HRTF CUSTOMIZATION USING MULTIWAY ARRAY ANALYSIS

Martin Rothbucher, Marko Durkovic, Hao Shen and Klaus Diepold

Institute for Data Processing, Technische Universität München, 80290 München, Germany
email: {martin.rothbucher, durkovic, hao.shen, kldi}@tum.de

ABSTRACT

Recently, Head Related Transfer Function (HRTF) based techniques have been promising approaches for 3D sound synthesis. To ensure high quality of 3D sound synthesis, it requires to utilize personal HRTFs of the listener, which are usually obtained by a complicated and time consuming procedure. To personalize HRTFs, one possibility is to construct multiple linear regression models between anthropometric data and features of HRTFs.

Due to high dimensionality of HRTF datasets, it is inefficient to use the original set for the purpose of personalization. To avoid such inefficiency, Principal Component Analysis (PCA) has been proposed to reduce dimensionality of HRTF datasets before customization. Based on the fact that HRTF datasets can be considered as three way data arrays, in this paper we propose three multi-way array analysis methods for HRTF customization. Performance of these three methods is compared with PCA based approaches by several experiments.

1. INTRODUCTION

Head Related Transfer Functions (HRTFs) describe spectral changes of sound waves when they enter the ear canal, caused by diffraction and reflection off the human body, e.g. the head, shoulders, torso and ears [2]. In the last decades, HRTF based techniques have become prominent in various applications of human related audio processing, e.g. binaural sound localization and synthesis [13].

As each individual has in general their unique body geometry, the corresponding personal HRTFs are naturally different from person to person. Usually, HRTFs are obtained from recorded Head Related Impulse Responses (HRIRs), which are the time domain representations of the HRTFs. Unfortunately, HRIRs have to be measured by a cumbersome procedure with expensive equipment in an anechoic chamber [1], which is not commonly accessible to many researchers. As a result, there are increasing research efforts in customization of HRTFs [4, 9], which aims to estimate HRTFs based only on geometric information of the individual without measuring their HRIRs. Such a process requires usually a collection of HRTF datasets of various subjects, which result in huge amount of data.

Since the pioneering work [6], Principal Component Analysis (PCA) has become a popular tool for HRTF reduction [8]. Recently, application of PCA in HRTF customization [4, 9], which reduces the dimension of the original dataset before customization, has demonstrated promising performance of being more efficient. Such an approach enables possibilities of implementing HRTF customization even on storage restricted systems, e.g. mobile phones or telepresence systems.

A collection of HRTFs of individuals can be considered as a three-way data array, whose three directions represent subject, location and frequency, respectively. Applying PCA to HRTF datasets requires in general a vectorization process of the original dataset. As a consequence, some useful information within the structure of the HRTF dataset is disregarded. To avoid such limits, the so-called Tensor-Singular Value Decomposition (T-SVD) method, which was originally introduced in the community of multiway array analysis [7], has been recently applied into HRTF customization successfully [3].

The authors are aware of existence of two recently proposed techniques of multiway array analysis in competition with the standard PCA in the community of image processing. They are two dimensional PCA (2DPCA), originally a direct generalization of PCA for image analysis, and the so-called Generalized Low Rank Approximations of Matrices (GLRAM) [15], a further generalized form of 2DPCA. Recently, the authors have demonstrated that GLRAM and T-SVD outperform the standard PCA in the task of dimensionality reduction of HRTFs [10]. It is worth noticing that GLRAM method is essentially a simple form of Tensor-SVD [12]. In this paper, we study GLRAM, 2DPCA and Tensor-SVD methods for the purpose of customizing HRTF datasets and compare their performance with the standard PCA.

The paper is organized as follows. Section 2 gives a description of the multiple regression customization of HRTFs. Section 3 provides a brief introduction to three feature extraction methods, namely, 2DPCA, GLRAM and Tensor-SVD. In section 4, performance of the three methods is investigated by several numerical experiments. Finally, a conclusion is given in section 5.

2. HRTF CUSTOMIZATION

In this section, we briefly describe a customization approach to estimate individual HRTFs. Given a set of measured HRTFs of different persons, a multiple linear regression seeks to match a set of anthropometric parameters to the characteristics of the individual's transfer functions [9].

In general, a collection of HRTFs can be represented as a three-way array $\mathscr{H} \in \mathbb{R}^{N_d \times N_f \times N_p}$, where the dimensions N_d is the spatial resolution of directions, N_f the frequency sample size and N_p is the number of persons in the training dataset. By a Matlab-like notation, in this work we denote $\mathscr{H}(i,j,k) \in \mathbb{R}$ the (i,j,k)-th entry of $\mathscr{H}$, $\mathscr{H}(l,m,:) \in \mathbb{R}^{N_p}$ the vector with a fixed pair of (l,m) of $\mathscr{H}$ and $\mathscr{H}(l,:,:) \in \mathbb{R}^{N_f \times N_p}$ the l-th slide (matrix) of $\mathscr{H}$ along the direction-dimension.

In order to receive only the direction dependent information between the different individuals, the mean of the subject's average log-HRTFs is subtracted from each log-HRTF [6]. It results in an interindividual direction transfer

 ISSN 2076-1465

functions between the subjects, denoted by $\mathscr{D} \in \mathbb{R}^{N_d \times N_f \times N_p}$, whose (i, j, k)-th entry is computed by

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

An idea of customizing unknown HRTFs is to firstly extract certain direction dependent main features out of the directional transfer functions $\mathscr{D}$, then to construct a multiple linear regression model between anthropometric features of subjects and the extracted directional dependent features. Let $W = [w_1, \ldots, w_{N_p}] \in \mathbb{R}^{r_p \times N_p}$ be a set of r_p chosen directional dependent features and $A = [a_1, \ldots, a_{N_p}] \in \mathbb{R}^{N_a \times N_p}$ be a collection of N_a anthropometric features of test subjects. For the k-th subject, a multiple linear regression model between anthropometric parameters and directional dependent features can be constructed as,

$$w_k = B\tilde{a}_k + \varepsilon, \quad (2)$$

where $B \in \mathbb{R}^{r_p \times (N_a+1)}$, $\tilde{a}_k = [1\ a_k^\top]^\top \in \mathbb{R}^{N_a+1}$, and $\varepsilon \in \mathbb{R}^{r_p}$ is the estimation error vector. Let us denote $\mathbf{1} \in \mathbb{R}^{N_p}$ the vector with all entries equal to one, and construct $\tilde{A} = [\mathbf{1}\ A^\top] \in \mathbb{R}^{N_p \times (N_a+1)}$. It is known that N_p is usually greater than N_a. We assume that matrix $\tilde{A}$ is full rank. Then, in terms of minimization of the error ε by least squares method, the regression coefficient matrix B in model (2) is computed by

$$B = W\tilde{A}(\tilde{A}^\top \tilde{A})^{-1}. \quad (3)$$

In this paper, we choose the set of anthropometric parameters for multilinear regression in accordance with [4], where anthropometric parameters are selected by applying correlation analysis.

Finally, given the anthropometric data $a_{new} \in \mathbb{R}^{N_a}$ of a person not in the training set, its transfer function features w_{new} can be constructed by

$$w_{new} = B\tilde{a}_{new} \in \mathbb{R}^{r_p}, \quad (4)$$

where $\tilde{a}_{new} = [1\ a_{new}^\top]^\top \in \mathbb{R}^{N_a+1}$.

3. HRTF-FEATURE EXTRACTION METHODS

In this section, we briefly describe three techniques of feature extraction for the dataset $\mathscr{D}$, namely, 2DPCA, GLRAM and Tensor-SVD.

3.1 Customization using 2DPCA

Similar to the popular approach of customizing HRTFs by using PCA, 2DPCA based HRTF customization can be described as follows. First of all, the so-called scatter matrix, instead of the covariance matrix, is computed by

$$S_p = \frac{1}{N_d}\sum_{i=1}^{N_d} \mathscr{D}(i,:,:)^\top \mathscr{D}(i,:,:), \in \mathbb{R}^{N_p \times N_p}. \quad (5)$$

Then we compute r_p eigenvectors $W = [w_1, \ldots, w_{r_p}] \in \mathbb{R}^{N_p \times r_p}$ corresponding to the r_p largest eigenvalues. The so-called principal components of 2DPCA for the i-th slides of $\mathscr{D}$ is computed as follows:

$$\widehat{\mathscr{D}}(i,:,:) = \mathscr{D}(i,:,:)W \in \mathbb{R}^{N_f \times r_p}. \quad (6)$$

Note that, the storage space for the reduced dataset depends on the value of r_p.

The direction dependent regression coefficient matrix B is then calculated as given in (3). A set of customized direction transfer functions $D_{new} \in \mathbb{R}^{N_d \times N_f}$ for an unknown person is obtained with its i-th slide computed by:

$$D_{new}(i,:) = \widehat{\mathscr{D}}(i,:,:)w_{new}^\top, \quad (7)$$

where w_{new} is computed in accordance with (4). We refer to [5] for further information on PCA and to [14] for further discussions on 2DPCA.

3.2 Customization using Tensor-SVD

Unlike customization using PCA, Tensor-SVD keeps the structure of the original 3D dataset intact and computes the customized dataset for every direction at once. Given a dataset $\mathscr{D} \in \mathbb{R}^{N_d \times N_f \times N_p}$, Tensor-SVD computes its best multilinear $rank-(r_d, r_f, r_p)$ approximation $\widehat{\mathscr{D}} \in \mathbb{R}^{N_d \times N_f \times N_p}$ [7], where $N_d > r_d$, $N_f > r_f$ and $N_p > r_p$, by solving the following minimization problem

$$\min_{\widehat{\mathscr{D}} \in \mathbb{R}^{N_d \times N_f \times N_p}} \left\| \mathscr{D} - \widehat{\mathscr{D}} \right\|_{\mathrm{F}}, \quad (8)$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm of tensors. The $rank-(r_d, r_f, r_p)$ tensor $\widehat{\mathscr{D}}$ can be decomposed as a *trilinear* multiplication of a $rank-(r_d, r_f, r_p)$ core tensor $\mathscr{C} \in \mathbb{R}^{r_d \times r_f \times r_p}$ 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

$$\widehat{\mathscr{D}} = (X, Y, W) \cdot \mathscr{C} \quad (9)$$

where the (i, j, k)-th entry of $\widehat{\mathscr{D}}$ is computed by

$$\widehat{\mathscr{D}}(i,j,k) = \sum_{\alpha=1}^{r_d}\sum_{\beta=1}^{r_f}\sum_{\gamma=1}^{r_p} x_{i\alpha} y_{j\beta} w_{k\gamma} \mathscr{C}(\alpha,\beta,\gamma). \quad (10)$$

Thus without loss of generality, the minimization problem as defined in (8) is equivalent to the following

$$\min_{X,Y,W,\mathscr{C}} \|\mathscr{D} - (X,Y,W)\cdot\mathscr{C}\|_{\mathrm{F}}, \quad \text{s.t.} \quad X^\top X = I_{r_d}, Y^\top Y = I_{r_f} \text{ and } W^\top W = I_{r_p}. \quad (11)$$

Finally, with the regression model built in (3) and (4), a new set of direction transfer functions can be retrieved by

$$D_{new} = (X, Y, w_{new}^\top) \cdot \mathscr{C} \in \mathbb{R}^{N_d \times N_f} \quad (12)$$

We refer to [11] for Tensor-SVD algorithms and further discussions.

3.3 Customization using GLRAM

Similar to Tensor-SVD, GLRAM methods do not require destruction of 3D tensors. Given a dataset $\mathscr{D} \in \mathbb{R}^{N_d \times N_f \times N_p}$, the task of GLRAM is to approximate slides (matrices) $\mathscr{D}(:,i,:)$, for $i = 1, \ldots, N_f$ of $\mathscr{D}$ along the second direction by a set of low rank matrices $\{XM_iW^\top\} \subset \mathbb{R}^{N_d \times N_p}$, for $i = 1, \ldots, N_f$, where the matrices $X \in \mathbb{R}^{N_d \times r_d}$ and $W \in \mathbb{R}^{N_p \times r_p}$ are of full rank, and the set of matrices $\{M_i\} \subset \mathbb{R}^{r_d \times r_p}$ with $N_d > r_d$ and

	PCA	*2DPCA*	*GLRAM*	*TSVD*
Subject 153	L:4.74dB R:5.19dB	L:4.67dB R:5.15dB	L:4.65dB R:5.14dB	L:4.65dB R:5.14dB
Subject 154	L:6.11dB R:6.05dB	L:5.63dB R:5.67dB	L:5.59dB R:5.66dB	L:5.59dB R:5.66dB
Subject 155	L:5.27dB R:5.06dB	L:5.25dB R:5.06dB	L:5.24dB R:5.05dB	L:5.24dB R:5.05dB
Subject 156	L:5.34dB R:5.73dB	L:5.17dB R:5.64dB	L:5.16dB R:5.63dB	L:5.16dB R:5.63dB
Subject 162	L:6.32dB R:5.60dB	L:6.15dB R:5.28dB	L:6.12dB R:5.24dB	L:6.12dB R:5.24dB
Subject 163	L:5.45dB R:5.20dB	L:5.29dB R:5.08dB	L:5.27dB R:5.07dB	L:5.27dB R:5.07dB
Subject 165	L:5.03dB R:6.04dB	L:4.81dB R:5.68dB	L:4.80dB R:5.66dB	L:4.80dB R:5.66dB

Table 1: Average spectral distortion values over all angles over the whole frequency spectrum.

$N_p > r_p$. This can be formulated as the following optimization problem

$$\min_{X,W,\{M_i\}_{i=1}^{N_p}} \sum_{i=1}^{N_p} \left\| \left(\mathscr{D}(:,:,i) - XM_iW^\top\right) \right\|_F, \quad \text{s.t.} \quad X^\top X = I_{r_d} \text{ and } W^\top W = I_{r_p}. \tag{13}$$

Let us construct a 3D array $\mathscr{M} \in \mathbb{R}^{r_d \times N_f \times r_p}$ by assigning $\mathscr{M}(:,i,:) = M_i$ for $i = 1,\ldots,N_f$. The minimization problem as defined in (13) can be reformulated in a Tensor-SVD style, i.e.

$$\min_{X,W,\mathscr{M}} \left\| \mathscr{D} - (X, I_{N_f}, W) \cdot \mathscr{M} \right\|_F, \quad \text{s.t.} \quad X^\top X = I_{r_d} \text{ and } W^\top W = I_{r_p}. \tag{14}$$

Instead of reducing the dataset $\mathscr{D}$ along all three directions as Tensor-SVD, GLRAM methods work with two pre-selected directions of a 3D data array. The storage space for the GLRAM-reduced dataset depends on the values of r_d and r_p.

Finally, a new set of direction transfer functions can be retrieved by

$$D_{new}(:,i,:) = X\mathscr{M}(:,i,:)w_{new}. \tag{15}$$

We refer to [15] for more details on GLRAM algorithms.

4. EXPERIMENTAL COMPARISON

In this section, we apply PCA, 2DPCA, GLRAM and Tensor-SVD to reduce dimensionality of direction transfer functions. Performance of HRTF customization using the corresponding regression model is investigated and discussed.

4.1 Experimental Setting

In the experiment, the CIPIC database [1] is used for the HRTF customization application. The database contains 37 Head Related Impulse Response (HRIR) tensors with the corresponding anthropometric data for both left and right ears. The CIPIC HRIRs are recorded in spatial resolution of $N_d = 1250$ points ($N_e = 50$ in elevation and $N_a = 25$ in azimuth), spaced uniformely around the head, with $N_t = 200$ time samples. To obtain the HRTFs, the Discrete Fourier Transformation (DFT) was applied on each HRIR.

The direction transfer functions $\mathscr{D}$ of the left and right ear HRTF magnitude of the first 30 persons in the CIPIC dataset, together with their anthropometry were taken as a training set for multiple linear regression, as explained in section 2.

For the regression model (2), we select the anthropometric parameters in accordance to [4]. It is demonstrated that eight selected parameters out of 27 from the original CIPIC database provide good regression performance. These eight paramters are: head width, head depth, shoulder width, cavum concha height, cavum concha width, fossa height, pinna height and pinna width.

4.2 Experimental Results

In each experiment, we construct a new set of HRTFs for persons not in the training set with one of the introduced feature extraction methods. To investigate the performance of the different feature extraction approaches in a HRTF customization application, the spectral distortion for every angle over the whole frequency spectrum is computed. The spectral distortion is defined as follows:

$$SD = \sqrt{\frac{1}{N_f}\sum_{i=1}^{N_f}\left(20\log_{10}\frac{|H_i|}{|H_{new_i}|}\right)^2}, \tag{16}$$

where H_i is the magnitude of the CIPIC-measured HRTF in the dataset and H_{new_i} is the magnitude of the HRTF constructed via regression at the i-th frequency. N_f is the number of frequency samples for each HRTF (200 in this case).

First of all, PCA is applied to the task. We use that as a reference for comparison with the other three multilinear methods. We select $r_d = 10$ dominant eigenvectors, which result in a data reduction of 83%. For seven subjects, who are not in the CIPIC training dataset, Table 1 summarizes the average spectral distortion values for estimation of both left and right ear HRTFs.

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. This might be caused by the possibility, that there exists a subject in the training set that is physically similar to these two subjects. Consequently the estimation of these subjects works better due to a more precise regression model.

Furthermore, the spectral distortion values are shown to be different for left and right ears. It indicates that precise determination of anthropometric parameters as well as measurement of HRIRs is sensitive to many other parameters, e.g. placement of the microphones or head movement during the measurement. Such inaccuracy in the training set can consequently lead to different estimation values for left and right ear HRTFs of the same person.

Direction dependency of the estimated HRTFs can be seen in Figure 1 and Figure 2. For different directions, the

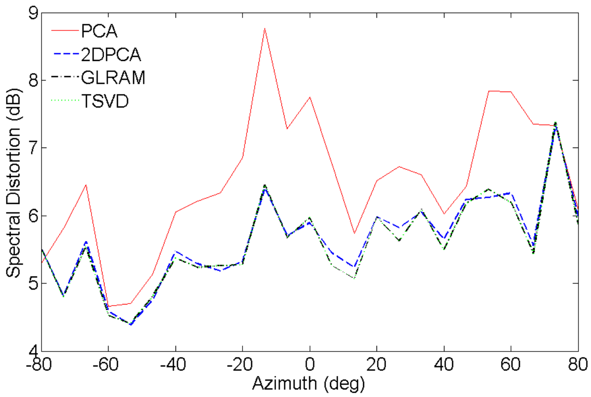


Figure 1: Spectral distortion values of subject 154 in the horizontal plane (elevation=0).

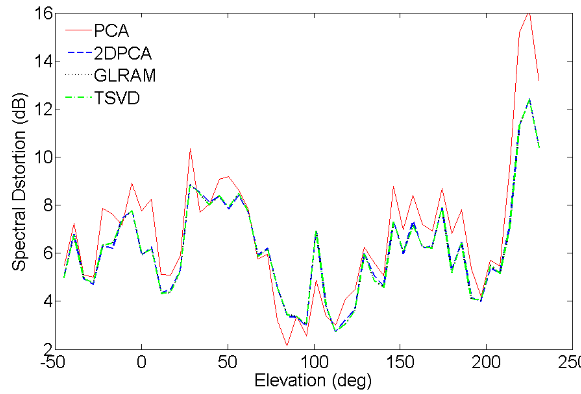


Figure 2: Spectral distortion values of subject 154 in the vertical plane (azimuth=0).

quality of estimation of the individual HRTFs slightly varies in terms of spectral distortion.

In Figure 3(a), the log-magnitude response of the CIPIC-measured and regression-estimated HRTF of a certain angle is shown. Especially for low frequencies, the estimated HRTF approximates the measured one well and could lead to a good result in sound synthesis applications.

Extractions of the direction dependent features with PCA disregard similarities between neighbouring angles. To take also the 3D structure of the dataset into account, 2DPCA is applied also using $r_d = 10$ eigenvectors, leading also to a data reduction rate of 83%. Results in terms of spectral distortions of HRTF customizations at two particular planes, shown in Figure 1 and Figure 2, indicate that 2DPCA extracted features lead to a better estimation of the HRTFs than PCA at the same data reduction rate. In Figure 3(b) one can see, that the log-magnitude response of the estimated HRTF is closer to the CIPIC-measured one than the PCA estimated one.

Finally, GLRAM and Tensor-SVD are applied to estimate the individual HRTFs of the test subjects not in the training set. The results achieved by GLRAM and Tensor-SVD are similar to the 2DPCA feature extraction, but the data reduction by GLRAM and Tensor-SVD is higher than by PCA and 2DPCA. The regression using GLRAM and Tensor-SVD with $r_p = 10$, $r_f = 200$ and $r_d = 100$ leads to data reduction of 97%.

As a final remark, it is worth noticing that, for regression using 2DPCA, GLRAM and Tensor-SVD, we only need to construct the multiple linear regression model once, while for PCA based approach, it requires to compute the multiple linear regression model for each direction.

5. CONCLUSION AND FUTURE WORK

In this paper, we address the problem of customization of HRTF dataset using PCA, 2DPCA, Tensor-SVD and GLRAM. Our experiments demonstrate that 2DPCA, Tensor-SVD and GLRAM outperform the standard PCA approach with respect to the spectral distortion values. Meanwhile GLRAM and Tensor-SVD approaches lead to a higher reduction rate than PCA and 2DPCA.

To evaluate the performance also in HRTF based 3D-sound applications with respect to the listener, we plan to investigate performance of multiway array analysis also on test subjects in listening experiments.

Acknowledgements

This work was fully supported by the German Research Foundation (DFG) within the collaborative research center SFB-453 ”High Fidelity Telepresence and Teleaction“.

REFERENCES

[1] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano. The CIPIC HRTF database. In *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 21–24, New Paltz, NY, 2001.

[2] J. Blauert. An introduction to binaural technology. In *Binaural and Spatial Hearing*, pages 593–609, R. Gilkey, T. Anderson, Eds., Lawrence Erlbaum, Hilldale, NJ, USA, 1997.

[3] G. Grindlay and M. Vasilescu. A multilinear (tensor) framework for HRTF analysis and synthesis. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 1, pages 161–164, Honolulu, Hawaii, USA, 2007.

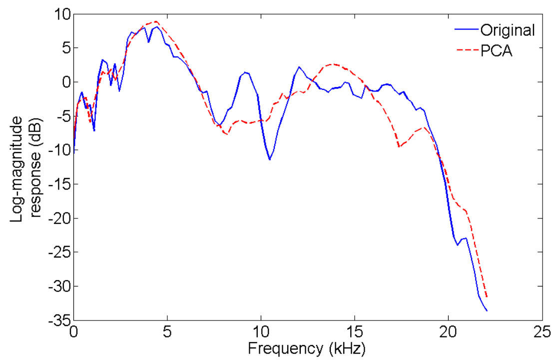
[4] H. Hu, L. Zhou, H. Ma, and Z. Wu. Head related transfer function personalization based on multiple regression analysis. In *Proceedings of the 2006 International Conference on Computational Intelligence and Security*, pages 1829–1832, Guangzhou, China, 2006.

[5] I. T. Jolliffe. *Principal Component Analysis*. Springer, second edition, 2002.

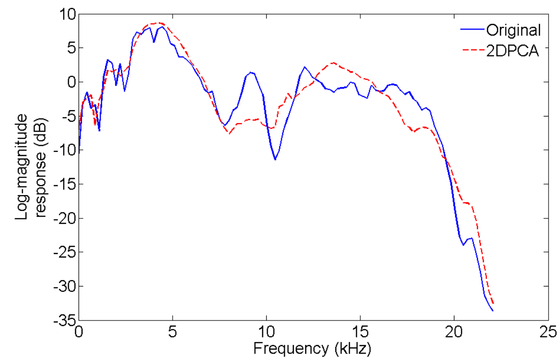
[6] D. J. Kistler and F. L. Wightman. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *Journal of the Acoustical Society of America*, 91(3):1637–1647, 1992.

[7] L. Lathauwer, B. Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1253–1278, 2000.

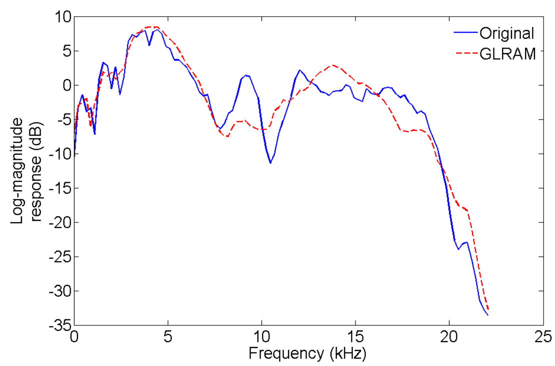
[8] J. Middlebrooks and D. Green. Observations on a principal components analysis of head-related transfer



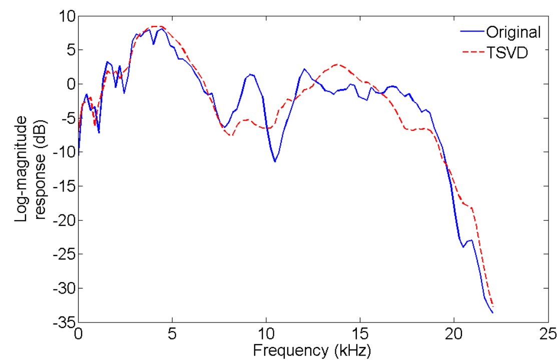
(a) PCA reconstructed HRTF



(b) 2DPCA reconstructed HRTF



(c) GLRAM reconstructed HRTF



(d) T-SVD reconstructed HRTF

Figure 3: Log-magnitude response of the measured CIPIC-HRTF (Original) and the estimated HRTF (subject 153, azimuth = 5°, elevation = 16.875°).

functions. *Journal of the Acoustical Society of America*, 92(1):597–599, 1992.

[9] T. Nishino, N. Inoue, K. Takeda, and F. Itakura. Estimation on HRTFs on the horizontal plane using physical features. *Applied Acoustics*, 68(8):897–908, 2007.

[10] M. Rothbucher, H. Shen, and K. Diepold. Dimensionality reduction in HRTF by using multiway array analysis. In *Proceedings of the 3rd International Workshop on Human-Centered Robotic Systems (HCRS'09)*, pages 102–108, Bielefeld, Germany, November 2009.

[11] B. Savas and L. Lim. Best multilinear rank approximation of tensors with quasi-Newton methods on Grassmannians. Technical Report LITH-MAT-R-2008-01-SE, Department of Mathematics, Linkpings University, 2008.

[12] B. N. Sheehan and Y. Saad. Higher order orthogonal iteration of tensors (HOOI) and its relation to PCA and GLRAM. In *Proceedings of the 2007 SIAM International Conference on Data Mining*, pages 355–366, Minnenpolis, Minnesota, USA, 2007.

[13] D. L. Wang and G. J. Brown. *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*. Wiley-Interscience, IEEE Press, 2006.

[14] J. Yang, D. Zhang, A. F. Frangi, and J.-Y. 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.

[15] J. Ye. Generalized low rank approximations of matrices. *Machine Learning*, 61(1-3):167–191, 2005.