PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Independent modes and dimensionality reduction of head-related transfer functions based on tensor decomposition

Tong ZHAO[1]; Bosun XIE[2]

[1] Acoustic lab, School of Physics and Optoelectronics, South China University of Technology, Guangzhou, China

[2] Acoustic lab, School of Physics and Optoelectronics, South China University of Technology, Guangzhou, China

ABSTRACT

Head-related transfer functions (HRTFs) are acoustical transfer functions from a source to two ears in the free-field and vital to spatial audio. Generally, far-field HRTFs vary with source direction, frequency and subject. Accordingly, the dimensionality of a full set of HRTFs is very large. Linear decomposition has been used to reduce the dimensionality of HRTFs. Due to the multiple variable-dependent characteristics of HRTFs, in present work, tensor decomposition is applied to analyze the independent modes of HRTF variation so as to achieve a more efficient representation of them. The measured HRTF magnitudes are first smoothed with a frequency bandwidth of 1 ERB. Then the smoothed HRTF magnitudes of 52 Chinese subject, 493 directions for each subjects and 42 discrete frequencies (up to 14 kHz) are combined as a 3rd-order tensor and then undergone a Tucker decomposition. The results indicate that 17 subject-related modes, 13 directional-related modes and 7 frequency-related modes account for more than 95% variations of HRTFs at corresponding dimension respectively, and thus the HRTF dataset can be reduced by truncating the modes to corresponding orders. Based on the results, it is also deduced that 17 anthropometric parameters at least are needed for the anthropometry-based HRTF customization.

Keywords: HRTF, Tensor decomposition, Dimensionality reduction

1. INTRODUCTION

Head-related transfer functions (HRTFs) are the acoustical transfer functions from a sound source to two ears in the free field, which characterize the comprehensive filtering effects of anatomical structures on sound waves (1, 2). HRTFs contain sound source localization information such as binaural time difference (ITD), binaural sound level difference (ILD), and high-frequency spectral cue. They play a very important role on the researches of binaural hearing and applications in virtual auditory display (VAD).

Generally, at the far-field source distance r > 1.0 m with respect to head centre, HRTFs vary as complicated functions of source direction and frequency. Due to the differences in the anthropometric geometries of head, pinnae and torso etc., HRTFs also depend on individual. Due to the multiple variable-dependent characteristics of HRTFs, the dimensionality of a full set of HRTFs is very large.

In researches and applications, it is often required to reduce the dimensionality of HRTFs. For example, an accurate VAD needs individualized or matched HRTFs (3). Measurement and numerical calculation are two common methods to acquire HRTFs. However, it is difficult to acquire a full set of HRTFs for each individual by measurement or numerical calculation. Alternatively, a full set of matched HRTFs of a new subject can be customized from an existing HRTF database and anthropometric-based measurement or from a few HRTF measurements on the new subject (4, 5). Prior to customization, dimensionality reduction of the existing HRTF database is needed. In addition, a low-dimensional representation of HRTFs reduces the required memory space in practical VAD.

Actually, there exist some similarities or correlations among the HRTFs of different directions, frequencies and subjects. The dimensionality of HRTFs can be reduced by removing these correlations.

[1] 371870906@qq.com
[2] Corresponding to: phbsxie@scut.edu.cn

Linear decomposition is often used to derive the low-dimensional representation of HRTFs. Principal component analysis (PCA), or equivalently, singular value decomposition (SVD), is a common example of linear decomposition (5-7). By using PCA, HRTFs are decomposed into a weighted sum of spectral shape basis functions or spatial basis functions. Each basis function represents an independent mode of HRTF variation in the frequency domain or spatial domain. The low-dimensional representation of HRTFs is derived by truncating the basis functions decomposition to some preceding orders that represent the most important variance of HRTFs.

Tensor is a multi-dimensional extension of conventional vector and matrix (8). Tensor decomposition is a multilinear modelling technique and an extension of traditional PCA or SVD calculation. Due to the multiple variable-dependent characteristics of HRTFs, a more efficient representation of HRTFs can be achieved by tensor decomposition. One advantage of tensor decomposition is that the interactions between multiple variables are considered. Grindlay and Vasilescu first presented a tensor framework for HRTF decomposition and used it to customize individualized HRTFs from anthropometric measurement (9). Huang and Li also explored the individualized HRTF customization using tensor decomposition (10). However, the independent dimensionalities or modes required for HRTF reconstruction were not analyzed in detail in aforementioned works, especially for subject-related modes.

In present work, tensor decomposition is further applied to analyse the independent modes of HRTF variation, so as to achieve a more efficient representation of the original HRTF datasets.

2. PRINCIPLE of Tensor decomposition of HRTFs

In present work, spatial direction is specified by elevation -90º ≤ Φ ≤90º and azimuth 0º ≤ θ < 360º, where Φ = -90°, 0° and 90° represents the bottom, horizontal and top direction, respectively; In the horizontal plane, θ = 0°, 90° and 180° represents the front, right and back direction, respectively.

Suppose that there are a full set of HRTFs from S subjects, M directions for each subject, and N discrete frequencies at each direction. The data can be HRTF magnitudes or complex-valued HRTFs with magnitude and phase. Then the full set of HRTFs are represented by a $S \times M \times N$ tensor $\boldsymbol{H}_{S\times M\times N}$ with entries $H(s, m, n) = H_{smn}$. Where $s = 1,2\ldots S$, $m = 1,2\ldots M$ and $n = 1,2\ldots N$ represent the indexes of subject, direction and frequency, respectively. To decompose the HRTF efficiently, the mean HRTF across S subjects and M directions are subtracted from the entries of $\boldsymbol{H}_{S\times M\times N}$, yielding a zero-mean HRTF tensor $\boldsymbol{H}'_{S\times M\times N}$ with entries:

$$H'_{smn} = H(s,m,n) - \frac{1}{SM}\sum_{s'=1}^{S}\sum_{m'=1}^{M} H(s',m',n) \tag{1}$$
$$s = 1,2\ldots S, \quad m = 1,2\ldots M, \quad n = 1,2\ldots N$$

The Tucker decomposition of tensor $\boldsymbol{H}'_{S\times M\times N}$ is expressed as the product of a core tensor and three unitary matrixes (8):

$$\boldsymbol{H}'_{S\times M\times N} = \boldsymbol{W}_{S\times M\times N} \times_s \boldsymbol{U}^s_{S\times S} \times_m \boldsymbol{U}^m_{M\times M} \times_n \boldsymbol{U}^n_{N\times N} \tag{2}$$

Where $\boldsymbol{W}_{S\times M\times N}$ is a $S \times M \times N$ core tensor with entries W_{smn}, $s = 1,2\ldots S$, $m = 1,2,\ldots M$, and $n = 1,2\ldots N$. $\boldsymbol{U}^s_{S\times S}$, $\boldsymbol{U}^m_{M\times M}$ and $\boldsymbol{U}^n_{N\times N}$ are three unitary matrixes with entries $U^s_{ss'}$, $U^m_{mm'}$ and $U^n_{nn'}$ respectively. Subscript in each matrix indicates the dimensionality of matrix. These unitary matrixes satisfy following orthonormal properties:

$$\mathbf{U}^{s+}_{S\times S}\mathbf{U}^{s}_{S\times S} = \mathbf{I}_{S\times S} \qquad \mathbf{U}^{m+}_{M\times M}\mathbf{U}^{m}_{M\times M} = \mathbf{I}_{M\times M} \qquad \mathbf{U}^{n+}_{N\times N}\mathbf{U}^{n}_{N\times N} = \mathbf{I}_{N\times N} \tag{3}$$

Where, the superscript "+" denotes the transpose and complex conjugation of the matrix. $\boldsymbol{I}$ is identity matrix.

The products of core tensor and three unitary matrixes are calculated so that the entries of tensor $\boldsymbol{H}'_{S\times M\times N}$ are related to the entries of core tensor $\boldsymbol{W}_{S\times M\times N}$ and entries of three unitary matrixes as follows:

$$H'_{smn} = \sum_{s'=1}^{S}\sum_{m'=1}^{M}\sum_{m'=1}^{N} W_{s'm'n'} U^s_{ss'} U^m_{mm'} U^n_{nn'} \tag{4}$$

Similar to the procedures on PCA (2), to find the unitarity matrix $\boldsymbol{U}^s_{S\times S}$, an $S \times (MN)$ matrix $\boldsymbol{H}^1_{S\times(MN)}$ is constructed. The rows of $\boldsymbol{H}^1_{S\times(MN)}$ represent the mean-subtracted HRTFs for different

subjects, and the columns of $\boldsymbol{H}^1_{S\times(MN)}$ represent the mean-subtracted HRTFs at various directions and frequencies. An $S\times S$ Hermitian matrix $\boldsymbol{R}^1_{S\times S}$ is constructed from $\boldsymbol{H}^1_{S\times(MN)}$ as,

$$\boldsymbol{R}^1_{S\times S} = \frac{1}{NM}\boldsymbol{H}^1_{S\times(MN)}\boldsymbol{H}^{1+}_{(MN)\times S} \tag{5}$$

The preceding S' columns of unitarity matrix $\boldsymbol{U}^s_{S\times S}$ are constructed from the orthonormal eigenvectors of matrix $\boldsymbol{R}^1_{S\times S}$ associated with the $S' \le S$ positive eigenvalues in a descending order:

$$\lambda^s_1 > \lambda^s_2 > \ldots > \lambda^s_{S'} > 0 \tag{6}$$

Similarly, to find the unitarity matrix $\boldsymbol{U}^m_{M\times M}$, an $M \times (SN)$ matrix $\boldsymbol{H}^2_{M\times(SN)}$ is constructed. The rows of $\boldsymbol{H}^2_{M\times(SN)}$ represent the mean-subtracted HRTFs at different directions, and the columns of $\boldsymbol{H}^2_{M\times(SN)}$ represent the mean-subtracted HRTFs for different subjects and at various frequencies. An $M\times M$ Hermitian matrix $\boldsymbol{R}^2_{M\times M}$ is constructed from $\boldsymbol{H}^2_{M\times(SN)}$ as,

$$\boldsymbol{R}^2_{M\times M} = \frac{1}{SN}\boldsymbol{H}^2_{M\times(SN)}\boldsymbol{H}^{2+}_{(SN)\times M} \tag{7}$$

The preceding M' columns of unitarity matrix $\boldsymbol{U}^m_{M\times M}$ are constructed from the orthonormal eigenvectors of matrix $\boldsymbol{R}^2_{M\times M}$ associated with the $M' \le M$ positive eigenvalues in a descending order.

Finally, to find the unitarity matrix $\boldsymbol{U}^n_{N\times N}$, an $N \times (SM)$ matrix $\boldsymbol{H}^3_{N\times(SM)}$ is constructed. The rows of $\boldsymbol{H}^3_{N\times(SM)}$ represent the mean-subtracted HRTFs at various frequencies, and the column of $\boldsymbol{H}^3_{N\times(SM)}$ represents the mean-subtracted HRTFs for different subjects and at various directions. An $N\times N$ Hermitian matrix $\boldsymbol{R}^3_{N\times N}$ is constructed from $\boldsymbol{H}^3_{N\times(SM)}$ as,

$$\boldsymbol{R}^3_{N\times N} = \frac{1}{SM}\boldsymbol{H}^3_{N\times(SM)}\boldsymbol{H}^{2+}_{(SM)\times N} \tag{8}$$

The preceding N' columns of unitarity matrix $\boldsymbol{U}^n_{N\times N}$ are constructed from the orthonormal eigenvectors of matrix $\boldsymbol{R}^3_{N\times N}$ associated with the $N' \le N$ positive eigenvalues in a descending order.

Once the three unitarity matrixes are found, using the orthonormal properties of Eq.(3), the core tensor for decomposition is calculated by:

$$\mathbf{W}_{S\times M\times N} = \mathbf{H}'_{S\times M\times N} \times_s \mathbf{U}^{s+}_{S\times S} \times_m \mathbf{U}^{m+}_{M\times M} \times_n \mathbf{U}^{n+}_{N\times N} \tag{9}$$

or

$$W_{s'm'n'} = \sum_{s=1}^{S}\sum_{m=1}^{M}\sum_{n=1}^{N} H'_{smn} U^{s*}_{ss'} U^{m*}_{mm'} U^{n*}_{nn'} \tag{10}$$

The superscript "*" denotes complex conjugation.

The nature of tensor decomposition in Eq.(2) is to decompose the mean-subtracted HRTFs into a combination of modes of independent variables. The preceding S' columns of matrix $\boldsymbol{U}^s_{S\times S}$ represent $S' \le S$ subject-related modes, each row of matrix $\boldsymbol{U}^s_{S\times S}$ represents a specified subject. The preceding M' columns of matrix $\boldsymbol{U}^m_{M\times M}$ represent $M' \le M$ direction-related modes, each row of matrix $\boldsymbol{U}^m_{M\times M}$ represents a specified direction. The preceding N' columns of matrix $\boldsymbol{U}^n_{N\times N}$ represent $N' \le N$ frequency-related modes, each row of matrix $\boldsymbol{U}^m_{N\times N}$ represents a specified frequency. The orthogonal condition given by Eq.(3) indicates that each mode is independent. In addition, the core tensor $\boldsymbol{W}_{S\times M\times N}$ is only relevant to the modes of subject, direction and frequency, but independent from the variations of direction, frequency and subject.

When the mean-subtracted HRTFs are approximately reconstructed from the modes or eigenvectors associated with the preceding eigenvalues, the columns or modes of matrixes $\boldsymbol{U}^s_{S\times S}$, $\boldsymbol{U}^m_{M\times M}$ and $\boldsymbol{U}^n_{N\times N}$ are truncated to $S'' < S' \le S$, $M'' < M' \le M$, and $N'' < N' \le N$, respectively, yielding $S \times S''$ matrix $\hat{\boldsymbol{U}}^s_{S\times S''}$, $M \times M''$ matrix $\hat{\boldsymbol{U}}^m_{M\times M''}$ and $N \times N''$ matrix $\boldsymbol{U}^n_{N\times N''}$. Then, the mean-subtracted HRTFs can be approximated as:

$$\hat{\boldsymbol{H}}'_{S\times M\times N} \approx \hat{\boldsymbol{W}}_{S''\times M''\times N''} \times_s \hat{U}^s{}_{S\times S''} \times_m \hat{U}^m{}_{M\times M''} \times_n \hat{U}^n{}_{N\times N''} \tag{11}$$

Where $\hat{\boldsymbol{W}}_{S''\times M''\times N''}$ is a 3-order core tensor with dimensionality of $S'' \times M'' \times N''$, which is also constructed from $S\times M \times N$ tensor $\boldsymbol{W}_{S\times M\times N}$ by truncating it to corresponding dimensionality.

Eq.(11) can also be written as:

$$\hat{H}'_{smn} \approx \sum_{s'=1}^{S''}\sum_{m'=1}^{M''}\sum_{n'=1}^{N''} W_{s'm'n'} U^{s}_{ss'} U^{m}_{mm'} U^{n}_{nn'} \tag{12}$$

To reconstruct the complete HRTFs, the mean in Eq.(1) should be supplemented into Eq.(11) or Eq.(12).

Similar to the case of PCA (2, 5, 6), the cumulative percentage variance of the energy represented by the preceding modes is used to evaluate the accuracy or error in HRTF reconstruction. For subject-related modes, it is calculated by corresponding eigenvalues:

$$\eta_s = \frac{\sum_{s=1}^{S''} \lambda_s}{\sum_{s=1}^{S'} \lambda_s} \times 100\% \tag{13}$$

Similar equations are available to the direction-related and frequency-related modes, respectively.

The error in reconstruction can also be evaluated by relative error $er(s, m, n)$.

$$er(s,m,n) = 10\log_{10}\frac{|H(s,m,n) - \hat{H}(s,m,n)|^2}{|H(s,m,n)|^2} \quad (dB) \tag{14}$$

Where $H(s,m,n)$ and $\hat{H}(s,m,n)$ are original and reconstructed HRTFs, respectively.

It can also be evaluated by the mean relative error $er_{av1}(m, n)$ across subjects:

$$er_{av1}(m,n) = \frac{1}{S}\sum_{s=1}^{S} er(s,m,n) \quad (dB) \tag{15}$$

and the mean relative error er_{av2} across subjects, directions and frequencies:

$$er_{av2} = \frac{1}{SMN}\sum_{s=1}^{S}\sum_{m=1}^{M}\sum_{N=1}^{N} er(s,m,n) \quad (dB) \tag{16}$$

3. HRTF data and pre-processing

The aforementioned tensor decomposition method is applicable to complex-valued HRTFs, linear or logarithmic HRTF magnitudes or even head-related impulse responses (HRIRs) in the time domain. Here, the tensor decomposition method is applied to linear HRTF magnitudes.

The measured HRTF database used for analysis contains 52 Chinese subjects (26 male and 26 female) and 493 source directions at source distance of 1.5 m for each subject (11). The measured directions are uniformly distributed in each elevation. Table 1 lists the number of measured directions and azimuthal intervals in each elevation. The data are binaural HRIRs measured using the blocked-ear-canal technique (12). The length of each HRIR is 512 points at the sampling frequency of 44.1 kHz.

Table 1 – Numbers and intervals of azimuthal HRIR measurement in each elevation

Elevation Φ	-30°	-15°	0°	15°	30°	45°	60°	75°	90°
Number of azimuthal measurement	72	72	72	72	72	72	36	24	1
Interval of azimuthal measurement $\Delta\theta$	5°	5°	5°	5°	5°	5°	10°	15°	360°

The measured HRIRs are first padded zeros, and undergone discrete time Fourier transform (DFT), yielding HRTF magnitudes of 8192 points. To adapt the auditory resolution of frequency, the 8192-point HRTF magnitudes are smoothed above 500 Hz by using a moving frequency window with a bandwidth of 1 ERB (below 500 Hz, smoothing is not needed), yielding the 128-point HRTF

magnitude in the linear frequency scale. Due to the conjugate symmetry of HRTFs along discrete frequency axes, HRTF magnitudes at the preceding 42 discrete frequencies in the linear scale are used for the tensor decomposition. Preceding 42 discrete frequencies correspond to a frequency range from 0 Hz to 14 kHz at a sampling frequency of 44.1 kHz and thus include most important localization cues and other perceivable attributes. Moreover, because the measured HRTFs exhibit larger measured error above 14 kHz, the data at this frequency range should not be included in the analysis.

Therefore, aforementioned pre-processing yields a tensor of HRTF magnitudes for S = 52 subjects, M = 493 directions and N = 42 discrete frequencies, which is used as the original HRTF magnitudes for subsequent tensor decomposition.

4. RESULTS

Because the results of left-ear HRTFs and right-ear HRTFs are similar, the results of left-ear HRTFs are mainly discussed here. Table 2 lists the relationship between the cumulative percentage variance of energy (Var-s) and the number *S*" of subject-related modes (basis vectors). As seen, the Var-s increases with the number *S*" of modes. The first mode accounts for nearly 80% of the subject-related variance. Therefore, there exists basic similarity among the HRTF magnitudes of different subjects. The contributions of other modes descend with the order. The preceding 7 and 17 modes account for approximately 90 % and 95 % of subject-related variation, respectively.

Table 2 – Cumulative percentage variance of energy for various numbers of subject-related modes (left-ear HRTFs)

S"	1	2	4	7	17	31	39
Var-s (%)	79.7	82.4	86.9	90.2	95.2	98.1	99.1

Similarly, Table 3 lists the relationship between the cumulative percentage variance of energy (Var-m) and the number *M*" of subject-related modes. Var-m also increases with the number *M*" of modes. The first mode accounts for nearly 70% of the direction-related variance. Therefore, there exists basic similarity among the HRTF magnitudes of different directions. The contributions of other modes descend with the order. The preceding 6 and 13 modes account for approximate 90 % and 95 % of directions-related variation, respectively.

Table 3 – Cumulative percentage variance of energy for various numbers of direction-related modes (left-ear HRTFs)

M"	1	3	4	6	13	28	49
Var-m(%)	69.5	81.3	85.9	90.1	95.3	98.1	99.0

Table 4 lists the relationship between the cumulative percentage variance of energy (Var-n) and the number *N*" of frequency-related modes. Var-n also increases with the number *N*" of modes. The first mode accounts for approximate 73% of the frequency-related variance. Therefore, there exists correlation among the HRTF magnitudes of different frequencies. The contributions of other modes descend with the order. The preceding 4 and 7 modes account for approximate 90 % and 95 % directions-related variation, respectively.

Table 4 – Cumulative percentage variance of energy for various numbers of frequency-related modes (left-ear HRTFs)

N"	1	2	3	4	7	12	16
Var-n(%)	73.3	80.9	87.1	90.2	95.4	98.1	99.1

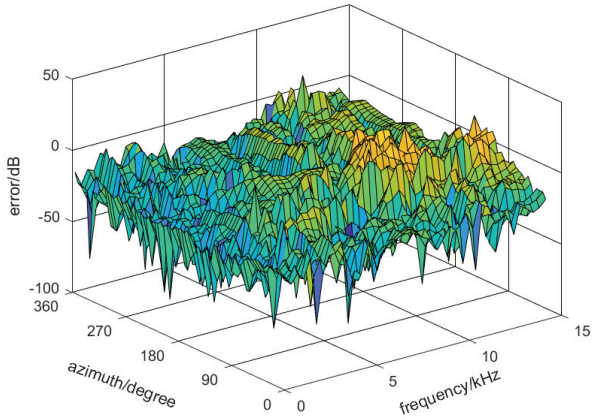

Figure1 – Mean $er_{av1}(m, n)$ across subjects for the left-ear HRTF magnitudes in the horizontal plane

Calculations indicate that, for construction with S" = 17 subjected-related modes, the mean relative error er_{av2} across subjects, directions and frequencies in Eq. (16) are -14.5 dB for the left-ear HRTF magnitudes and -13.6 dB for the right-ear HRTF magnitudes. In most cases, the mean relative error $er_{av1}(m, n)$ across subjects is less than -10 dB, which is basically sufficient for the HRTF reconstruction. Errors mainly occur at high frequencies and source directions contralateral to the ear. Actually, the attenuation and interference caused by head shadow reduce the contralateral HRTF magnitudes (therefore reduce the signal-to-noise ratio) and make them variate complicatedly at high frequency, which makes the reconstruction of contralateral HRTF magnitudes difficult. However, the contributions of contralateral HRTFs to localization are less than those of ipsilateral HRTFs (13), the errors in contralateral HRTF reconstruction are not so significant as those of ipsilateral HRTFs.

As an example, Figure 1 illustrate the mean relative error $er_{av1}(m, n)$ across subjects for the left-ear HRTF magnitudes in 72 horizontal direction (Φ = 0°) and up to frequency of 14 kHz. Calculation from the results in Figure 1 indicate that the mean $er_{av1}(m, n)$ across frequencies are -19.1 dB at the ipsilateral direction (θ = 265°, Φ = 0°) and -5.1 dB at the contralateral direction (θ = 85º, Φ = 0°).

5. Discussion

The results in Section 4 indicate that 4 and 7 frequency-related modes are needed to represent 90 % and 95 % frequency-related variation of HRTF magnitudes, respectively; 6 and 13 direction-related modes are needed to represent 90 % and 95 % directions-related variation of HRTF magnitudes. Kistler and Wightman applied PCA to the logarithmic HRTF magnitudes of two ears for S = 10 human subjects at M = 265 directions and N = 150 discrete frequencies (ranging from 0.2 to 15 kHz). The results indicated that 5 spectral shape basis functions or frequency-related modes were needed to represent 90% frequency-related variation in logarithmic HRTF magnitudes (6). Xie applied spatial principal component analysis (SPCA) to the HRTF magnitudes of two ears for S = 52 human subjects at M = 493 source directions and N = 59 discrete frequencies (from DC to 20 kHz) (5). The results indicated that 16 spatial basis functions (direction-related modes) were needed to represent more than 95 % direction-related variation in HRTF magnitudes. Therefore, the results in present work are comparable to previous works.

Xie et al. analyzed the correlations among HRTF magnitudes of different subjects and applied a cluster analysis to them (14). The HRTF magnitudes used in that analysis were identical to those used in present work. The results indicated that HRTF magnitudes for most of 52 subjects could be classified into seven clusters. Some singleton clusters were also observed on a few subjects, which reflected the diversity in individualized HRTF magnitudes. The results in Section 4 also indicate that 7 subject-related modes are needed to represent 90 % subject-related variations in HRTF magnitudes. To represent the diversity in HRTF magnitudes more accurately, more subject-related modes are needed. For example, 17 subject-related modes are needed to represent 95 % subject-related variations in HRTF magnitudes. Therefore, the results of subject-related modes analysis in present work are also comparable to previous works.

It should be noted that the subject-related variations of HRTF magnitudes from a Chinese-based database are analyzed in present work. Due to the statistical difference among the HRTF magnitudes of different nationalities (such as Chinese and western subjects), it is expected that more subject-

related modes are needed to represent the diversity of HRTF database including subjects of multiple nationalities. From this point of view, it is beneficial to analyze the subject-related modes for the HRTFs of each special nationality separately.

In anthropometry-based HRTF customization, individualized HRTFs are approximately estimated from anthropometric measurements by linear or nonlinear mapping technique (2, 4). The mapping methods are derived from the statistical analysis on a baseline HRTF and anthropometric measurement database with sufficient number of samples. To make the mapping efficiently, decomposition and dimensionality reduction of HRTFs are needed. Because 17 subject-related modes are needed to represent 95 % subject-related variations in HRTF magnitudes, it is deduced that 17 anthropometric parameters at least are needed for the mapping in anthropometry-based HRTF customization. Grindlay and Vasilescu suggested a tensor decomposition of HRTFs and anthropometry-based HRTF customization method (9). They chose 5 subject-related modes to represent the subject-related variations in HRTF, which seems not enough.

6. CONCLUSIONS

Far-field HRTFs vary with subject, direction and frequency. Tensor decomposition is an effect tool for analyzing the independent modes of HRTF variation and reducing the dimensionality of HRTFs. Analysis on the HRTF magnitudes of 52 Chinese subjects, 493 directions and up to a frequency of 14 kHz indicates that 17 subject-related modes, 13 directional-related modes and 7 frequency-related modes account for more than 95% variations of HRTFs at corresponding dimension respectively. The reconstructed individualized HRTF magnitudes from 17 subject-related modes exhibit a reasonable accuracy but some errors occur at high-frequency above 10 kHz and source direction contralateral to ear. Based on the results, it is also deduced that 17 anthropometric parameters at least are needed for the anthropometry-based HRTF customization, and the number of anthropometric parameters used for HRTF customization in some previous was insufficient. The future works include HRTF customization, simplifying HRTF data and signal processing in virtual auditory display using tensor decomposition.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (11674105) and State Key Lab of Subtropical Building Science, South China University of Technology.

REFERENCES

1. Blauert J. Spatial hearing: the psychophysics of human sound localization. MIT press; 1997.
2. Xie B. Head-related transfer function and virtual auditory display. J. Ross Publishing; 2013 Jul 23.
3. Wenzel EM, Arruda M, Kistler DJ, Wightman FL. Localization using nonindividualized head-related transfer functions. The Journal of the Acoustical Society of America. 1993 Jul;94(1):111-23.
4. Jin C, Leong P, Leung J, Corderoy A, Carlile S. Enabling individualized virtual auditory space using morphological measurements. In Proceedings of the First IEEE Pacific-Rim Conference on Multimedia (2000 International Symposium on Multimedia Information Processing) 2000 Dec (pp. 235-238).
5. Xie BS. Recovery of individual head-related transfer functions from a small set of measurements. The Journal of the Acoustical Society of America. 2012 Jul;132(1):282-94.
6. Kistler DJ, Wightman FL. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. The Journal of the Acoustical Society of America. 1992 Mar;91(3):1637-47.
7. Larcher V, Warusfel O, Jot JM, Guyard J. Study and comparison of efficient methods for 3-d audio spatialization based on linear decomposition of HRTF data. In Audio Engineering Society Convention 108 2000 Feb 1. Audio Engineering Society.
8. Cichocki A, Mandic D, De Lathauwer L, Zhou G, Zhao Q, Caiafa C, Phan HA. Tensor decompositions for signal processing applications: From two-way to multiway component analysis. IEEE Signal Processing Magazine. 2015 Mar;32(2):145-63.
9. Grindlay G, Vasilescu MA. A multilinear (tensor) framework for HRTF analysis and synthesis. In2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07 2007 Apr 15 (Vol. 1, pp. I-161). IEEE.
10. Huang Q, Li L. Modeling individual HRTF tensor using high-order partial least squares. EURASIP Journal on Advances in Signal Processing. 2014 Dec 1;2014(1):58.

11. Xie B, Zhong X, Rao D, Liang Z. Head-related transfer function database and its analyses. Science in China Series G: Physics, Mechanics and Astronomy. 2007 Jun 1;50(3):267-80.
12. Møller H. Fundamentals of binaural technology. Applied acoustics. 1992 Jan 1;36(3-4):171-218.
13. Hofman P, Van Opstal A. Binaural weighting of pinna cues in human sound localization. Experimental brain research. 2003 Feb 1;148(4):458-70.
14. Xie B, Zhong X, He N. Typical data and cluster analysis on head-related transfer functions from Chinese subjects. Applied Acoustics. 2015 Jul 1;94:1-3.