

# HRTF Customization by Regression

Alexander Kuhn, Martin Rothbucher,  
Klaus Diepold





Technical Report

# HRTF Customization by Regression

Alexander Kuhn, Martin Rothbucher, Klaus Diepold

April 14, 2014



Institute for Data Processing  
Technische Universität München



Alexander Kuhn, Martin Rothbucher, Klaus Diepold. *HRTF Customization by Regression*.  
Technical Report, Technische Universität München, Munich, Germany, 2014.

Supervised by Prof. Dr.-Ing. K. Diepold ; submitted on April 14, 2014 to the Department of  
Electrical Engineering and Information Technology of the Technische Universität München.

# 1 Abstract

This thesis deals with HRTF customization by regression. Five different regression approaches were investigated, implemented and compared to one another. All the regression algorithms were executed with different variations and modifications in order to find ways to improve the results.

These results were then compared to both non-customized HRIRs and individually measured HRIRs, acquired at the LDV audio laboratory. The evaluation is based on spectral distortion, but the run time, the complexity and the effort of the different methods were taken into consideration as well.

The goal of this thesis is to provide a wide comparison between several different possible solutions in order to provide a good foundation for further research.

The results showed that the PCA's results were the worst and that changing it to a 2DPCA improved the results significantly. GLRAM and Tensor SVD, despite better SD values, are less suited for regression as they take much longer to execute than the PCA and the 2DPCA. The best results were achieved by using the PLSR.

\*\*\*

Diese Arbeit behandelt das Thema HRTF customization by regression. Es wurden fünf verschiedene Regressionsansätze untersucht, implementiert und miteinander verglichen. Alle Regressionsalgorithmen wurden mit verschiedenen Variationen und Modifikationen ausgeführt, um einen Weg zu finden, die Ergebnisse zu verbessern.

Diese Ergebnisse wurden dann sowohl mit nicht personalisierten HRIRS als auch mit individuell gemessenen HRIRs - aufgenommen im LDV Audio Labor - verglichen.

Die Evaluierung beruht auf spectral distortion, jedoch wurden ebenfalls die Laufzeit, die Komplexität und der Aufwand der verschiedenen Methoden in Betracht gezogen.

Das Ziel dieser Arbeit ist es, einen breiten Vergleich zwischen mehreren verschiedenen möglichen Lösungen zu bieten, um eine gute Grundlage für weitere Forschung zu bilden.

Die Ergebnisse haben gezeigt, dass die Ergebnisse der PCA die schlechtesten waren und dass durch das Ändern in eine 2DPCA die Ergebnisse deutlich verbessert wurden.

GLRAM und Tensor SVD sind trotz der besseren SD Werte zur Regression weniger geeignet, da sie deutlich mehr Zeit zur Ausführung benötigen als die PCA und die 2DPCA. Die besten Ergebnisse wurden mit der PLSR erzielt.



# Contents

<b>1</b>	<b>Abstract</b>	<b>3</b>
<b>2</b>	<b>Introduction and Objectives</b>	<b>7</b>
<b>3</b>	<b>Definitions</b>	<b>9</b>
<b>4</b>	<b>State of the Art</b>	<b>11</b>
4.1	KEMAR HRIRs . . . . .	11
4.2	Selection . . . . .	11
4.3	HRIR Measurement . . . . .	12
4.4	HRTF Customization by Regression . . . . .	14
<b>5</b>	<b>HRTF Customization by Regression</b>	<b>15</b>
5.1	Regression Methods . . . . .	16
5.2	Dimensionality Reduction . . . . .	22
5.3	Removing the Initial Time Delay . . . . .	22
<b>6</b>	<b>Experimental Set Up</b>	<b>25</b>
6.1	Procedure . . . . .	25
6.2	Time Domain . . . . .	26
6.3	Frequency Domain . . . . .	26
6.4	Removing the Initial Time Delay . . . . .	28
6.5	Measuring the Quality . . . . .	29
6.6	Selective Algorithm . . . . .	31
<b>7</b>	<b>Results</b>	<b>33</b>
7.1	Customization with CIPIC Data . . . . .	33
7.2	Customization with Data from the LDV Data Base . . . . .	43
7.3	Listening Tests . . . . .	44
<b>8</b>	<b>Discussion and Future Work</b>	<b>45</b>
<b>9</b>	<b>Appendix</b>	<b>51</b>





## 2 Introduction and Objectives

In recent years the number of scientists dealing with the generation of 3D sound has been growing rapidly. With 3D television becoming more and more dominant, it is only logical to take sound to the next level as well. Especially with our world becoming increasingly connected, the importance of teleconferences has risen significantly. As of now, there are several systems for this kind of communication, but they all lack the feeling of the person at the other end being in the same room. This would of course change if there was 3D sound. People would be able to talk one another as though they were standing next to each other.

Another area that takes great interest in three dimensional sound is robotics. Especially when it comes to tele-operating a robot it can be useful to know where the perceived sound originates.

Since it is clear to see where 3D sound can be of use the next logical step is to ask how it can be accomplished.

In order to make sound seem three-dimensional for the human ear one has to be in possession of the Head Related Impulse Response. Since every individual has its own Head Related Impulse Response, there are two possibilities: you can either take a Head Related Impulse Response from a different person or the subject's own Head Related Impulse Response. In the first case the feeling of surround sound will not be as good as in the latter. However, since no two humans are the same, each and every individual will perceive sound differently. So in order to produce the perfect 3D illusion for everyone you have to take the individually measured Head Related Impulse Response of every single person who intends to enjoy three dimensional sound which takes a lot of time and professional equipment.

The preferred way to obtain one's Head Related Impulse Response is to put small microphones into a person's ears and measure how a signal played from a speaker is perceived by the test subject and thus the microphone. This approach entails several problems. Not only is it extremely time consuming to measure the Head Related Impulse Response for every individual but it is also complicated to find microphones that are both exact and small enough so they will neither contort the signal nor change the shape of the ear. The latter point is crucial as the Head Related Impulse Responses are mainly influenced by the shape and size of the different body parts. Which is why it is of utmost importance that during the measurement there are no changes to the actual proportions caused by equipment such as microphones. And last but not least there is the problem of the subject not being allowed to move the slightest bit during the measurements because otherwise it would ruin

## *2 Introduction and Objectives*

the measurements.

So it has been tried to find a way to produce surround sound for everyone as good as when using individual HRIRs without the effort of measuring them separately. Ideally, a person would only have to know some of his or her anthropometric measurements and be able to calculate the perfect, customized Head Related Impulse Response (HRIR). Currently, there exists a theory of how to achieve this. How they work in reality and whether the results will be able to compare to those from any Head Related Impulse Response (HRIR) from the CIPIC database will be investigated in this thesis. Rothbucher et al. [17] have already done some research in this field. Their paper will be the starting point of this thesis. Its main goal then is to research the topics above even further and compare different approaches.

## 3 Definitions

In order to understand the basic terms and abbreviations of this thesis, the most important ones are explained beforehand.

### HRIR

Short for Head Related Impulse Response.

The HRIR describes how sound is perceived by a human ear. On its way from the source to the ear the sound is influenced by parts of the ear, the head, the torso and other anthropometric features that function as filters. The HRIR varies depending on the direction of the sound. The location of the sound source influences the amplitude and the initial time delay. Just like most impulse responses, a HRIR is close to zero until the sound hits the ear which then results in a peak. The length of this delay will be shorter if the sound source is close by and longer if for instance the sound source is on the far side of the head. The amplitude behaves similarly. The closer the source, the greater the amplitude. As the human ear processes sound much the way a microphone does, HRIRs can be measured with small microphones inside the ear. The HRIRs can then be used in a convolution to create three-dimensional sound. In this thesis, the CIPIC data base was used. Refer to [28] for more information on this data base.

### HRTF

Short for Head Related Transfer Function. It is the Fourier Transform of the HRIR.

### 3 Definitions

## DTF

Short for Directional Transfer Function.

It is a method to remove the logarithmic mean of a HRTF and thus center it according to

$$\mathcal{D}(i, j, k) = 20 \log_{10} |\mathcal{H}(i, j, k)| - \frac{1}{N_p} \sum_{k=1}^{N_p} 20 \log_{10} |\mathcal{H}(i, j, k)|, \quad (3.1)$$

where  $\mathcal{H} \in \mathbb{R}^{N_d \times N_r \times N_p}$  is an HRTF and  $\mathcal{D}$  is the resulting DTF. The subtracted mean has to be added to the newly calculated HRTF after regression before using it to generate 3D sound.

## Regression

Regression is used to investigate and express the influence of one or more independent variables on one or more dependent ones. In this thesis, it is used to find the influence of a persons' anthropometric data on his or her personal HRIRs.

## Spectral Distortion

A method to evaluate the difference between two signals in the frequency domain. It is computed according to

$$SD = \sqrt{\frac{1}{N_f} \sum_{i=1}^{N_f} \left( 20 \log_{10} \frac{|\mathcal{H}_i|}{|\mathcal{H}_{new_i}|} \right)^2}, \quad (3.2)$$

where  $\mathcal{H}_i \in \mathbb{R}^{N_d \times N_r \times N_p}$  is an HRTF at its  $i$ -th frequency from a data base and  $\mathcal{H}_{new_i}$  is a HRTF calculated with a regression algorithm at its  $i$ -th frequency.  $N_f$  denotes the number of frequencies.

## 4 State of the Art

This chapter deals with what has already been researched in the field of HRTFs and surround sound creation. There is for instance the option to simply use KEMAR HRTFs which were measured for a dummy. This approach assumes that the KEMAR is close enough to most people and that most people will not hear any difference.

The second approach is to select suited HRTFs from existing data bases like in [5] and [26].

The third and probably most complex solution is measuring HRTFs individually, which though very precise, is also very time consuming.

A rather new method is HRTF customization by regression. Even though some sources like [17] did some research in this field, it is not as well explored as the first three approaches. As of the time of this thesis there are no indications that the PLSR has ever been used for customizing HRTFs. This chapter includes a short description of this solution and of what has already been done in this field.

### 4.1 KEMAR HRIRs

The first and probably most basic approach to create 3D sound is to simply use KEMAR HRIRs. They were measured using a plastic dummy by G.R.A.S. with microphones in its ears. On the one hand, this means it certainly did not move and thus did not disturb the measurements. On the other hand, it also means that the HRTFs are highly unlikely to perfectly fit most people since hardly any people look exactly like the KEMAR manikin.

### 4.2 Selection

Some sources like [5] and [26] suggest to choose fitting HRTFs from existing data bases instead of creating new ones. The main idea of this approach is for the subject to listen to several sounds created with different HRIRs. Based on a tournament style comparison, the ones with better localization move on. Whenever the subject is not able to correctly localize a sound based on a HRIR, this HRIR is eliminated.

### 4.3 HRIR Measurement

Since hardly any two people look exactly the same, everyone needs an individual HRIR in order to experience the best 3D sound possible. A HRIR is influenced a great deal by the outer appearance of the individual which is why it does not suffice to simply take someone else's HRIR. Even though it most probably would create a 3D impression, the result is expected not to be as good as with the individual HRIR.

However, it is quite time consuming to measure a person's HRIR. Not only does it take a lot of time, but it also requires a lot of precise equipment and expertise. The exact way to measure HRIRs at the LDV audio laboratory is as follows:

1. Get a person's ears prints using a special silicone and let it dry and harden
2. Cut the prints into shape and drill a hole in both of them for the microphones
3. Set up the turning table, the speaker and pre load all the scripts
4. Play MLS while the subject sits on a chair placed on the turning table
5. Calculate the HRIRs from the recorded noise sounds
6. Cut and maybe post process the HRIRs to improve quality

This first picture shows the mold of an ear canal created using a special silicone. The



**Figure 4.1:** Mold of an ear canal

second picture shows the turn table on which the subject has to sit. It was built with the goal of altering the shape of a human body as little as possible in order to not distort or influence the HRIRs in any way.

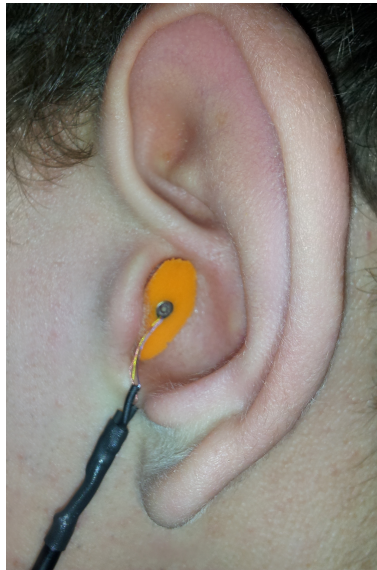
Obviously it is not the easiest procedure as it is rather uncomfortable to sit on the rotating chair in 4.2 without even the slightest movement and it requires a lot of equipment. Furthermore, the room in which the HRIRs are measured has to be absolutely anechoic for optimal results. Another problem is posed by the microphones. As they have to fit into

### 4.3 HRIR Measurement



**Figure 4.2:** Turn table

a person's ear canal, they have to be fairly small. This third picture 4.3 shows how the microphones were placed in the ears.



**Figure 4.3:** Microphone and mold placed inside an ear for measurement

#### **4.4 HRTF Customization by Regression**

This approach has not been investigated as thoroughly as the other ones yet. Its main idea is to use existing HRTF from data bases, train a regression model with that data and obtain a personalized HRTF by including the new subject's anthropometric data. The aim of this thesis is to investigate this approach, to implement and compare different regression algorithms and compare them to the existing approaches.



## 5 HRTF Customization by Regression

With the traditional approach of measuring individual HRIRs being very complex and the goal still being perfect three-dimensional sound, one is tempted to ask: Isn't there any other way to achieve this goal?

And the answer is yes. It is the HRTF Customization based on regression algorithms. It simply requires a person's anthropometric data - or just 8 of all the features to be precise - and a few seconds of computer run time in order to result in a personalized, individual HRIR. And this method opens a whole new world of possibilities. When someone wants to have a conference in 3D sound, the user can not take the time to have his or her HRIR measured in an anechoic room. But having his or her anthropometric features measured only takes a few second and maybe could even be done by a laser scanner or a webcam based application. A method to accomplish this was proposed in [18].

As for the math behind this approach, it basically comes down to this:

Since every HRIR is influenced by anthropometric data, it can be described as

$$w_k = B\tilde{a}_k + \varepsilon, \quad (5.1)$$

where  $\tilde{a}_k$  includes the anthropometric data and  $\varepsilon$  is an error as Rothbucher et al. have shown in [17]. The result  $w_k$  is called features. The anthropometric data  $\tilde{a}_k$  includes several sizes of the human body, such as length and width of the ears. According to [9] it is sufficient to use only 8, namely head width, head depth, shoulder width, cavum concha height, cavum concha width, fossa height, pinna height and pinna width. In [19] slightly different sizes of the human head were used which were less simple to measure as they included radial distances whereas the measurements used by [17] only required linear distances. As a main aspect of this thesis is to make HRTF personalization as easy as possible, the easier way was chosen. Furthermore, the used method maybe allows for future work to develop a camera based algorithm that automatically measures the relevant values based on a picture or a laser scanner.

The result of (5)  $w_k$  is a matrix of  $r_p$  chosen features with  $w \in \mathbb{R}^{r_p \times N_p}$  [17].

These features can either be calculated using PCA, 2DPCA, Tensor SVD, GLRAM, PLSR or potentially another regression algorithm.

The PCA's and the 2DPCA's features are the eigenvectors - also called principal components or PCs - whereas the Tensor SVD uses singular value decomposition and with the GLRAM, the features are the matrix R. Hwang et al. [22] found that the first ten principal components are sufficient to cover 86.4% of the variances, Hugeng et al. [29] found slightly

## 5 HRTF Customization by Regression

different results as can be seen in the appendix.

Rothbucher et al. [17] found that the first 10 largest PCs are sufficient to cover plenty of the information as well.

Since there were three separate findings, it was assumed that results in this thesis would be similar which is why it was not pursued any further, but rather used as a starting point and ground truth.

With the features extracted and by using the Pseudo-Inverse it is possible to find the matrix  $B$  that does not include any anthropometric data:

$$B = W\tilde{A}(\tilde{A}^T\tilde{A})^{-1} \quad (5.2)$$

where  $\tilde{A} = [\mathbf{1}A]^T$ . For more information on customization by regression please refer to [8], [9], [17], [19], [23], or [29].

### 5.1 Regression Methods

As explained in the definitions above, regression is used to find the influence of independent variables on one or more dependent ones. In this thesis, it is used to determine the connection between a person's anthropometric data and his or her HRTFs. There are several more or less common regression methods, some of which will be explained and applied in the following. Since the scope of this thesis is limited, not all existing regression algorithms could be tested. Instead, the ones described in [17] were explored. In addition to those algorithms, the PLSR was used as a fifth approach.

## PCA

The PCA is one of the most common regression methods and thus very well known. This is why it was used as a ground truth in this thesis. Even if the results turned out to be worse than those of the other approaches, the PCA still works great as a starting point to which other results can be compared. Furthermore, it is very easy to implement and to use. Rothbucher et al. also used it as a ground truth in [17].

The PCA uses the eigenvectors of the covariance matrix as regression weights. It can come in useful for achieving data reduction, because the most important principal components - according to Hwang et al. [22] the first 12 - are sufficient to cover a large percentage of the original variance while only using a small part of the original data's space.

The procedure once the eigenvectors - also called regression weights - are calculated is the same as with the 2DPCA. For more information on PCA, refer to [21], [23], [12], [17], [11], [30], [12] or [7]. The way the PCA was applied in this thesis is described in figure 5.1.

---



---

### Principal Component Analysis

---



---

Step 1: Calculate the DTFs

Step 2: Calculate the eigenvectors  $W$  from the covariance matrix  $\mathcal{C} = \mathcal{D}(i, :, :)^T \mathcal{D}(i, :, :)$

Step 3: Use eigenvectors  $W$  to solve  $B = W\tilde{A}(\tilde{A}^T\tilde{A})^{-1}$

Step 4: Multiply the new anthropometric data  $\tilde{a}_{new}$  on B to obtain  $w_{new} = B\tilde{a}_{new}$

Step 5: Calculate  $\hat{\mathcal{D}} = \mathcal{D}W$  where  $\mathcal{D}$  is the DTF of the CIPIC HRIRs.

Step 6: Reverse the DTF by adding the logarithmic mean:

$$H_{new}(i, j, k) = 10^{\frac{1}{20}(D_{new}(i, j, k) + \frac{1}{N_p} \sum_{k=1}^{N_p} 20 \log_{10} |\mathcal{H}(i, j, k)|)}$$

Step 7: Calculate ifft and add the initial delay to be able to listen to results

---



---

**Table 5.1:** Pseudo Code for the PCA

## 2DPCA

With the PCA being the most basic approach in this thesis, the step to improve it to a 2DPCA lies very close. As it was shown in [17], the 2DPCA should deliver better results while still being relatively simple regarding implementation and use. Another advantage of the 2DPCA is that the whole tensor is used in the eigenvalue decomposition whereas the PCA can only be applied to one matrix at a time. This should help reducing the errors. Just like the regular PCA, this method uses eigenvalue decomposition. The only difference is that it doesn't just use one covariance matrix but the scatter matrix  $S_p$  since the input data is three-dimensional. The scatter matrix is the mean of all covariance matrices. It is calculated as described by [13] using the equation

$$S_p = \frac{1}{N_d} \sum_{i=1}^{N_d} \mathcal{D}(i, :, :)^T \mathcal{D}(i, :, :), \in \mathbb{R}^{N_p \times N_p}, \quad (5.3)$$

with  $N_d$  being the number of directions, in case of the CIPIC data this means 1250, and  $N_p$  being the number of subjects used for training;  $\mathcal{D}$  is the DTF of the HRTF. For more information on 2DPCA, refer to [13] or the sources on PCA. Based on this, the algorithm can be described as follows:

---



---

### 2D Principal Component Analysis

---



---

Step 1: Calculate the DTFs (or use HRTFs or HRIRs respectively)

Step 2: Calculate the eigenvectors  $W$  from the Scatter matrix

$$S_p = \frac{1}{N_d} \sum_{i=1}^{N_d} \mathcal{D}(i, :, :)^T \mathcal{D}(i, :, :), \in \mathbb{R}^{N_p \times N_p}$$

Step 3: Use eigenvectors  $W$  to solve  $B = W\tilde{A}(\tilde{A}^T\tilde{A})^{-1}$

Step 4: Multiply the new anthropometric data  $\tilde{a}_{new}$  on B to obtain  $w_{new} = B\tilde{a}_{new}$

Step 5: Calculate  $\hat{\mathcal{D}} = \mathcal{D}W$  where  $\mathcal{D}$  is the DTF of the CIPIC HRIRs.

Step 6: Reverse the DTF by adding the logarithmic mean:

$$H_{new}(i, j, k) = 10^{\frac{1}{20}(D_{new}(i,j,k) + \frac{1}{N_p} \sum_{k=1}^{N_p} 20 \log_{10} |\mathcal{H}(i,j,k)|)}$$

Step 7: Calculate ifft and add the initial delay to be able to listen to results

---



---

**Table 5.2:** Pseudo Code for the 2DPCA

## Tensor SVD

Short for Tensor Singular Value Decomposition. The SVD - much like the PCA - decomposes a signal into its main components. Instead of eigenvalue decomposition (PCA) it uses singular value decomposition. Just like with the PCA, data reduction can be achieved by using only the most important singular values to represent the signal.

The main advantage of performing calculations using tensors is that the three dimensional data from the CIPIIC data base does not have to be split up into two dimensional matrices only to be put back together again later on. This entails the risk of errors being made in the course of splitting it up and rebuilding it which can be avoided by using tensors.

The Tensor SVD is probably the worst of the demonstrated algorithms in terms of usability and especially in terms of availability. The user first needs to download the n-way tool box [6] before being able to start the actual work. The run time is not great either. While the other algorithms take less than a minute, the Tensor SVD takes at least several minutes, depending on the chosen parameters.

There is an algorithm from the n-way Toolbox [6] called *tucker()* that calculates the desired singular values of a tensor which was used in this thesis. Apart from it using tensors instead of matrices, its main functionality is still a singular value decomposition. For more information on Tensors and the Tensor SVD refer to [6], [27], [4] and [24]. The following algorithm was adapted from [17].

---



---

### Tensor SVD

---



---

Step 1: Solve the minimization problem:  $\min_{\hat{\mathcal{D}} \in \mathbb{R}^{N_d \times N_f \times N_p}} \|\mathcal{D} - \hat{\mathcal{D}}\|_F$

Step 2: Decompose the rank- $(r_d, r_f, r_p)$  tensor  $\hat{\mathcal{D}}$  as a trilinear multiplication of a rank- $(r_d, r_f, r_p)$  core tensor  $\mathcal{C} \in \mathbb{R}^{r_d \times r_f \times r_p}$  with three full-rank matrices  $X = (x_{ij}) \in \mathbb{R}^{N_d \times r_d}$ ,  $Y = (y_{ij}) \in \mathbb{R}^{N_f \times r_f}$  and  $W = (w_{ij}) \in \mathbb{R}^{N_p \times r_p}$  which is defined by  $\hat{\mathcal{D}} = (X, Y, W) \cdot \mathcal{C}$

Step 3: Use eigenvectors  $W$  to solve  $B = W\tilde{A}(\tilde{A}^T\tilde{A})^{-1}$

Step 4: Multiply the new anthropometric data  $\tilde{a}_{new}$  on B to obtain  $w_{new} = B\tilde{a}_{new}$

Step 5: Calculate  $D_{new} = (X, Y, w_{new}^T) \cdot \mathcal{C} \in \mathbb{R}^{N_d \times N_f}$

Step 6: Reverse the DTF by adding the logarithmic mean:

$$H_{new}(i, j, k) = 10^{\frac{1}{20}(D_{new}(i, j, k) + \frac{1}{N_p} \sum_{k=1}^{N_p} 20 \log_{10} |\mathcal{H}(i, j, k)|)}$$

Step 7: Calculate ifft and add the initial delay to be able to listen to results

---



---

**Table 5.3:** Pseudo Code for the Tensor SVD

## GLRAM

Short for Generalized Low Rank Approximation. It uses three lower rank matrices  $X \in \mathbb{R}^{N_d \times r_d}$ ,  $M \in \mathbb{R}^{r_d \times r_p}$  and  $W \in \mathbb{R}^{N_p \times r_p}$  to approximate a matrix of higher rank. The left matrix  $X$  is then - just like the eigenvectors before - used to solve the equation  $w_k = B\tilde{a}_k +$  where in this case  $w_k = R \in \mathbb{R}^{N_p \times r_p}$ . After multiplication with new anthropometric data  $M$  and  $W$  are multiplied to the result from the right side in order to reconstruct the new personalized HRIR.

According to [17] it fares well in comparison with the 2DPCA. It is relatively easy to use and even the run time is good. Finding the right parameters is not easy, but [17] found a combination that provided good results. There are of course other possible combinations as well, but this thesis' main goal is to further explore the methods and results of [17].

Regarding the difference between GLRAM and SVD, Ye [14] says:

*"Both GLRAM and SVD aim to minimize the reconstruction error. The essential difference is that GLRAM applies a bilinear transformation on the data. Such a bilinear transformation is particularly appropriate for data in matrix representation and often leads to lower computational cost compared to SVD [14]"*.

For more information on GLRAM refer to [14].

The following algorithm was adapted from [17]:

---



---

### GLRAM

---



---

- Step 1: Solve the optimization problem:  $\min_{X, W, \{M_i\}_{i=1}^{N_p}} \sum_{i=1}^{N_p} \| (\mathcal{D}(:, :, i) - XM_i W^T) \|_F$
- Step 2: Construct a 3D array  $\mathcal{M} \in \mathbb{R}^{r_d \times N_f \times r_p}$  with  $\mathcal{M}(:, i, :) = M_i$  for  $i = 1, \dots, N_f$
- Step 3: Use eigenvectors  $W$  to solve  $B = W\tilde{A}(\tilde{A}^T\tilde{A})^{-1}$
- Step 4: Multiply the new anthropometric data  $\tilde{a}_{new}$  on  $B$  to obtain  $w_{new} = B\tilde{a}_{new}$
- Step 5: Calculate new direct transfer functions  $D_{new}(:, i, :) = X\mathcal{M}(:, i, :)w_{new}$
- Step 6: Reverse the DTF by adding the logarithmic mean:  

$$H_{new}(i, j, k) = 10^{\frac{1}{20}(D_{new}(i, j, k) + \frac{1}{N_p} \sum_{k=1}^{N_p} 20 \log_{10} |\mathcal{H}(i, j, k)|)}$$
- Step 7: Calculate ifft and add the initial delay to be able to listen to results
- 
- 

**Table 5.4:** Pseudo Code for the GLRAM

## PLSR

The PCA bears the problem that even if the found PCs describe the original data very well, it does not necessarily also provide a good prediction for new data. The PLSR on the other hand uses the new anthropometric data and tries to find a connection between this data and the HRTF data base. It calculates the weights from both the new anthropometric data and the DTF training data. This means it takes the new information already into account which should make for a better prediction than the one from the PCA or 2DPCA.

In addition to the promising approach the run time of the PLSR algorithm is very short.

The PLSR was implemented using the Matlab Help while referring to [25], [20] and [10].

For more information on PLSR refer to [20], [10] or [25].

---



---

### PLSR

---



---

Step 1: Center  $X$  and  $Y$  thus obtain  $X_0$  and  $Y_0$

Step 2: Calculate the weights  $W$  and  $C$  by applying SVD on  $R = X^T Y$  to obtain  $W$  and  $C$  with  $R = W \Delta C^T$

Step 3: Calculate  $t_1 = X_0 w_1$  and  $u_1 = Y_0 c_1$

Step 4: Calculate  $X_1 = X_0 - t_1(t_1^T X_0)/(t_1^T t_1)$  and  $Y_1 = Y_0 - t_1(t_1^T Y_0)/(t_1^T t_1)$

Step 5: Calculate the loadings  $P$  of factor  $t_1$  on  $X$  with  $p_1 = X_0^T t_1/(t_1^T t_1)$

Step 6: Calculate the estimated regression coefficient  $b_1$ :  $b_1 = u_1^T t_1/(t_1^T t_1)$  where  $u_1$  are the  $Y$  factor scores

Step 7: Calculate the new DTF with  $\hat{Y}_0 = T \text{diag}(b) C'$

Step 8: Reverse the DTF by adding the logarithmic mean:

$$H_{new}(i, j, k) = 10^{\frac{1}{20}(D_{new}(i, j, k) + \frac{1}{N_p} \sum_{k=1}^{N_p} 20 \log_{10} |\mathcal{H}(i, j, k)|)}$$

Step 9: Calculate ifft and add the initial delay to be able to listen to results

---



---

**Table 5.5:** Pseudo Code for the PLSR

## Summary and Comparison

While 2DPCA, GLRAM, TSVD and PLSR only need to be performed once per subject, the PCA needs to be calculated for every direction or for every frequency, depending on which plane it was performed on. This means the PCA needs to split up the tensor, whereas the other algorithms use the whole tensor, which is likely to reduce computational errors.

Another big difference is between PCA, 2DPCA, GLRAM and Tensor SVD on the one side and PLSR on the other one. While the first four approaches only use existing data to calculate features, the PLSR uses both training data and the new subject's anthropometric data. This should ensure that the prediction is closer to the measured HRTF than when using the other algorithms.

Besides the differences regarding the math behind the algorithms, they also differ regarding run time. GLRAM and Tensor SVD take considerably longer than the other methods, with Tensor SVD being the slowest.

## 5.2 Dimensionality Reduction

By not using all the Principal Components dimensionality reduction can be achieved. As shown by [16] some dimensions can be reduced significantly while still retaining the original quality while others could not be cut without resulting in a loss of quality. This reduction results in both storage and run time reduction, which while not being the main aspect of this thesis could prove to be useful nevertheless. According to [17], using only the 10 largest eigenvectors in PCA or 2DPCA also improves the SD values compared to the results that used all the PCs.

## 5.3 Removing the Initial Time Delay

Before the main calculations can begin, there is preprocessing that needs to be done. There are two suggestions on how to improve signal quality. One is to remove the noticeable time delay between the beginning of the sound being played and the time when the sound hits the ear drum which is visible in the HRIR as a close to zero part before the first peaks. The other approach calculates a minimum phase approximation of the training HRIRs which results in HRIRs without initial delay. The two methods proposed to eliminate this delay are explained in the following.



#### **Direct Approach**

The first way to approach this problem is very straight forward as it simply cuts off the initial delay at the point where the HRIR first reaches 12% of its maximum amplitude like it was proposed by [22].

Apart from these findings, the initial delay can be added after regression performed in the frequency domain and the following ifft which improves the sense of direction provided by the HRIRs. Without adding the delay, there would be no difference between the two channels besides the different amplitudes. This does create a 3D impression, but it should be improved by using the original delays. Since the initial delay should - in theory - be zero as there is no sound played yet and anything different from zero has to be noise it does not need to be stored. It is sufficient to simply append zeros of the same length as the original delay to the results of the different regression algorithms' results. A side effect of discarding the original delay is an improvement in the needed memory as well as a reduced run time. Experimental results will show if the approach really yields the expected improvements.

#### **Minimum Phase Approach**

The second way to get rid of the initial delay is less straight forward. According to Hugeng et al. [29] an approximation of the HRIR without the initial time delay can be achieved with a minimum phase HRIR.

Unlike the first approach that likely changes the amplitude by simply cutting the signal, this second method only changes the phase. This means that the SD values should remain the same as when not using the minimum phase approximation but the normal HRIRs for training.

Experiments will show if either one of the above attempts to improve quality is suitable and does in fact work properly.



## 6 Experimental Set Up

Having investigated and implemented the different approaches with the different modifications, an experimental set up was designed to test the developed algorithms.

### 6.1 Procedure

While some sources like [8] and [2] also use the so-called snowman or HAT (Head And Torso) model along with other approaches, this thesis will focus on the comparison between individually measured HRTFs, the most similar HRTFs from the CIPIC data base - as done by [8] - and the customized HRTFs using different regression methods.

With the algorithms implemented and ready to use, they were applied to create HRIRs and HRTFs with the CIPIC data used for training. The subject whose HRIR was to be created was not included in the training. The goal is to create a new HRIR simply from known training data and a person's anthropometric features.

This excluded subject's HRIR was produced using PCA, 2DPCA, Tensor SVD, GLRAM and PLSR in both time and frequency domain. The result was then compared to the subject's HRIR from the CIPIC data base. In order to evaluate the results objectively, the MSE for results in the time domain and the spectral distortion for results in the frequency domain are used.

There are several options from which it is possible to create a variety of combinations. The options are listed below:

- time domain, with and without removal of the initial delay
- time domain, with and without removal of the mean over all subjects
- time domain, PCA and the 2DPCA with different numbers of principal components
- time domain, GLRAM and Tensor SVD with different parameters
- time domain, with the PCA applied to different planes
- frequency domain, with and without removal of the initial delay
- frequency domain, with and without using the DTFs

## 6 Experimental Set Up

- frequency domain, PCA and the 2DPCA with different numbers of principal components
- frequency domain, GLRAM and Tensor SVD with different parameters
- frequency domain, with the PCA applied to different planes

### 6.2 Time Domain

Even though in [17] all the experiments were performed in the frequency domain using DTFs, this thesis also includes results from algorithms in the time domain. However, in time domain, the DTFs can not be calculated, which means that the mean over all subjects had to be subtracted from the training data in order to center it.

### 6.3 Frequency Domain

Having performed the various transformations in the time domain, the algorithms were then applied to data in the frequency domain. In the frequency domain, there are two options: one is to simply use the HRTF, the other is to first calculate the DTF and use it. Both approaches were tested thoroughly.

#### DTF

According to Rothbucher et al. [17] and [7] it is suitable to use the DTFs instead of the regular HRTFs. The directional transfer function is calculated using the equation

$$\mathcal{D}(i, j, k) = 20 \log_{10}(|\mathcal{H}(i, j, k)|) - \frac{1}{N_P} \sum_{k=1}^{N_P} 20 \log_{10}(|\mathcal{H}(i, j, k)|), \quad (6.1)$$

where  $\mathcal{H} \in \mathbb{R}^{N_d \times N_r \times N_p}$  is an HRTF.

#### Regression Set Up

The basic idea is to solve the equation  $w_k = B\tilde{a}_k + \epsilon$  for  $B$  where  $w_k$  are the eigenvectors of a HRIR's covariance matrix and  $\tilde{a}_k$  is an array consisting of the subjects' anthropometric data. Once  $B$  is calculated, it can then be multiplied with the anthropometric data of any person. This will - via a few steps - result in a personalized HRIR. At the beginning of this thesis there were no individually measured HRIRs which is why the ones from the CIPIC

data base were used for training. In this set up, one subject from the CIPIC data base was left out of the training and later on its anthropometric data was used to create a new HRIR. This was then compared to the measured one from the CIPIC data base.

### PCA and 2DPCA

Since the HRIRS are three-dimensional and the input to a PCA has to be  $\in \mathbb{R}^2$  the *squeeze()* command has to be used. Naturally, there are three options when picking 2 dimensions out of 3. However, in reality you can only perform it in two different ways, because one would be to perform it  $N_p$  times on matrices of  $\in \mathbb{R}^{N_f \times N_d}$ . This contradicts the idea of using several subjects in the regression which only leaves two variations. You could for instance perform the PCA on matrices  $\in \mathbb{R}^{N_d \times N_p}$  for each frequency or time sample. In [17], however the PCA was performed on matrices  $\in \mathbb{R}^{N_f \times N_p}$  where  $N_p$  is the number of subjects included in the training which means it had to be performed  $N_d$  times.

### Tensor SVD

The Tensor SVD was implemented using the n-way toolbox [6]. The parameters of the tucker algorithm were varied and several combinations were tested. One set up was the one used by [17] using  $r_p = 10$ ,  $r_d = 100$  and  $r_f = 200$ . The other ones were randomly chosen. The best combination would then be used in the comparison.

### GLRAM

According to [14] it is suitable to choose  $l_1 = l_2$  because this yields both a small reconstruction error and a fairly good compression rate while at the same time keeping the computational costs small enough.

On the other hand, [17] used  $r_p = 10$ ,  $r_d = 100$  and  $r_f = 200$  which was implemented as well in order to get [17].

Since the choice of the initial low  $L_0$  did not make much of a difference, it was not varied after the first tests. See figure 7.6 for the comparison of the results using different parameters.

### PLSR

The PLSR does not offer many parameters that can be changed. As an input it needs the CIPIC HRTFs and the anthropometric data of the new subject. The only thing left to be varied is the dimension of the input variables. They have to be identical which means that ones had to be appended to the anthropometric data similar to  $\tilde{a}_{new} = [1 \ a_{new}^T]^T$ .

## 6.4 Removing the Initial Time Delay

As mentioned earlier, it might be suitable to remove the initial time delay. As the delay of the to be customized HRIR is unknown before the regression, the mean over all subjects from the training data was calculated. This resulted in one HRIR for every direction. Similar to the HAT model, which adds a certain delay depending on the location, this approach adds a delay based on the mean of the training data. This could be more precise than the HAT model, seeing as it is more likely that a person looks like the mean of several other subjects than like a snowman. In the following, the initial delay refers to the delay of this set of HRIRs. The summary of how it was implemented and used is described in the following.

### Direct Approach

This method allows to regain the directional information stored in the time delay which would otherwise be lost due to the fft and the following DTF. With this approach the initial time delay is cut off, fft, DTF and regression are performed and after all is done, the result is transformed back into time domain using ifft where the removed delay can be added again. The initial delay was not saved directly to be added again. Instead - to improve run time - the length of the initial delay was stored. After the ifft, zeros of the length of the original delay were added in front of the HRIR.

The first time the left and the right HRIR exceeded an amplitude of 12% of their maximum values, the part of the vector from the beginning to this 12% mark was cut off and its length was stored separately. It is important that the left and the right channel were treated separately - otherwise, the delay would be the same and the adding after the ifft would not make any difference.

For instance: let the sound source be to the right of the subject's head. Then the signal will arrive at the right ear noticeably earlier than at the left ear. This difference has to be preserved when preparing the HRIRs. Before using the regressions' results, the initial time delays have to be added again.

In order for the regression to still work after removing the initial time delay, matrix dimensions have to agree. Meaning that if the signal without the initial delay is shorter than the

original one the PCA will not work. To avoid these kinds of complications, the shortened HRIRs had zeros added at the end. As most if not all HRIRs are close to zero towards the end, this should not make a big difference.

### Minimum Phase Approach

Since simply cutting off a part of a vector can lead to errors, another method was implemented. In this version, `rceps()` was used to create a minimum phase approximation of the HRIR. This effectively leads to a HRIR without the initial delay. In order to regain the directional information after the regression, the length to the point of the HRIR where it first exceeds 12% of its maximum value was stored. This allowed for zeros to be appended at the beginning of the personalized HRIR. All in all, this algorithm was a combination of the minimum phase approach and the cutting off approach. It thus combines the preservation of the directional information from the first approach and the smoothness of the cutting from the second one.

## 6.5 Measuring the Quality

How to measure quality and furthermore the question of what defines quality is a wide stretched field. This is why two very common methods of evaluation were used: MSE and SD.

The method of choice in this case is cross validation: The regression was applied to the CIPIC data without one of the 37 subjects. This subject's anthropometric data was used later on to produce a customized HRIR for this individual. For evaluation purposes this personalized HRIR was compared to the subject's measured HRIR from the CIPIC data base using the two different algorithms mentioned above.

### MSE

The Mean Square Error is a common tool of evaluating the difference between two signals. However, the error might be caused only by a shift of a few samples. Knowing this, the results of an MSE evaluation are not too strong an indicator. Another problem with the MSE is that with small amplitudes of the compared signals, the difference and thus the MSE will be small. Even if the two signals are not very similar, the error will not be much larger than the maximum of the amplitudes. This would lead to the illusion of a good result where really the signals are just too small to cause a large error.

So to make sure the error was not caused by a small shift or looked small because of small amplitudes, only the spectral distortion was used.

## Spectral Distortion

The customized HRTFs were compared to the CIPIC HRTFs using the spectral distortion which is calculated in the frequency domain. For regression methods using the DTFs this is not a problem. Those, however, that are performed in the time domain need their results to be transformed using the fft first. This is why the MSE - even though it might not be the perfect evaluation tool - was originally used as well.

When performing regression on DTFs, the subtraction of the logarithmic mean has to be reversed by calculating

$$H_{new}(i, j, k) = 10^{\frac{1}{20}(D_{new}(i,j,k) + \frac{1}{N_p} \sum_{k=1}^{N_p} 20 \log_{10} |\mathcal{H}(i,j,k)|)}, \quad (6.2)$$

where  $D_{new}$  is the customized DTF,  $\mathcal{H}$  is the original HRTF from the training data and  $H_{new}$  is the resulting HRTF.

Once this is done, the spectral distortion can be calculated and it is the equivalent in the frequency domain to the MSE in the time domain. It calculates the logarithmic difference between two signals. A comparison between the 2DPCA result and the HRTF from the CIPIC data base can be found in figure 6.1. The setting was Azimuth =  $-80^\circ$ , elevation =  $0^\circ$  and it shows the left channel.

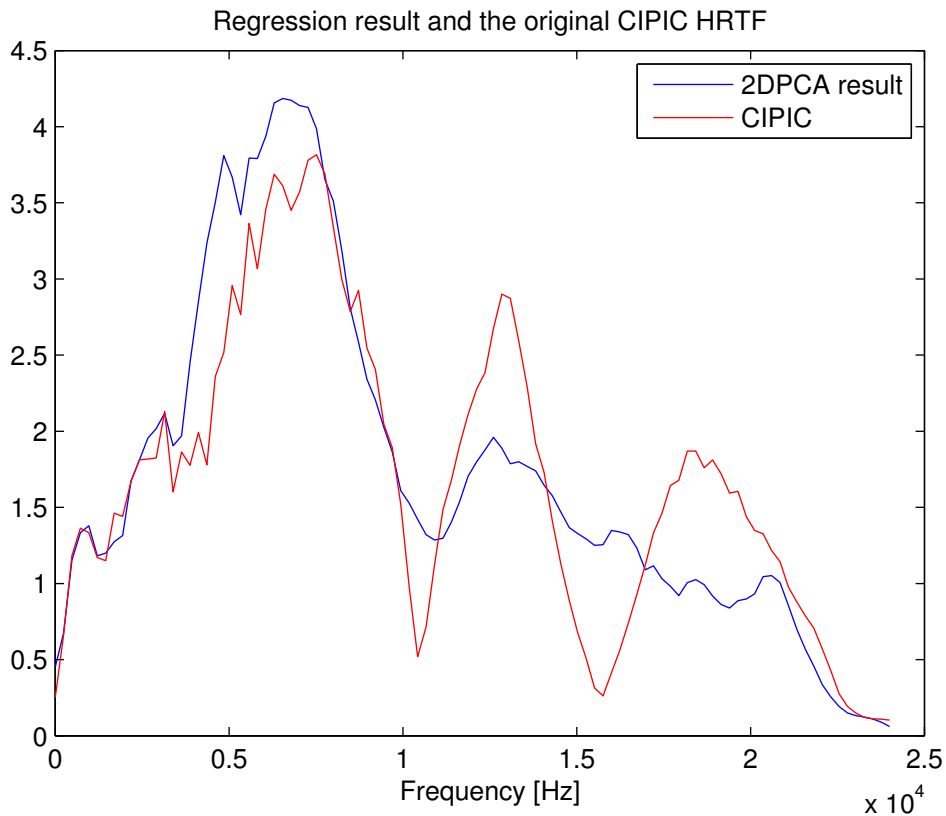
## Listening Tests

In some cases it might be necessary to use a third, less mathematical method of evaluation: listening tests. In case the personalized HRIR does not look much like the measured HRIR from the data base, the SD and the MSE results will be bad. But just because it does not look exactly like the measured one does not automatically mean it does not work properly. The measured HRIRs are not the one and only perfect solution which means that even if the regression results are different, they can still produce very good sound. This is why this third approach is needed.

Since at the time of this thesis there were only 14 individually measured HRIRs in the LDV data base that could be compared to customized ones, it was not possible to perform a listening test with a larger survey group.

This is why preliminary listening tests were only done by the author. The tests included CIPIC HRIRs, customized HRIRs and the individually measured ones. They were compared regarding both localization and overall sound quality. The direction was measured listening to pink noise, while the quality was evaluated using sound samples. The direction is harder to evaluate when listening to sound samples as the song or text might distract from the direction, which is why for this part, noise is used.





**Figure 6.1:** CIPIC HRTF and the 2DPCA HRTF of Subject 31.

## 6.6 Selective Algorithm

Rothbucher et al. found in [17] that the results vary depending on the subject. This was found to be true in this thesis as well.

Based on this observation, a new algorithm was implemented whose task it is to find the subjects with anthropometric data from the CIPIC data base that is closest to the one of the new, to be synthesized subject. The algorithm has a variable threshold which enables the user to choose how many subjects are to be included in the training. Results have to show whether this algorithm is able to improve the SD values.



## 7 Results

This chapter includes the results to the previously explained algorithms and methods. The first part of the experiments was aiming to create a HRIR/HRTF from anthropometric features from the CIPIC data base with the regression being trained with CIPIC HRTFs. The result was then compared to the corresponding measured HRIR/HRTF from the CIPIC data base to see how well the customization algorithms worked.

The second part of this chapter includes the results of the regression algorithms being trained with the HRTFs that were measured at the LDV audio laboratory. The relevant anthropometric features of several people were measured. They then went on to get a personally measured HRIR created. This way, a direct comparison would be possible to see if HRIR personalization can replace individually measured HRIRs or not. However, it would not be possible to train with CIPIC data and compare the results to the newly acquired HRTFs as they were measured under rather different conditions.

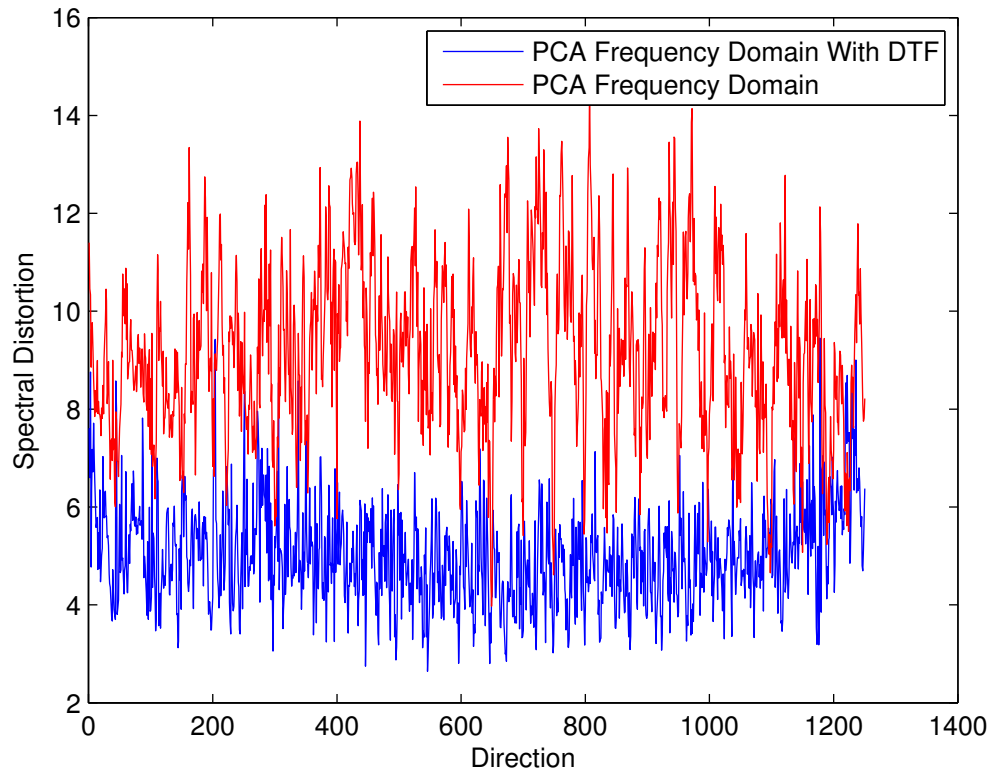
### 7.1 Customization with CIPIC Data

In this first part, the CIPIC data was used for training to customize a new HRIR from the anthropometric data of the CIPIC data base. This newly created HRIR was then compared to the measured one from the CIPIC data. Most of the results were gathered using spectral distortion. The CIPIC data base includes 37 HRIRs of 37 people whose anthropometric data is included as well. So as soon as an algorithm produced a customized HRIR it could be compared directly to the one that was measured and stored in the data base. Since Rothbucher et al. [17] had done something similar, there were existing spectral distortion results which served as an orientation and a benchmark. The different approaches showed results that were sometimes very different from one another.

However, the results were very similar to the ones found by [17].

In first comparisons, the results based on the directional transfer functions were indeed better than the ones with the regular HRTFs from the CIPIC data base. The spectral distortion was significantly smaller. Figure 7.1 shows clearly that it is important to use DTFs and not HRTFs for training as it improves the spectral distortion significantly.

## 7 Results



**Figure 7.1:** PCA results based on HRTFs and on DTFs

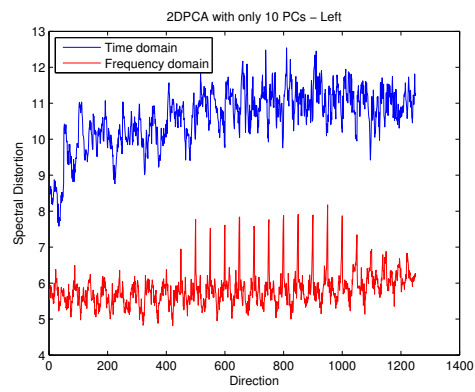
### Regression Results

The regression was performed in both time and frequency domain. The results of these two parts are listed in the following.

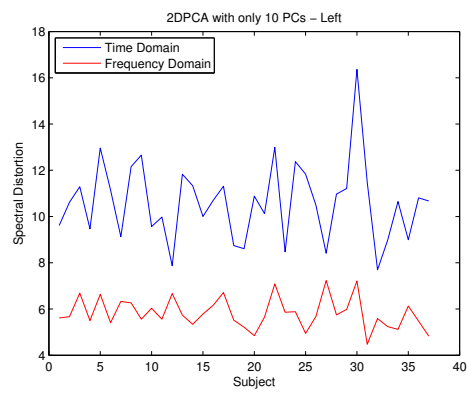
#### Time Domain

Results in time domain were significantly worse than the ones achieved in frequency domain even though the mean over all subjects was removed before the actual regression. This means the data was centered similar to like the DTFs were. More results can be found in table 9.3. The following plots in figure 7.4 show why the MSE was not used as the main tool for evaluation: As the top plot in figure 7.4 shows, the two versions are often shifted by a few samples, which would result in high MSE values even though the two signals are rather similar. The bottom plot in figure 7.4 shows that when a sound is played from the far

## 7.1 Customization with CIPIC Data



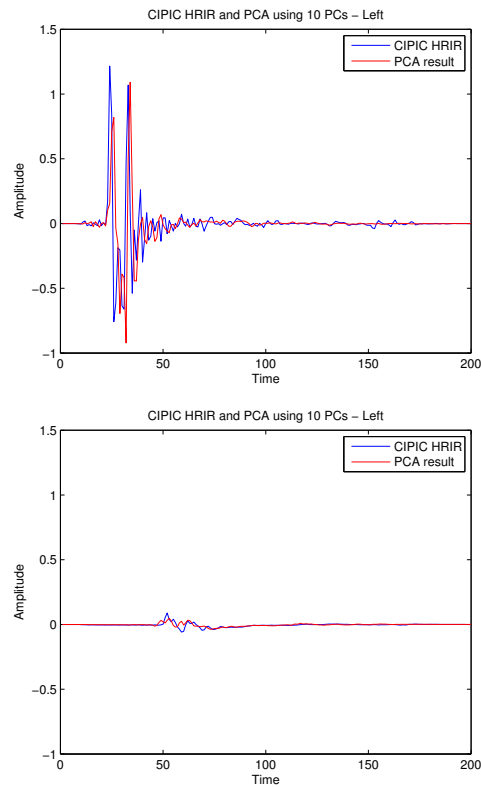
**Figure 7.2:** Spectral Distortion, mean over all 37 subjects - DTF vs time domain



**Figure 7.3:** Spectral Distortion, mean over all directions - DTF vs time domain

side of the ear, the amplitude gets smaller. And even if there are differences in the signal, the MSE will not be very high.

## 7 Results



**Figure 7.4:** Top: sound played from the left. Bottom: sound played from the right

### Frequency Domain

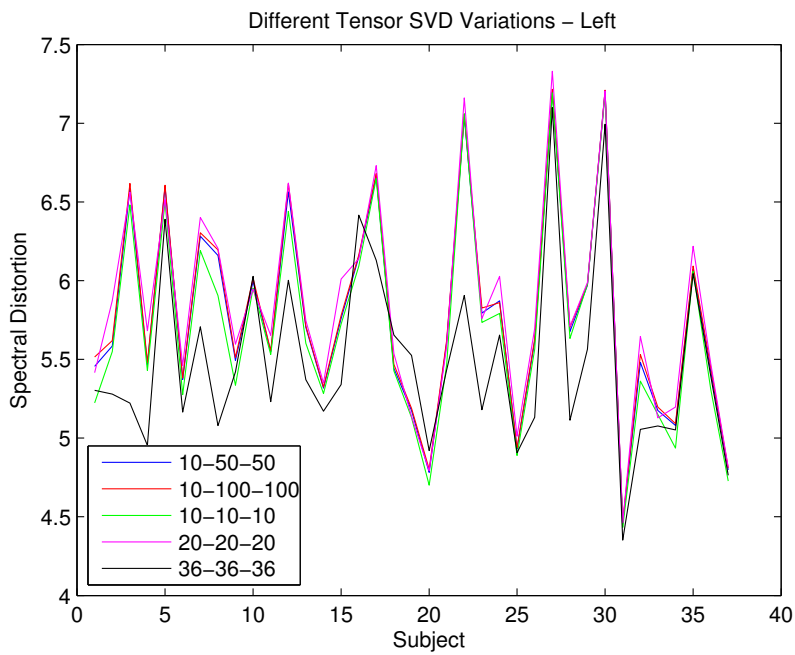
There are several options when performing the PCA or the 2DPCA and the GLRAM and the Tensor SVD can be called with different sets of parameters. The table 7.3 contains the best results of the algorithms. The differences between the different GLRAM and Tensor SVD versions can be found in the following to show that indeed the best one was chosen. See table 9.2 in the appendix for the results of PCA and 2DPCA with all PCs as well as the GLRAM with the configuration that [17] used and figure 9.1 for results of the PCA performed on a different plane.

### Tensor SVD

The following figure 7.5 shows the different spectral distortion values based on different sets of parameters and the table 7.1.

## 7.1 Customization with CIPIC Data

The first number is  $r_p$ , the second one is  $r_d$  and the last one is  $r_f$ .



**Figure 7.5:** Difference in SD values based on different parameters

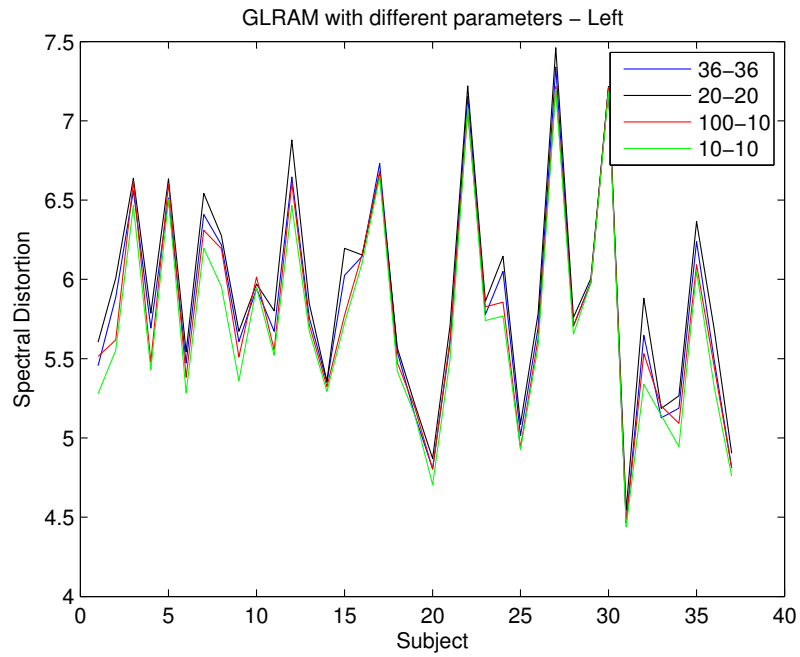
Used Parameters	10,100,100	10,50,50	20,20,20	36,36,36	10,10,10
Total mean	L: 5.79 R: 5.87	L: 5.77 R: 5.85	L: 5.83 R: 5.92	L: 5.49 R: 5.56	L: 5.70R: 5.78

**Table 7.1:** Tensor SVD SD values for different parameters

## 7 Results

### GLRAM

This next figure 7.6 shows the results of the GLRAM being called with different parameters. It shows the mean of the SD values of the left channel over all directions. The first number is  $r_p$ , the second one is  $r_d$  and the last one is  $r_f$ . Since  $r_f$  was always 200, it was left out of the legend in the figure. The numbers shown there are  $r_d$  and  $r_p$ .



**Figure 7.6:** Difference in SD values based on different parameters

Used Parameters	10,10,200	20,20,200	36,36,200	10,100,200
Total mean	L: 5.71 R: 5.79	L: 5.84 R: 5.93	L: 5.93 R: 6.01	L: 5.79 R: 5.87

**Table 7.2:** GLRAM SD values for different parameters



**All Algorithms Compared**

Table 7.3 contains the results of the different regression algorithms. The PCA values were achieved by performing it  $N_d$  times on matrices of  $\in \mathbb{R}^{N_f \times N_p}$  with only the largest 10 principal components. The 2DPCA SD values are the result of calling the 2DPCA and using only the 10 largest eigenvectors. The GLRAM results were achieved by calling *GLRAM()* with  $r_p = 10$ ,  $r_f = 200$ ,  $r_d = 10$ .

The results of the Tensor SVD in this table were achieved by calling *TSVD()* with  $r_p = 36$ ,  $r_f = 36$ ,  $r_d = 36$ .

All results were achieved using DTFs, all the values are in dB.

## 7 Results

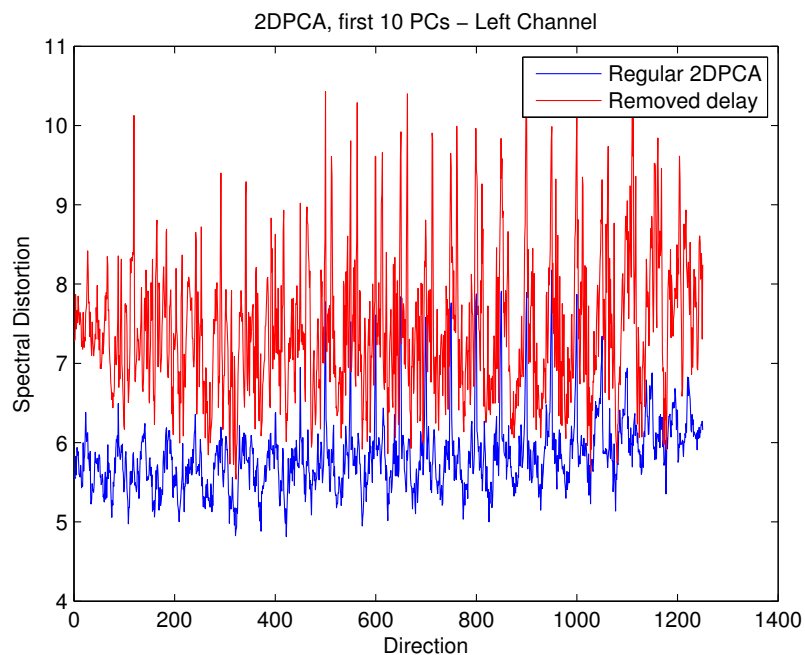
Subject	PCA	2DPCA	GLRAM	PLSR	Tensor SVD
Subject 1	L: 5.94R: 5.65	L: 5.61R: 5.17	L: 5.26R: 4.93	L: 5.32R: 5.09	L: 5.30R: 5.13
Subject 2	L: 6.18R: 6.85	L: 5.67R: 6.53	L: 5.54R: 6.36	L: 5.23R: 5.83	L: 5.28R: 5.87
Subject 3	L: 6.85R: 6.54	L: 6.69R: 6.36	L: 6.46R: 6.20	L: 5.20R: 5.26	L: 5.22R: 5.29
Subject 4	L: 6.02R: 6.92	L: 5.50R: 6.55	L: 5.43R: 6.44	L: 4.82R: 5.54	L: 4.95R: 5.55
Subject 5	L: 6.81R: 6.87	L: 6.64R: 6.50	L: 6.51R: 6.37	L: 6.54R: 6.92	L: 6.39R: 6.83
Subject 6	L: 5.63R: 5.28	L: 5.41R: 5.20	L: 5.29R: 5.10	L: 5.16R: 4.97	L: 5.16R: 5.02
Subject 7	L: 6.78R: 6.13	L: 6.33R: 5.64	L: 6.18R: 5.41	L: 5.70R: 5.31	L: 5.71R: 5.43
Subject 8	L: 6.47R: 6.59	L: 6.26R: 6.49	L: 5.96R: 6.18	L: 5.04R: 5.28	L: 5.08R: 5.29
Subject 9	L: 5.84R: 5.84	L: 5.56R: 5.64	L: 5.36R: 5.39	L: 5.40R: 4.98	L: 5.42R: 5.01
Subject 10	L: 6.02R: 6.45	L: 6.03R: 6.40	L: 5.95R: 6.39	L: 6.02R: 6.17	L: 6.03R: 6.20
Subject 11	L: 5.99R: 7.78	L: 5.56R: 7.59	L: 5.52R: 7.45	L: 5.19R: 6.73	L: 5.23R: 6.75
Subject 12	L: 7.24R: 7.03	L: 6.67R: 5.97	L: 6.47R: 5.83	L: 5.82R: 5.72	L: 6.00R: 5.70
Subject 13	L: 5.93R: 6.32	L: 5.73R: 6.14	L: 5.67R: 6.10	L: 5.37R: 5.59	L: 5.37R: 5.60
Subject 14	L: 5.44R: 5.61	L: 5.34R: 5.46	L: 5.29R: 5.43	L: 5.17R: 5.27	L: 5.17R: 5.33
Subject 15	L: 6.36R: 6.79	L: 5.78R: 6.56	L: 5.72R: 6.45	L: 5.34R: 6.26	L: 5.34R: 6.32
Subject 16	L: 6.20R: 6.99	L: 6.17R: 6.93	L: 6.11R: 6.91	L: 6.40R: 7.20	L: 6.42R: 7.25
Subject 17	L: 6.76R: 6.90	L: 6.71R: 6.55	L: 6.65R: 6.52	L: 6.12R: 6.19	L: 6.13R: 6.18
Subject 18	L: 5.67R: 5.82	L: 5.52R: 5.68	L: 5.40R: 5.56	L: 5.58R: 6.00	L: 5.65R: 5.96
Subject 19	L: 5.33R: 5.69	L: 5.22R: 5.62	L: 5.14R: 5.56	L: 5.56R: 5.50	L: 5.53R: 5.51
Subject 20	L: 5.03R: 5.13	L: 4.85R: 4.74	L: 4.70R: 4.68	L: 4.92R: 4.83	L: 4.92R: 4.85
Subject 21	L: 5.88R: 6.09	L: 5.65R: 5.81	L: 5.50R: 5.70	L: 5.43R: 5.54	L: 5.43R: 5.54
Subject 22	L: 7.36R: 6.26	L: 7.09R: 6.07	L: 7.08R: 5.99	L: 6.08R: 5.76	L: 5.91R: 5.87
Subject 23	L: 6.01R: 5.95	L: 5.86R: 5.75	L: 5.74R: 5.59	L: 5.19R: 5.31	L: 5.18R: 5.32
Subject 24	L: 6.36R: 6.21	L: 5.88R: 5.94	L: 5.77R: 5.81	L: 5.64R: 5.46	L: 5.65R: 5.47
Subject 25	L: 5.19R: 5.12	L: 4.95R: 4.88	L: 4.93R: 4.84	L: 4.86R: 4.63	L: 4.91R: 4.62
Subject 26	L: 5.93R: 5.82	L: 5.67R: 5.66	L: 5.60R: 5.57	L: 5.08R: 4.99	L: 5.1R: 5.12
Subject 27	L: 7.55R: 7.00	L: 7.23R: 6.55	L: 7.20R: 6.45	L: 7.25R: 6.08	L: 7.10R: 6.11
Subject 28	L: 5.84R: 6.32	L: 5.75R: 6.18	L: 5.66R: 6.15	L: 5.10R: 5.42	L: 5.11R: 5.43
Subject 29	L: 6.05R: 5.13	L: 5.98R: 5.02	L: 5.98R: 4.99	L: 5.61R: 4.90	L: 5.56R: 4.91
Subject 30	L: 7.24R: 6.60	L: 7.21R: 6.52	L: 7.19R: 6.52	L: 7.05R: 6.39	L: 6.99R: 6.37
Subject 31	L: 4.57R: 5.32	L: 4.48R: 5.18	L: 4.44R: 5.16	L: 4.34R: 4.83	L: 4.35R: 4.85
Subject 32	L: 6.21R: 6.26	L: 5.59R: 5.60	L: 5.34R: 5.40	L: 5.09R: 5.33	L: 5.06R: 5.26
Subject 33	L: 5.33R: 5.22	L: 5.24R: 5.12	L: 5.12R: 5.05	L: 5.10R: 4.90	L: 5.08R: 4.89
Subject 34	L: 5.42R: 5.89	L: 5.12R: 5.52	L: 4.95R: 5.45	L: 5.01R: 5.36	L: 5.05R: 5.34
Subject 35	L: 6.49R: 5.73	L: 6.13R: 5.68	L: 6.05R: 5.39	L: 6.02R: 5.17	L: 6.04R: 5.17
Subject 36	L: 5.85R: 5.53	L: 5.48R: 5.27	L: 5.32R: 5.09	L: 5.33 R: 4.72	L: 5.39R: 4.80
Subject 37	L: 5.08R: 6.15	L: 4.82 R: 5.96	L: 4.76 R: 5.88	L: 4.79R: 5.68	L: 4.76R: 5.73
<b>Total mean</b>	L: 6.08 R: 6.16	L: 5.82 R: 5.90	L: 5.71 R: 5.79	L: 5.48 R: 5.54	L: 5.49 R: 5.56

**Table 7.3:** SD values in dB

## Removing Initial Delay

Having excluded a subject from the training, this subject's HRIR can then be customized using regression. As it is suitable to perform the calculations in the frequency domain, the initial time delay until the HRIR's first peak gets lost. So even if the regression worked well, in the end, there is no initial delay after the ifft. As mentioned earlier, there are some suggestions on how to get rid of this problem like cutting of the initial delay. However, since the new HRIR is unknown before the regression, it is not possible to store the HRIRs own delay. This is why the length of the initial delay of the mean over all training subjects was stored. Zeros of that length were appended to the customized HRIR later on.

In first listening tests, this did improve the sense of direction and thus the overall quality. However, the SD values are worse than without removing it: The 2DPCA with only the 10



**Figure 7.7:** Difference in SD values with and without removing the delay

largest PCs was used as an example here, but it is similar with all the regression methods. The difference between the results without initial delay and the ones that did not have it removed varies a bit, but it is always worse with the removal algorithm applied beforehand than without it.

### **Minimum Phase**

As removing the delay by cutting it off made the SD worse, the second approach was then tested. When using the minimum phase approximation of the DTF, the SD does not change since the spectral distortion only includes the magnitude of a signal and the phase does not matter. When adding zeros at the beginning of the customized HRIRs of the stored length, the sound quality was improved significantly in comparison to the results that just used the normal DTFs.

### **Selective Algorithm**

Rothbucher et al. [17] found that the results vary depending on the subject. This was found to be true in this thesis as well.

Based on this observation, a new algorithm was implemented whose task it is to find the anthropometric data from the CIPIC data base that is closest to the one of the new, to be synthesized subject. However, there is one drawback. Even if for one subject an ideal number of training data can be found, the same number will probably not work with the next subject. There are two possibilities: One is to only select the subjects from the CIPIC data that look very similar to the new subject - this is done using a threshold. This entails the problem that there is almost no way of finding a threshold that will work equally good for everybody.

The second option is to select only the  $n$  data sets whose anthropometric data is closest to the subject at hand. This approach has a less obvious problem. For a person who looks a lot like several subjects from the training data 20 might be ideal. A person however who does not look like most of the people in the training set will need more subjects to help improve regression results.

Regardless of which method of selection is used, it is - at least at this stage - not possible to automatically find the perfect number of training data for every individual.

And if the chosen number is not the ideal one, the algorithm is at risk of choosing a worse training set than the one that includes all the CIPIC data as figure 7.1 shows.

In conclusion this algorithm does bear some potential but in its current implementation can not be used to automatically improve customized HRIRs.

The following plot shows the SD values for 3 different subjects' customized HRIRs with different numbers of subjects included in the training data. The regression algorithm used in this case is the 2DPCA using the 10 largest PCs as it provided decent results. The figure 7.1 shows that the best results were not necessarily achieved by using the most training data. However, the number of training data that results in the best SD values for one subject does not always work as well for other subjects. Based on these findings, it is suitable to use as much training data as possible.

## 7.2 Customization with Data from the LDV Data Base

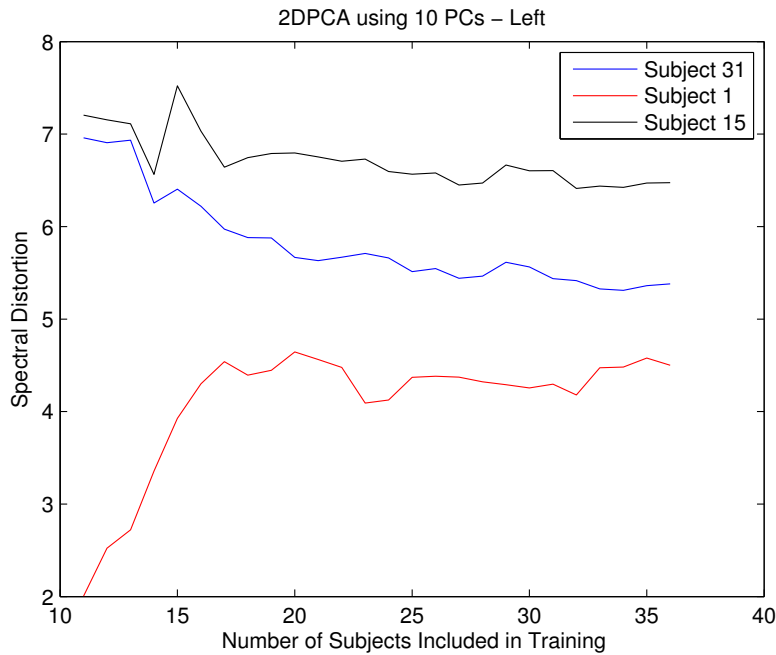


Figure 7.8: SD values of 2DPCA with 10 PCs with various numbers of training data - Left

## 7.2 Customization with Data from the LDV Data Base

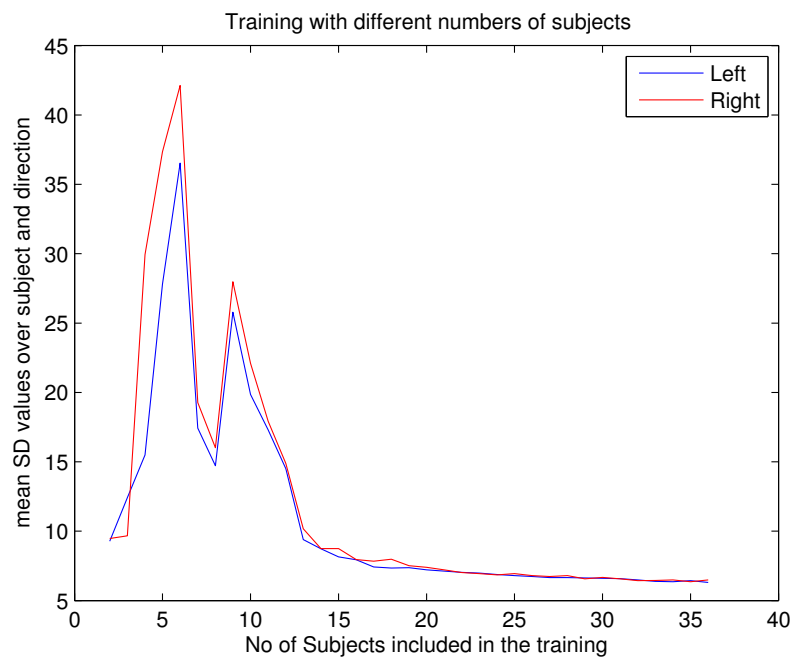
After the algorithms had been proven fully functional it was time to create new HRIRs for test subjects.

### Regression Results

Since the individually measured HRIRs were not acquired in the same way as the CIPIC data, the difference between a CIPIC-trained HRIR and the measured HRIR could be possibly rather large even if the customized HRIR sounds very good to the listener. This is why in this part of the evaluation, the algorithms were trained with the newly measured HRIRs with their anthropometric data.

The SD values were significantly worse than when training with CIPIC data. However, this does not necessarily mean that the data is worse than the CIPIC data, because in the LDV data base there are only 14 subjects whereas the CIPIC data base contains the HRIRs of 36 subjects. This is why the 2DPCA was then trained with CIPIC data with varying numbers of training data ranging from 1 to 36. The results can be seen in the following figure 7.2.

## 7 Results



**Figure 7.9:** 2DPCA with different numbers of training data - Left

### 7.3 Listening Tests

In subjective listening tests, the algorithms all fared well. The overall quality of the original sound samples was not diminished by the convolution. Localization worked well, too. Apart from front-back confusion, the author was able to determine the source of the sound with only a small margin of error.

A small testing group of subjects that did not have any experience with 3D sound was not able to distinguish between the customized HRIRs trained with CIPIC data and the CIPIC HRIRs.

The difference between the CIPIC HRIRs and the newly measured HRIRs was audible, as the resulting convolutions sounded a little bit different. However, neither one of the sounded worse or better than the other one. Both provided good localization results.

When training with the newly acquired data - with only 14 sets of HRIRs - the sound was as bad as the SD values already suggested. It sounded dull and the localization was flawed.

## 8 Discussion and Future Work

It was shown in this thesis that HRTF customization by regression does provide usable results - a conclusion that is based mainly on spectral distortion values. Even though the customized HRIRs are not exactly like the measured ones, the surround impression as well as the overall quality they can generate is decent.

When there is more data of more test subjects, a larger listening test can be used to further investigate how well the personalized HRIRs fare against the measured ones subjectively. The more people there are in such a testing group, the more reliable the results will be, which would be a good conclusion to the mathematical evaluation in this thesis.

The common method to evaluate most surround sound related algorithms is to play either pink or white noise convoluted with different HRIRs. The person doing the test then has to determine where the sound came from. However, since this thesis' aim is to produce the best sound for each individual, the directional sense is not quite enough. The subjective quality matters as well. There could be differences between the different regression methods regarding the sound quality. In order to account for this, the evaluation included a few song samples to determine the overall impression provided by the different HRIRs. As the song might distract from the localization problem, there will be both noise and song tests - one for localization the other one for overall quality.

### Regression Algorithms

There were some differences between the regression algorithms regarding run time and spectral distortion. There was no noticeable difference between the results in terms of localization and sound quality.

### PCA

The first regression algorithm, the PCA, was the most basic of all of them. Already beforehand, this approach was expected to not provide the best of all results. Nevertheless, even with it being not the most complex or refined algorithm, the results were not much worse than the other ones. So it has been proven that the PCA indeed works well as a starting point and ground truth - but for better results other approaches need to be used.

## **2DPCA**

The 2DPCA showed that a small modification to the PCA results in a noticeable improvement regarding the SD values. By using the Scatter matrix instead of the covariance matrix, all the data was taken into account in one step whereas the PCA needed to calculate the eigenvectors of many different matrices. Apparently this is enough to make a basic algorithm almost as good as the more refined ones.

## **GLRAM**

Having delivered the best results in [17], the GLRAM was expected to behave similarly in this thesis, which it did. It was even better than the 2DPCA. To these results there is however also a down side: the GLRAM's run time. While not being terribly bad, it does take noticeably longer than the 2DPCA which could turn into a real issue when a whole set of HRTFs is to be customized for a conference within a couple of seconds. Admittedly this is a small price to pay when the results are in fact better than those of the the 2DPCA.

## **Tensor SVD**

As the Tensor SVD had returned the same results as the GLRAM in [17] its results were expected to be equally good in this thesis. And once again, it did as it was expected to. And it even provided slightly better results than the GLRAM with one of the tested sets of parameters. It's implementation was based on the n-way toolbox [6]. However, it takes even longer than the GLRAM to compute the HRIRs.

## **PLSR**

Not only does the PLSR provide the best results, its run time, unlike with GLRAM and Tensor SVD, is also very short - in most cases it is even faster than the 2DPCA. The reason for the results being better than the other algorithms lies within the concept of the PLSR. While the other regression algorithms only use the existing training data to find the eigenvectors, the PLSR takes the eigenvectors - the weights - from a matrix that includes both the training data and the new anthropometric data. Like this, the prediction is not only based on existing data, but also on part of the new data. This allows for a better prediction which becomes visible in the SD values.



## **Dimensionality Reduction**

The results showed that the regression algorithms were able to perform better when not all the PCs were used. This means that while achieving data reduction, the results even get better. Refer to [16] for more information.

## **Removing the Initial Delay**

The results showed that simply cutting the delay off resulted in worse SD values. Calculating the minimum phase approximation of a HRIR did not change the SD values, but it did improve the listening quality and the localization of the results compared to the results based on the normal DTFs. Without the delay being added after the regression, the localization works only based on differences in the amplitudes of the left and the right HRIR. So a combination of the two approaches was used, where the minimum phase approximation was calculated, the length of the initial delay was stored and zeros of that length were added after the regression. This method increased the run time a little bit, but without it, the results would not be suited for 3D sound generation. So all in all, this combined method is necessary for this thesis's results.

## **Training with Different Data**

The first results were achieved by training the regression algorithms with CIPIC data and comparing it to the HRIRs from the CIPIC data base. The results were good, regarding both SD values and listening tests.

Training the regression algorithms with the newly measured data from the LDV data base did not provide good results - neither in listening tests nor in terms spectral distortion values. One reason for this is that there was not enough data: 14 subjects are not enough for training as 7.2 shows. With only 14 sets of HRTFs even the CIPIC data does not provide good enough training data. Another reason for the insufficient results could be that the measured data from the LDV data base was simply cut to a certain length and was not filtered in any way. If they were processed it could help improve the results.

## **Outlook and Future Work**

In further work, when there is more data available it might be useful to use the individually measured HRIRs to make the LDV data base larger and use it instead of the CIPIC data base for training the regression algorithms. Even now, the CIPIC data base is not

## 8 Discussion and Future Work

the only one of its kind. Maybe other HRIRs with different subjects are more suited for HRTF customization. Or all the existing data bases could be fused together to provide a maximum of possible training data. The more subjects there are, the higher the probability that there are subjects similar to the newly to be synthesized one.

Since the quality of the customization algorithms' results depends strongly on the anthropometric data of the new subject it could be possible to use this circumstance:

*"It can be seen that, customization procedure leads to different spectral distortion values from subject to subject. For subject 153 and subject 165, estimation of the HRTF works quite well in comparison with subject 162" [17].*

This thesis was able to reproduce those findings. So in addition to the realization that the quality of the customized HRIR depends on the number of used principal components there now is a new dependency that has to be taken into account.

It is possible that using more data does not automatically make the results better. In this case, the additional less similar subjects would only mean a distortion while also increasing the run time. This is why a new algorithm is proposed. It selects only those subjects whose anthropometric features are very similar to the person whose HRIR is to be synthesized. This method entails both data and run time reduction. However, it also means a trade-off. If the number of selected subjects is too small it might at some point be too few whilst on the other hand there must not be too many subjects in order to avoid a faulty training. Future work could try and develop a way of combining regression with the tournament style selection.

The described regression algorithms all vary depending on several parameters. The possible combinations are almost limitless and only a very small fraction was tested in this thesis. The aim of future work could be to refine the algorithms by finding better suited parameters.

This thesis only investigated five regression algorithms, which means there are still some regression algorithms that have not been tested. Seeing as the results vary from algorithm to algorithm there is still a chance to find a new approach that might produce even better results. Future work could also investigate the HAT model to see if this approach can provide better results than the one found in this thesis.

### **HAT model**

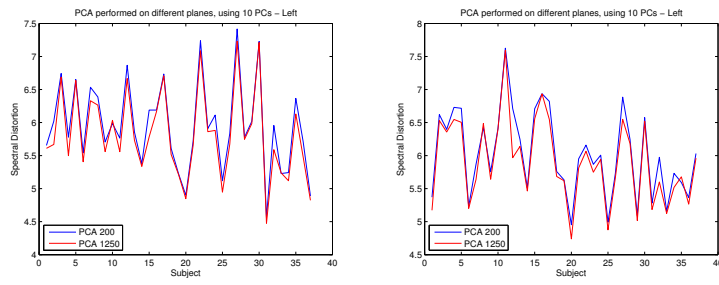
Several authors have found this model to be very effective. According to [8], the HAT model performs better regarding localization than regression. *"We suggest that researchers and software developers use the HAT model to both improve subjective quality of the audio scene and help correct localization of the acoustic signals in virtual audio environments [8]"*.

It is based on a numerical approximation of the HRTFs instead of regression algorithms. It uses a snowman model that represents a human, but only consists of two balls: a large

one representing the body and a smaller one representing the head. However, this thesis focuses on customization by regression which is why the HAT model was not investigated. It shall however not go unmentioned as it is a valid approach and it is this thesis' goal to be as thorough and complete as possible. Please refer to [2] for more information on the HAT model.



# 9 Appendix



**Figure 9.1:** Difference in SD values based on which plane the PCA was performed on

Times the PCA was called	1250	200
Total mean	L: 5.82 R: 5.90	L: 5.95 R: 6.03

**Table 9.1:** Mean over all spectral distortion values in dB

**Table 9.2:** Spectral Distortion Values Frequency Domain. All the values are in dB.

Subject	PCA-10	PCA-full	2DPCA-10	2DPCA-full	GLRAM-10	GLRAM-[17]
1	L: 5.9R: 5.7	L: 6.12 R: 5.84	L: 5.6R: 5.2	L: 6.61R: 7.88	L: 5.3R: 4.9	L: 5.51R: 5.14
2	L: 6.2R: 6.9	L: 6.37 R: 6.98	L: 5.7R: 6.5	L: 5.82 R: 6.43	L: 5.5R: 6.4	L: 5.62R: 6.48
3	L: 6.9R: 6.5	L: 6.92 R: 6.60	L: 6.7R: 6.4	L: 6.24R: 6.26	L: 6.5R: 6.2	L: 6.62R: 6.30
4	L: 6.0R: 6.9	L: 6.19 R: 7.06	L: 5.5R: 6.6	L: 6.19R: 6.77	L: 5.4R: 6.4	L: 5.48R: 6.52
5	L: 6.8R: 6.9	L: 6.88 R: 6.97	L: 6.6R: 6.5	L: 6.45R: 7.24	L: 6.5R: 6.4	L: 6.61R: 6.48
6	L: 5.6R: 5.3	L: 5.72 R: 5.39	L: 5.4R: 5.2	L: 5.82R: 5.78	L: 5.3R: 5.1	L: 5.38R: 5.18
7	L: 6.8R: 6.1	L: 6.93 R: 6.29	L: 6.3R: 5.6	L: 7.05R: 6.58	L: 6.2R: 5.4	L: 6.31R: 5.60
8	L: 6.5R: 6.6	L: 6.61 R: 6.70	L: 6.3R: 6.5	L: 6.25R: 6.03	L: 6.0R: 6.2	L: 6.19R: 6.44
9	L: 5.8R: 5.8	L: 5.92 R: 5.94	L: 5.6R: 5.6	L: 6.21R: 5.90	L: 5.4R: 5.4	L: 5.51R: 5.57
10	L: 6.0R: 6.5	L: 6.04 R: 6.47	L: 6.0R: 6.4	L: 6.17R: 6.75	L: 6.0R: 6.4	L: 6.02R: 6.39
11	L: 6.0R: 7.8	L: 6.13 R: 7.87	L: 5.6R: 7.6	L: 5.75R: 6.88	L: 5.5R: 7.5	L: 5.56R: 7.57
12	L: 7.2R: 7.0	L: 7.48 R: 7.34	L: 6.7R: 6.0	L: 8.25R: 7.57	L: 6.5R: 5.8	L: 6.61R: 5.93
13	L: 5.9R: 6.3	L: 6.02 R: 6.38	L: 5.7R: 6.1	L: 5.86R: 6.03	L: 5.7R: 6.1	L: 5.71R: 6.12
14	L: 5.4R: 5.6	L: 5.47 R: 5.66	L: 5.3R: 5.5	L: 6.20R: 6.17	L: 5.3R: 5.4	L: 5.32R: 5.46
15	L: 6.4R: 6.8	L: 6.47 R: 6.91	L: 5.8R: 6.6	L: 6.18R: 6.72	L: 5.7R: 6.5	L: 5.77R: 6.53
16	L: 6.2R: 7.0	L: 6.24 R: 7.01	L: 6.2R: 6.9	L: 6.23R: 7.50	L: 6.1R: 6.9	L: 6.16R: 6.92
17	L: 6.8R: 6.9	L: 6.81R: 6.97	L: 6.7R: 6.6	L: 6.82R: 6.43	L: 6.7R: 6.5	L: 6.68R: 6.54
18	L: 5.7R: 5.8	L: 5.75R: 5.88	L: 5.5R: 5.7	L: 6.03R: 6.50	L: 5.4R: 5.6	L: 5.47R: 5.66
19	L: 5.3R: 5.7	L: 5.39R: 5.75	L: 5.2R: 5.6	L: 6.22R: 6.00	L: 5.1R: 5.6	L: 5.19R: 5.59
20	L: 5.0R: 5.1	L: 5.14R: 5.24	L: 4.9R: 4.7	L: 5.43R: 5.55	L: 4.7R: 4.7	L: 4.80R: 4.70
21	L: 5.9R: 6.1	L: 5.95R: 6.18	L: 5.7R: 5.8	L: 5.89R: 6.56	L: 5.5R: 5.7	L: 5.62R: 5.78
22	L: 7.4R: 6.3	L: 7.44R: 6.39	L: 7.1R: 6.1	L: 6.12R: 6.23	L: 7.1R: 6.0	L: 7.06R: 6.03
23	L: 6.0R: 6.0	L: 6.10R: 6.07	L: 5.9R: 5.8	L: 5.82R: 5.75	L: 5.7R: 5.6	L: 5.83R: 5.72
24	L: 6.4R: 6.2	L: 6.46R: 6.32	L: 5.9R: 5.9	L: 6.97R: 6.40	L: 5.8R: 5.8	L: 5.86R: 5.88
25	L: 5.2R: 5.1	L: 5.28R: 5.21	L: 5.0R: 4.9	L: 6.02R: 5.29	L: 4.9R: 4.8	L: 4.94R: 4.86
26	L: 5.9R: 5.8	L: 6.02R: 5.92	L: 5.7R: 5.7	L: 5.90R: 6.53	L: 5.6R: 5.6	L: 5.63R: 5.63
27	L: 7.6R: 7.0	L: 7.70R: 7.14	L: 7.2R: 6.6	L: 8.40R: 6.46	L: 7.2R: 6.5	L: 7.22R: 6.53
28	L: 5.8R: 6.3	L: 5.88R: 6.36	L: 5.8R: 6.2	L: 5.83R: 6.12	L: 5.7R: 6.2	L: 5.70R: 6.16
29	L: 6.1R: 5.1	L: 6.08R: 5.16	L: 6.0R: 5.0	L: 6.02R: 5.28	L: 6.0R: 5.0	L: 5.98R: 5.01
30	L: 7.2R: 6.6	L: 7.26R: 6.62	L: 7.2R: 6.5	L: 7.45R: 6.76	L: 7.2R: 6.5	L: 7.21R: 6.52
31	L: 4.6R: 5.3	L: 4.63R: 5.36	L: 4.5R: 5.2	L: 4.92R: 5.15	L: 4.4R: 5.2	L: 4.47R: 5.17
32	L: 6.2R: 6.3	L: 6.41R: 6.49	L: 5.6R: 5.6	L: 5.79R: 8.28	L: 5.3R: 5.4	L: 5.53R: 5.57
33	L: 5.3R: 5.2	L: 5.40R: 5.30	L: 5.2R: 5.1	L: 5.82R: 5.52	L: 5.1R: 5.1	L: 5.20R: 5.11
34	L: 5.4R: 5.9	L: 5.49R: 5.96	L: 5.1R: 5.5	L: 5.87R: 6.19	L: 5.0R: 5.5	L: 5.09R: 5.50
35	L: 6.5R: 5.7	L: 6.61R: 5.83	L: 6.1R: 5.7	L: 6.70R: 6.04	L: 6.1R: 5.4	L: 6.09R: 5.60
36	L: 5.9R: 5.5	L: 5.98R: 5.67	L: 5.5R: 5.3	L: 6.51R: 5.65	L: 5.3R: 5.1	L: 5.45R: 5.23
37	L: 5.1R: 6.2	L: 5.16R: 6.24	L: 4.8R: 6.0	L: 5.37R: 6.40	L: 4.8R: 5.9	L: 4.81R: 5.93
<b>Mean</b>	L: 6.1R: 6.2	L: 6.18R: 6.26	L: 5.8R: 5.9	L: 6.25R: 6.37	L: 5.7R: 5.8	L: 5.79R: 5.87

**Table 9.3:** SD values - time domain compared to DTF results. All values are in dB.

Subject	PCA10 - DTF	2DPCA10 - DTF	PCA10 - Time	2DPCA10 - Time
1	L: 5.94 R: 5.65	L: 5.61 R: 5.17	L: 8.03 R: 7.60	L: 9.61 R: 8.21
2	L: 6.18 R: 6.85	L: 5.67 R: 6.53	L: 8.58 R: 8.76	L: 10.61 R: 9.82
3	L: 6.85 R: 6.54	L: 6.69 R: 6.36	L: 9.10 R: 9.00	L: 11.29 R: 10.01
4	L: 6.02 R: 6.92	L: 5.50 R: 6.55	L: 7.66 R: 7.72	L: 9.47 R: 8.03
5	L: 6.81 R: 6.87	L: 6.64 R: 6.50	L: 11.54 R: 11.41	L: 12.96 R: 12.91
6	L: 5.63 R: 5.28	L: 5.41 R: 5.20	L: 9.11 R: 8.73	L: 11.15 R: 10.23
7	L: 6.78 R: 6.13	L: 6.33 R: 5.64	L: 7.89 R: 8.40	L: 9.13 R: 9.81
8	L: 6.47 R: 6.59	L: 6.26 R: 6.49	L: 8.99 R: 9.09	L: 12.15 R: 10.77
9	L: 5.84 R: 5.84	L: 5.56 R: 5.64	L: 9.91 R: 9.41	L: 12.66 R: 12.18
10	L: 6.02 R: 6.45	L: 6.03 R: 6.40	L: 8.55 R: 9.30	L: 9.57 R: 9.34
11	L: 5.99 R: 7.78	L: 5.56 R: 7.59	L: 8.45 R: 8.60	L: 9.98 R: 9.35
12	L: 7.24 R: 7.03	L: 6.67 R: 5.97	L: 7.66 R: 7.79	L: 7.87 R: 8.29
13	L: 5.93 R: 6.32	L: 5.73 R: 6.14	L: 10.22 R: 9.39	L: 11.83 R: 10.33
14	L: 5.44 R: 5.61	L: 5.34 R: 5.46	L: 10.38 R: 11.30	L: 11.34 R: 13.26
15	L: 6.36 R: 6.79	L: 5.78 R: 6.56	L: 8.81 R: 7.88	L: 10.0 R: 8.40
16	L: 6.20 R: 6.99	L: 6.17 R: 6.93	L: 9.66 R: 7.92	L: 10.7 R: 8.24
17	L: 6.76 R: 6.90	L: 6.71 R: 6.55	L: 10.28 R: 10.99	L: 11.31 R: 11.36
18	L: 5.67 R: 5.82	L: 5.52 R: 5.68	L: 8.15 R: 7.44	L: 8.75 R: 7.76
19	L: 5.33 R: 5.69	L: 5.22 R: 5.62	L: 8.15 R: 9.87	L: 8.61 R: 11.01
20	L: 5.03 R: 5.13	L: 4.85 R: 4.74	L: 8.91 R: 9.09	L: 10.88 R: 9.79
21	L: 5.88 R: 6.09	L: 5.65 R: 5.81	L: 8.37 R: 8.10	L: 10.12 R: 9.73
22	L: 7.36 R: 6.26	L: 7.09 R: 6.07	L: 11.47 R: 10.57	L: 12.99 R: 11.09
23	L: 6.01 R: 5.95	L: 5.86 R: 5.75	L: 7.86 R: 8.49	L: 8.47 R: 10.18
24	L: 6.36 R: 6.21	L: 5.88 R: 5.94	L: 10.62 R: 10.13	L: 12.38 R: 11.52
25	L: 5.19 R: 5.12	L: 4.95 R: 4.88	L: 9.04 R: 8.82	L: 11.84 R: 11.23
26	L: 5.93 R: 5.82	L: 5.67 R: 5.66	L: 9.00 R: 9.97	L: 10.50 R: 12.08
27	L: 7.55 R: 7.00	L: 7.23 R: 6.55	L: 8.24 R: 8.14	L: 8.41 R: 8.45
28	L: 5.84 R: 6.32	L: 5.75 R: 6.18	L: 10.35 R: 10.89	L: 10.97 R: 11.74
29	L: 6.05 R: 5.13	L: 5.98 R: 5.02	L: 9.81 R: 11.12	L: 11.21 R: 12.51
30	L: 7.24 R: 6.60	L: 7.21 R: 6.52	L: 14.84 R: 14.51	L: 16.36 R: 15.28
31	L: 4.57 R: 5.32	L: 4.48 R: 5.18	L: 10.37 R: 9.24	L: 11.48 R: 10.15
32	L: 6.21 R: 6.26	L: 5.59 R: 5.60	L: 7.50 R: 8.07	L: 7.69 R: 9.77
33	L: 5.33 R: 5.22	L: 5.24 R: 5.12	L: 8.37 R: 9.33	L: 8.99 R: 11.45
34	L: 5.42 R: 5.89	L: 5.12 R: 5.52	L: 9.03 R: 9.18	L: 10.65 R: 10.35
35	L: 6.49 R: 5.73	L: 6.13 R: 5.68	L: 8.24 R: 8.57	L: 9.00 R: 9.68
36	L: 5.85 R: 5.53	L: 5.48 R: 5.27	L: 9.80 R: 7.44	L: 10.81 R: 8.06
37	L: 5.08 R: 6.15	L: 4.82 R: 5.96	L: 9.09 R: 10.36	L: 10.67 R: 11.71
<b>Mean</b>	L: 6.08 R: 6.16	L: 5.82 R: 5.90	L: 9.24 R: 9.26	L: 10.61 R: 10.38





# List of Figures

4.1	Mold of an ear canal . . . . .	12
4.2	Turn table . . . . .	13
4.3	Microphone and mold placed inside an ear for measurement . . . . .	14
6.1	CIPIC HRTF and the 2DPCA HRTF of Subject 31. . . . .	31
7.1	PCA results based on HRTFs and on DTFs . . . . .	34
7.2	Spectral Distortion, mean over all 37 subjects - DTF vs time domain . . . . .	35
7.3	Spectral Distortion, mean over all directions - DTF vs time domain . . . . .	35
7.4	Top: sound played from the left. Bottom: sound played from the right . . . . .	36
7.5	Difference in SD values based on different parameters . . . . .	37
7.6	Difference in SD values based on different parameters . . . . .	38
7.7	Difference in SD values with and without removing the delay . . . . .	41
7.8	SD values of 2DPCA with 10 PCs with various numbers of training data - Left . . . . .	43
7.9	2DPCA with different numbers of training data - Left . . . . .	44
9.1	Difference in SD values based on which plane the PCA was performed on . . . . .	51



## Bibliography

- [1] A.Kulkarni, S.K.Isabelle, and H.S.Colburn. On the minimum-phase approximation of head-related transfer functions. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 84–87, 1995.
- [2] V.R. Algazi, R.O.Duda, and D.M.Thompson. The use of head-and-torso models for improved spatial sound synthesis. In *113<sup>th</sup> Audio Engineering Society Convention*, pages 1–18, 2002.
- [3] A.O.Sykes. An Introduction to Regression Analysis. [online] <http://www.law.uchicago.edu/node/1309> [last access 19/02/2013].
- [4] B.Savas and L.-H.Lim. Best multilinear rank approximation of tensors with quasi-newton methods on grassmannians. Technical Report LITH-MAT-R-2008-01-SE, Department of Mathematics, Linköpings Universitet, 2008.
- [5] B.Seeber and H.Fastl. Subjective selection of nonindividual head-related transfer functions. pages 1–4, 2003.
- [6] C.Andersson and R.Bro. The N-way Toolbox for MATLAB. *Chemometrics and Intelligent Laboratory Systems*, 52(1):1–4, 2000.
- [7] 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*, 91(3):1637–1647, 1992.
- [8] D.N.Zotkin, J.Hwang, R.Duraiswami, and L.S.Davis. Hrtf personalization using anthropometric measurements. In *Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 157–160, 2003.
- [9] H.Hu, L.Zhou, J.Zhang, H.Ma, and Z.Wu. Head related transfer function personalization based on multiple regression analysis. In *International Conference on Computational Intelligence and Security*, pages 1829 –1832, 2006.
- [10] H.Martens and T.Naes. *Multivariate Calibration*. John Wiley & Sons Ltd., Hoboken (New Jersey), 1989.
- [11] I.T.Jolliffe. *Principal Component Analysis*. Springer-Verlag, New York, Inc., second edition, 2002.

## Bibliography

- [12] J.Shlens. A tutorial on principal component analysis [online] <http://www.sn1.salk.edu/shlens/> [last access 19/02/2013]. 2009.
- [13] J.Yang, D.Zhang, A.F.Frangi, and J.Yang. Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1):131–137, 2004.
- [14] J.Ye. Generalized low rank approximations of matrices. *Machine Learning Journal*, 61:167–191, 2005.
- [15] K.R.Rao and J.Ben-Arie. Optimal head related transfer functions for hearing and monaural localization in elevation: a signal processing design perspective. *IEEE Transactions on Biomedical Engineering*, 43(11):1093–1105, 1996.
- [16] M.Rothbucher, H.Shen, and K.Diepold. Dimensionality reduction in HRTF by using multiway array analysis. In *Human Centered Robot Systems*, volume 6 of *Cognitive Systems Monographs*, pages 103–110. Springer Berlin Heidelberg, 2009.
- [17] M.Rothbucher, M.Durkovic, H.Shen, and K.Diepold. HRTF customization using multiway array analysis. In *Proceedings of the European Signal Processing Conference*, pages 229–233, 2010.
- [18] M.Rothbucher, T.Habigt, J.Habigt, T.Riedmaier, and K.Diepold. Measuring anthropometric data for hrtf personalization. *International IEEE Conference on Signal-Image Technologies and Internet-Based System*, 0:102–106, 2010.
- [19] N.Inoue, T.Kimura, T.Nishino, K.Itou, and K.Takeda. Evaluation of hrtfs estimated using physical features. *Acoustical Science and Technology*, 26(5):199–202, 2005.
- [20] S.de Jong. SIMPLS: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 18(3):251–263, 1993.
- [21] S.Dong, R.Hu, W.Tu, X.Zheng, J.Jiang, and S.Wang. Enhanced principal component using polar coordinate pca for stereo audio coding. In *International Conference on Multimedia and Expo*, pages 628–633, 2012.
- [22] S.Hwang and Y.Park. Interpretations on principal components analysis of head-related impulse responses in the median plane. *Journal of The Acoustical Society of America*, 123:65–71, 2008.
- [23] S.Hwang, Y.Park, and Y.-S.Park. Modeling and customization of head-related transfer functions using principal component analysis. In *International Conference on Control, Automation and Systems*, pages 227–231, 2008.
- [24] S.Weiland and F.van Belzen. Singular value decompositions and low rank approximations of tensors. *IEEE Transactions on Signal Processing*, 58(3):1171–1182, 2010.

- [25] S.Wold, M.Sjöström, and L.Eriksson. Pls-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58:109–130, 2001.
- [26] S.Yairi, Y.Iwaya, and Y.Suzuki. Individualization feature of head-related transfer functions based on subjective evaluation. *Proceedings of the 14th International Conference on Auditory Display*, pages 340–343, 2008.
- [27] T.G.Kolda and B.W.Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009.
- [28] V.R.Algazi, R.O.Duda, D.M.Thompson, and C.Avendano. The cipic hrtf database. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 99–102, 2001.
- [29] W.Hugeng and D.Gunawan. Improved method for individualization of head-related transfer functions on horizontal plane using reduced number of anthropometric measurements. *Journal of Telecommunications*, 2(2):31–41, 2010.
- [30] W.W.Hugeng and D.Gunawan. Effective preprocessing in modeling head-related impulse responses based on principal components analysis. *Signal Processing: An International Journal*, 4(4):201–212, 2010.