

RESEARCH

Open Access

Modeling individual HRTF tensor using high-order partial least squares

Qinghua Huang* and Lin Li

Abstract

A tensor is used to describe head-related transfer functions (HRTFs) depending on frequencies, sound directions, and anthropometric parameters. It keeps the multi-dimensional structure of measured HRTFs. To construct a multi-linear HRTF personalization model, an individual core tensor is extracted from the original HRTFs using high-order singular value decomposition (HOSVD). The individual core tensor in lower-dimensional space acts as the output of the multi-linear model. Some key anthropometric parameters as the inputs of the model are selected by Laplacian scores and correlation analyses between all the measured parameters and the individual core tensor. Then, the multi-linear regression model is constructed by high-order partial least squares (HOPLS), aiming to seek a joint subspace approximation for both the selected parameters and the individual core tensor. The numbers of latent variables and loadings are used to control the complexity of the model and prevent overfitting feasibly. Compared with the partial least squares regression (PLSR) method, objective simulations demonstrate the better performance for predicting individual HRTFs especially for the sound directions ipsilateral to the concerned ear. The subjective listening tests show that the predicted individual HRTFs are approximate to the measured HRTFs for the sound localization.

Keywords: Head-related transfer function; Tensor; Multi-linear model; Laplacian score; High-order partial least squares

1 Introduction

The generation of virtual three-dimensional (3D) audio based on head-related transfer functions (HRTFs) becomes important in many applications, such as PC entertainment, hearing aids, multimedia, and virtual reality, with spatial and immersive feelings in 3D auditory space. Its key technique is to recover the location information of a sound source by HRTFs. HRTFs describe the spectral changes of sound waves from the sound positions to a listener's ear canal, due to the diffraction and reflection properties of the anthropometric structures. The corresponding representations in the time domain are head-related impulse responses (HRIRs). HRTFs not only vary with sound source locations (elevations and azimuths) and frequencies but also depend on external physiological structures uniquely from one listener to another. The tiny difference of anthropometric structures can create a significant influence on HRTFs for sound localization.

Perceptual distortions may occur in spatial sound localization with non-individual HRTFs. Unfortunately, individual HRIR measurements for each listener are very time consuming and difficult to implement with specific instruments. So, it is not practical and economical for various applications. It is essential to obtain individual HRTFs fast and effectively.

Some theoretical calculation methods were used to generate individual HRTFs based on a snowman model [1] and the boundary element method (BEM) [2], which are unsuitable for the individual HRTF estimation at high frequencies. They also need a large amount of calculations and have a high request for computers [3]. To customize HRTFs more easily and effectively, some researchers attempted to explore the effect resulting from the difference of anthropometric structures on HRTFs [4-6] and further study the relationship between HRTFs and anthropometric parameters [7,8]. Zotkin et al. measured the pinna size of a new subject and used the similarity of pinna structures between the new subject and the listeners in a known HRTF database to select the best matching HRTFs as the individual HRTFs of that subject [9]. Zeng

* Correspondence: qinghua@shu.edu.cn
School of Communication and Information Engineering, Shanghai University,
Shanghai 200072, People's Republic of China

et al. applied a hybrid algorithm for selecting the individual HRTFs based on the similarity of anthropometric structures [10]. These HRTF personalization methods based on database matching were limited by the sizes of the existing databases and the matching criteria. Consequently, individual modeling based on anthropometric measurements is a big breakthrough in HRTF customization. HRTF estimation has three indispensable parts including the dimension reduction of original HRTFs, a reasonable selection of anthropometric measurements, and the mapping relation between the compacted HRTFs and the selected anthropometric measurements.

HRTFs depend on frequencies, sound directions (azimuths and elevations), and listeners. The collection of high-spatial-resolution HRTFs of each listener makes up a large dataset with a multi-dimensional structure and complex characteristic. It is difficult to directly apply the original HRTFs into learning and storing. Due to the high dimensionality, it is necessary to extract the individual factors with lower dimension from the original HRTFs and get rid of non-individual features. Principal component analysis (PCA) was popularly applied to get individual weight coefficients and basis vectors before the HRTF customization [11-15]. Sodnik et al. found a suitable representation for the weight variations of the HRTF amplitudes by PCA [16]. Wang et al. applied PCA for the HRTF compression [17] and interpolation [18], respectively. Xie presented to recover the high-spatial-resolution HRTFs from the individual weight coefficients by a small set of measurements [19]. Kistler and Wightman modeled the HRTF matrix by PCA [20]. PCA successfully reduces the dimensionality of HRTF datasets and it is based on the so-called vector space model. Under this model, the HRTFs of a subject at different source locations are modeled as a vector and the collection of individual HRTFs is modeled as a matrix. It cannot capture the variations among different sound source directions and the interaction of multiple variables in HRTFs. To overcome the weakness of the vector model, Grindlay and Vasilescu modeled HRTFs using a multi-factor (tensor) representation [21]. The tensor framework is used to learn the interaction of the multi-variable for HRTF representation and can achieve the dimensionality reduction of the HRTF dataset along each variable separately. Rothbacher et al. used multi-way array analysis to customize HRTFs [22]. There is no specific selection of anthropometric parameters in combination with the HRTF tensor. However, the key parameter selection is important for the accuracy of HRTF model estimation.

A number of anthropometric parameters (head, torso, pinna, etc.) affect the immersion of listener's spatial hearing and show different influence on HRTFs. It is undesirable to apply all the anthropometric measurements for modeling the individual HRTFs. A reasonable selection

of the anthropometric measurements is necessary to the individual HRTF customization. Xu et al. [5] found that ear parameters were significantly correlated with the magnitudes of HRTFs at high frequencies. Chen et al. studied the influence of the neck and torso parameters on the near-field HRTFs [23]. Rothbacher et al. [7] developed a measurement system that was capable to scan a human body for the anthropometric measurements. Hu et al. used correlation coefficients twice for the parameter selection and retained eight parameters for the HRTF estimation [12,24,25]. Zeng et al. [10] also selected 13 reference physiological parameters utilizing correlation analyses between the measured HRTFs and all the anthropometric parameters for matching the best HRTFs. Xu et al. [26] used a weighted correlation method and selected eight significant parameters. Hugeng and Gunawan [27] analyzed the correlation among anthropometric parameters and three physical quantities (interaural time difference, interaural level difference, and pinna notch frequency). However, correlation analyses only describe the linear relationship among anthropometric parameters and cannot evaluate the significance of a single parameter. These existing methods for the parameter selection did not consider the relation between the anthropometric parameters and the multi-dimensional HRTFs. In order to avoid the HRTF vector modeling, it is necessary to keep the multi-dimensional structure of the original HRTFs and on this basis to select the key parameters.

Once the individual HRTF factors with lower dimension and a few key parameters were obtained, some methods were widely used to construct the relationship between them. The anthropometric parameters were treated as the inputs and lower-dimensional HRTFs as the outputs. Many researchers constructed the HRTF prediction model based on an assumption of a linear relation between the HRTF vectors and anthropometric parameters [11-13,24,28,29]. In [12,24,28,29], the relation between the HRTFs and physical sizes of the head and ear was investigated by the multiple regression analysis and optimized by the least squares method. The performance of the estimated HRTFs was evaluated by objective and subjective ways. The results indicated that good performance was obtained with no significant difference between the measured and estimated HRTFs with respect to perception when the bandwidth ranged from 0 to 8 kHz. To get rid of trivial anthropometric measurements and improve the performance of the HRTF estimation, Hu et al. used partial least squares regression (PLSR) to model the linear relation [24]. Subsequently, to further describe the scattering of the incident sound by the physical structures, many researchers explored some non-linear multivariable statistical estimation methods to improve the performance of HRTF customization [25,30-32]. In [25], a three-layer back-

propagation artificial neural network (ANN) was used to HRTF personalization. Huang et al. applied support vector regression (SVR) to model personalized HRIRs [30,31]. All these linear or non-linear HRTF personalization methods were based on the vector model to customize HRTFs. They are unable to establish the mapping from anthropometric parameters to a HRTF tensor. In order to predict the individual HRTFs of different sound directions, it requires a separate regression model for each sound direction. A high-spatial-resolution HRTF prediction based on the tensor model requires only one regression. Therefore, it is necessary to consider the HRTF data involving multiple variables as multi-way data structures [33] and a multi-linear extension for modeling a HRTF tensor using a small set of anthropometric measurements.

The above considerations motivate us to construct a multi-linear model for predicting an individual HRTF tensor. To learn HRTFs of any listener from a measured database, we present a HRTF customization through three steps. First, to keep the inherent interaction of different variables in HRTFs, a tensor is used to describe measured HRTFs, and the individual core tensor (ICT) with lower dimension in tensor subspace is extracted by high-order singular decomposition (HOSVD). Then, combining with the lower-dimensional ICT, a few anthropometric parameters are selected in consideration of the local geometric structure and global information in the parameter data space. Last, we use a multi-linear subspace regression to model the multi-linear relationship between the ICT and the selected key parameters. Section 2 presents the data processing including the dimension reduction of the HRTF tensor and the selection of key anthropometric parameters. The multi-linear subspace regression between the compacted HRTF tensor and the selected key parameters is developed in Section 3. The proposed method can realize the direct multi-linear mapping from the anthropometric parameters to the HRTF tensor. In Section 4, we give the results and discussions of the proposed HRTF personalization method. The conclusions are given in the last section.

2 Data processing

2.1 Notations and basic multi-linear algebra

In order to facilitate the distinctions of scalars, matrices, and tensors, the following notations are used. Scalars are denoted by italic letters, e.g., a ; vectors by lowercase boldface letters, e.g., \mathbf{a} ; matrices by uppercase boldface letters, e.g., \mathbf{A} ; and tensors by boldface calligraphic letters, e.g., \mathcal{A} . The i th entry of the vector \mathbf{a} is denoted by a_i , the (i, j) -element of the matrix \mathbf{A} is denoted as a_{ij} , the column- n vector of the matrix \mathbf{A} as \mathbf{a}_n , and the element (i_1, i_2, \dots, i_N) of an N th-order tensor $\mathcal{A} \in$

$\mathbb{R}^{I_1 \times \dots \times I_n \times \dots \times I_N}$ by $a_{i_1 i_2 \dots i_N}$. The indices range from 1 to their corresponding capital versions, e.g., $i_N = 1, 2, \dots, I_N$. The n -mode unfolding operation of the tensor \mathcal{A} is denoted by $\mathbf{A}_{(n)} \in \mathbb{R}^{I_n \times (I_1 \dots I_{n-1} I_{n+1} \dots I_N)}$. The n th factor matrix is denoted by $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times I_n}$. The $I_n \times I_n$ unitary matrix is denoted by \mathbf{U}_{I_n} .

The superscripts T and + are used for representing the matrix transposition and the Moore-Penrose pseudoinverse, respectively. \otimes represents the Kronecker product. Since the following sections in this paper mainly focus on the data of a third-order HRTF tensor, let us introduce the fundamental of a third-order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$. It has three modes with size I_n along mode n . The Frobenius norm of \mathcal{A} is computed as

$$\|\mathcal{A}\|_F = \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \sum_{i_3=1}^{I_3} |a_{i_1 i_2 i_3}|^2} \quad (1)$$

The n -mode vectors of \mathcal{A} are obtained by varying the n th index and keeping all other indices fixed. The n -mode product of the tensor \mathcal{A} and a matrix $\mathbf{U} \in \mathbb{R}^{I_n \times I_n}$ is denoted by $\mathcal{B} = \mathcal{A} \times_n \mathbf{U}$. It is calculated by multiplying all n -mode vectors from the left-hand side by the matrix \mathbf{U} . For example, if $n = 2$, $\mathcal{B} = \mathcal{A} \times_2 \mathbf{U} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ is calculated as

$$b_{i_1 j_2 i_3} = \sum_{i_2=1}^{I_2} a_{i_1 i_2 i_3} u_{j_2 i_2} \quad (2)$$

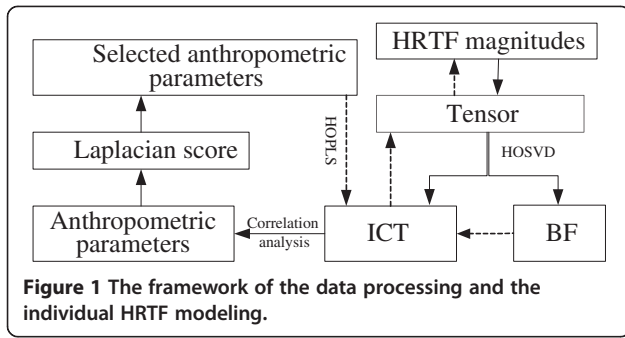
The HOSVD of a third-order tensor \mathcal{A} is denoted as follows:

$$\mathcal{A} = \mathcal{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)} = [\mathcal{S}; \mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \mathbf{U}^{(3)}] \quad (3)$$

where $\mathcal{S} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ is the core tensor [34]. $\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times I_n}$ is the unitary matrix and can be calculated by performing a matrix singular value decomposition (SVD) on the $\mathbf{A}_{(n)}$ [35]. The last term is the simplified notation [36].

2.2 Data processing

The structure of the data processing and the individual HRTF modeling are shown in Figure 1 along the solid arrows. BF represents basis function. Firstly, a tensor is used to model the measured HRTFs. The individual core tensor can be extracted from the HRTF tensor. Secondly, the key parameters are selected in combination with the individual core tensor by correlation analyses and Laplacian scores. The selected anthropometric parameters and the individual core tensor are prepared for the latter multi-linear learning by high-order partial least squares (HOPLS) in Section 3. The prediction of HRTF magnitudes can be achieved after the data processing along the dotted arrows.



2.2.1 Dimension reduction of the HRTF tensor

Each measured HRIR can be transformed into the corresponding complex HRTF by fast Fourier transform (FFT). The HRTF is defined as the ratio of the sound pressure

$$H(p, r, \theta, \phi, f) = \frac{P(p, r, \theta, \phi, f)}{P_0(r, f)} \quad (4)$$

where f is the frequency and r is the distance from the sound source to the center of a listener's head. The spatial direction of the sound source is marked by azimuth θ and elevation ϕ . The individual factors are embodied in the first variable p on behalf of different subjects. $P(p, r, \theta, \phi, f)$ represents the sound pressure at the left or right ear, and $P_0(r, f)$ is the sound pressure at the center of the listener's head with the listener absent. In the following section, r is omitted because HRTFs are asymptotically independent of distance in the far field ($r > 1$ m) [37].

Even in the far field, HRTFs are functions of frequencies and sound directions uniquely from one person to another. To keep the interaction of different variables, a third-order tensor is used to describe the HRTFs of different subjects. Due to the high dimension of each mode, a core tensor in a lower-dimensional subspace is extracted from the original HRTF tensor by HOSVD. It still keeps the multi-dimensional structure and contains the individual characteristic of the measured HRTFs.

A tensor $\mathcal{H} \in \mathbb{R}^{P \times D \times F}$ denotes the HRTF magnitudes without phases of P subjects at D sound directions, where F is the number of frequencies. In order to extract

the individual factors, the dimensions of the frequency and the direction should be reduced and the subject mode is unchanged. HOSVD is the extension of conventional SVD for higher-order tensor decomposition [34]. Using HOSVD on the latter two modes, we can get the decomposition of the high-order tensor \mathcal{H} as

$$\mathcal{H} = \mathcal{W} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)} \quad (5)$$

where $\mathbf{U}^{(m)}$ is the left singular matrix of the m -mode unfolding matrix of \mathcal{H} , m equals 2 or 3 corresponding to the 2- (direction) or 3-mode (frequency) of \mathcal{H} , respectively, $\mathbf{U}^{(2)} \in \mathbb{R}^{D \times D}$, $\mathbf{U}^{(3)} \in \mathbb{R}^{F \times F}$, and $\mathcal{W} \in \mathbb{R}^{P \times D \times F}$ has the same sizes as those of \mathcal{H} . The main variations of the HRTF tensor can be explained by parts of the basis vectors of $\mathbf{U}^{(m)}$. So, a truncation on $\mathbf{U}^{(m)}$ can achieve an approximate representation of the original HRTF tensor. The projection on the HRTF tensor subspace spanned by the truncated matrices [38] is obtained as

$$\tilde{\mathcal{W}} = \mathcal{H} \times_2 \tilde{\mathbf{U}}^{(2)T} \times_3 \tilde{\mathbf{U}}^{(3)T} \quad (6)$$

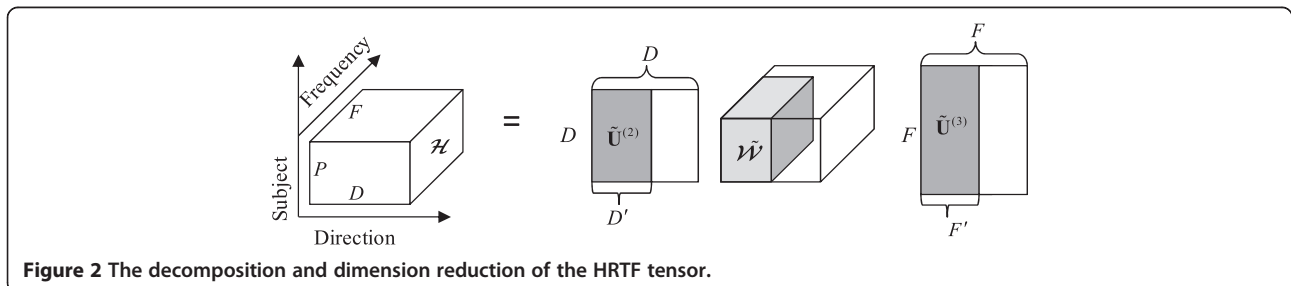
where $\tilde{\mathcal{W}} \in \mathbb{R}^{P \times D' \times F'}$ ($D' < D$ and $F' < F$) is the ICT and $\tilde{\mathbf{U}}^{(m)}$ denotes the truncated matrix of $\mathbf{U}^{(m)}$ called the basis function. The ICT has lower dimension for each mode excluding the subject mode compared to the original HRTFs. The characteristic of the individual core tensor will be discussed in Section 4. Figure 2 shows the HRTF tensor decomposition. The approximate reconstruction of the original HRTFs is through

$$\tilde{\mathcal{H}} = \tilde{\mathcal{W}} \times_2 \tilde{\mathbf{U}}^{(2)} \times_3 \tilde{\mathbf{U}}^{(3)} \quad (7)$$

with a controllable error, where $\tilde{\mathcal{H}}$ is the approximation of the original HRTF tensor. The error analysis of the HRTF approximation is introduced in the following experiments of Section 4.1.

2.2.2 Selection of key anthropometric parameters combining with the ICT

HRTFs describe the responses resulting from the diffraction and reflection of listener's anthropometric structures and are related to anatomy concentrating on the head, torso, and pinna. Each listener has his specific anthropometric shape and size. The parameter data space can be obtained by anthropometric measurements of the



physiological structures from each subject [7,39]. There is correlation among all the different anthropometric parameters. The correlation coefficients are not equal or close to one. So, the anthropometric parameters cannot replace each other, and it is better to select a group of necessary anthropometric measurements for approximately reflecting the fundamental property of HRTFs. How to select such a group of anthropometric parameters is a key work. We can take the following three procedures to select the key anthropometric parameters for the latter multi-linear personalization modeling.

First, in order to measure the influence of parameters on HRTFs, the correlation analyses are performed between all the measured parameters and the individual core tensor. The parameters with larger correlation coefficients are reserved as the results of the first selection procedure.

Then, we use a Laplacian score to further select appropriate parameters. It can measure the importance of each anthropometric parameter. In order to avoid the unbalance selection of similar parameters, the reserved parameters are divided into three classes before calculating the Laplacian score. We examine the intrinsic properties of the parameter space to evaluate each parameter after the correlation analysis. For each parameter, its Laplacian score is computed to reflect the locality preserving power. Laplacian score is based on the local observation and an assumption that two parameters are probably related to the same topic if they are close to each other [40].

Suppose there are P subjects and K parameters of each subject. Let a_{pk} denote the k th anthropometric parameter of the p th subject, $k = 1, 2, \dots, K, p = 1, 2, \dots, P$. All the parameters can be denoted by $\mathbf{A}_{all} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_P]^T \in \mathbb{R}^{P \times K}$. The vector \mathbf{a}_p contains all the elements a_{pk} with $k = 1, 2, \dots, K$. It is treated as a data point which represents all the anthropometric measurements of the p th subject and corresponds to the p th node of a graph. In order to model the local geometric structure of the measured parameters, a nearest neighbor graph with P nodes is established. If \mathbf{a}_p is close to another parameter vector $\mathbf{a}_{p'} (p' = 1, 2, \dots, P, p' \neq p)$, an edge is put between these two nodes p and p' . If nodes p and p' are connected, a weight is assigned to the edge as

$$S_{pp'} = \frac{\|\mathbf{a}_p - \mathbf{a}_{p'}\|^2}{t} \quad (8)$$

where t is an appropriate constant. Otherwise, $S_{pp'} = 0$. All the weights on the graph model the local structure of the parameter space. For a parameter we choose, it is reasonable to minimize the following objective function

$$L_k = \frac{\sum_{pp'} (a_{pk} - a_{p'k})^2 S_{pp'}}{\text{var}(\mathbf{b}_k)} \quad (9)$$

where $\mathbf{b}_k = [a_{1k}, a_{2k}, \dots, a_{Pk}]^T$ consists of the k th parameter of all P subjects and $\text{var}(\cdot)$ is the variance computation. L_k is the Laplacian score of the k th parameter, which concerns two aspects of the reserved anthropometric parameter. One is the variance of the parameter that reflects its representative power. The other relates to the local geometric structure of the parameter data space. It seeks the anthropometric parameters which best reflect the underlying manifold structure and are probably better for predicting HRTFs. Thus, we select these parameters with lower Laplacian scores, which have significant influence on HRTFs at the same time.

Last, correlation analysis is applied to delete some of the above selected parameters that have strong correlation with the others. Through the above selection process, K' key parameters for P subjects are selected as the inputs of the multi-linear HRTF model. It is denoted by a matrix $\mathbf{A} \in \mathbb{R}^{P \times K'}$.

3 Multi-linear personalization modeling by HOPLS

When the individual core tensor and a few key parameters are obtained, a multi-linear HRTF personalization model can be learned by HOPLS regression. HOPLS is a generalized multi-linear regression model with the aim to predict a tensor from a tensor through projecting the data onto the latent space and performing regression on corresponding latent variables [41]. Moreover, it is particularly suitable for small sample sizes [42]. HOPLS regression is used to explore the multi-linear subspace approximation for both the selected parameters and the HRTF tensor. It is employed to learn the relation between the parameter matrix and the individual core tensor. The complexity of the regression model is controlled by the hyperparameters which are the numbers of orthogonal loadings denoted by J_2, J_3, I and latent vectors denoted by R in Figure 3. Figure 3 shows the framework of a joint subspace approximation for the ICT and the anthropometric parameters by the HOPLS model. After the regression model is constructed from training data, the individual HRTFs for a new subject can be predicted by his anthropometric measurements.

Consider a second-order tensor $\mathbf{A} \in \mathbb{R}^{P \times K'}$ containing the selected parameters and a third-order tensor $\tilde{\mathbf{W}} \in \mathbb{R}^{P \times D \times F}$ including all the individual factors of the HRTFs, having the same size in the first mode. The objective is to find an optimal joint subspace approximation for the anthropometric parameters and the individual core tensor based on the latent variables obtained from maximizing the tensor covariance of \mathbf{A} and $\tilde{\mathbf{W}}$. Back to the PLS method, it is to search for the common latent variables from dependent variables \mathbf{X} and independent variables \mathbf{Y} with the constraint that these common components explain as much as possible the covariance between \mathbf{X} and \mathbf{Y}

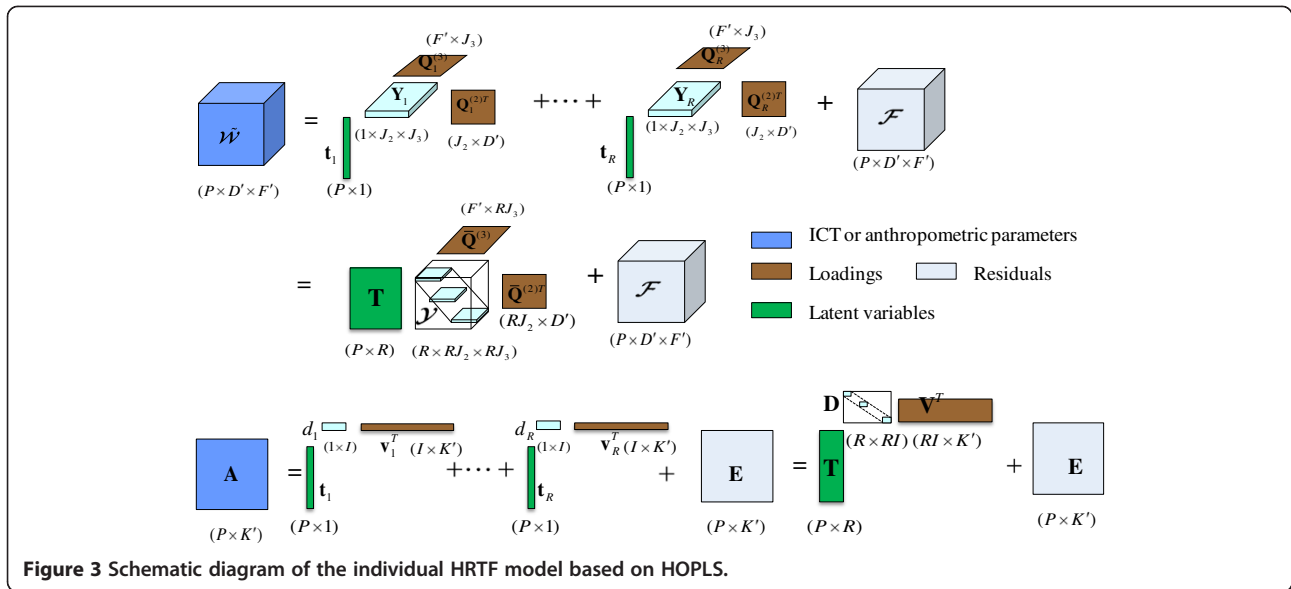


Figure 3 Schematic diagram of the individual HRTF model based on HOPLS.

[24,43]. Here, based on the property, do a high-order extension to find common latent variables for explaining the covariance between $\tilde{\mathbf{W}}$ and \mathbf{A} , as illustrated in Figure 3. The mathematical model can be expressed as [41,42]

$$\begin{aligned} \tilde{\mathbf{W}} &= \sum_{r=1}^R \mathbf{Y}_r \times_1 \mathbf{t}_r \times_2 \mathbf{Q}_r^{(2)} \times_3 \mathbf{Q}_r^{(3)} + \mathcal{F} \\ \mathbf{A} &= \sum_{r=1}^R \mathbf{d}_r \times_1 \mathbf{t}_r \times_2 \mathbf{v}_r + \mathbf{E} \end{aligned} \quad (10)$$

where R is the number of latent vectors, $\mathbf{t}_r \in \mathbb{R}^P$ is the r th latent vector, and $\mathbf{Q}_r^{(2)} \in \mathbb{R}^{D' \times J_2}$ ($J_2 \leq D'$) and $\mathbf{Q}_r^{(3)} \in \mathbb{R}^{F' \times J_3}$ ($J_3 \leq F'$) are the loading matrices corresponding to the latent vector \mathbf{t}_r on the mode of the sound direction and the frequency, respectively; similarly, $\mathbf{v}_r \in \mathbb{R}^{K' \times I}$ ($I \leq K'$) is the loading vector of the parameter matrix \mathbf{A} , and \mathcal{F} and \mathbf{E} are the residuals. Use the rank-(1, J_2 , J_3) decomposition of the individual core tensor $\tilde{\mathbf{W}}$ to get the core tensor $\mathbf{Y}_r \in \mathbb{R}^{1 \times J_2 \times J_3}$ corresponding to the r th latent vector. $\mathbf{d}_r \in \mathbb{R}^{1 \times I}$ is the core of the anthropometric parameter tensor by the rank-(1, I) decomposition. The model in (10) is boiled down to a concise form as

$$\begin{aligned} \tilde{\mathbf{W}} &= \mathcal{Y} \times_1 \mathbf{T} \times_2 \bar{\mathbf{Q}}^{(2)} \times_3 \bar{\mathbf{Q}}^{(3)} + \mathcal{F} \\ \mathbf{A} &= \mathbf{D} \times_1 \mathbf{T} \times_2 \mathbf{V} + \mathbf{E} \end{aligned} \quad (11)$$

where $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_R] \in \mathbb{R}^{P \times R}$ is the latent matrix, $\mathcal{Y} \in \mathbb{R}^{R \times R J_2 \times R J_3}$ is a block-diagonal tensor containing the tensor \mathbf{Y}_r ($r = 1, 2, \dots, R$) on the diagonal line; similarly, the core \mathbf{d}_r ($r = 1, 2, \dots, R$) is contained in a block-diagonal matrix $\mathbf{D} \in \mathbb{R}^{R \times R I}$, direction loading matrix $\bar{\mathbf{Q}}^{(2)} = [\mathbf{Q}_1^{(2)}, \mathbf{Q}_2^{(2)}, \dots, \mathbf{Q}_R^{(2)}] \in \mathbb{R}^{D' \times R J_2}$, frequency loading matrix $\bar{\mathbf{Q}}^{(3)} = [\mathbf{Q}_1^{(3)}, \mathbf{Q}_2^{(3)}, \dots, \mathbf{Q}_R^{(3)}] \in \mathbb{R}^{F' \times R J_3}$, and

the anthropometric parameter loading matrix $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_R] \in \mathbb{R}^{K' \times R I}$.

Observing Figure 3, how to choose R and estimate the loading matrices from $\tilde{\mathbf{W}}$ and \mathbf{A} is the key optimization of the multi-linear subspace regression for individual HRTF customization. There are two different ways for extracting the latent variables: sequential and simultaneous methods. We choose to obtain the latent vectors in sequence since it provides better performance [42]. If the first latent vector is obtained, the other latent vectors can be estimated by the deflation of $\tilde{\mathbf{W}}$ and \mathbf{A} . Therefore, we firstly find the latent vector \mathbf{t}_1 and the corresponding loading matrices $\mathbf{Q}_1^{(m)}$ ($m = 2, 3$) and \mathbf{v}_1 . The subscript r is omitted to simplify the notations in the following discussions. The whole optimization is based on the strategy for the simultaneous minimization of the Frobenius norm of residuals \mathcal{F} and \mathbf{E} , while keeping a common latent vector \mathbf{t} . Assume that $\mathbf{Q}^{(2)}$, $\mathbf{Q}^{(3)}$, \mathbf{v} , and \mathbf{t} are given; then, the cores in (10) can be calculated as

$$\begin{aligned} \mathbf{Y} &= \tilde{\mathbf{W}} \times_1 \mathbf{t}^T \times_2 \mathbf{Q}^{(2)T} \times_3 \mathbf{Q}^{(3)T} \\ \mathbf{d} &= \mathbf{A} \times_1 \mathbf{t}^T \times_2 \mathbf{v}^T \end{aligned} \quad (12)$$

In [42], minimization of the Frobenius norm of the residuals \mathcal{F} and \mathbf{E} under the orthonormality constraint is converted to maximize a cross-covariance tensor. Zhao et al. defined a cross-covariance tensor of independent variables and dependent variables. Then, the optimization problem for loading matrices can be finally formulated as

$$\begin{aligned} \max_{\{\mathbf{Q}^{(2)}, \mathbf{Q}^{(3)}, \mathbf{v}\}} & \|\mathcal{C}; \mathbf{Q}^{(2)T}, \mathbf{Q}^{(3)T}, \mathbf{v}^T\|_F^2 \\ \text{s.t.} & \mathbf{Q}^{(2)T} \mathbf{Q}^{(2)} = \mathbf{I}_{J_2}, \mathbf{Q}^{(3)T} \mathbf{Q}^{(3)} = \mathbf{I}_{J_3}, \mathbf{v}^T \mathbf{v} = \mathbf{1}, \end{aligned} \quad (13)$$

where $\mathbf{C} = \text{cov}(\mathbf{A}, \tilde{\mathbf{W}}) \in \mathbb{R}^{K' \times D' \times F'}$ is a 1-mode cross-covariance tensor [41]. We try to find a rank- (L, J_2, J_3) tensor decomposition of \mathbf{C} by employing HOSVD [34]. When the loading matrices of the parameters and the individual core tensor are estimated, the latent vector \mathbf{t} should explain the variance of the anthropometric parameters as much as possible estimated by

$$\mathbf{t} = \arg \min_{\mathbf{t}} \|\mathbf{A} - \mathbf{D}; \mathbf{t}, \mathbf{v}\|_F^2 \quad (14)$$

When the latent vector \mathbf{t} is fixed, the cores \mathbf{Y} and \mathbf{d} are obtained by (12).

Once the latent vectors and loading matrices are estimated, the prediction of the individual core tensor for a new subject using the corresponding anthropometric measurements \mathbf{a}^{new} can be predicted as

$$\begin{aligned} \mathbf{T}^{\text{new}} &= \mathbf{a}^{\text{new}} \mathbf{V} \mathbf{D}^+ \\ \tilde{\mathbf{W}}^{\text{new}} &= \mathbf{T}^{\text{new}} \mathbf{y}_{(1)} \left(\bar{\mathbf{Q}}^{(3)} \otimes \bar{\mathbf{Q}}^{(2)} \right)^T \end{aligned} \quad (15)$$

4 Experimental results

In the section, the performance of the proposed method is measured by objective evaluation and subjective sound localization based on a large number of HRTF measurements. The Center for Image Processing and Integrated Computing (CIPIC) database provides high-spatial-resolution HRIR measurements of 45 different subjects. It contains measured HRIRs for both left and right ears at 1,250 sound directions (25 azimuths and 50 elevations) [44]. The azimuths vary from -80° to 80° , and the elevations range from -45° to $+230.625^\circ$. Moreover, 27 anthropometric parameters of 45 subjects are measured in the CIPIC database including 17 for the head and the torso from x_1 to x_{17} and 10 for the pinna expressed by $d_1 - d_8$, θ_1 , and θ_2 [44]. The CIPIC database is used to evaluate the performance of our proposed regression model based on HOPLS.

4.1 Simulations of the data processing

4.1.1 HRTF tensor compaction

In the simulations, the HRTFs of the left ears are chosen to construct the model. The HRIRs in CIPIC of each subject are transformed into HRTFs by a 200-point FFT. Collect the HRTFs of random 30 subjects acting as the training samples denoted by a third-order tensor $\mathcal{H} \in \mathbb{R}^{P \times D \times F}$, where P is the number of subjects (30), D is the number of the sound source elevations (50), and F is the frequency points (100). The other subjects are used for testing the model. When the high-dimensional HRTF tensor is acquired, an individual core tensor $\mathcal{W} \in \mathbb{R}^{P \times D' \times F'}$ can be extracted from the original data via HOSVD. The following discussions focus on how to determine the

dimensions D' and F' of the ICT and the performance of the HRTF tensor subspace approximation:

1. The selections of D' and F' depend on the energy loss of the original HRTF tensor in each mode, respectively. The energy contained in the HRTF tensor is calculated by the squared Frobenius norm of \mathcal{H} . It also equals the sum of the squared m -mode singular values [34] expressed by

$$\|\mathcal{H}\|_F^2 = \sum_{i=1}^P \left(\sigma_i^{(1)} \right)^2 = \sum_{i=1}^D \left(\sigma_i^{(2)} \right)^2 = \sum_{i=1}^F \left(\sigma_i^{(3)} \right)^2 \quad (16)$$

where $\sigma_i^{(m)}$ ($m = 1, 2, 3$) is the singular value corresponding to the m -mode of the HRTF tensor. The square of the m -mode singular value is called the m -mode eigenvalue denoted by $\lambda_i^{(m)}$. The eigenvalue magnitudes and their cumulative distributions are shown in Figure 4. The loss of energy with the selected D' and F' is proportional to the sum of the corresponding singular values of the discarded singular vectors contained in $\mathbf{U}^{(m)}$. The ratios of the retained energy to the total energy for different modes are $E_{\text{ratio}}^{(2)} = \sum_{i=1}^{D'} \lambda_i^{(2)} / \sum_{i=1}^D \lambda_i^{(2)}$ and $E_{\text{ratio}}^{(3)} = \sum_{i=1}^{F'} \lambda_i^{(3)} / \sum_{i=1}^F \lambda_i^{(3)}$. D' and F' are less than the original D and F . The dimension reduction of the original HRTF tensor $\mathcal{H} \in \mathbb{R}^{P \times D \times F}$ brings corresponding compression ratio (CR) as

$$\text{CR} = \frac{D \times F \times P}{D \times D' + F \times F' + D' \times F' \times P} \quad (17)$$

According to Figure 4, the numbers of singular vectors are kept in each mode, and the different CRs are shown in Table 1. Similar amount of energy is kept in each mode with the same E_{ratio} , but the dimension reduction of each mode is quite different. This indicates that the redundancy of these two modes is different. The eigenvalue cumulative distributions in Figure 4 show different redundancy between the frequency mode and the sound direction mode, resulting in different selections of D' and F' . Although with the same E_{ratio} , similar amount of variations is kept in each mode, the amount of dimension reduction in the frequency mode (80% for the azimuth -80° , 64% for the azimuth 0° , and 64% for the azimuth 80°) is different from that in the sound direction mode (90%, 48%, and 70% for those three azimuths -80° , 0° , and 80° , respectively). After dimension reduction of \mathcal{H} , the individual core tensor with lower dimension still captures most of the variations in the original

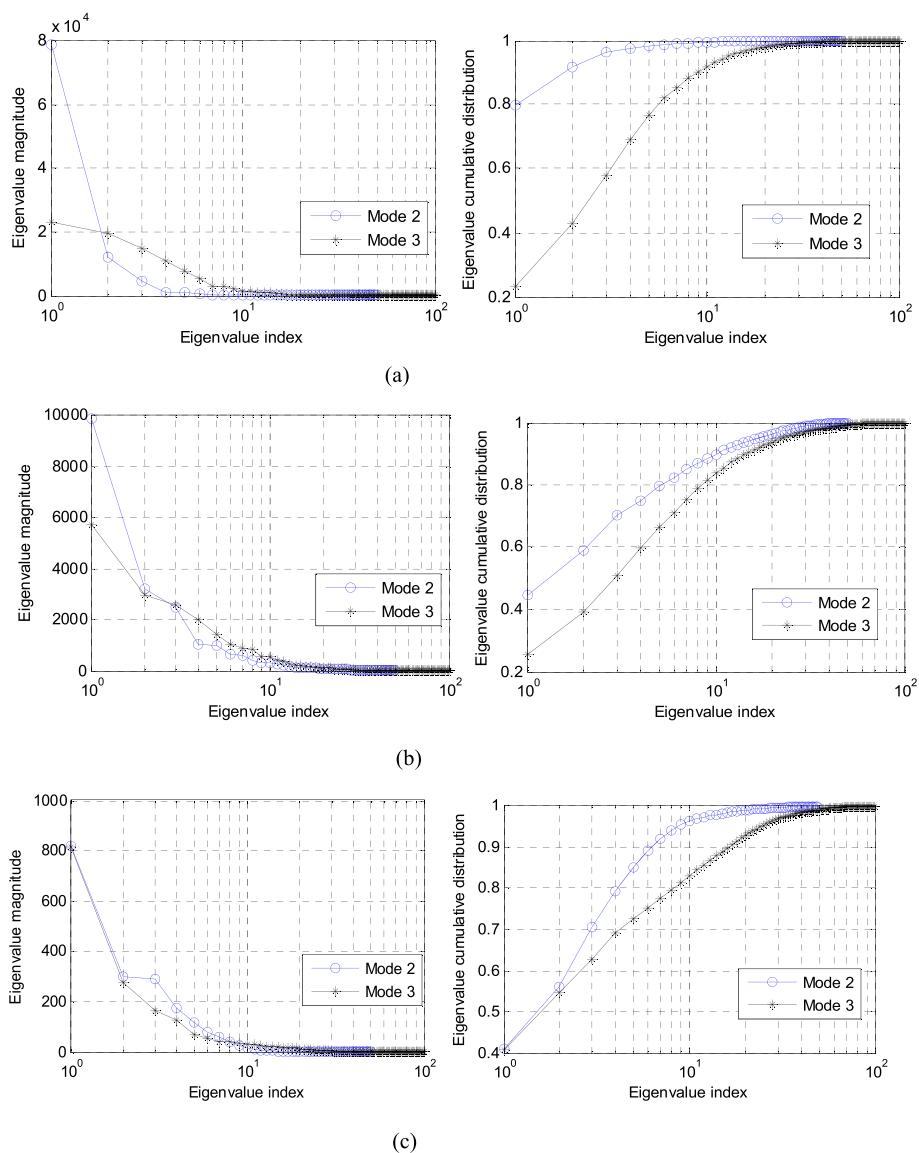


Figure 4 Eigenvalue magnitudes and their cumulative distributions for the HRTF tensor at different azimuths. (a) $\theta = -80^\circ$, (b) $\theta = 0^\circ$, and (c) $\theta = 80^\circ$.

HRTFs. Figure 5 shows the ICT for some subjects. It can be seen that the main energy of the ICT is concentrated in the upper left corner areas.

- To measure the quantitative error of the reconstruction using the basis functions and the

individual core tensor, the signal-to-distortion ratio (SDR) is defined in decibels as

$$SDR(p, \theta, \phi) = 10 \lg \frac{\sum_{f=1}^F |\mathcal{H}(p, \theta, \phi, f)|^2}{\sum_{f=1}^F |\mathcal{H}(p, \theta, \phi, f) - \hat{\mathcal{H}}(p, \theta, \phi, f)|^2} \quad (18)$$

where $\mathcal{H}(p, \theta, \phi, f)$ and $\hat{\mathcal{H}}(p, \theta, \phi, f)$ represent the original and the reconstructed or predicted HRTF, respectively. The average of SDR (ASDR) defined as

Table 1 The selections of D' and F' for three azimuths

Azimuth	D'	F'	CR	$E_{ratio}^{(2)}$	$E_{ratio}^{(3)}$
-80°	5	20	28.6		
0°	26	37	4.4	0.98	0.98
80°	15	37	7.1		

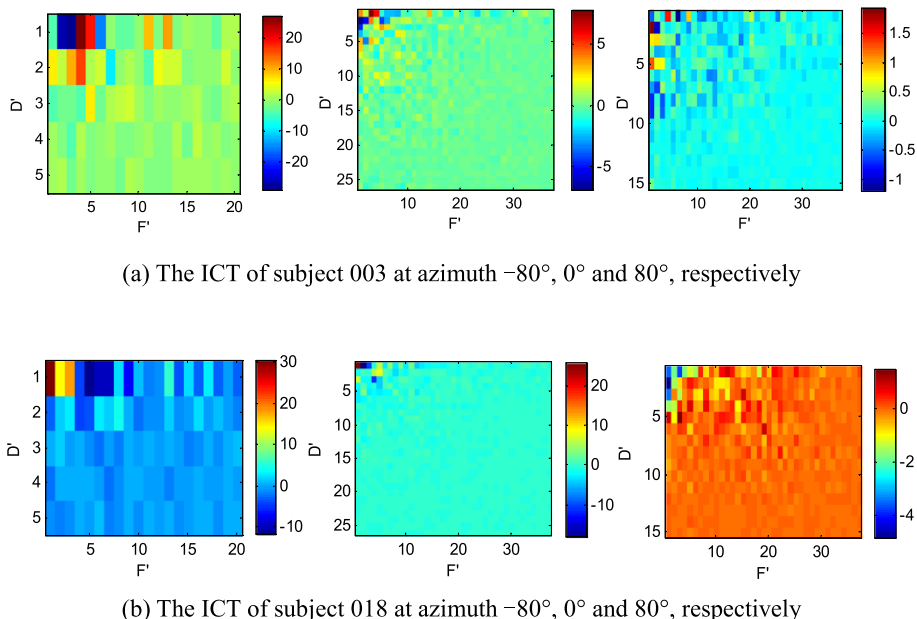


Figure 5 The ICT of different subjects (a) 003 and (b) 018 at azimuths -80° , 0° , and 80° , respectively, obtained by HOSVD.

$ASDR(\theta, \phi) = \frac{1}{P} \sum_{p=1}^P SDR(p, \theta, \phi)$ is used to measure the mean performance for the reconstruction or prediction in the following discussions.

Figure 6a shows the ASDR for $P = 30$ subjects and the reconstruction with the corresponding D' and F' selection. In most cases, ASDR exceeds 20 dB. The average of ASDR over all the azimuths and elevations is 24.3 dB. The results imply that the reconstructed HRTFs can approximate the original HRTFs accurately via selecting appropriate D' and F' . For example, the reconstruction of the subject 003 at three sound directions ($-80^\circ, 0^\circ$), ($0^\circ, 0^\circ$), and ($80^\circ, 0^\circ$) are shown in Figure 6b,c,d compared to the original measured HRTFs. Moderate deviations between the original HRTFs and the reconstructed HRTFs occur at the frequencies of the spectral notches. These reconstruction errors imply that the information loss of the lower-dimensional HRTF tensor may affect the subsequent modeling.

4.1.2 Selecting anthropometric parameters

There are 27 parameters measured in the CIPIC database. The detailed definitions of these parameters can be referred in [44]. To avoid the loss of some important parameters, a mass of correlation analyses are done between all the parameters and the ICT instead of the original HRTFs. Three steps for selecting parameters are used in the following simulation:

1. In order to reduce the amount of computation and make correlation analyses more effectively, it is

desirable to sample the upper left corner areas of the ICT for correlation analyses. In this procedure, the compacted ICT denoted by $\tilde{\mathcal{W}}_c \in \mathbb{R}^{P \times D' \times F'}$ ($D'' < D'$ and $F'' < F'$) is reshaped to a matrix $\tilde{\mathbf{W}}_c \in \mathbb{R}^{P \times (D'F')}$. Then, the absolute values of Pearson correlation coefficients are calculated and stored in a matrix $\mathbf{R}_{\tilde{\mathbf{W}}_c} \in \mathbb{R}^{27 \times (D'F')}$. There are 25 compacted ICTs corresponding to 25 azimuths, so 25 correlation analyses are constructed for the different azimuths. The significance of all the anthropometric parameters on the HRTFs can be shown by the correlation coefficient matrices with elements larger than 0.35 and plotted in Figure 7. The results in Figure 7 show that all the anthropometric measurements affect the HRTFs with different levels. It is necessary to delete unimportant parameters. After 25 correlation analyses, 22 parameters shown more important to the HRTFs are reserved for the next selection step. The parameters x_2, x_4, x_5, x_7 , and d_2 have the weak correlation with the HRTFs and they are deleted in this step.

2. After the correlation analyses, we model the intrinsic geometric structure of the reserved parameter space by the nearest neighbor graph. These reserved parameters ($x_1, x_3, x_6, x_8 - x_{17}, d_1, d_3 - d_8, \theta_1, \theta_2$) are arranged to three different classes shown in Table 2. Combining (8) and the graph, each parameter is evaluated by a Laplacian score. These parameters of each class are arranged by their corresponding scores in an ascending sequence. By this means, 17 parameters are reserved as the results

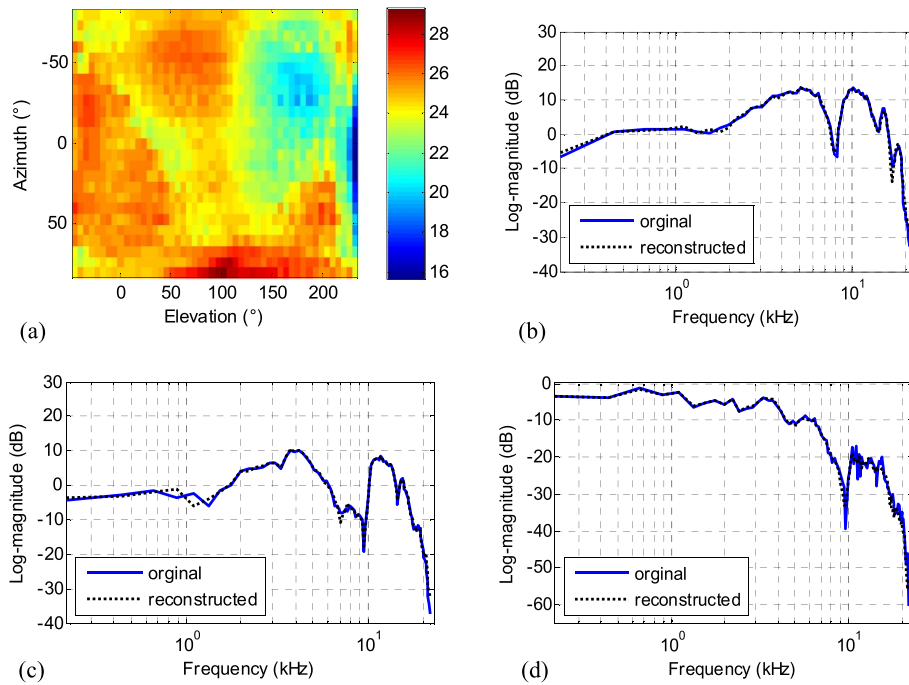


Figure 6 The HRTF magnitude reconstruction by HOSVD. (a) ASDR in HRTF reconstruction for $P = 30$ subjects. (b) $\theta = -80^\circ$, $D' = 5$, $F' = 20$ for subject 003. (c) $\theta = 0^\circ$, $D' = 26$, $F' = 37$ for subject 003. (d) $\theta = 80^\circ$, $D' = 15$, $F' = 37$ for subject 003.

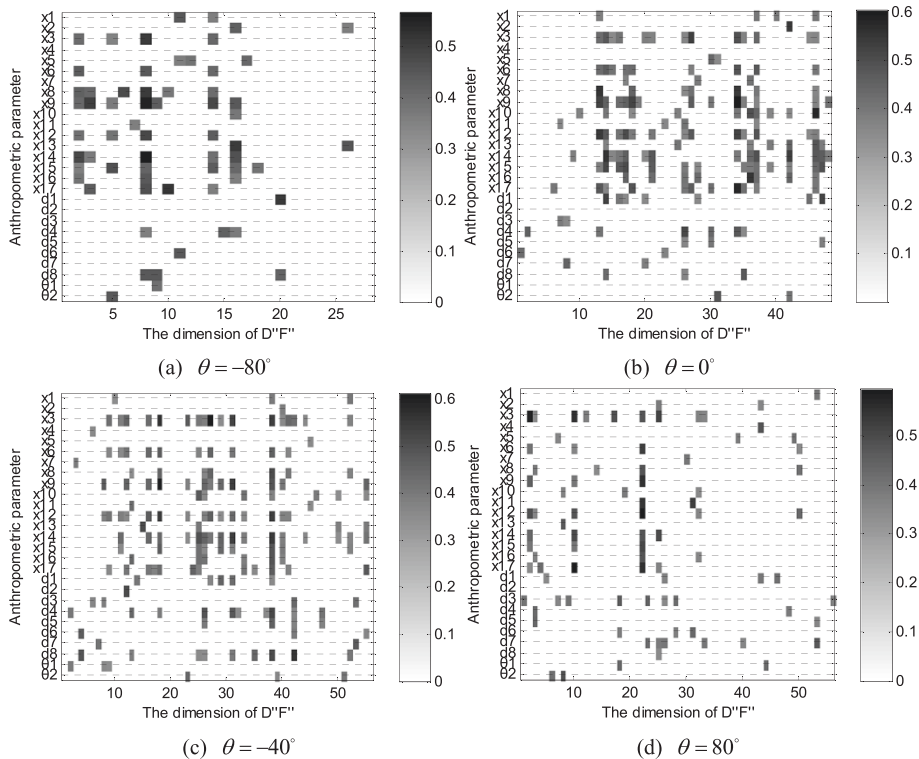


Figure 7 Correlation coefficients between the compacted ICT and the anthropometric parameters.

Table 2 Reserved parameters arranged by Laplacian scores in an ascending order

Class	Reserved anthropometric parameter
Class 1 (head and torso)	$x_{17}, x_{15}, x_{12}, x_6, x_{14}, x_9, x_{13}, x_{11}, x_3, x_{10}, x_8, x_{16}, x_1$
Class 2 (pinna)	$d_3, d_8, d_4, d_7, d_1, d_6, d_5$
Class 3 (pinna angle)	θ_1, θ_2

of the Laplacian score procedure. They are $x_3, x_6, x_9 - x_{15}, x_{17}, d_1, d_3, d_4, d_6 - d_8$, and θ_1 with the Laplacian score less than 0.4.

- The selected parameters are fed into the training of the individual HRTF modeling by HOPLS. The last step performs the correlation analysis among the reserved parameters. Similarly in order to show the dependent relation among those parameters, the gray image in Figure 8 presents the correlation coefficients of the reserved parameters larger than 0.5. From Figure 8, x_6, x_9, x_{12}, x_{14} , and x_{17} have strong correlation with others and are deleted. Thus, the parameters $x_3, x_{10}, x_{11}, x_{13}, x_{15}, d_1, d_3, d_4, d_6, d_7, d_8$, and θ_1 are selected as the final necessary measurements for the individual HRTF prediction. All the final reserved parameters are selected by the procedures of the correlation analysis and the Laplacian score. We select these 12 parameters as the key parameters. However, the significance of each selected parameter on the HRTFs is still not clear. Since measurements of the anthropometric parameters need special instruments, we cannot implement the anthropometric measurements at present.

4.2 Objective evaluation and subjective localization experiment

Through the simulations in Section 4.1, we can obtain the individual core tensor and key anthropometric parameters.

The goal of our proposed HRTF personalization is to model the multi-linear relation between the key parameters and the individual core tensor. In this section, the experiments are implemented to evaluate the feasibility of the proposed individual HRTF customization by objective evaluation and subjective perception. It is important to select the appropriate hyperparameters for preventing overfitting and controlling the complexity of the HRTF estimation model.

4.2.1 Selecting the numbers of loadings and latent vectors for the individual HRTF prediction

The different selections for the numbers of loadings and latent vectors can control the personalization model complexity and improve the predicting performance. In order to simplify the selection, we define $J_2 = J_3 = I = \lambda$. λ and R are chosen based on cross-validation [43]. The results of the optimal hyperparameters are shown in Table 3.

The optimal R and λ of five predicted subjects at three azimuths are different. These optimal R and λ bring good predicted performance. The ASDR is larger than 12.46 dB, but lower ASDR is obtained by other selections of R and λ . This implies that the performance of the individual HRTF prediction model can be adjusted by these two hyperparameters.

Compared with the PLSR method, the same 12 selected parameters are treated as the inputs and the ICT unfolded in 1-mode as the output. The optimal number of the latent variables in PLSR for the individual HRTF linear model is also chosen by cross-validation. The SDRs of the individual HRTF prediction for subject 124 at all the measured elevations of three different azimuths are shown in Figure 9. It can be seen that the proposed HRTF model has achieved larger SDRs than the PLSR method in all the elevations at azimuth -80° and 0° , excluding the high elevations of azimuth 80° . The complex

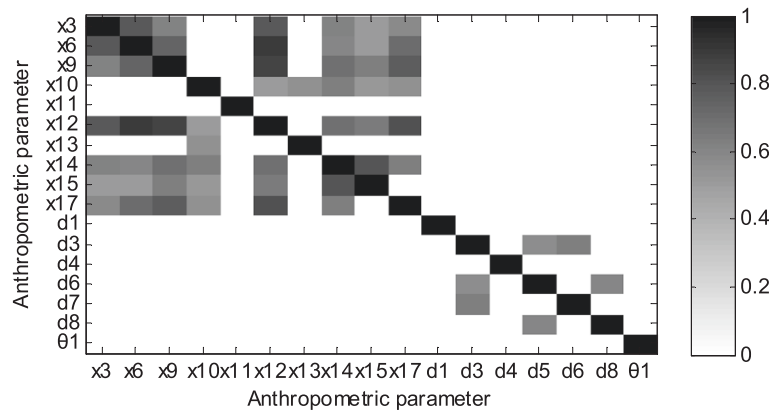


Figure 8 Correlation coefficients of the reserved parameters.

Table 3 The selected numbers of latent vectors and loadings of HOPLS

Azimuth	Subject	HOPLS	
		R	λ
-80°	003	1	1
	033	4	1
	124	5	4
	134	4	2
	153	3	5
0°	003	1	3
	033	3	1
	124	10	3
	134	2	1
	153	1	7
80°	003	1	3
	033	7	5
	124	10	7
	134	4	3
	153	1	9

property of the measured HRTFs for the contralateral ear especially at the rear directions near the horizontal plane leads to predict the individual HRTFs more difficultly. The performance for predicting the individual HRTFs by HOPLS model is much better than that of the PLSR method especially for the sound directions

ipsilateral to the concerned ear. In Figure 10, the discrepancy between the original HRTFs and the individual HRTFs may be caused by the information loss in the dimension reduction for the HRTF tensor and the inherent defect of the HOPLS model. In general, the predicted HRTFs can approximate the measured HRTFs based on the HOPLS method more accurately than the predicted HRTFs by PLSR.

4.2.2 Subjective localization experiment

The desirable individual HRTF modeling provides the accurate sound localization by the predicted HRTFs. The purpose of the subjective hearing experiment in this section is to compare the sound localization performance of the original HRTFs and the predicted HRTFs. Subjective tests, using five pink noises repeated five times, with 0.5-s silence between each repetition are constructed by headphone listening binaural signals. The used pink noises have 22.05-kHz bandwidth and 44.1-kHz sample ratio. Five test subjects participate in the subjective listening experiment with five test stimuli. The five test stimuli are pink noise samples of duration 1 s with 50 ms onset and offset time [45]. Each pink noise sample is rendered using the predicted HRTFs as well as the measured ones at randomly chosen azimuths in the horizontal plane. Then, each rendering testing stimulus is played back through a headphone. The participating subjects are asked to mark the level of sound localization using the grades in Table 4. Figure 11 shows the results of the listening tests with the virtual sounds

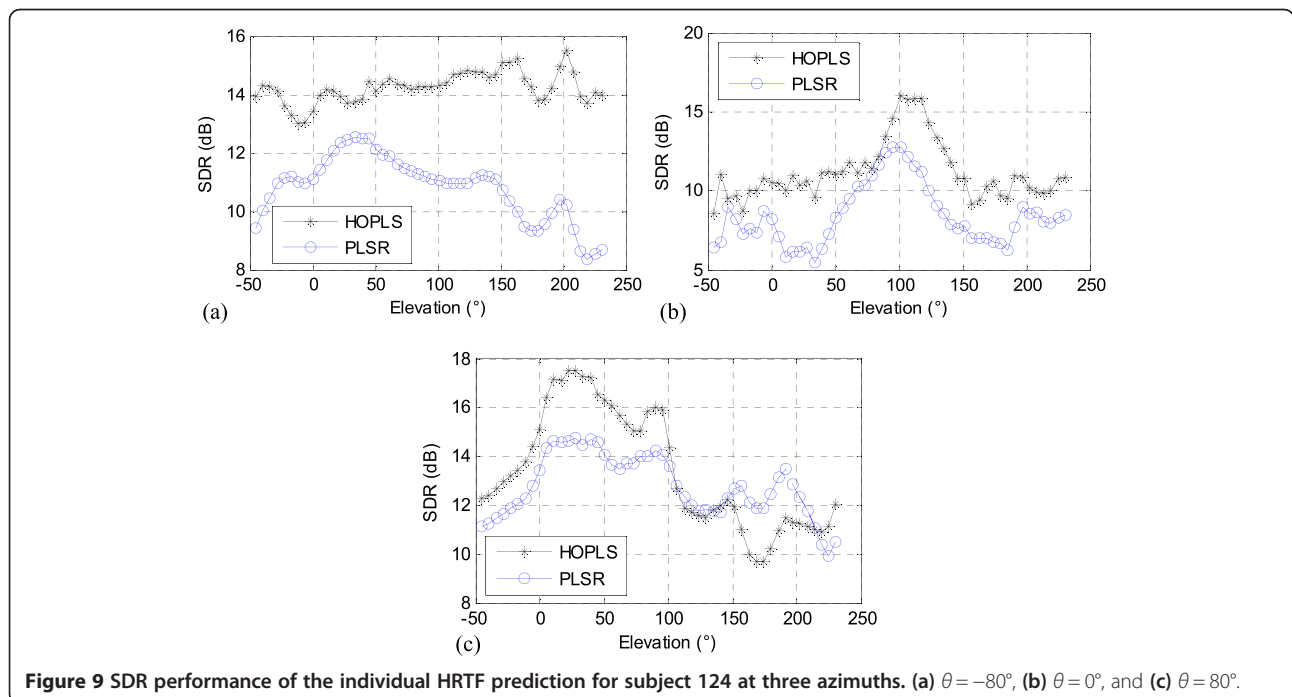


Figure 9 SDR performance of the individual HRTF prediction for subject 124 at three azimuths. (a) $\theta = -80^\circ$, (b) $\theta = 0^\circ$, and (c) $\theta = 80^\circ$.

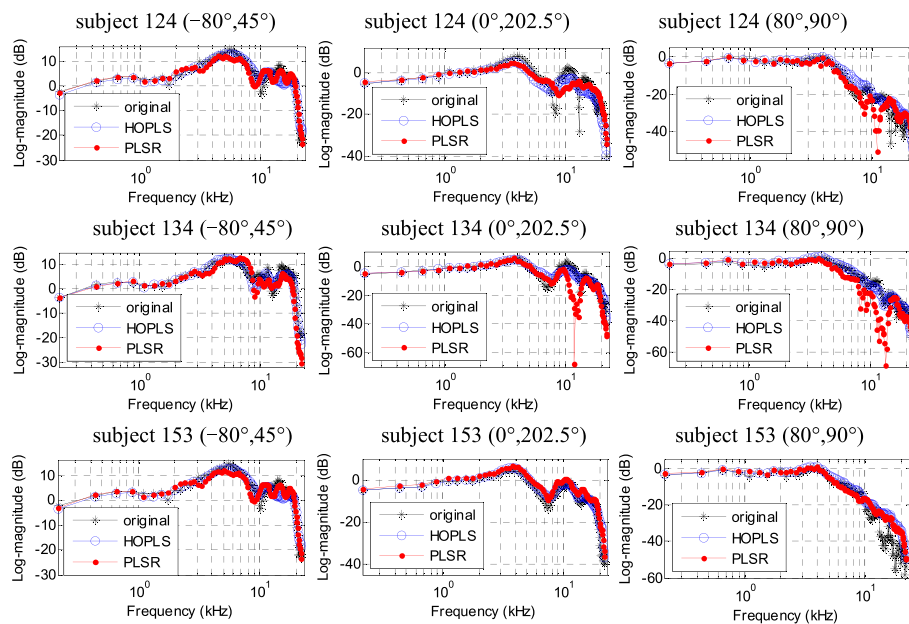


Figure 10 The individual HRTF prediction by HOPLS and PLSR compared with original HRTFs.

by convolving the stimuli with the predicted HRTFs and the measured HRTFs of the subjects 003, 033, 124, 134, and 153. For the sound localization, the predicted HRTFs are approximate to the original HRTFs.

5 Conclusions

High-dimensional HRTFs and redundant anthropometric parameters greatly affect the individual HRTF customization. We construct a multi-linear regression model between the HRTFs and the anthropometric parameters. The individual core tensor as the output variable of the regression model is firstly extracted from the measured HRTFs. Then, the key parameters are selected as the input variables of the multi-linear model based on the individual core tensor. The appropriate hyperparameter selection can achieve good prediction performance for the multi-linear model. Experimental results demonstrate the better performance for predicting the

individual HRTFs in comparison to the PLSR method especially for the sound directions ipsilateral to the concerned ear. The listening tests show that the predicted HRTFs are approximate to the original ones for the sound localization. The performance of the individual HRTF prediction is relatively not good in the region of the high elevations to the contralateral ear. In our future work, we will further implement the anthropometric measurements to predict the individual HRTFs and focus on the improvement of the prediction performance of the contralateral HRTF personalization. At the same time, the non-linear methods for the HRTF tensor estimation will be our future task based on the current work.

Table 4 Localization impairment scale in the subjective tests

Grade	Localization similarity
1	Very different
2	Slightly different
3	Slightly similar
4	Very similar
5	No difference

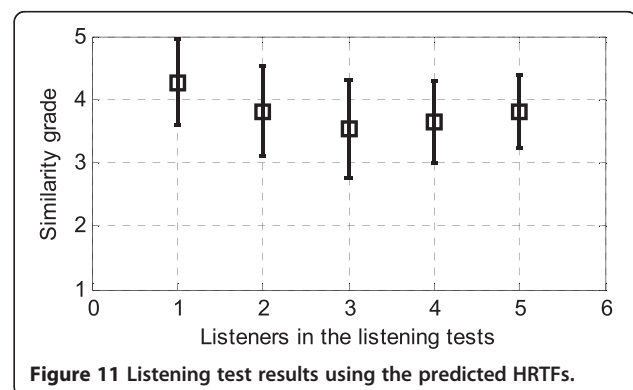


Figure 11 Listening test results using the predicted HRTFs.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

The authors would like to thank the editor and anonymous reviewers for their valuable comments. This work was supported by the National Natural Science Foundation (61001160) and Innovation Program of Shanghai Municipal Education Commission (12YZ203) of China.

Received: 31 August 2013 Accepted: 3 April 2014

Published: 3 May 2014

References

1. NA Gumerov, R Duraiswami, ZH Tang, Numerical study of the influence of the torso on the HRTF, in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2 (Orlando, 2002), pp. 111965–11968
2. NA Gumerov, AE O'Donovan, R Duraiswami, DN Zotkin, Computation of head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation. *J. Acoust. Soc. Am.* **127**(1), 370–386 (2010)
3. Y Kahana, PA Nelson, Boundary element simulations of the transfer function of human heads and baffled pinnae using accurate geometric models. *J. Sound Vib.* **119**(5), 552–579 (2007)
4. V Algazi, R Duda, R Duraiswami, N Gumerov, Z Tang, Approximating the head-related transfer function using simple geometric models of the head and torso. *J. Acoust. Soc. Am.* **112**(5), 2053–2064 (2002)
5. S Xu, Z Li, L Zeng, G Salvendy, A study of morphological influence on head-related transfer functions, in *IEEE International Conference on Industrial Engineering and Engineering Management, Singapore, 2007*, pp. 472–476
6. J Fels, M Vorlander, Anthropometric parameters influencing head-related transfer functions. *Acta Acustica united with Acustica* **95**(2), 331–342 (2009)
7. M Rothbucher, T Habigt, JL Habigt, T Riedmaier, K Diepold, Measuring anthropometric data for HRTF personalization, in *Processing of the 6th International Conference on Signal-Image Technology and Internet-Based Systems, Kuala Lumpur, 2010*, pp. 102–106
8. M Zhang, RA Kennedy, TD Abhayapala, W Zhang, Statistical method to identify key anthropometric parameters in HRTF individualization, in *2011 Joint Workshop on Hands-Free Speech Communication and Microphone Arrays, HSCMA'11 (Edinburgh, 2011)*, pp. 213–218
9. DN Zotkin, J Hwang, R Duraiswami, LS Davis, HRTF personalization using anthropometric measurements, in *IEEE ASSP WASPAA'2003, New Paltz, 2003*, pp. 157–160
10. XY Zeng, SG Wang, LP Gao, A hybrid algorithm for selecting head-related transfer function based on similarity of anthropometric structures. *J. Sound Vib.* **329**(19), 4093–4105 (2010)
11. N Inoue, T Kimura, T Nishino, K Itou, K Takeda, Evaluation of HRTFs estimated using physical features. *Acoust. Sci. Technol.* **26**(5), 453–455 (2005)
12. HM Hu, L Zhou, J Zhang, H Ma, ZY Wu, Head related transfer function personalization based on multiple regression analysis, in *IEEE International Conference on Computational Intelligence and Security, Guangzhou 2, 1829–1832 (2006)*
13. S Xu, ZZ Li, G Salvendy, Improved method to individualize head-related transfer function using anthropometric measurements. *Acoust. Sci. Technol.* **29**(6), 388–390 (2008)
14. K Matsui, A Akio, Estimation of individualized head-related transfer function based on principal component analysis. *Acoust. Sci. Technol.* **30**(5), 338–347 (2009)
15. S Hwang, YJ Park, YS Park, Modeling and customization of head-related transfer functions using principal component analysis, in *IEEE International Conference on Control, Automation and Systems (ICCAS) (Seoul, 2008)*, pp. 227–231
16. J Sodnik, A Umek, R Susnik, G Bobojevic, Representation of head related transfer functions with principal component analysis, in *Proceedings of the Annual Conference of the Australian Acoustical Society, NSW, 2004*, pp. 603–607
17. L Wang, FL Yin, Z Chen, HRTF compression via principal components analysis and vector quantization. *IEICE Electron Express* **5**(9), 321–325 (2008)
18. L Wang, FL Yin, Z Chen, Head-related transfer function interpolation through multivariate polynomial fitting of principal component weights. *Acoust. Sci. Technol.* **30**(6), 395–403 (2009)
19. BS Xie, Recovery of individual head-related transfer functions from a small set of measurements. *J. Acoust. Soc. Am.* **132**(1), 282–294 (2012)
20. DJ Kistler, FL Wightman, A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *J. Acoust. Soc. Am.* **91**(3), 1637–1647 (1992)
21. G Grindlay, MAO Vasilescu, A multilinear (tensor) framework for HRTF analysis and synthesis. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 1 (Honolulu, 2007), pp. 1161–1164
22. M Rothbucher, M Durkovic, H Shen, K Diepold, HRTF customization using multiway array analysis, in *EUSIPCO'2010 (Denmark, 2010)*, pp. 229–233
23. ZW Chen, GZ Yu, BS Xie, SQ Guan, Calculation and analysis of near-field head-related transfer functions from a simplified head-neck-torso model. *Chin. Phys. Lett.* **29**(3), 034302 (2012)
24. HM Hu, L Zhou, H Ma, ZY Wu, Head-related transfer function personalization based on partial least square regression. *J. Electron. Inform. Technol.* **30**(1), 154–158 (2008)
25. HM Hu, L Zhou, H Ma, ZY Wu, HRTF personalization based on artificial neural network in individual virtual auditory space. *J. Appl. Acoust.* **69**(2), 163–172 (2008)
26. S Xu, ZZ Li, S Gavriel, Individual head-related transfer functions based on population grouping. *J. Acoust. Soc. Am.* **124**(5), 2708–2710 (2008)
27. W Hügeng, D Gunawan, Improved method for individualization of head-related transfer functions on horizontal plane using reduced number of anthropometric measurements. *J. Telecommun.* **2**(2), 31–41 (2010)
28. T Nishino, N Inoue, K Takeda, F Itakura, Estimation of HRTFs on the horizontal plane using physical features. *Appl. Acoust.* **68**(8), 897–908 (2007)
29. T Nishino, Y Nakai, K Takeda, F Itakura, Estimating head related transfer function using multiple regression analysis. *IEICE Trans. A* **84**, 260–268 (2001)
30. QH Huang, Y Fang, Modeling personalized head-related impulse response using support vector regression. *J. Shanghai Univ (English edition)* **13**, 428–432 (2009)
31. QH Huang, QL Zhuang, HRIR personalisation using support vector regression in independent feature space. *Electron. Lett.* **45**(19), 1002–1003 (2009)
32. L Li, QH Huang, HRTF personalization modeling based on RBF neural network, in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), Vancouver, 2013*, pp. 3707–3710
33. M Rothbucher, H Shen, K Diepold, Dimensionality reduction in HRTF by using multiway array analysis, in *Human Centered Robot Systems (Springer, Berlin, 2009)*, pp. 103–110
34. LD Lathauwer, LD Moor, J Vandewalle, A multilinear singular value decomposition. *SIAM J Matrix Anal Appl* **21**(4), 1253–1278 (2000)
35. G Bergqvist, EG Larsson, The higher-order singular value decomposition: theory and an application [lecture notes]. *IEEE Signal Process. Mag.* **27**(3), 151–154 (2010)
36. TG Kolda, BW Bader, Tensor decompositions and applications. *SIAM Rev.* **51**(3), 455–500 (2009)
37. BS Xie, XL Zhong, D Rao, ZQ Liang, Head-related transfer function database and its analyses. *Sci. China, Ser. G* **50**, 267–280 (2007)
38. LD Lathauwer, BD Moor, J Vandewalle, On the best rank-1 and rank-(R1, R2, ..., RN) approximation of higher-order tensors. *SIAM J Matrix Anal Appl* **21**(4), 1324–1342 (2000)
39. N Gupta, A Barreto, M Joshi, JC Agudelo, HRTF database at FIU DSP Lab, in *IEEE International Conference on Acoustic Speech and Signal Processing (ICASSP), Dallas, 2010*, pp. 169–172
40. X He, D Cai, N Partha, Laplacian score for feature selection, in *Proceedings of Advances in Neural Information Processing Systems (Vancouver, 2005)*, pp. 507–514
41. QB Zhao, CF Caiafa, DP Mandic, L Zhang, T Ball, A Schulze-Bonhage, A Cichocki, Multilinear subspace regression: an orthogonal tensor decomposition approach, in *Advances in Neural Information Processing Systems 24 (NIPS) (Granada, 2011)*, pp. 1269–1277
42. QB Zhao, CF Caiafa, DP Mandic, ZC Chao, Y Nagasaka, N Fujii, L Zhang, A Cichocki, Higher-order partial least squares (HOPLS): a generalized multi-linear regression method. *IEEE Trans Pattern Anal Mach Intell* **35**(7), 1660–1673 (2013)
43. HW Wang, *Partial Least Square Regression-Method and Application* (National Defense Industry Press, Beijing, 2009), pp. 150–170
44. CIPIC HRTF database files, release 1.0. <http://interface.cipic.ucdavis.edu/>. Accessed 28 July 2012
45. PS Chanda, S Park, TI Kang, A binaural synthesis with multiple sound sources based on spatial features of head-related transfer functions, in *IEEE ICNN'06 (Vancouver, 2006)*, pp. 1726–1730

doi:10.1186/1687-6180-2014-58

Cite this article as: Huang and Li: Modeling individual HRTF tensor using high-order partial least squares. *EURASIP Journal on Advances in Signal Processing* 2014 **2014**:58.