held pointer on a sphere. For each of the four sound directions, the best-matched stimulus was selected. Comparison with KEMAR stimulus indicated significantly lower front-back confusions from 29 to 10%.

Lindau *et al.* [LEW10] proposed an approach by real time manipulation of the IDT,

because non-individualized binaural data can degrade localization accuracy. Onset detection was used because previous listening tests confirmed the robustness of this method. IDT individualization was performed in real time by using a head tracking system. By separating the processing of magnitude spectrum and phase, it is possible to apply different spatial resolution and interpolation methods. Comb filter effects could be minimized because cross fading was performed on time-aligned signals. Moreover, the author suggested an anthropometry-based prediction model for an individual ITD correction factor.

Individualization of HRTFs is an iterative tuning process that can be very time consuming and exhausting. In summary, all studies tend to reduce the parameters as much as possible, but many of them work only in small regions (e.g. median plane) and few source positions. The HRTF model in the next section should overcome this limitation and provide a customization procedure including the whole hemisphere.

## 5 HRTF Model

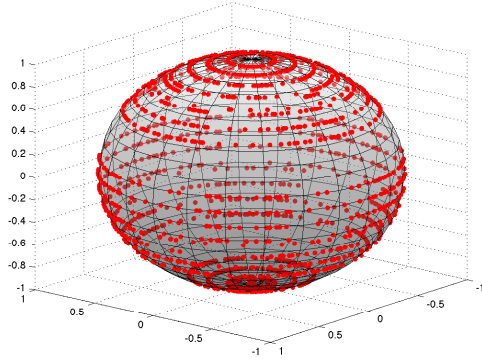
### 5.1 HRTF Database Analysis

Five different HRTF databases were used for calculations, listed in Table 4. Figure 5 indicates the different spatial resolutions. The public domain CIPIC database includes 45 subjects with 1250 source positions and anthropometric measurements of all subjects. Due to easy access and large amount of data that database is very well known and often used by researchers. A convenient side effect of the strong use of CIPIC is the comparability of various scientific works.

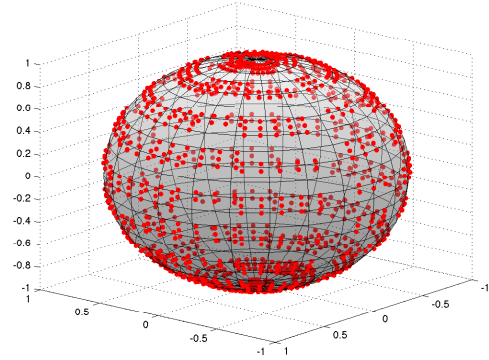
Name	Department	Subjects	Positions
IEM	Institute of Electronic Music and Acoustics	30	24
IRCAM	Institut de Recherche et Coordination Acoustique/Musique	50	187
CIPIC*	University of California at Davis	45	1250
ARI*	Acoustics Research Institute	66	1550
KEMAR*	MIT Media Laboratory	1	710
GLOBAL	IRCAM, ARI	116	44

Table 4: HRTF databases used in this report. Marked with asterisk include anthropometric measurements.

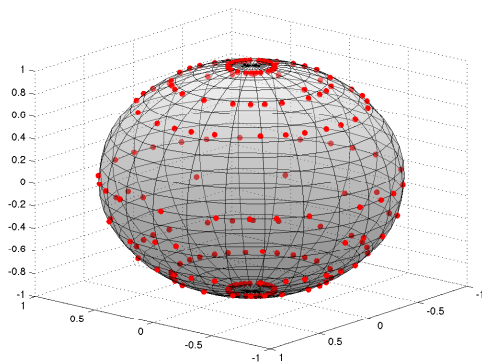
The ARI HRTF database consists of 66 normal hearing subjects including several anthropometric data of 15 subjects, like in the CIPIC database. 1550 source positions were measured for each listener including the full azimuthal-space ( $0^\circ$  to  $360^\circ$ ) and elevations from  $-30^\circ$  to  $+80^\circ$ . The measurement process for one person takes only 20 minutes, because of the usage of *multiple exponential sweep method (MESM)*. The goal is to play a sweep before the end of a previous one. As described in [MB07], the methods allows



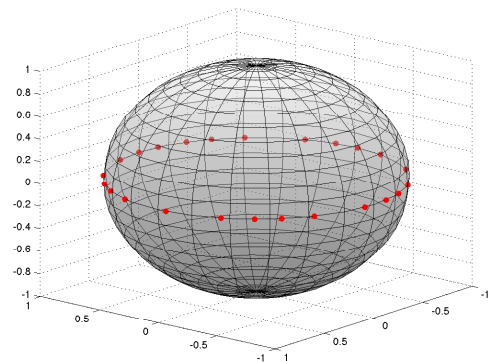
(a) ARI, 66 subjects.



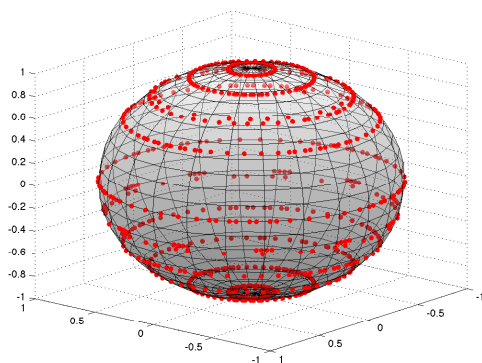
(b) CIPIC, 45 subjects.



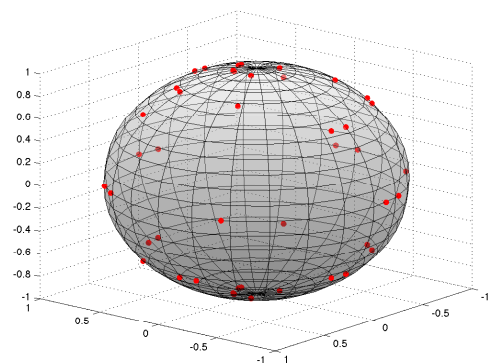
(c) IRCAM, 50 subjects.



(d) IEM, 30 subjects.



(e) KEMAR, 1 subject.



(f) GLOBAL, 116 subjects.

Figure 5: Spatial resolution in existing HRTF databases.

the interleaving of three sweeps and overlapping eight groups of the interleaved sweeps. For this reason the measurement time can be drastically reduced without artifacts in HRTF reconstruction. In-Ear-microphones (Sennheiser KE-4-211-2) were located inside the subject's ear canal. During HRTF measurement, the head position is monitored constantly. If it is outside a valid range, the recording stops automatically and the subject gets an acoustical feedback to return into the valid area. For more information about measurement procedure and recording equipment, please go to ARI website<sup>1</sup>.

In order to obtain a larger dataset, a *global* database was formed with 44 source positions and 116 subjects. It contains positions that coincide in IRCAM and ARI database.

### 5.1.1 PCA Compression Efficiency

According to Leung [LC09], the optimal format for the PCA operation is the linear amplitude form in frequency domain, because the compression efficiency is on the highest level. Analysis of our data set confirms this statement and shows that this is consistent over all databases. Figure 6 indicates the compression efficiency of different input data and databases. When using the linear amplitude in frequency domain, total variance over 90% can be achieved by only 6-7 components. Contrary, more than 20 components are essential when HRIRs (raw data) are applied. However, it is to be noted, that only minor differences in compression can lead to significant changes in localization performance, such as accuracy or front-back confusions.

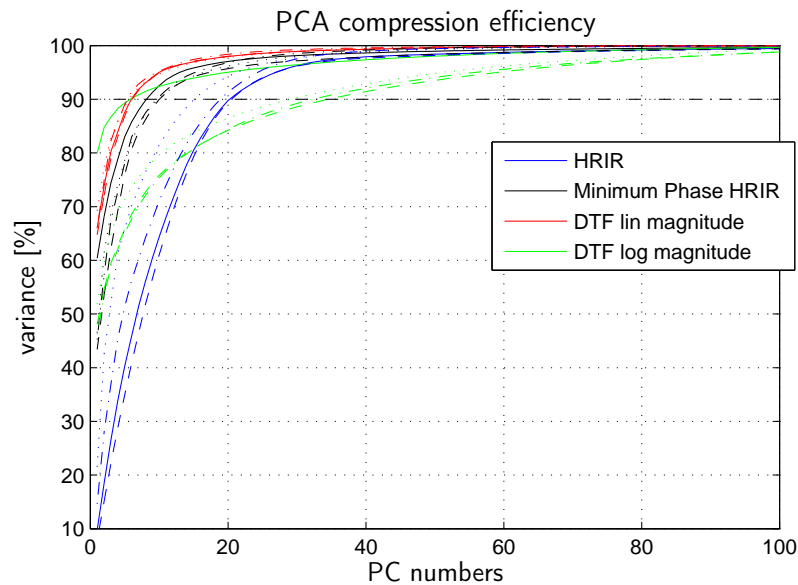


Figure 6: PCA compression efficiency of different input data in four different databases: IRCAM (dashed), ARI (dotted), CIPIC (dash-dot) and GLOBAL (solid line). Four different colors indicate the input data are indicated: HRIR (blue), minimum phase HRIR (black), DTF with linear spectrum (red) and DTF with logarithmic spectrum (green). PCA includes all subjects and positions in each database.

1. <http://www.kfs.oeaw.ac.at/content/view/608/606/>

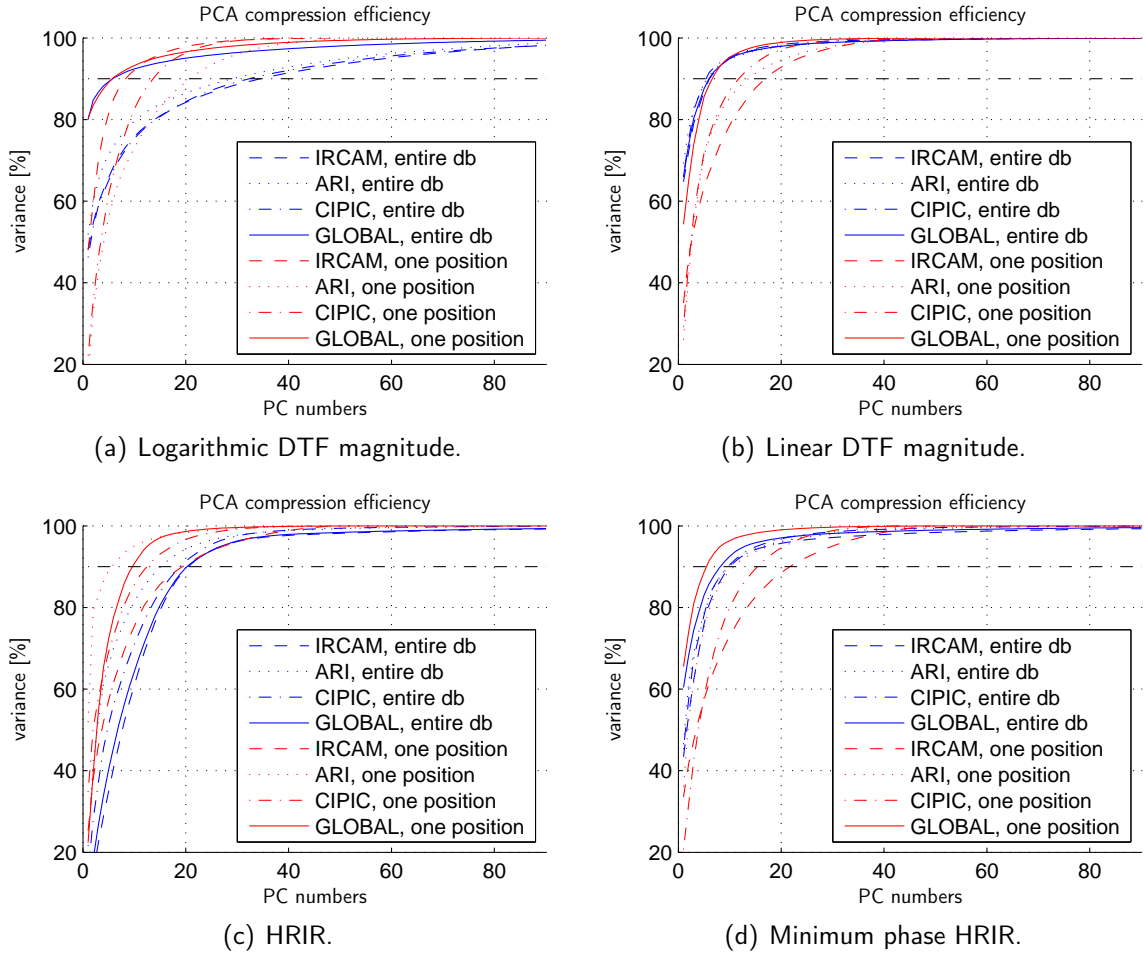


Figure 7: PCA compression efficiency for different input data when using entire database (blue) or only one source position (red) in four different databases (IRCAM, IEM, CIPIC and GLOBAL).

There are studies which perform PCA on each source position separately ([Shi08], [Ram05], [QE98]), while the remaining ones use the whole database including all directions for calculating the principal components. One could argue that the main advantage of processing PCA separately in each position could be that higher variance is explained by the first components, because of smaller variation in the input dataset. This is only true, if the logarithmic measure is used (Figure 7a). The opposite is the case for the linear amplitude. Figure 7b-c shows the amount of variance explained as a function of the number of principal components used, for logarithmic vs. linear amplitude DTFs used as input to PCA performed either on a typical single position or at all positions in the database simultaneously, for three different HRTF databases. PCA of only one position requires an average of 5 components more than the calculation of the entire database. No discernible differences were found with calculations using HRIRs or minimum phase versions (Figure 7c-d). Generally, the more data available, the better will do the PCA for the linear magnitude. In addition, analysis of structure and fluctuations of PCs and corresponding PCWs in a database is easier when involving the entire database because

the same principal components can be used to model all sound directions.

### 5.1.2 Correlation of PCs

	ARI	CIPIC	IEM	GLOBAL	IRCAM
ARI	1.00	0.94	0.78	0.99	0.95
CIPIC	0.92	1.00	0.71	0.96	0.98
IEM	0.86	0.96	1.00	0.78	0.74
GLOBAL	0.97	0.96	0.92	1.00	0.97
IRCAM	0.82	0.95	0.96	0.91	1.00

Table 5: Correlation in percent of left ear **PC1** (lower left side) and **PC2** (upper right side) among different HRTF databases.

	ARI	CIPIC	IEM	GLOBAL	IRCAM
ARI	1.00	0.36	0.55	0.86	0.75
CIPIC	0.85	1.00	-0.38	0.73	0.78
IEM	0.34	0.66	1.00	0.11	0.01
GLOBAL	0.95	0.92	0.51	1.00	0.92
IRCAM	0.96	0.88	0.40	0.97	1.00

Table 6: Correlation in percent of left ear **PC3** (lower left side) and **PC4** (upper right side) among different HRTF databases.

	ARI	CIPIC	IEM	GLOBAL	IRCAM
ARI	1.00	0.80	-0.18	0.96	0.55
CIPIC	0.92	1.00	-0.21	0.86	0.73
IEM	-0.14	0.01	1.00	-0.17	-0.17
GLOBAL	0.71	0.69	-0.48	1.00	0.72
IRCAM	0.85	0.96	0.07	0.72	1.00

Table 7: Correlation in percent of left ear **PC1** left (lower left side) and **PC2** (upper right side) in the **median plane** among different HRTF databases. The results of IEM database are greyed out, because there are no source positions in elevation plane.

According to Middlebrooks *et al.* [MG92], the differences in subjects and measurement of various HRTF database should have little effect on the principal components. To verify this, correlation coefficients of the first 4 PCs between all HRTF databases were computed. PCA including all subjects and source positions of each database was calculated using the logarithmic DTF spectrum in order to include only direction dependent information. Tables 5-6 show the correlation coefficients for each component across the different databases.

The first PCs correlate strongly across the different databases (mean percentage PC1: 92%, PC2: 88%, PC3: 75%). This is an expected result and is consistent with the theory



	ARI	CIPIC	IEM	GLOBAL	IRCAM
ARI		0.84	0.90	0.97	0.72
CIPIC	0.90		0.88	0.82	0.73
IEM	0.85	0.96		0.88	0.69
GLOBAL	0.97	0.95	0.91		0.81
IRCAM	0.82	0.95	0.97	0.90	

Table 8: Correlation in percent of left ear PC1 (lower left side) and PC2 (upper right side) in the **horizontal plane** among different HRTF databases.

	ARI	IRCAM	GLOBAL
ARI		0.82 (0.63)	0.98 (0.06)
IRCAM	0.82 (0.05)		0.90 (0.99)
GLOBAL	0.97 (0.06)	0.92 (0.99)	

Table 9: Correlation in percent of left ear PC1 (lower left side) and PC2 (upper right side) of the same 44 source positions among 3 different HRTF databases. Logarithmic spectrum was used for PCA, values in parentheses indicate the correlation coefficients when linear spectrum was used.

of PCA. However, closer inspection reveals that PC3 and PC4 in the IEM database have almost no correlation with corresponding PCs of other databases. The reason for this is that IEM database only contains 24 positions in azimuth plane and the correlation of PCW2-4 to all other databases is very low, because these principal components mainly contain information about elevation. When the calculations are performed without logarithmic scale of the magnitude, the results are significantly lower. The reason could be that the logarithm compresses the HRTF and therefore hides the detail in the HRTF function, thus making it easier to model.

It has to be noted that the spatial resolution of the databases is different. For this reason, the correlation of PCs was calculated again using three selected direction sets. These were: the set of common directions across all databases, the set of sound directions in the median plane and the set of sound directions in the horizontal plane in the three databases. Table 7 and 8 shows the results for the median and the horizontal plane, which are in general sampled differently in each database. Generally, the values have not changed compared to the previous case. Figures 8 and 9 indicate the resulting PCs with logarithmic and linear spectrum respectively. In addition, PCs of 44 specific source positions that coincide in IRCAM, ARI and GLOBAL database were calculated. Table 9 shows that the correlation coefficients are increasing when only the same source positions are used. It can be seen, that the resulting PCs of positions in horizontal plane are almost the same for all positions and for 44 specific positions. Consequently, the first component describes the variance in the horizontal plane, no matter what positions are used for PCA.

It has to be noted that in some cases the resulting PCs and PCWs are mirrored. The reason for this is that PCA returns a principal component basis that is rotation invariant.

It can be that depending on the input set, that different rotations version of basis emerge. However, the reconstruction matrix remains still the same, because the signs and potentially the magnitude of the corresponding PCWs is also changing. In our case a rotation of 180 degrees occurred which was corrected in the following figures to ensure better visibility.

### 5.1.3 Variation of PCWs

Hwang and Park [HP08] investigated the characteristics of PCWs for the left-ear HRIRs in the median plane. Only the first 1.5 ms after the direct pulse were used to include the effects on pinna, head, shoulder and torso. PCA of HRIRs in the CIPIC database revealed that PC1 had positive and negative mean PCWs at certain elevation values, thus this component provides sound cues for vertical perception. PC2 indicated cues for front-back discrimination, because the mPCWs were positive in frontal region and negative in the rear section. We also inspected our databases but we took the logarithmic DTF magnitude of all source positions for further analysis. Figures 10 and 11 indicate the distribution of the first six left ear PCWs in the median and horizontal plane respectively. It is clear to see that the range of weights in ascending number is getting smaller, because the variance of the corresponding PCs decreases.

**Median Plane** PCW1 left has positive and negative values at certain position in elevation plane, this might be a major localization cue for up-down discrimination. PC1 has positive mean PCWs (mPCWs) above  $60^\circ$  and negative values below  $80^\circ$ . This is almost consistent with the results by Hwang and Park. PCW2 is positive in lower frontal region and negative above the head. This could be a cue for up-down discrimination.

**Horizontal Plane** It is obvious that left PCW1 amplifies the corresponding component on the ipsilateral side ( $0 - 180^\circ$  azimuth) and decreases it on the contralateral side. PCW2 tend to have negative values in frontal and positive values in rear positions. This could be a cue for front-back localization. The variation of the remaining weights is more complex and need to be considered in more detail.

**Ear Symmetry** Some of the previous models focusing on the median plane were based on the simplification of ear symmetry ([Shi08], [HPP08], [HPP10]), although the left-ear and right-ear HRTFs are slightly different, particularly in high frequencies. Thus, PCWs of each ear in all databases were examined. Figure 12a indicates the weights for left and right ears across all source positions in the IRCAM database. Obviously, the weights across almost all positions appear to be symmetrical. Closer inspection of the PCWs in the median plane (Figure 12b-f) reveals that all left and right ear PCWs are almost identical. This is also true for the remaining weights with lower variance. Moreover, Morimoto [Mor01] confirmed that the perception in elevation mainly depends on the monaural characteristics, consequently assuming symmetrical ears when modelling in median plane can be adequately.

#### 5.1.4 Least Squares Reconstruction of HRTFs

In order to study the effect of different numbers of weights on the reconstruction error, the method of least squares can be applied. For each frequency  $f$  ( $0, \dots, \frac{f_s}{2}$ ) the reconstructed HRTF is calculated by

$$\sum_{j=1}^N \mathbf{X}_{fj} \cdot w_j = h_f, \quad (7)$$

with  $\mathbf{X}$  as principal component matrix,  $w$  as vector of principal weights and  $h$  as the reconstructed HRTF. Index  $j$  goes through the  $N$  principal components and there are  $\frac{f_s}{2}$  equations. The error  $e$  can be minimized by

$$\min_e ||h - \mathbf{X}w||^2. \quad (8)$$

By introducing  $\mathbf{X}^T$  in Equation 7, the weights for reconstruction with smallest error is obtained:

$$(\mathbf{X}^T \mathbf{X})w = \mathbf{X}^T H \quad (9)$$

$$w = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T H. \quad (10)$$

However, it must be ensured that the matrix  $(\mathbf{X}^T \mathbf{X})$  is invertible. First, it was investigated how the number of observations has influence on the reconstruction error. To this, PCA was applied on data sets with different numbers of subjects.

*Spectral distortion (SD)* is introduced which has been used to evaluate speech recognition, but it can also describe the errors of HRTF estimation:

$$SD_{\phi, \theta} = \sqrt{\frac{1}{N} \sum_{j=1}^N |h_{\phi, \theta}(f_j) - \hat{h}_{\phi, \theta}(f_j)|^2} [dB], \quad (11)$$

with  $h_j$  and  $\hat{h}_j$  as measured and estimated HRTF log magnitudes (in dB) and  $N$  as the number of points in frequency domain. Takanori [TSK99] suggested that the SD of an estimated HRTF should not be greater than 5.7 dB.

Two groups of subjects were defined, a training and testing set. For example, if a database consists of 50 subjects, the training set was constructed calculating the reconstruction error of the first four subjects and altering the number of training subjects from 1 to 50. The testing set was applied by getting the reconstruction error of the last 4 subjects in the database and altering the number of training subjects from 1 to 46. Thus, all HRTFs in the testing set were excluded from the training set.

For a start, the influence of whether a HRTF belongs to the training set on the reconstruction error was investigated. Although PCA provides an orthogonal basis that

in principle would allow the projection of arbitrary datasets, we wanted to verify that reconstruction error remains low even for HRTFs outside the training set. Even if only two data sets are used for training, the PCWs of any other person in the database was predicted correctly. Indeed, the projection upon the basis yields similar error for people that belong to the training and the test set. Apparently the least squares prediction of the PCWs in Equation 10 works so well that it is not necessary to use a larger training sets. This also confirms the thesis that for any arbitrary HRTF, a set of correct principal weights exists and the HRTF reconstruction can be ideal when using all PCs.

In the next step, the PCA reconstruction was limited, so e.g. only the first 10 PCs and corresponding estimated PCWs were used for HRTF reconstruction. Unlike before, it was shown that the number of observations has influence on the reconstruction error. Figure 13 indicates the fluctuation of the reconstruction error, when different numbers of PCs are used in the CIPIC database. The less principal components are used for reconstruction, the more training data is required to minimize the reconstruction error. When only one principal component is used, the amount of the training set has almost no influence on the error. The same is true if all components are used, because the error is already minimal. However, analysis of all databases shows that the error is not always monotonically decreasing. In some cases, there are local minima, consequently involving more training data does not contribute to the HRTF reconstruction. Further investigation should be done here.

Figures 14 and 15 show the same calculations in the ARI and IRCAM database. Contrary to CIPIC, the amount of training data has also influence when only 1 PC is used. Particularly striking is the sudden decrease of the error when training subject number 28 is included in the IRCAM database (Figure 15d). This phenomenon was only found in this database. Perhaps the dataset of this subject is so extremely important for the least-squares algorithm and so decreases the error for all other test subjects.

In the final step, individual source positions were excluded from the training data. This was accomplished by introducing a density grid that can be modified from very close to far apart. Figure 16 shows the results of the reconstruction error as a function of density. It reveals that the smaller the spatial resolution for the training, the higher is the error. As already mentioned, when using more PCs, the error can be reduced.

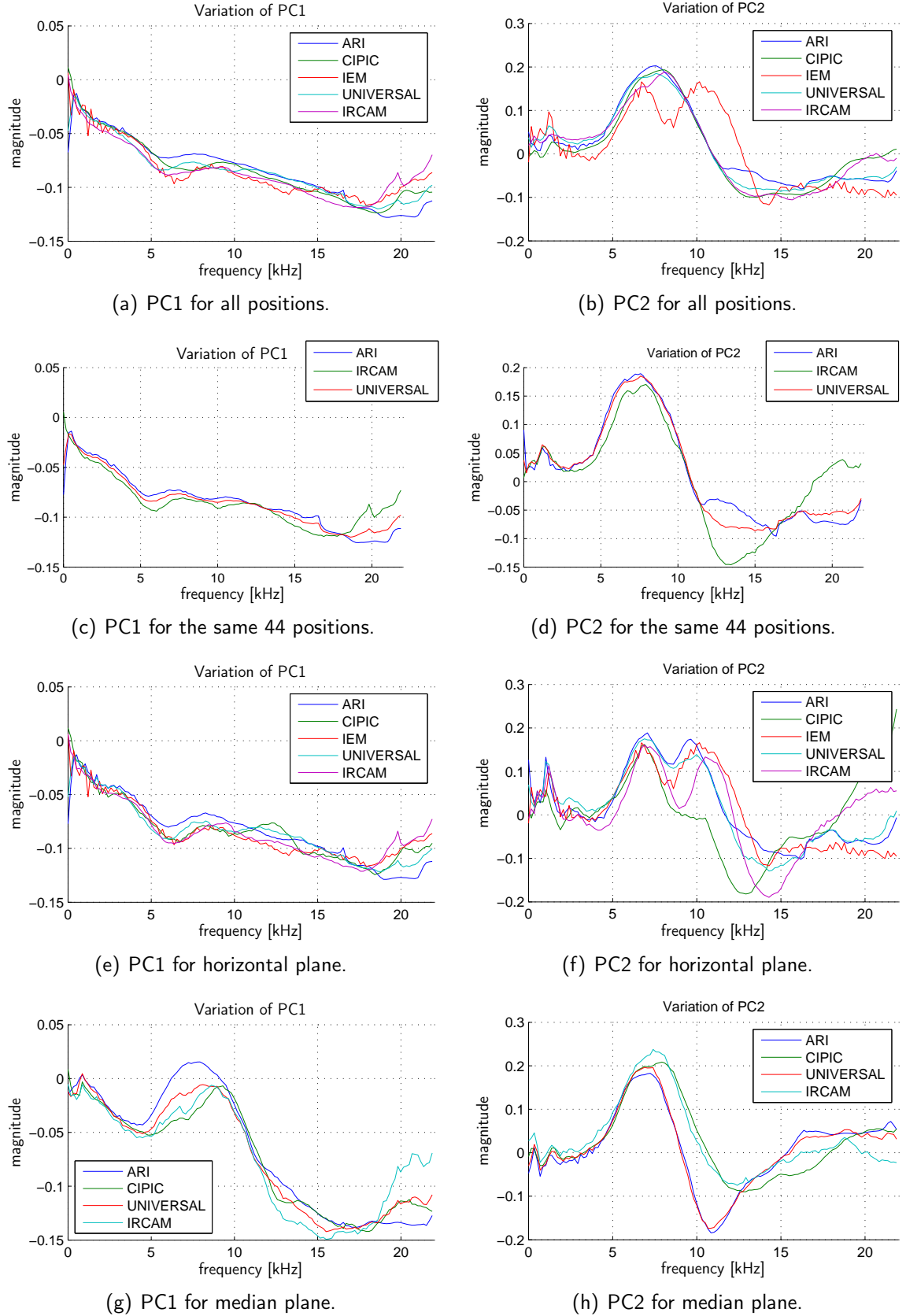


Figure 8: Variation of left ear PC1 and PC2 in each database when calculating PCA with source positions of the entire database (a,b), at specific 44 positions (c,d), in horizontal (e,f) and in median (g,h) plane respectively. **Logarithmic** magnitude spectrum was used for PCA.

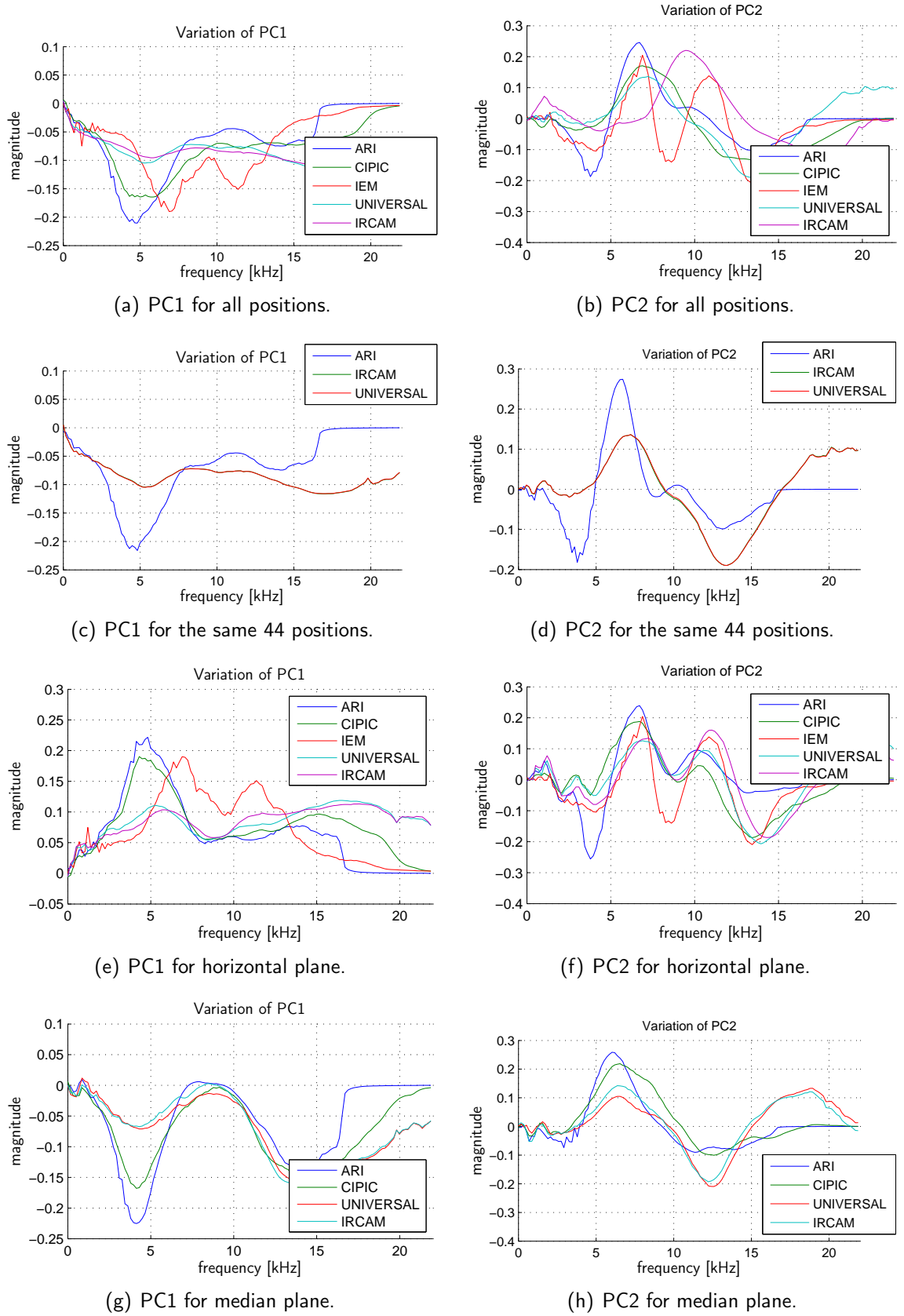


Figure 9: Variation of left ear PC1 and PC2 in each database when calculating PCA with source positions of the entire database (a,b), at specific 44 positions (c,d), in horizontal (e,f) and in median (g,h) plane respectively. **Linear** magnitude spectrum was used for PCA.

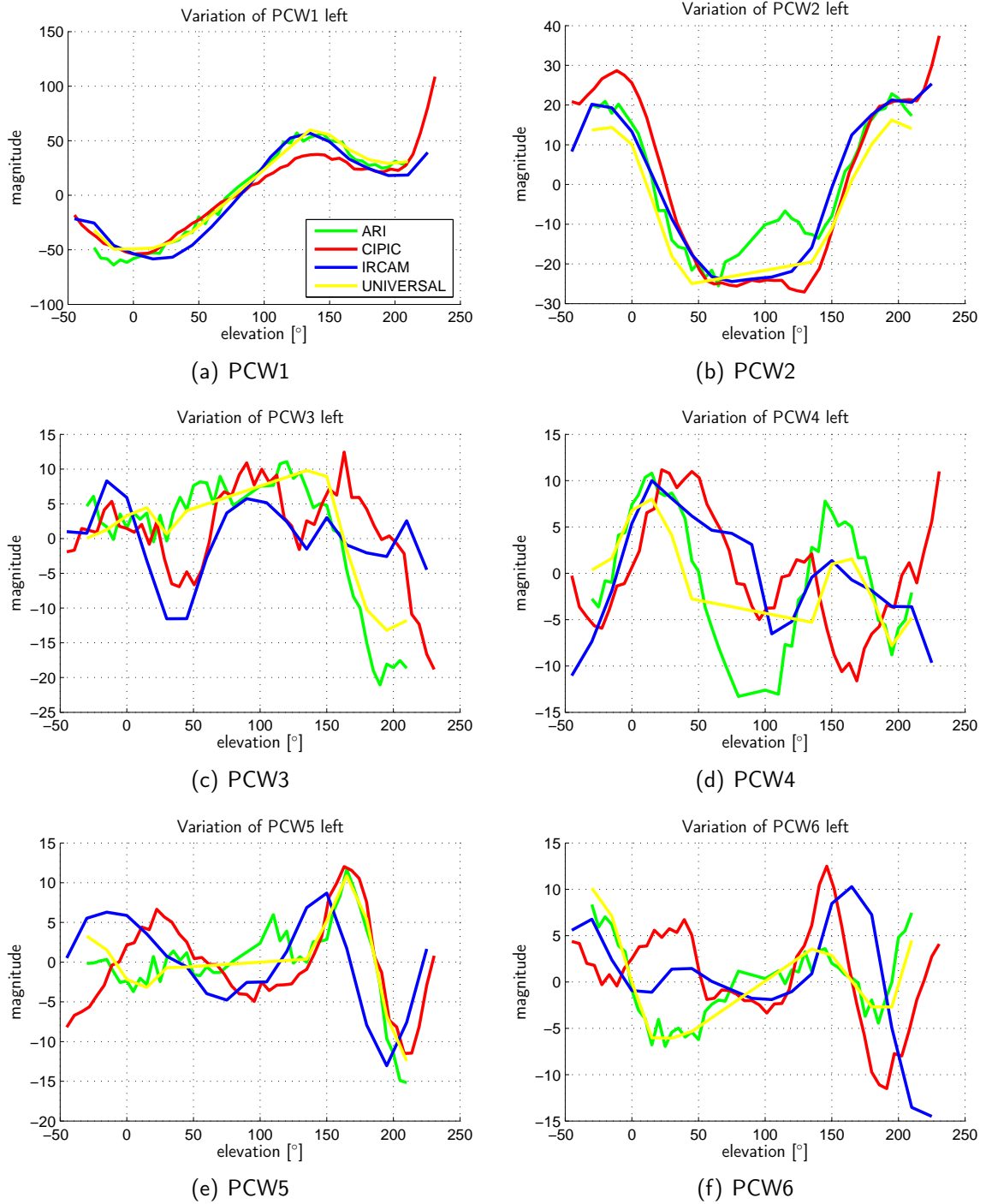


Figure 10: Variation of first six left ear PCWs in the **median plane** of four different databases: CIPIC (red), ARI (green), IRCAM (blue) and GLOBAL (yellow).

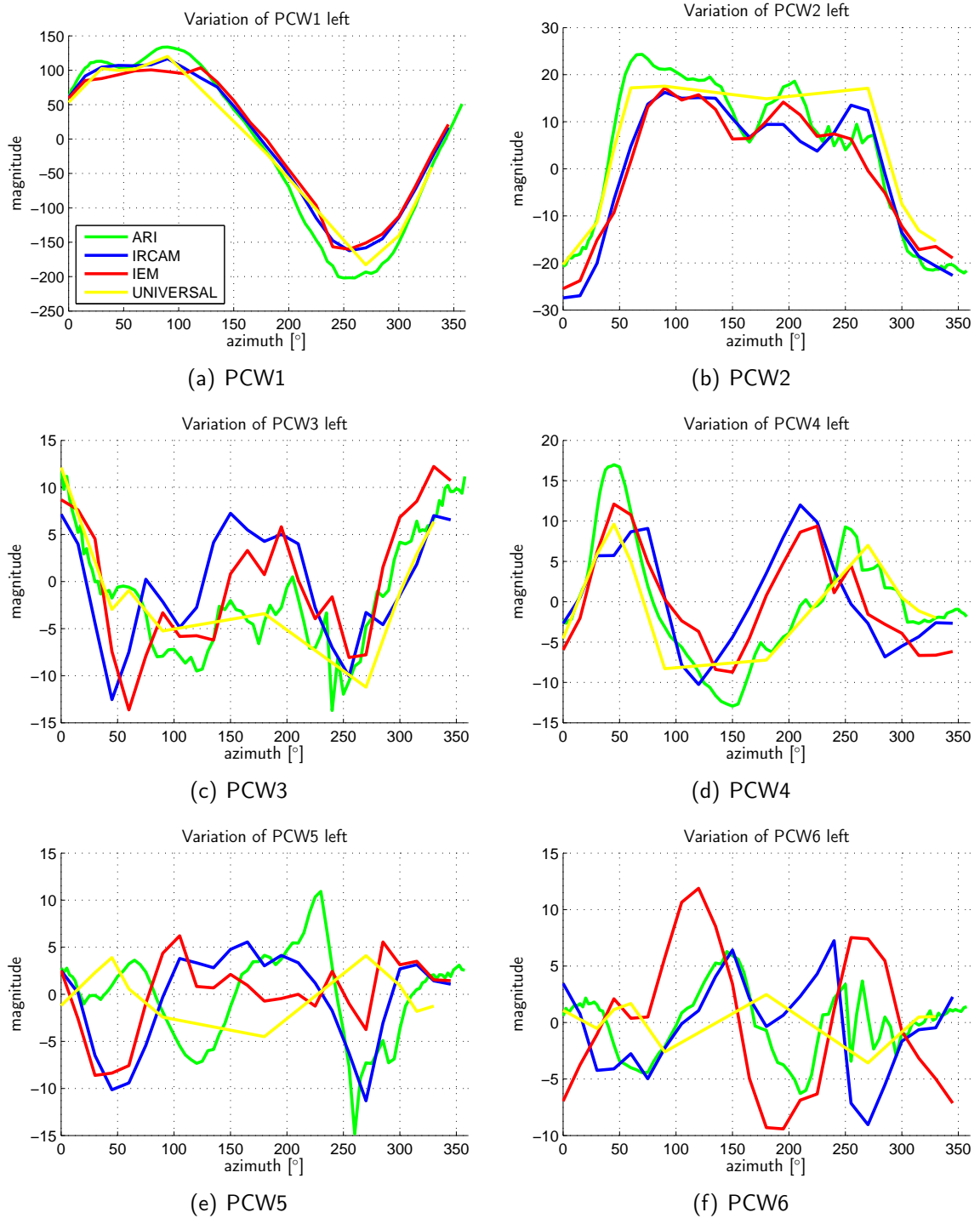
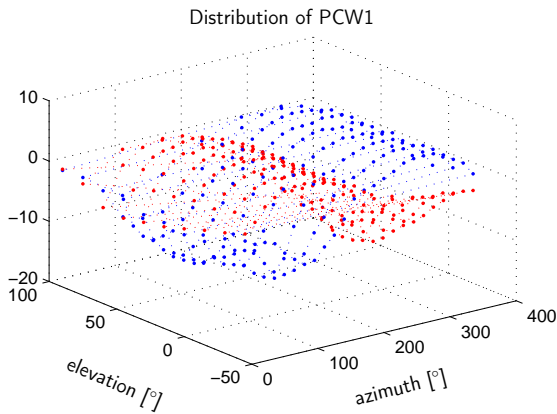
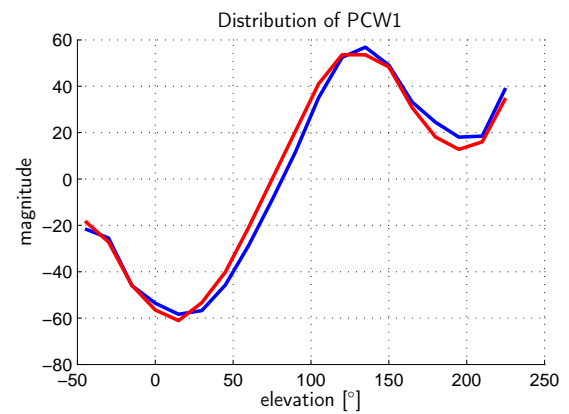


Figure 11: Variation of first six left ear PCWs in the **horizontal plane** in four different databases: ARI (green), IRCAM (blue), IEM (red) and GLOBAL (yellow).

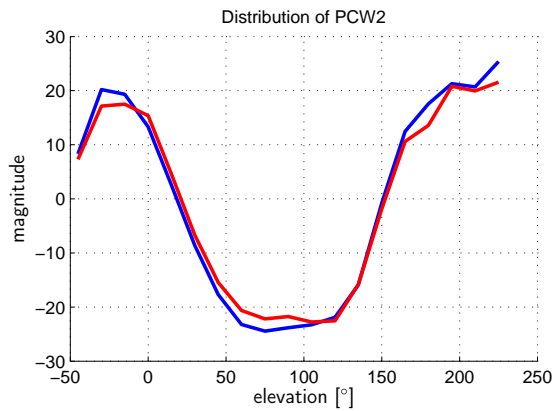




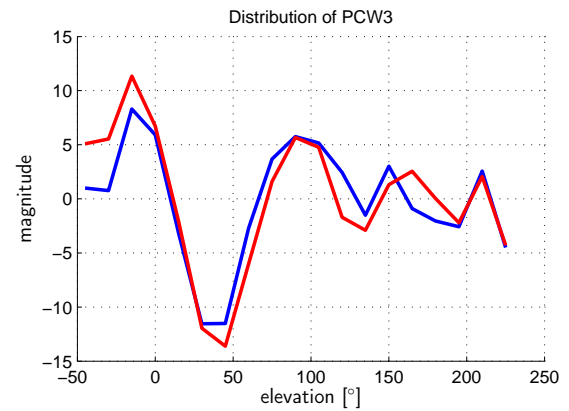
(a) Left and right ear PCW1 across all azimuths and elevations in IRCAM database.



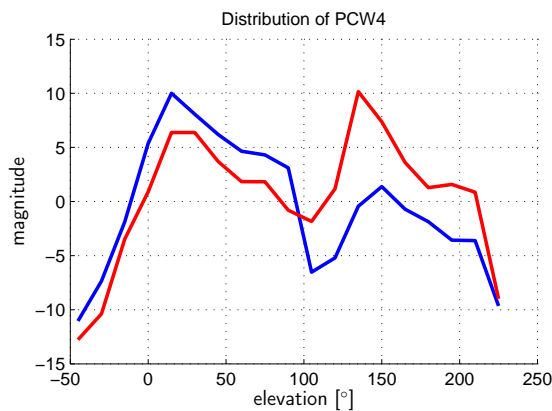
(b) Distribution of left and right PCW1 in the median plane.



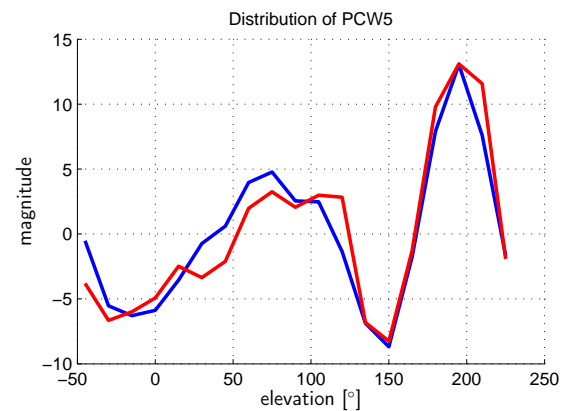
(c) Distribution of left and right PCW2 in the median plane.



(d) Distribution of left and right PCW3 in the median plane.



(e) Distribution of left and right PCW4 in the median plane.



(f) Distribution of left and right PCW5 in the median plane.

Figure 12: Ear symmetry: Left (blue) and right ear (red) PCWs in the median plane. IRCAM database with logarithmic DTF magnitude was used for PCA.

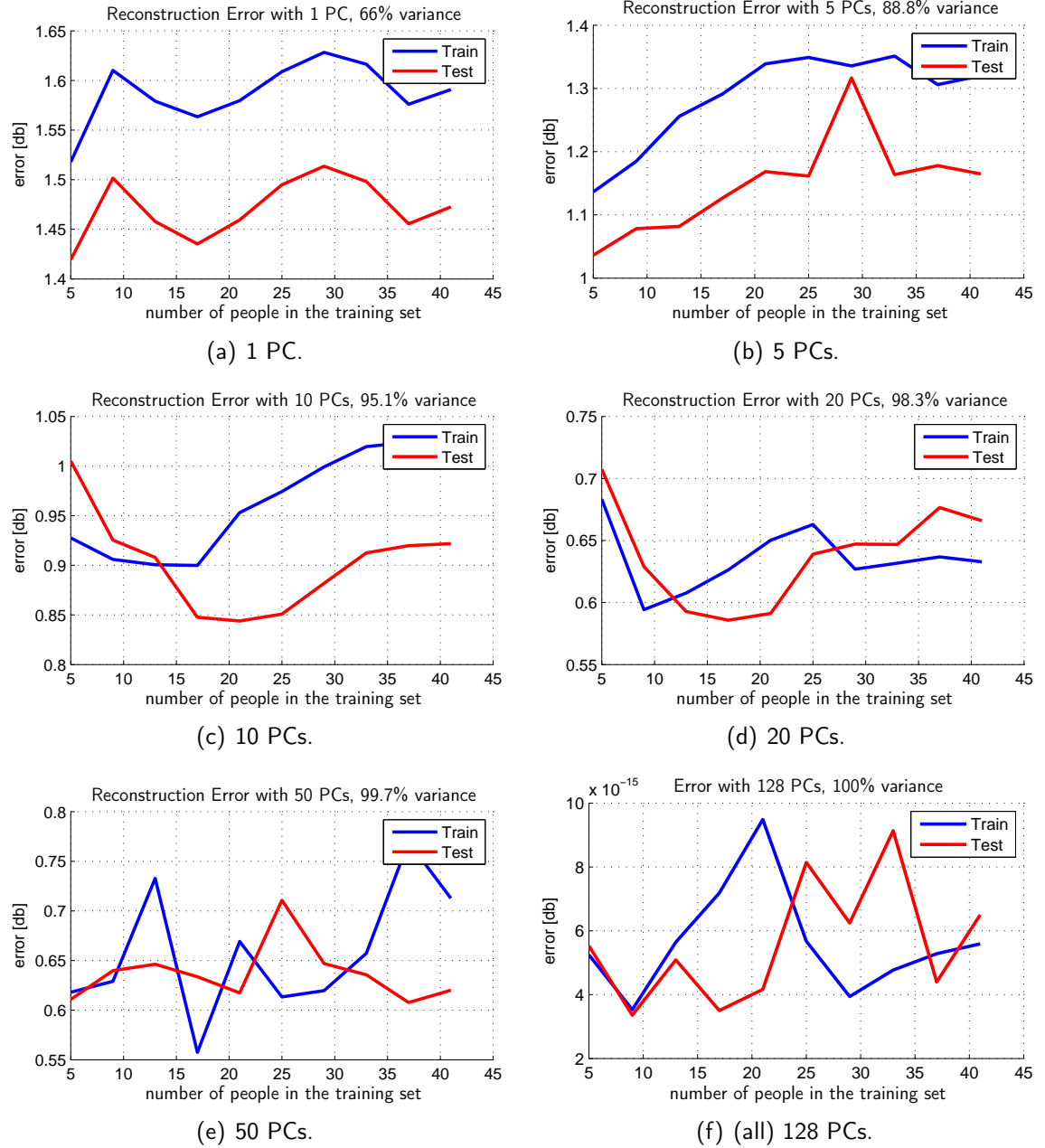


Figure 13: Mean error overall source positions when using different numbers of PCs for HRTF reconstruction in **CIPIC** database. Blue and red lines indicate training and testing set respectively.

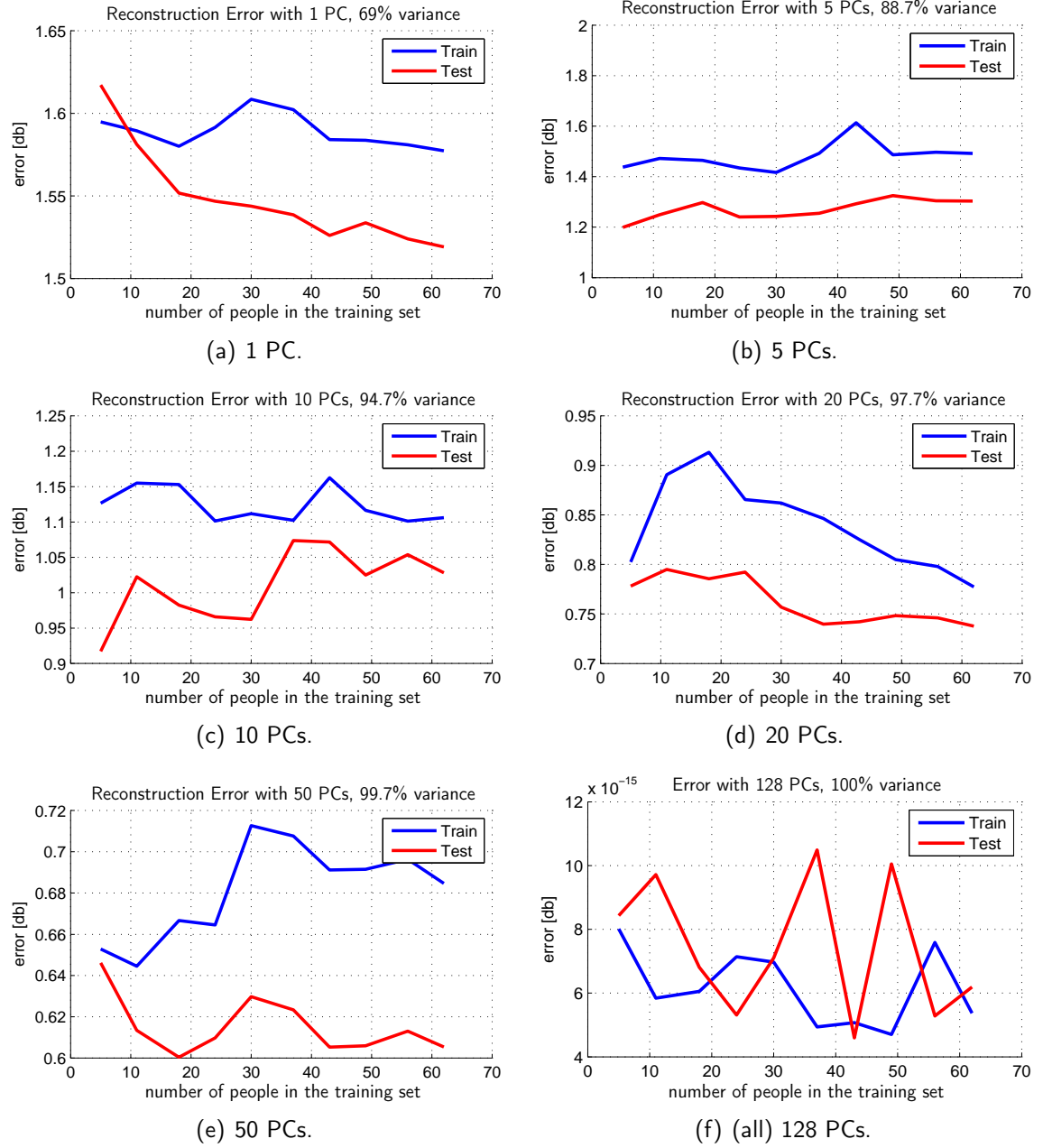


Figure 14: Mean error overall source positions when using different numbers of PCs for HRTF reconstruction in **ARI** database. Blue and red lines indicate training and testing set respectively.

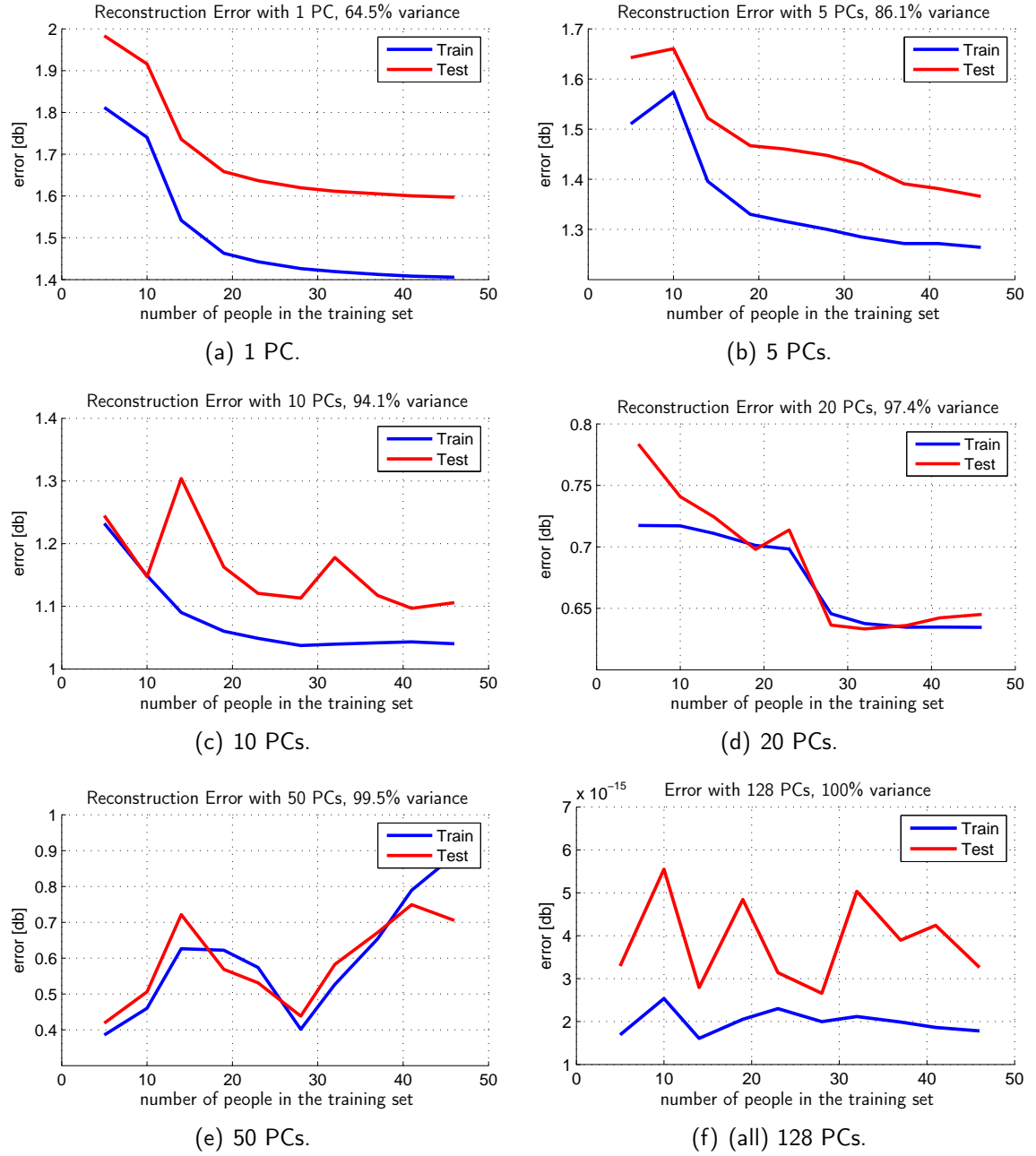


Figure 15: Mean error overall source positions when using different numbers of PCs for HRTF reconstruction in IRCAM database. Blue and red lines indicate training and testing set respectively.

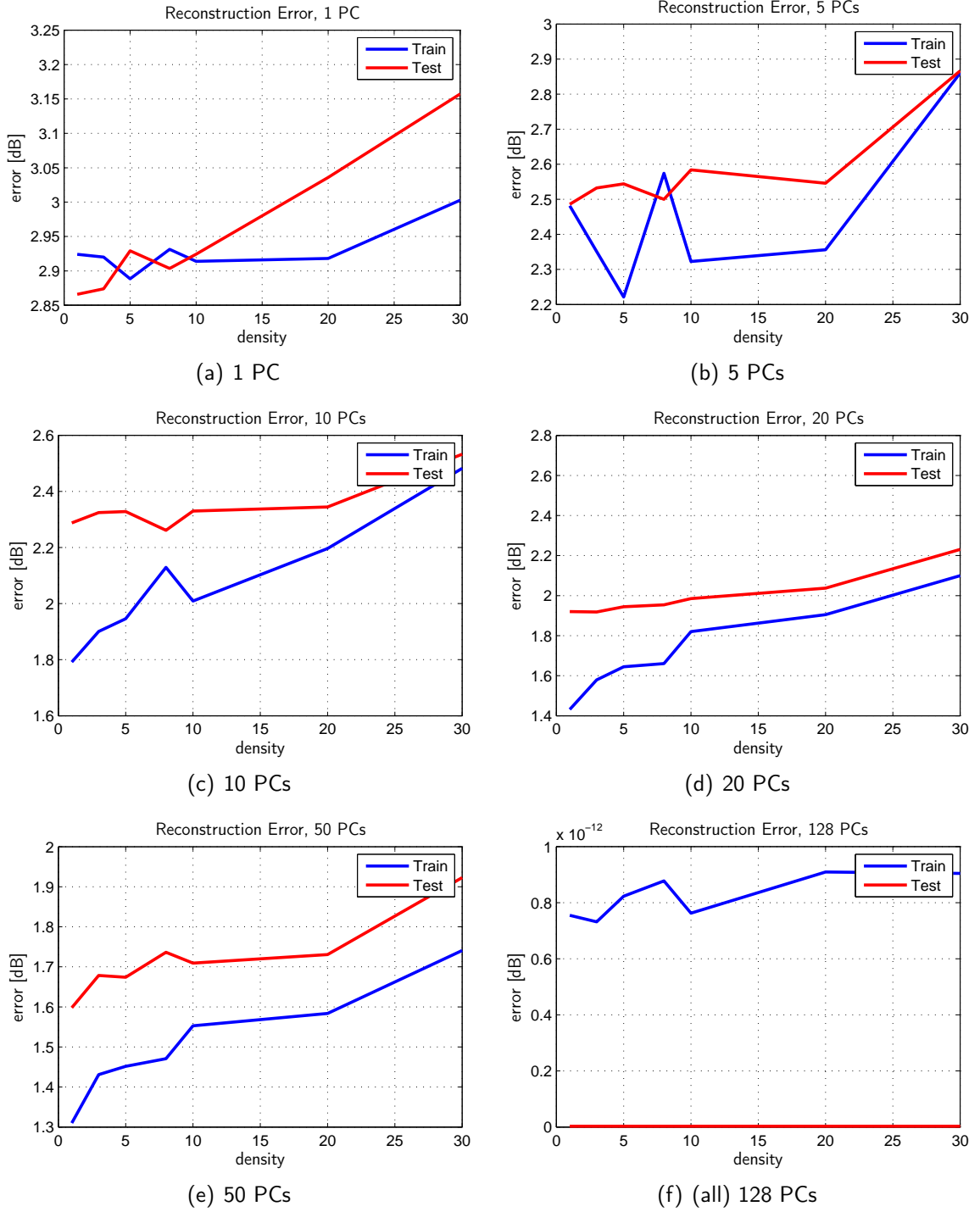


Figure 16: HRTF reconstruction error as a function of density that is used for the training data in **GLOBAL** database. Blue and red lines indicate training and testing set respectively.

## 5.2 Evaluation of Input Matrices

It has been shown that different structures of input matrices lead to different results. Consequently, three structures for input matrices including three different input data were compared and evaluated.

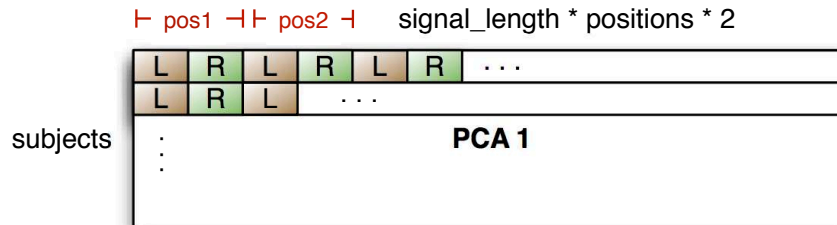


Figure 17: **PCA1** input matrix with corresponding dimensions.

In the first approach, a matrix **PCA1** was formed with dimension presented in Figure 17. The columns define the number of people and the lines show their HRTF sets. All HRTF pairs are stringed together for each position. The length of each signal depends on the choice of the input signal. When using HRIRs, the length is database-specific (200-512 samples, but when choosing DTFs, the length depends on the number of Fourier coefficients used for FFT.

In Figure 18, the standard deviation of the weights of the first principal component across all subjects is shown. Because the input data was centered before PCA, the mean of these weights is zero. Ideally, the graph shows a Gaussian normal distribution.

It is necessary to prove that the weights are within a proper range. An outliers can adulterate the mean value and increase the standard deviation. This results in a more difficult adjustment, because the tuning range has increased dramatically. An outlier can be caused by measurement errors or unique anatomical characteristics of a person. In this implementation an outlier detection was added that automatically deletes unusual weights for the first principal component. Afterwards, PCA is performed again without the detected entity.

The major advantage of this structure is that PCA returns only one weight per subject for each principal component. Consequently, only one weight can alter all positions of a person. However, the more positions are given for one person, the lower the variance of the first components. When using 10 PCs for reconstruction, only 50-60 percent of variance can be achieved. In Table ??, the total variance of the first 10 components for different input data of ARI database is listed.

In order to enhance variance for the first components, band-limiting can be used. The HRTF is smoothed in frequency domain by calculating the Fourier transform of the frequency spectrum but using only limited Fourier coefficients for inverse transformation. According to Kulkarni [KC98], the localization performance is not significantly affected, even when the number of Fourier coefficients is reduced from 256 to 16. However, the result is not overwhelming. When using a database with 116 subjects and a trajectory with 6 source positions, the components are increased only by up to 10 percent.

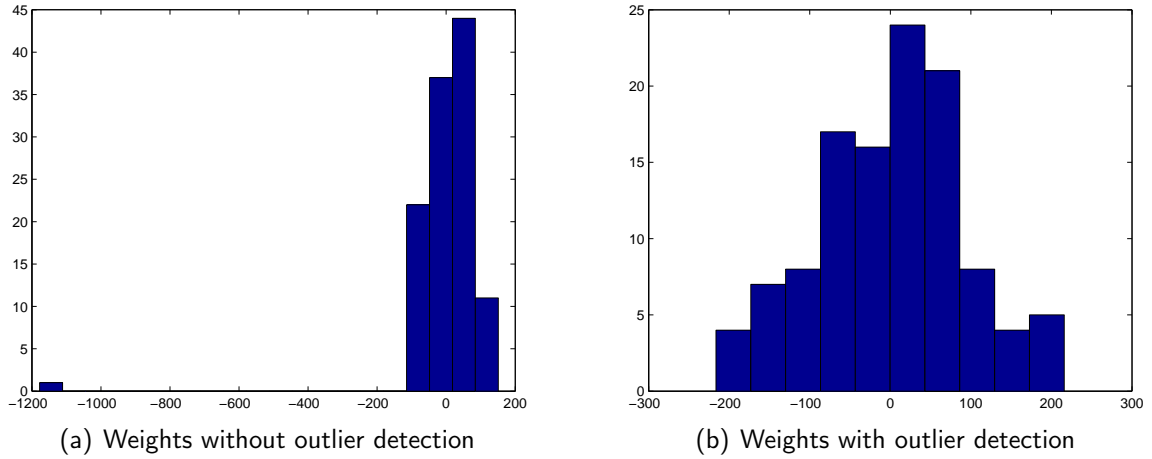


Figure 18: Histogram of the weights for the first principal component. Global database with 116 subjects and DTFs of 7 source positions ( $15^\circ$  azimuth,  $-30 - 45^\circ$  elevation) was used for PCA input data.

As shown in Figure 19, the second structure has a simpler construction. Each row of the matrix **PCA2** consists of one signal. For that reason, the matrix has a lot of more rows as columns. This turns out to be a good choice for PCA, because the variance of the first PCWs is much more higher than using the structure of **PCA1**. However, it is not possible to adjust the weights as easy as in the first format. Instead of receiving one weight for each person, each column has its own principal weight. Therefore the mean and standard deviation of the first PCs can not longer used as a dimension for adjusting one subject. In fact, each ear, position and subject has its own weights. In order to obtain a suitable distribution of weights, the matrix has to be grouped by subjects or positions.

**PCA3** has almost the same structure as **PCA2**, except that it generates weights for both HRTF pairs. This could lead to an improvement in the adjustment process because changing a particular weight has impact on both ear signals simultaneously.

### 5.3 Methodology

The model for individualization of HRTFs presented in this work, is based on the studies of Kistler *et al.* [KW92] and Rodriguez *et al.* [Rod05]. HRTFs are assumed to be minimum-phase functions [OS75] and ITD will be estimated as a constant, thus frequency independent delay. In previous experiments headphone equalization was a critical issue. Only small variations in positions of the headphone can lead to serious artifacts and perceptual distortions. According to [CJ98], a possible solution for this problem is to *short circuit* the pinna by using insert earphones. For this model, the headphone transfer function was equalized because this function is closely related to localization performance.

The major disadvantage of the discussed models in Section 3.1.1 is that tuning can only

be performed on one position or on a small region of positions, such as trajectories. If another position should be adjusted, the PCA must be recalculated with new HRTFs. By using a *global* model, all positions in the database are included for PCA. Then the subject can choose one position of interest for further tuning. Only the weights for the selected position are adjusted.

In order to allow more scope and flexibility for adjustment, weights of surrounding source positions can be involved. The subject controls the expansion of the included weights by drawing a virtual rectangular which enlarges the source position with maximum  $\pm 180$  degrees (Figure 20). Consequently, if the slider for azimuth extension is set as a maximum, all positions in azimuth are involved. According to Kistler and Wightman [KW92], the first PC mainly contains information about interaural intensity differences. Using the structure **PCA2** or **PCA3** with all azimuth positions for adjustment, this statement can be easily reproduced. When looking at the values of PCW1 more precisely, one can see that the weights for left and right ear have almost the same values, but with different signs. In summary, the first component provides the biggest variance between the source positions, so in azimuth plane this means mostly level differences through head shadowing and diffraction.

PCs that do not contribute significantly to the reconstruction are excluded. Thus only the first 10 PCs were used with a variance of 70-90 percent, but this very much depends on the database and source positions. This coincides with previous studies ([KW92], [MG92]).



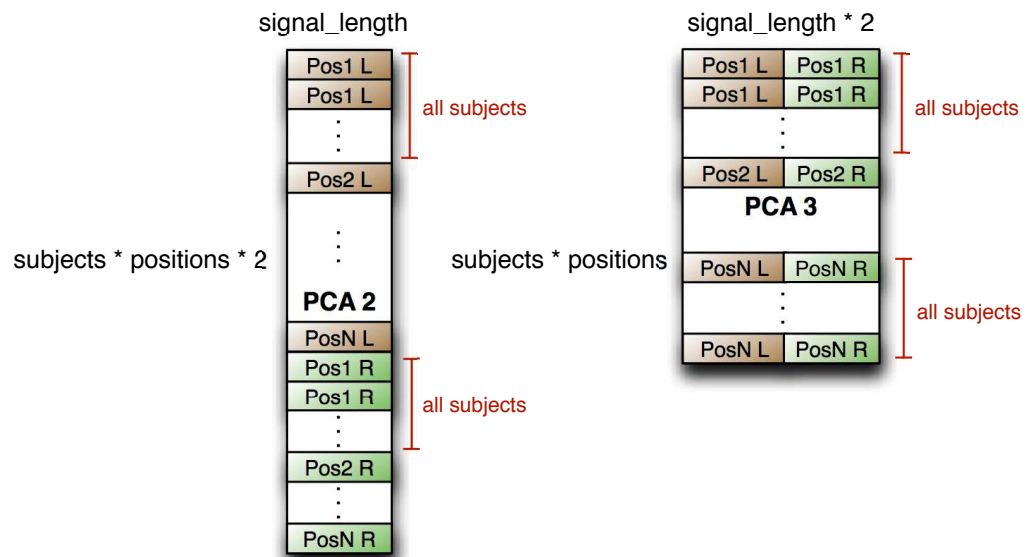
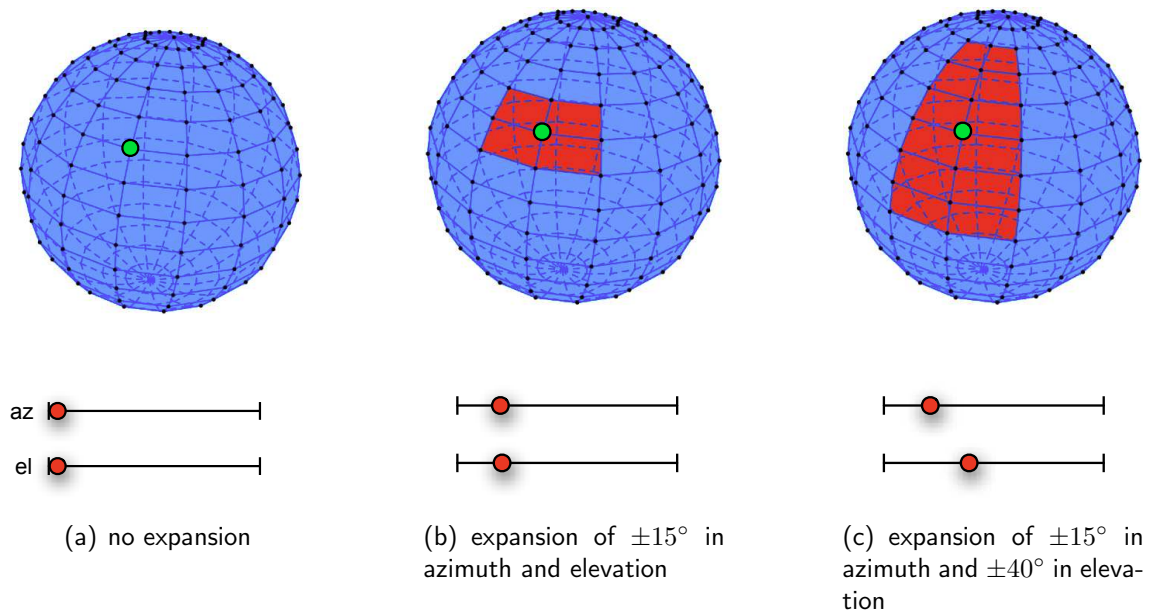
Figure 19: Input matrices **PCA2** and **PCA3** with corresponding dimensions.

Figure 20: Extension of selected weights (red plane) used to adjust an individual source position (green point).

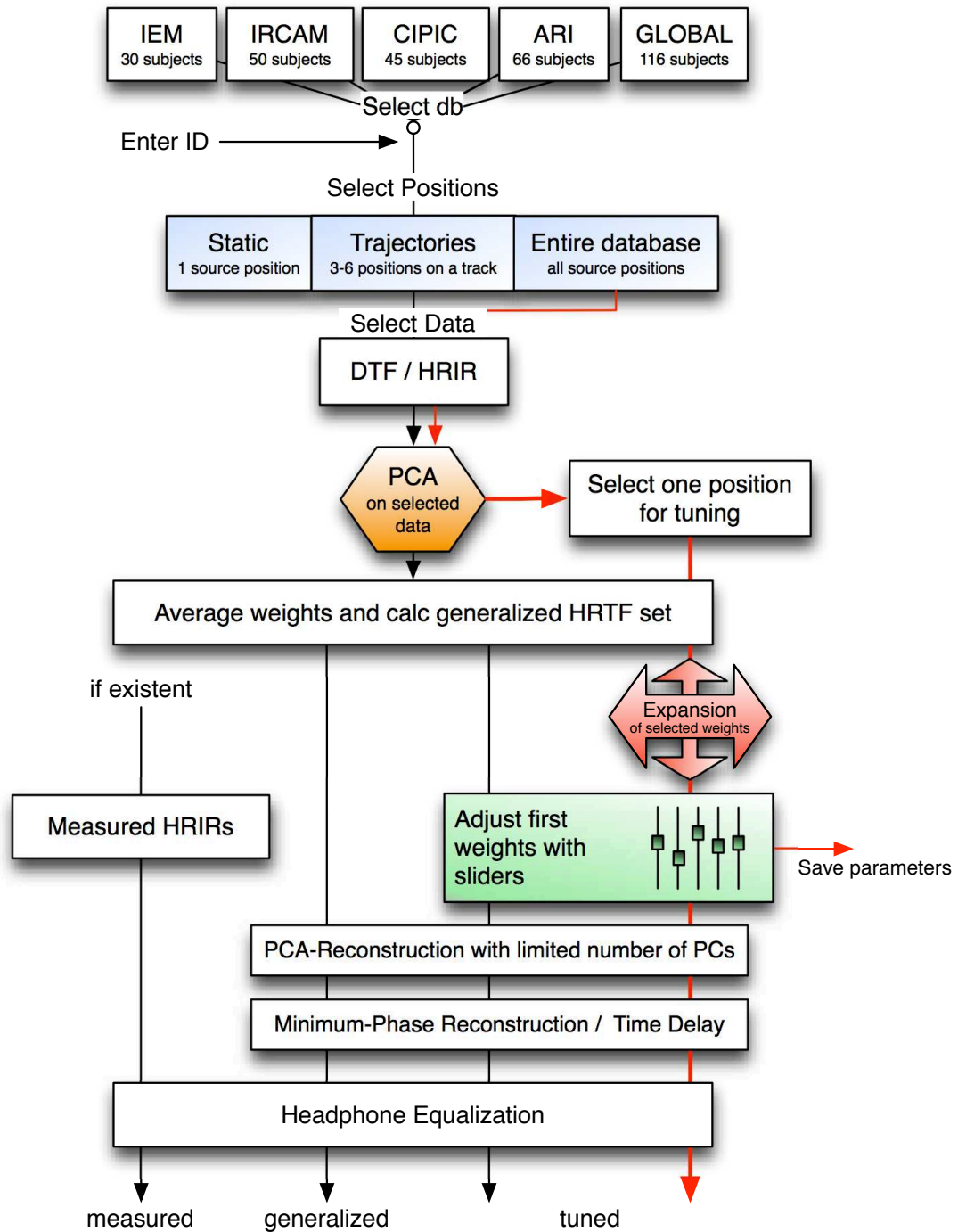


Figure 21: Overview of tuning process with various input data. Red path indicates the global model.

## 5.4 Self Tuning of PCWs

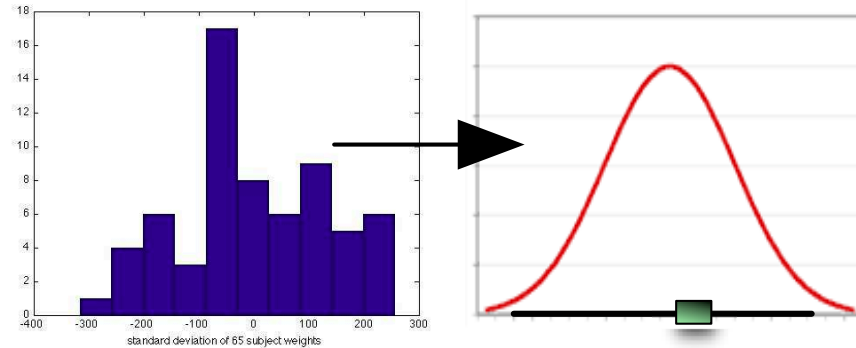


Figure 22: Methodology of adjusting the principal weights by sliders.

The principal task is to provide the test person as few information as possible while having the possibility to tune the HRTF as much. Thus it is a major decision which and how many PCs the subject should adjust. The PC corresponding to a large standard-deviation of PCWs contributes significantly to the inter-subject variation [HPP08]. For that reason, only the first 10 PCs with the largest standard-deviation for each elevation is displayed to the test person.

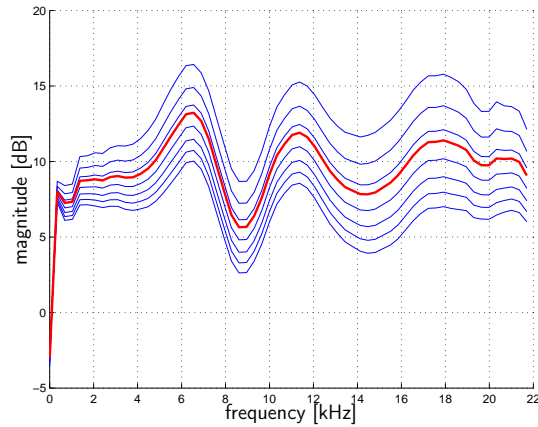
In the test procedure the listener could tune the weights for reconstructing the HRTF on his own through a MATLAB GUI (Figure 25). In order to prevent the subject from stress, there is no time limit for adjusting the weights. However, the process is time-consuming and exhausting, because of the high concentration is required.

In order to adapt the weight of a component, a slider has to be changed. At the beginning, the slider position is in the middle (mean across all weights of the component). The minimum and maximum values are set to be  $\text{mean} \pm 3$  standard deviation of each PCW. While changing the slider position, different weights are used for reconstruction. A major advantage of using **PCA1** is that all source positions in a sound trajectory can be altered by changing only one slider. Consequently adjustment process can speed up. In **PCA2** and **PCA3**, one source position can be tuned. If the slider is left in the middle position, an averaged weight across all subjects is used for the corresponding component. In this case, the reconstruction returns only a generalized HRTF set.

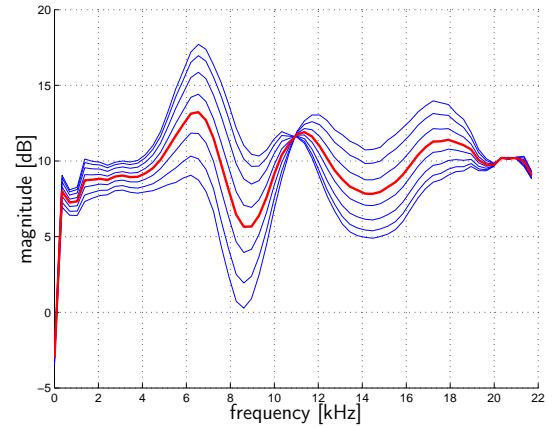
Every time when the subject changes the slider positions, a new HRIR pair are calculated through Inverse Fourier Transform of the adapted HRTFs. The prior subtracted mean is added again to the DTF to obtain the HRTF. The phase information is obtained by calculating the Hilbert transform of the corresponding logarithmic magnitude spectrum.

In Figure 23, the changes in logarithmic magnitude spectrum for each component is shown. It is easy to see that for example the first component represent just a constant scaling over the whole frequency range. The changes in the other weights are more complex.

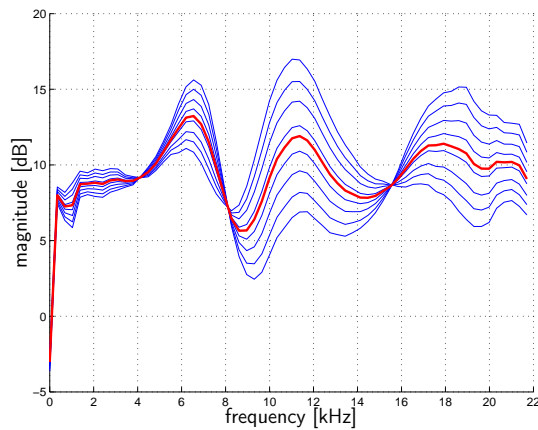
The subject can listen to tuned HRIRs on every change or press the "Play" buttons to



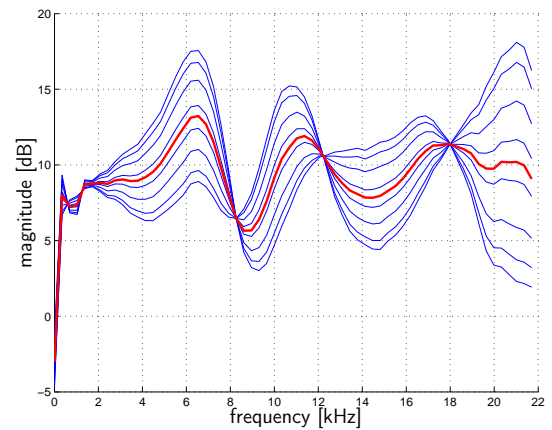
(a) Variation of PCW 1



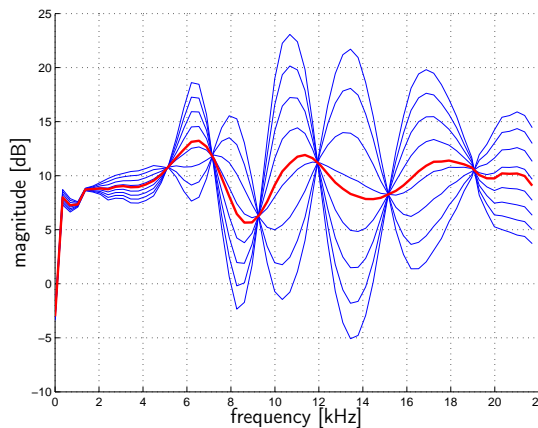
(b) Variation of PCW 2



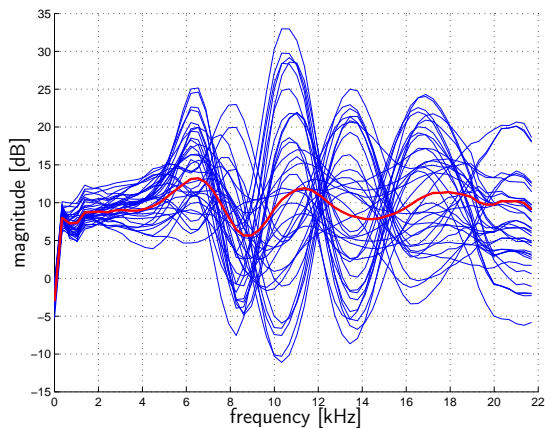
(c) Variation of PCW 3



(d) Variation of PCW 4



(e) Variation of PCW 5



(f) Variation of PCW 1-5

Figure 23: Mean value (red line) and minima/maxima amplitudes (blue lines) of left ear magnitude spectrum when changing the first five PCWs separately (a-e) or simultaneously (f). IRCAM database with source position  $90^\circ$  azimuth and  $0^\circ$  elevation was used for PCA.

compare the various stimuli or listen to the original HRTF as a reference. A broad band noise is used as stimulus, but several other sounds can be selected.

If the subjects measurement data are available in the database, the button "My Solution" appears on the interface. This function sets the slider positions to the subjects weights obtained from PCA. After finding appropriate slider positions, the customized data can be saved through pushing the "Save" button and is stored in the filesystem.

There are some discrepancies between tuned and individual HRTFs. This might be resulted in the limit of parameters (incompleteness of the customization process) and limited principal components for reconstruction. When calculating the DTFs by removing subjects mean, in the reconstruction the mean of the test person is not known. Thus, the mean of all subjects mean values is added to obtain the HRTF. Moreover, in **PCA2** only the left ear HRTFs are used for PCA, so the error of the right ear HRTFs increases dramatically. However, the purpose of this project is not to recover the exact measured HRIRs but to customize generalized HRTFs for improved localization performance. The error increases as the number of basis functions is reduced. When using only few PCs (above 5), significant differences in the spectral details are visibly.

## 5.5 HRTF Tuning Tool

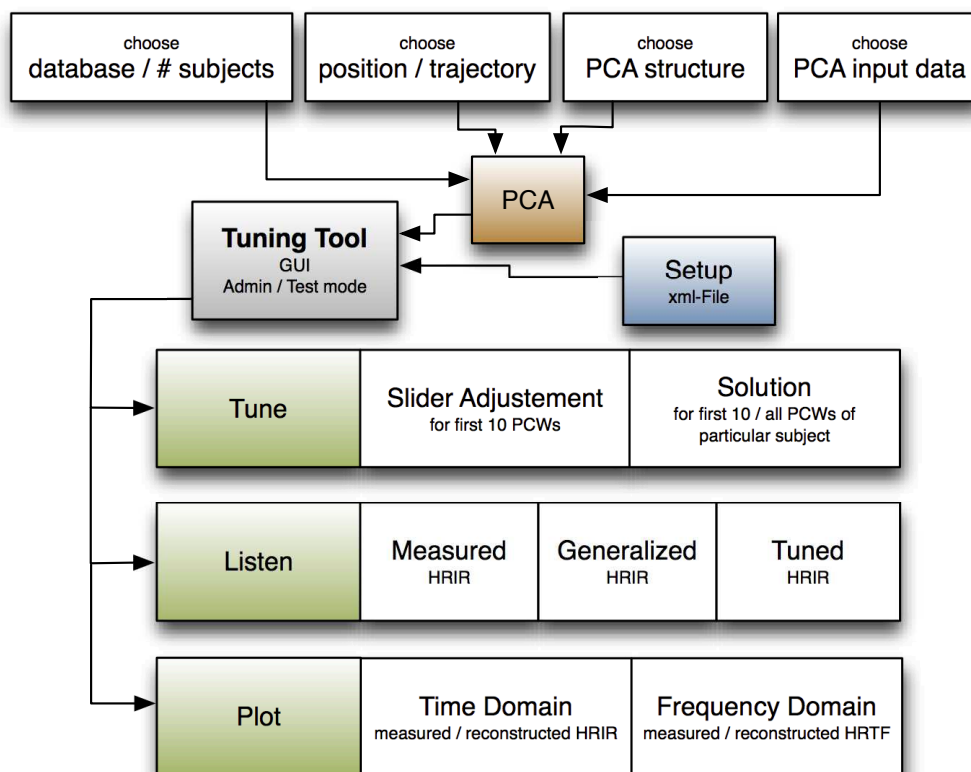


Figure 24: HRTF Tuning Tool Tool: Overview of functionality.

The HRTF tool works with three modes. Either a single static source position, a sound trajectory or the entire database can be chosen for tuning process. The trajectories consists of 5-20 static positions in a coherent way e.g. all available elevations for a specified azimuth position or the other way around.

Figure 24 provides an overview of the main functions of the HRTF tuning tool. At startup, the ID of the participant has to be specified. The subject has to choose the database and a static source position or trajectory of interest. Afterwards the first 10 significant PCWs are calculated and selected for the slide-bar automatically. Now the test person can begin listening to stimuli and adjust sliders.

Three different structures and data for the PCA input matrix can be chosen, described in section 5.2. Additionally, following features can be enabled:

- Number of FFT points for HRTF
- Band limiting of the PCA input matrix
- Outlier detection for the first weights
- Expansion of PCA weights (in azimuth and elevation plane) for tuning a source position
- Selection of different sound stimuli
- Headphone equalization (for AKG "K271 Studio" and "K272 HD")
- Plots of HRIRs, HRTFs and PCWs

After PCA processing, the variance of each of the first 10 components is indicated on the right side. By changing the number of components for reconstruction, the total variance of the components can be obtained. The integration of HRTF databases can be configured in an external setup file with in xml structure.

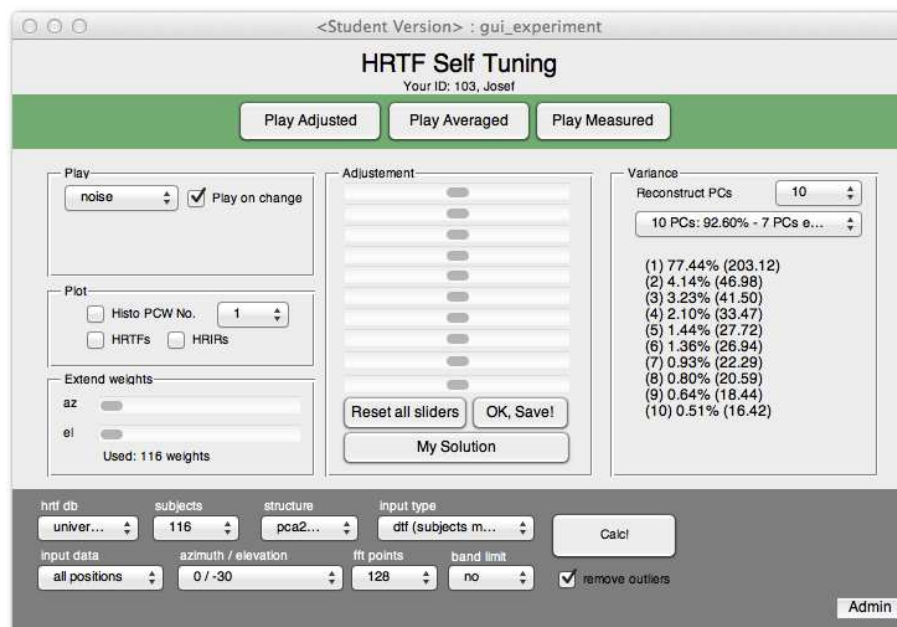


Figure 25: MATLAB GUI for Experiment (Admin view). The various settings can be faded out during listening tests, so only the slider box in the middle is visible.



## 6 HRTF Analysis Tool



Figure 26: HRTF Analysis Tool in MATLAB: GUI in standard view.

For visual and aural inspection of HRTF data, a tool in MATLAB was built. In general, the software supports people who are listening and analyzing HRTF databases. The implementation provides measurement data of five different departments that are listed in Table 4 on Page 28.

Figure 27 gives an overview of the main functions. The tool is designed so that new databases can be added with little effort. A configuration file contains the relevant information for each database in xml structure. It is necessary to add the import function and adapt database specific code, such as anthropometric dimension and measurement positions. Two matrixes have to be formed with the structure

- **DB** as a 4D-matrix ( $subjects \times sourcepositions \times ears \times hrirs$ ),
- **ANGLES** as a 2D-matrix ( $sourcepositions \times 2$ ) that specifies the source positions with values for azimuth and elevation.

If a new database fits this structure, all operations in the GUI are working without adaption. A brief instruction for adding a new database can be found in readme.txt file.

### 6.1 Basic Operations

In the first step, a database has to be imported. The mat-files of the entire measured data are read from the filesystem and stored as a single matrix in global workspace.

After import, it is possible to choose a subject and pick elevation and azimuth positions

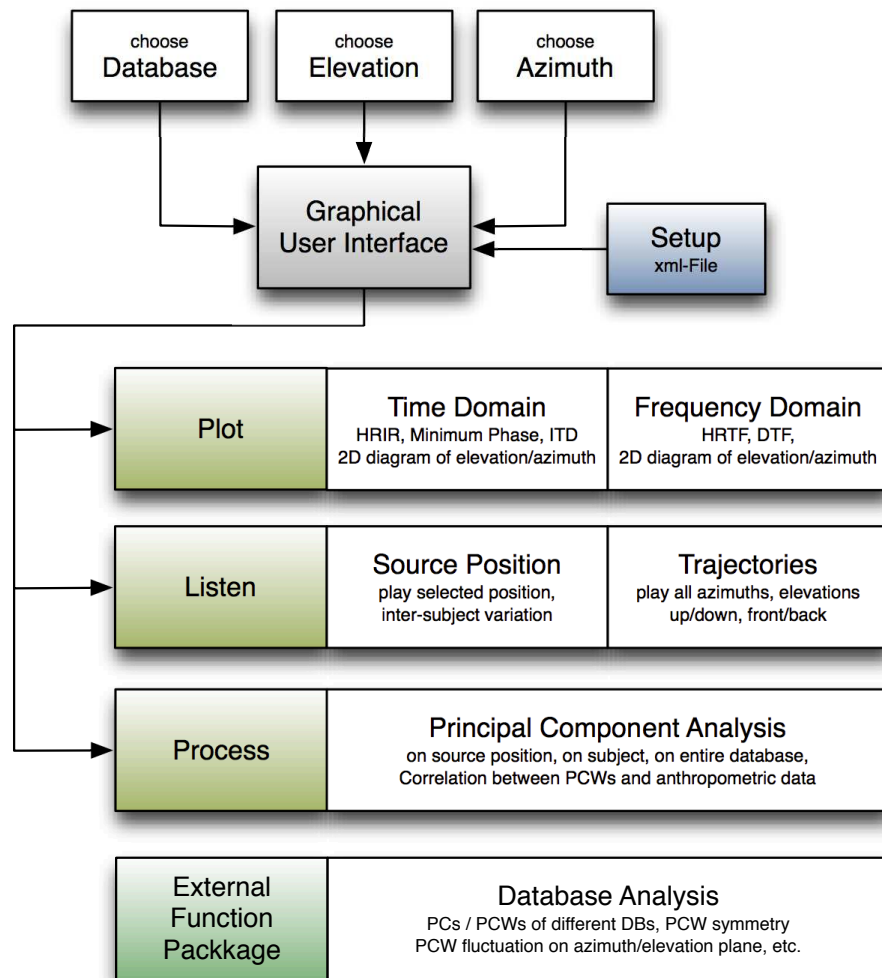


Figure 27: HRTF Analysis Tool: Overview of functionality.

from a listbox. On the right side several graphs are available for visualizing HRIRs and HRTFs. By default, the left and right HRIRs are presented in time domain and the calculated HRTFs are visualized in frequency domain. The FFT is processed with 1024 points. Because of the real input signal, the FFT output magnitude returns symmetry, thus a 512 point HRTF magnitude is obtained.

These diagrams are available for head-related impulse responses:

- Overall interaural time difference of all azimuth positions of selected subject
- 2D diagram of all azimuth HRIRs of chosen subject: a red line indicates the current azimuth position
- 2D diagram of all elevation HRIRs of chosen subject: a red line indicates the current elevation position
- Minimum phase version of HRIR

In addition these diagrams are available for head-related transfer functions:

- Directional transfer function (DTF) and HRTF



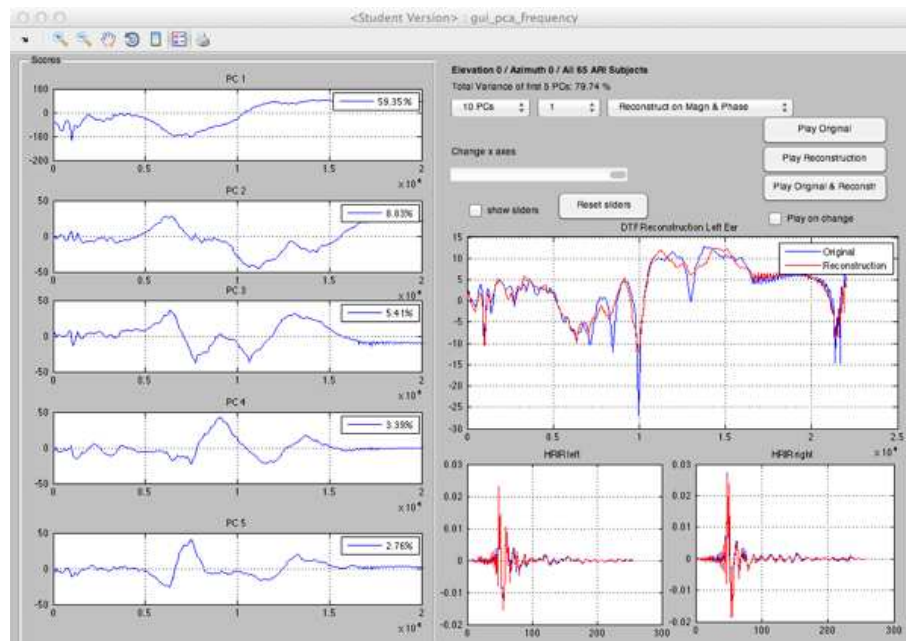


Figure 28: Matlab GUI for analyzing PCA data for a source position.

- 2D diagram of all azimuth HRTFs / DTFs of chosen subject: a red line indicates the current azimuth position
- 2D diagram of all elevation HRTFs / DTFs of chosen subject: a red line indicates the current elevation position

It is possible to expand each of the four diagrams and switch between linear and logarithmic view. By selecting one position and changing the subject, it is easy to discover the inter-subject differences of head-related transfer functions.

Peaks and notches of the magnitude of HRTF or DTF can be marked with an algorithm that finds local maxima and minima. Moreover, the threshold for detection can be adjusted. The magnitude can also be smoothed by reducing the Fourier coefficients in the reconstruction.

Several listening tasks can be executed for the selected subject:

- Play HRIR of chosen position
- Play HRIRs of chosen position and all subjects
- Play all available azimuth HRIRs of select elevation angle
- Play all available elevation HRIRs of select azimuth angle
- Play front-back HRIRs
- Play up-down HRIRs

## 6.2 PCA Operations

The implemented PCA functions are based on the current selection of database, subject and source position. Four different data analysis are described below:



Figure 29: Matlab GUI for visualizing correlation between principal component weights and anthropometric data for left and right ear respectively. The yellow background color indicates a correlation of more than 60 percent.

- PCA on entire database
- PCA on entire measurement data of subject
- PCA on all subjects of selected position
- PCA on all subjects of selected position and visualization of correlation with existing anthropometric data (only for CIPIC and ARI)

### 6.3 Visualizing Correlations

- Correlation of weights and principal components
- Correlation of weights and anthropometric data
- Fluctuations of PCWs of all subjects across all elevations

## 7 Conclusion

Based on existing studies about synthesizing HRTFs, an objective model for HRTF was presented. The report reviews previous methods for HRTF individualization and discusses the importance of using adapted models for accurate localization. Despite current limitations of such models, the research of 3D virtual auditory display has a promising future.

In order to understand the behavior and variations of PCWs, existing HRTF databases

were analyzed. Further investigation how certain PCWs affect localization would be appropriate. The Least-Squares method confirms that any arbitrary HRTF can be modelled by adjusting PCWs. For a better statistical validity, an algorithm such as *Bootstrapping* should be used.

A HRTF model that uses PCA was proposed. It was shown, that individual HRTFs for each source position or sound trajectories (4-6 source positions) can be modeled by simple tuning of preselected general basis functions. A subjective localization test should be carried out to assess the performance of different conditions. By using the algorithm of LyTTE project<sup>2</sup> which simulate the acoustic of rooms, externalization may well be improved. An important step would be an interpolation between the measured source positions.

A toolbox for analyzing and listening HRTF data in MATLAB was introduced. A key benefit would be a larger HRTF database of different ethnic groups in order to analyze cultural and gender differences.

Another way to learn HRTFs from other persons or to improve the own one would be a mobile app which should simply present stimuli and track subjects performance over a long period time.

---

2. LyTTe is an open-source project which deals with architectural acoustics. More information on <http://sourceforge.net/projects/lytteproject>

## A Principal Component Analysis (PCA)

In some problems it is advisable to use a method that classifies the various dimensions on a given amount of data according to their relevance. The Principal Component Analysis (PCA) exactly returns this result [Jan04]. It is a useful statistical technique to find specific patterns in mass of data with high dimension. Basically, correlated variables are transformed into uncorrelated ones, called *principal components*. The aim of PCA is to reduce dimensions and calculate components with largest and lowest impact on the data record. Often, several components of a data set are irrelevant, because they do not provide more information or are almost constant. If the new components are found through transformation, the data set can be represented in another way, usually by fewer dimensions. Thus, PCA compresses the data without discarding significant information.

PCA is generally an analysis of the variance in the data set. It highlights the directional information, so that the first PCA component has the greatest variance, the other components have a decreasing variance with respect to the orthogonality to all other components.

If the record has been released from mean before the transformation, the PCA components are uncorrelated, thus normally distributed. For example, the origin of a coordinate system including a three-dimensional data, is set to the focus of the data before transformation. In a data matrix, the column means must be subtracted. Only this way ensures that the first components are along the largest variance.

A few new definitions are introduced: *Component scores* are the transformed variables and *loadings* describe the weights to multiply the normed original variable to get the component score.

The important decision criterion in PCA is the deviation of the components of a vector to its arithmetic mean, also known as the *variance*. Similarly, the *covariance* is defined as the difference between the variances of two vectors. If the value is positive, it indicates that the two dimensions are increasing together. If the covariance is zero, the measured dimensions are *independent* of each other [Lin07].

The *correlation coefficient* describes the linear relationship between two variables, giving a value between -1 and 1:

$$r = \frac{Cov(x, y)}{\sqrt{Var(x)Var(y)}}. \quad (12)$$

There are two common methods to calculate the principal components of a data set. PCA can be performed by eigen value decomposition of a covariance matrix or singular value decomposition.

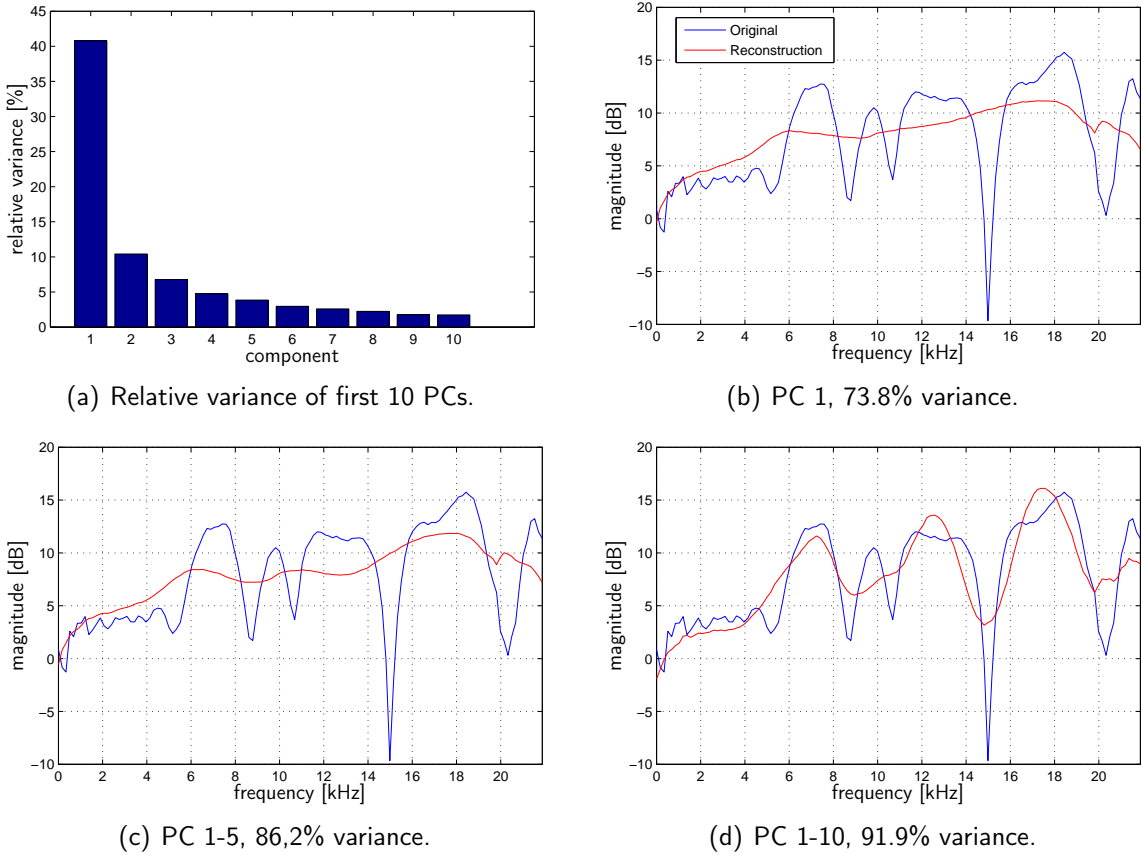


Figure 30: Reconstruction of one left ear DTF (90° azimuth, 0° elevation) in IRCAM database by comparing different numbers (1, 5, 10) of principal components. Red and blue lines indicate reconstruction and original data respectively.

## A.1 Covariance Matrix

The first step is to subtract the column mean of each column of the input data matrix  $\mathbf{X}$  ( $N \times M$ ).  $N$  denotes the sample length and  $M$  indicates the number of observations.

$$\mathbf{B} = \mathbf{X} - u * h, \quad (13)$$

where  $u$  indicates the mean of the input matrix and  $h$  is a  $1 \times M$  row vector of all 1's.

The subtracted means should be buffered, because it is essential for reconstruction and has to be added again. Next, the covariance matrix  $\mathbf{C}$  ( $M \times M$ ) is calculated as

$$\mathbf{C} = \frac{1}{N} \sum \mathbf{B} * \mathbf{B}^*. \quad (14)$$

If the variables are not centered, this matrix can not longer be regarded as the the covariance matrix. Consequently the variables are not statistically decorrelated, but remain orthogonal [LJMG00].

As each vector in  $\mathbf{C}$  is compared to all others, it describes how much the dimensions vary from the mean with respect to each other. The matrix is square and symmetrical about the main diagonal. It consists of all the covariances of different dimension. If the non-diagonal elements have positive values, the data variables are increasing together.

The eigenvectors  $\mathbf{V}$  of the covariance matrix are unit vectors and linearly independent. They are obtained by

$$\mathbf{V}^{-1} \mathbf{C} \mathbf{V} = \mathbf{D}, \quad (15)$$

with  $\mathbf{D}$  as a diagonal matrix including the eigenvalues of  $\mathbf{C}$ . By sorting the matrix  $\mathbf{D}$  and  $\mathbf{V}$  according to the largest eigenvalue in descending order, the corresponding part of the variance of an eigenvector relating to the total variance can be calculated.

The eigenvector with the highest eigenvalue is called the first principal component, the second component indicates the direction of the second biggest variance, and so on. In [QE98], it is described that the first component mainly contains azimuth information, and the second and third components are more associated with high and low elevations.

The  $q$  eigenvectors of  $\mathbf{C}$  are the *basis functions*  $v_i$ . If  $q = N$ , the original data can be fully reconstructed. However, the objective of PCA is to reduce the dimensionality, therefore a reduced number of dimensions  $L$  ( $1 \leq L \leq N$ ) can be chosen for a specific situation. For example, thus much components, such as 90% of the data is presented. Madsen [MH03] explains that also the stability of the singular values can be a criterion to set the number of components. Later, it is shown that only 5-10 components (depends on the input data) are sufficient for a close approximation of a HRTF. Note that the distribution of the PCWs becomes smaller as the eigenvalues decreases.

The scores  $\mathbf{Z}$  are calculated as

$$\mathbf{Z} = \frac{\mathbf{B}}{s * h}, \quad (16)$$

with  $s$  as roots of the diagonal values of covariance matrix  $\mathbf{C}$ . The projected score

$$\mathbf{Y} = \mathbf{W}^* * \mathbf{Z}, \quad (17)$$

with  $\mathbf{W}^*$  as conjugate transposed matrix of  $\mathbf{W}$ . Finally, the approximation of the original data can be expressed as

$$\hat{\mathbf{X}} = \hat{\mathbf{B}} + \mathbf{u} * \mathbf{h}. \quad (18)$$

## A.2 Singular Value Decomposition (SVD)

The *Singular value decomposition* is a powerful data analysis method and for that reason relevant to PCA. First of all, a real ( $n \times m$ ) matrix  $\mathbf{X}$  where  $n \geq m$  can also be written as

$$\mathbf{X} = \mathbf{U} \mathbf{\Gamma} \mathbf{V}^T, \quad (19)$$

with

- $\mathbf{U}$  as a  $n \times m$  matrix with  $n$  observations and  $m$  variables,
- $\mathbf{\Gamma}$  as a  $m \times m$  diagonal matrix with nonnegative and real values (singular values), also known as the square roots of the eigenvalues,
- $\mathbf{V}$  as a  $m \times m$  matrix, the eigenvectors,
- $r$  as the rank of  $\mathbf{X}$  (number of linear independent rows of the matrix).

Matrices  $\mathbf{U}$  and  $\mathbf{V}$  have orthonormal columns so that  $\mathbf{U}^T \mathbf{U} = \mathbf{I}_r$  and  $\mathbf{V}^T \mathbf{V} = \mathbf{I}_r$ .

From  $\mathbf{X}$ , two positive-definite symmetric matrices can be formed:

$$\mathbf{X}\mathbf{X}^T = \mathbf{U}\mathbf{\Gamma}\mathbf{V}^T \mathbf{V}\mathbf{\Gamma}\mathbf{U}^T = \mathbf{U}\mathbf{\Gamma}^2\mathbf{U}^T, \quad (20)$$

$$\mathbf{X}^T\mathbf{X} = \mathbf{V}\mathbf{\Gamma}\mathbf{U}^T \mathbf{U}\mathbf{\Gamma}\mathbf{V}^T = \mathbf{V}\mathbf{\Gamma}^2\mathbf{V}^T. \quad (21)$$

Assuming  $n \geq m$ ,  $\mathbf{X}\mathbf{X}^T$  ( $n \times n$ ) and  $\mathbf{X}^T\mathbf{X}$  ( $m \times m$ ) share  $m$  eigenvalues, the remaining  $n - m$  eigenvalues will be zero [MH03]. The covariance matrix of the input data  $\mathbf{X}$  can be calculated as

$$\mathbf{C} = \frac{1}{n}\mathbf{X}\mathbf{X}^T = \frac{1}{n}\mathbf{U}\mathbf{\Gamma}^2\mathbf{U}^T \quad (22)$$

and the transformed data can be expressed as

$$\mathbf{Y} = \tilde{\mathbf{U}}^T \mathbf{X} = \tilde{\mathbf{U}}^T \mathbf{U} \mathbf{\Gamma} \mathbf{V}^T. \quad (23)$$

Mostly, the number of features is bigger than the number of samples ( $m \gg n$ ), like in image, text or sound processing [MH03]. Thus the covariance matrix  $\mathbf{C}$  becomes very large. In this case, it's sufficient to decompose the smaller  $m \times m$  matrix

$$\mathbf{D} = \frac{1}{m}\mathbf{X}^T\mathbf{X}. \quad (24)$$

In the HRTF model presented in this report, mostly the number of examples is smaller than the number of variables ( $m < n$ ), but if the PCA is performed with a total HRTF database as input matrix, the routines in MATLAB can slow down or interrupt. As in [MH03] mentioned, this problem can be avoided using the transposed input arguments on the matlab function `svd()`.

## References

- [ADT01] V. R. Algazi, R. O. Duda, and D. Thompson, "The cipc hrtf database," *Proc 2001 ieee ...*, 2001.
- [AS90] F. Asano and Y. Suzuki, "Role of spectral cues in median plane localization," *The Journal of the Acoustical Society of ...*, 1990.

- [BD98] C. Brown and R. O. Duda, "A structural model for binaural sound synthesis," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 5, pp. 476–488, 1998.
- [Beg94] D. R. Begault, *3-D sound for virtual reality and multimedia*. Morgan Kaufmann Pub, 1994.
- [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.
- [Bla83] ———, "Spatial hearing: the psychophysics of human sound localization," 1983.
- [Bro95] A. W. Bronkhorst, "Localization of real and virtual sound sources." *Journal of the Acoustical Society of America*, 1995.
- [Che99] C. I. Cheng, "Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space," *PREPRINTS-AUDIO ENGINEERING SOCIETY*, 1999.
- [CJ98] S. Carlile and C. Jin, "The generation and validation of high fidelity virtual auditory space," *Engineering in Medicine and . . .*, 1998.
- [CM10] J. A. S. Catarina Mendonça, "On the improvement of auditory accuracy with non-individualized HRTF-based sounds," in *Audio Engineering Society Convention Paper 8266*, 2010.
- [CvVH93] J. Chen, B. D. van Veen, and K. E. Hecox, "A spatial feature extraction and regularization model for the head-related transfer function," *J. Acoustic Soc.*, 1993.
- [DS07] R. O. Duda and P. Satarzadeh, "Physical and Filter Pinna Models Based on Anthropometry," . . . of the AES 122nd Convention, 2007.
- [EAT98] M. Evans, J. Angus, and A. Tew, *Analyzing head-related transfer function measurements using surface spherical harmonics*. JOURNAL- . . . , 1998.
- [Gie92] H. W. Gierlich, "The application of binaural technology," *Applied Acoustics*, vol. 36, no. 3-4, pp. 219–243, Jan. 1992.
- [GS10] M. Geronazzo and S. Spagnol, "Estimation and modeling of pinna-related transfer functions," 2010.
- [Heb74] J. Hebrank, "Spectral cues used in the localization of sound sources on the median plane," *The Journal of the Acoustical Society of America*, 1974.
- [HF98] R. Höldrich and M. Fellner, "Modellierung von HRTF - Kurven," Tech. Rep., 1998.
- [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.
- [Jan04] S. Jan, “Principal Component Analysis (PCA),” Tech. Rep., 2004.
- [KC98] A. Kulkarni and H. S. Colburn, “Role of spectral detail in sound-source localization,” *Nature*, 1998.
- [KC04] —, “Infinite-impulse-response models of the head-related transfer function,” *The Journal of the Acoustical Society of America*, vol. 115, no. 4, pp. 1714–1728, 2004.
- [Kuh77] G. Kuhn, “Model for the interaural time differences in the azimuthal,” *J Acoust Soc Am*, 1977.
- [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.
- [Kyr98] C. Kyriakakis, “Fundamental and technological limitations of immersive audio systems,” in *Proceedings of the IEEE*, 1998, pp. 941–951.
- [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.
- [LC09] J. Leung and S. Carlile, “PCA Compression of HRTF and localization performance,” in *International Workshop on the Principles 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.
- [Lin07] I. Lindsay, *A tutorial on principal components analysis*. Details available at [left angle bracket csnet. otago. ac. . . .](#), 2007.
- [LJMG00] V. Larcher, J. Jean-Marc, and J. Guyard, “Study and comparison of efficient methods for 3d audio spatialization based on linear decomposition of HRTF data,” *PREPRINTS-AUDIO . . .*, 2000.
- [Mar87] W. Martens, *Principal components analysis and resynthesis of spectral cues to perceived direction*. Proceedings of the 1987 International Computer Music . . . , 1987.
- [Mar10] —, “Evaluating Candidate Sets of Head-Related Transfer Functions for Control of Virtual Source Elevation,” in *AES 40th International Conference, Tokyo, Japan*, 2010, p. 12.
- [MB07] P. Majdak and P. Balazs, “Multiple exponential sweep method for fast measurement of head-related transfer functions,” *JOURNAL-AUDIO ENGINEERING . . .*, 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.
- [MH03] R. Madsen and L. Hansen, "Singular value decomposition and principal component analysis," *Class notes*, 2003.
- [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.
- [MM77] S. Mehrgardt and V. Mellert, *Transformation characteristics of the external human ear*. The Journal of . . . , 1977.
- [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.
- [Mor01] M. Morimoto, *The contribution of two ears to the perception of vertical angle in sagittal planes*. The Journal of the Acoustical Society of America, 2001.
- [Mus84] A. Musicant, "The influence of pinnae-based spectral cues on sound localization," *The Journal of the Acoustical Society of . . .*, 1984.
- [Mye89] P. Myers, "Three-dimensional auditory display apparatus and method utilizing enhanced bionic emulation of human binaural sound localization," 1989.
- [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.
- [OS75] A. V. Oppenheim and R. W. Schaffer, *Digital signal processing*. Prentice Hall, 1975.
- [POM00] J. Plogsties, S. Olesen, and P. Minnaar, "Audibility of all-pass components in head-related transfer functions," *PREPRINTS-AUDIO . . .*, 2000.
- [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.
- [Ram05] M. A. Ramirez, "HRTF Individualization by Solving the Least Squares Problem," *aes.org*, 2005.
- [RD05] V. C. Raykar and R. Duraiswami, "Extracting the frequencies of the pinna spectral notches in measured head related impulse responses," *The Journal of the . . .*, 2005.
- [RDD03] V. C. Raykar, R. Duraiswami, and L. Davis, "Extracting significant features from the hrtf," *Journal*, 2003.

- [Rod05] S. G. Rodríguez, "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 International 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.
- [Sha97] E. Shaw, "Acoustical features of the human external ear," Wiley-Interscience, 1997.
- [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.
- [SNH<sup>+</sup>10] R. H. Y. So, B. Ngan, A. Horner, J. Braasch, and J. Blauert, "Toward orthogonal non-individualised head-related transfer functions for forward and backward directional sound: cluster analysis and an experimental study," *Ergonomics*, 2010.
- [Sot99] R. Sottek, "Physical modeling of individual head-related transfer functions (HRTFs)," *The Journal of the Acoustical Society of America*, 1999.
- [Tan98] C. Tan, "User-defined spectral manipulation of HRTF for improved localisation in 3D sound systems," *Electronics Letters*, 1998.
- [TSK99] N. Takanori, K. Shoji, and T. Kazuya, *Interpolation of the head related transfer function on the horizontal plane*, ser. October 17-20, Mohonk Mountain House, New Paltz, New York. J. Acoust. Soc. Jpn.(J), 1999.
- [WAK93] E. Wenzel, M. Arruda, and D. J. Kistler, "Localization using nonindividualized head-related transfer functions," *The Journal of the . . .*, 1993.
- [Wen88] E. Wenzel, "Acoustic origins of individual differences in sound localization behavior," *The Journal of the Acoustical Society of America*, vol. 84, no. S1, p. S79, 1988.
- [WK89] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. II: Psychophysical validation," *The Journal of the Acoustical Society of . . .*, 1989.
- [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.

- [Xie02] B.-s. Xie, "Effect of head size on virtual sound image localization [J]," *Applied Acoustics*, 2002.
- [XLS07] S. Xu, Z. Li, and G. Salvendy, "Individualization of head-related transfer function for three-dimensional virtual auditory display: A review," *Virtual Reality*, 2007.
- [XLS09] —, "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.
- [XLZ07] S. Xu, Z. Li, and L. Zeng, "A study of morphological influence on head-related transfer functions," *Industrial Engineering and . . .*, 2007.
- [XZR07] B.-s. Xie, X. Zhong, and D. Rao, "Head-related transfer function database and its analyses," *Science in China Series G . . .*, 2007.
- [Zaa10] J. Zaar, "Vermessung von Außenohrübertragungsfunktionen mit reziproker Messmethode," Tech. Rep., Dec. 2010.
- [ZDG06] D. N. Zotkin, R. Duraiswami, and E. Grassi, *Fast head-related transfer function measurement via reciprocity*. The Journal of the Acoustical . . . , 2006.
- [ZDG09] D. N. Zotkin, R. Duraiswami, and N. 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.