Interspeech 2018
2-6 September 2018, Hyderabad

Sparsity-Constrained Weight Mapping for Head-Related Transfer Functions Individualization from Anthropometric Features

Xiaoke Qi[1], *Jianhua Tao*[1,2,3]

[1]National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, China
[2]CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, China
[3]School of Artificial Intelligence, University of Chinese Academy of Sciences, China

xiaoke.qi@nlpr.ia.ac.cn, jhtao@nlpr.ia.ac.cn

Abstract

Head-related transfer functions (HRTFs) describe the propagation of sound waves from the sound source to ear drums, which contain most of information for localization. However, HRTFs are highly individual-dependent, and thus because of the difference of anthropometric features between subjects, individualization of HRTFs is a great challenge for accurate localization perception in virtual auditory displays (VAD). In this paper, we propose a sparsity-constrained weight mapping method termed SWM to obtain individual HRTFs. The key idea behind SWM is to obtain optimal weights to combine HRTFs from the training subjects based on the relationship of anthropometric features between the target subject and the training subjects. To this end, SWM learns two sparse representations between the target subject and the training subjects in terms of anthropometric features and HRTFs, respectively. A non-negative sparse model is used for this purpose when considering the non-negative property of the anthropometric features. Then, we build a mapping between the two weight vectors using a nonlinear regression. Furthermore, an iterative data extension method is proposed in order to increase training samples for mapping model. The objective and subjective experimental results show that the proposed method outperforms other methods in terms of log-spectral distortion (LSD) and localization accuracy.

Index Terms: head-related transfer functions, anthropometric features, individualization, spatial hearing

1. Introduction

Head-related transfer functions (HRTFs) describe the propagation of sound wave in the form of reflection, and diffraction from the sound source to ear drums through body, head and pinna in free space [1], which contain all of localization information for spatial auditory perception. Because of its highly individual dependence, HRTF individualization is one of the challenges for immersive auditory perception.

The most accurate method for HRTF individualization is by directly measuring the impulse responses from the sound source to the human ears [2]. However, it is greatly time consuming, expensive, and not scalable. In light of this, several alternative methods have been proposed, including HRTF estimation from a small set of measurements [3], theoretical or numerical models to attempt to approximate the complicated human anatomy, such as spherical head model [4], snowman model [5], structural models [6], boundary element method [7], finite-difference time-domain method [8]. However, they require expensive acquisition hardware and are computationally intensive. Then, perceptual-based methods are proposed through listening tests by tuning some parameters until they achieve an acceptable spatial accuracy [9] [10]. Sunder et al. proposed an individualization method in the horizontal plane that uses a frontal projection headphone to introduce idiosyncratic pinna cues [11]. When considering the dependence relationship between human anthropometric characteristic and HRTFs, anthropometry-based regression methods are popular to predict individual HRTFs [12][13][14].

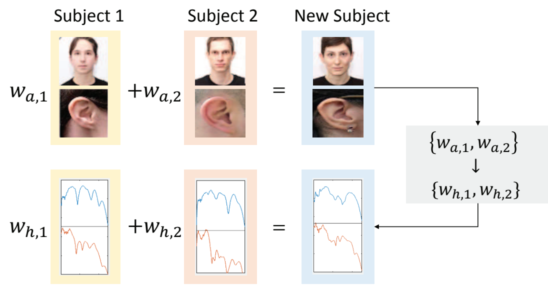


Figure 1: *The main idea in this paper.*

Anthropometry-based methods can be categorized into several approaches. One method constructs individual HRTFs in full space from a small set of measurements [3], which requires a prior measured data for a target subject. Another method builds a direct correlation relationship between anthropometric features and HRTFs with dimensional reduction by artificial neural network (ANN) [12] [15], support vector regression in conjunction with PCA, ICA or other nonlinear dimension reduction methods [16]. Then, given the anthropometric measurements of a target subject, HRTFs are obtained by passing the model. There is another method to generate a representation of anthropometric features between the known subjects and a target subject, with the assumption that a given HRTF set can be described by the same combination as the anthropometric data [17]. Based on this study, [18] investigates the influence of different preprocessing and postprocessing methods in terms of log-spectral distortion (LSD), and shows that some processing methods have a positive effect on individualization.

However, the assumption of the same combination weights for anthropometric features as HRTFs is not completely accurate, since there is still unknown for the relationship between the human anthropometric characteristic and HRTFs. Motivated by this observation, we propose a sparsity-constrained weight mapping method termed SWM to individualize HRTFs based on anthropometric features. As shown in Fig. 1, the main idea behind SWM is to generate a representation between the target subject and the training subjects in terms of anthropometric features and HRTFs, respectively. By considering the nonnegative characteristic, a nonnegative sparsity-constrained model is built for this purpose. Then, a weight mapping model is trained to make a mapping between the two weight vectors using a deep neural network (DNN) method with an iterative data extension.

10.21437/Interspeech.2018-1615

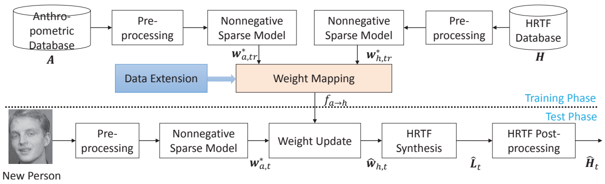


Figure 2: *The system architecture of the proposed sparsity-constrained weight mapping for HRTF individualization.*

2. System overview

In this paper, we propose a sparsity-constrained weight mapping method for HRTF individualization based on anthropometric features, termed SWM. The system architecture is shown as Fig. 2. The process of SWM can be divided into three steps: data preprocessing, model training and HRTF synthesis.

For the data processing process, we operate on the anthropometric features and HRTFs to normalize the values, respectively. After that, the weight vectors between the known subjects and the target subject in terms of the anthropometric features and HRTFs are learned by using a non-negative sparsity-constrained method, respectively. Then, a weight mapping model is trained to find a relationship between the two vectors. To this end, DNN is used because of its great feature representation and good performance in many tasks.

At the test phase, given anthropometric features of any target subject, the weight vector for anthropometric features is first generated from the non-negative sparsity-constrained model, and then by passing the weights to the trained weight mapping model, the weight vector for synthesizing individual HRTFs can be obtained. Finally, using the weights, individual HRTFs will be generated by combining HRTFs from the training subjects.

3. Proposed SWM Method

3.1. Data preprocessing

Assuming that the database contains N_s subjects, and for each subject N_a anthropometric features and HRTFs from N_d directions are measured. The database is divided into two parts by randomly allocating the data from $N_{s,tr}$ subjects to the training set, and others from $N_{s,t}$ subjects to the test set. Prior to the model training, the anthropometric features and HRTFs need to be normalized for the same scale and variance.

Anthropometric Features. By utilizing the characteristic that anthropometric measurements for human beings are always positive, a non-negative sparsity-constrained method is used to build the relationship between the training subjects and the target subject. In order to keep the property of the anthropometric features, min-max normalization is used to preprocess the anthropometric features to limit the values between 0 and 1. For the n-th anthropometric feature of the m-th subject, denoted as $a_{m,n}$, the preprocess is expressed as

$$\bar{a}_{m,n} = \frac{a_{m,n} - m_a(n)}{M_a(n) - m_a(n)}, \tag{1}$$

where $m_a(n)$ and $M_a(n)$ denote the minimum and maximum of the n-th anthropometric feature over the training set, respectively.

HRTFs. Because human is not sensitive to the fine details of the phase spectrum of HRTFs in localization and discrimination perception [19], and the logarithmic magnitude of HRTFs is more approaching human's auditory perception, which is experimentally verified in [20], the log-magnitude of HRTFs are chosen to model HRTFs in this paper. In order to obtain HRTFs in the same scale, we preprocess the log-magnitude of HRTFs by calculating the directional transfer function (DTF) followed by min-max normalization, which is expressed as

$$L_{m,(jN_b+i)} = H_{m,j,i} - \frac{1}{N_{s,tr}N_d}\sum_{q=1}^{N_{s,tr}}\sum_{p=1}^{N_d} H_{q,p,i}, \tag{2}$$

$$\bar{L}_{m,n} = \frac{L_{m,n} - m_h(n)}{M_h(n) - m_h(n)}, \tag{3}$$

where $H_{m,j,i}$ is the log-magnitude of the m-th subject's HRTFs for the i-th frequency bin from the j-th direction. $i = 1, ..., N_b$ with N_b of the number of the frequency bins. $m_h(n)$ and $M_h(n)$ denote the minimum and maximum of the n-th frequency bin of log-magnitude spectrum across $N_{s,tr}$ subjects, respectively. $\bar{L}_{m,n}$ denotes the normalized log-magnitude for the m-th subject, and $n = (jN_b + i) = 1, ..., N_dN_b$.

3.2. Weight generation

The key idea of this paper is to synthesize HRTFs given anthropometric features of a target subject by combining HRTFs from the training set, without the assumption as previous work in [17][18] that the combination weight vector is the same as those of anthropometric features. To this end, two sparse vectors, which respectively build a relationship in terms of anthropometric features and HRTFs between the training subjects and a target subject, are first generated, and then a weight mapping model is built to find a relationship between the two vectors.

Given the anthropometric features for a target subject, $\mathbf{w}_a^*$ can be obtained by solving a minimum problem of linear regression between the training subjects and a target subject with non-negative sparsity constraint, which is expressed as

$$\begin{aligned} &\mathbf{w}_a^* = \arg\min_{\mathbf{w}_a} ||\mathbf{A}_t - \mathbf{w}_a\mathbf{A}_{tr}||_2^2 + \lambda_a||\mathbf{w}_a||_1, \\ &\text{s.t., } \mathbf{w}_a \geq 0, \end{aligned} \tag{4}$$

where $\mathbf{A}_t$ denotes the preprocessed anthropometric features for a target subject. $\mathbf{A}_{tr} \in \mathbb{R}^{N_{s,tr}\times N_a}$ contains all the preprocessed anthropometric features across the training set, whose item of the m-th row and n-th column is $\bar{a}_{m,n}$. The shrinking parameter λ_a controls the sparsity level of the model.

Meanwhile, a weight vector $\mathbf{w}_h^*$ for combining HRTFs of a target subject is obtained by solving a nonnegative sparsity-constrained regression problem, which is formulated as

$$\begin{aligned} &\mathbf{w}_h^* = \arg\min_{\mathbf{w}_h} ||\mathbf{L}_t - \mathbf{w}_h\mathbf{L}_{tr}||_2^2 + \lambda_h||\mathbf{w}_h||_1, \\ &\text{s.t., } \mathbf{w}_h \geq 0, \end{aligned} \tag{5}$$

where $\mathbf{L}_t$ denote the preprocessed HRTFs for a target subject. $\mathbf{L}_{tr} \in \mathbb{R}^{N_{s,tr}\times N_dN_b}$ contains the preprocessed HRTFs, whose item of the m-th row and n-th column is $\bar{L}_{m,n}$. λ_h is the shrinking parameter.

Eq.(4) and (5) are two convex optimization problems. We solve them by l_1-regularized least squares problem solver discussed in [21], which estimates a vector of regression coefficients by minimizing the residual sum of squares subject to a constraint on the l_1-norm of the coefficient vector. To prevent overfitting, the K-fold cross-validation approach is used to train λ_a and λ_h[22]. Here, we use tenfold, and the parameters are chosen with the minimum cross-validation errors.

3.3. Weight mapping

The main idea of the presented SWM is to generate more accurate weights to superpose HRTFs of a target subject from the training subjects. To this end, we use weight mapping to build a match from the weights of anthropometric features to the weights of HRTFs. Based on the weights from Sec. 3.2, the mapping process can be formulated as

$$\text{Minimize } ||\hat{\mathbf{w}}_h^* - \mathbf{w}_h^*||_F^2, \text{ with } \hat{\mathbf{w}}_h^* = f_{a\to h}(\mathbf{w}_a^*). \quad (6)$$

We exploit DNN to solve this problem because of its great performance verified in other tasks. DNN is capable of approximating nonlinear functions of the inputs. In this paper, we use a back propagation DNN whose inputs and outputs are $\mathbf{w}_a^*$ and $\mathbf{w}_h^*$ from Sec. 3.2, respectively. Mean square error is used for cost function. Besides, an early stop approach is used to improve generalization.

DNN-oriented method significantly works well in solving problems in big data. In this paper, an iterative data extension method is proposed to increase the number of training samples. At the k-th iteration, we exhaustively exclude the data of k subjects from the training set, and calculate the weight vectors using (4) and (5). Following that, $\mathbf{w}_a^*$ and $\mathbf{w}_h^*$ are constructed by inserting 0 at the corresponding position of the excluded subjects. The illustration of the first two iterations is shown in Fig. 3. For the first iteration, one of $N_{s,tr}$ training subjects is excluded from weight calculation by brute-force method. Assuming that the j-th subject ($j = 1, ..., N_{s,tr}$) is chosen, the weights vector is obtain by calculating an initial weights using the data from the other subjects followed by inserting 0 into j-th item. Then, for the second iteration, any two of $N_{s,tr}$ training subjects are excluded, such as the i-th and j-th subjects in Fig. 3. After weight calculation, 0s are setting for the i-th and j-th positions of $\mathbf{w}_a^*$ and $\mathbf{w}_h^*$. Therefore, after K iterations, the total number of extended weight pairs is $[N_{s,tr}(N_{s,tr}+1)\cdots(N_{s,tr}+K-1)]/K$.

Finally, we synthesize HRTFs for a target subject by first applying $\hat{\mathbf{w}}_h^*$ to HRTFs of the training subjects as

$$\hat{\mathbf{L}}_t = \hat{\mathbf{w}}_h^* \mathbf{L}_{tr}/||\hat{\mathbf{w}}_h^*||_1, \quad (7)$$

and post-processing, which is calculated by inverting operations of Eqs. (2) and (3).

4. Performance evaluation

In this section, the performance of the proposed SWM method is evaluated in terms of objective and subjective experiments. CIPIC database is used for this purpose [23]. For 43 subjects in CIPIC, head-related impulse responses (HRIRs) are obtained from the directions of 25 azimuths and 50 elevations at a distance of 1m. Each HRIR has been windowed in about 4.5ms (200 points) with the sampling rate of 44.1kHz. Moreover, 37 anthropometric features, which contains 17 for the head and the torso, and 10 for each pinna, are measured from 35 subjects. We randomly choose data from 30 subjects as the training set, and the rest as the test set. First, each HRIR is transformed into an HRTF by using 256-point FFT followed by constant-Q filtering. Also, the frequency band is limited between 200Hz and 20kHz, resulting in 115 coefficients for each HRTF. The numbers of input nodes and output nodes in DNN for weight mapping both are set to 30, which is the same as the number of training subjects. Furthermore, we set the number of hidden layers to 2 with 10 nodes per layer. The dropout fraction is set to

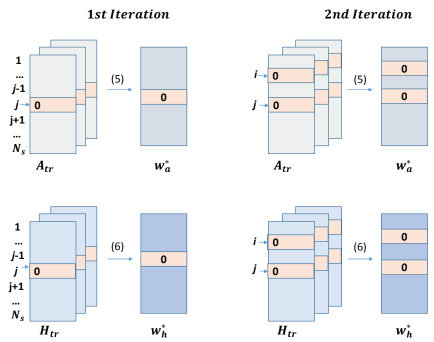


Figure 3: *The illustration for iterative data extension process.*

Table 1: *Objective performance comparison in terms of log-spectral distortion (LSD), root mean square error (RMSE)and relative RMSE (RRMSE).*

Method	LSD[dB]	RMSE	RRMSE
L2	13.2874	0.0809	0.1695
[17]	7.0294	0.1063	0.2234
[18]	5.6096	0.0765	0.1588
Proposed SWM	4.3040	**0.0715**	**0.1488**
OptWT	**3.8981**	0.0724	0.1510

0.5, and the sparsity target is 0.2. Moreover, we use the *Sigmoid* activation function because of its nonnegative property and great performance in other tasks. The number of the iteration is set to be 3 and thus we obtain the total of 4960 extended weight pairs.

4.1. Objective evaluation

Log-spectral distortion (LSD) in frequency domain, and root mean square error (RMSE) in time domain are used as the metrics for objective evaluation. LSD expresses the distortion between the estimated and the measured HRTFs as LSD $= \sqrt{\frac{1}{N_f}\sum_{m,d,k}\left(20\log_{10}\frac{|H_{m,k,d}|}{|\hat{H}_{m,k,d}|}\right)^2}$, RMSE defines the difference between the estimated and the measured HRIRs, which is expressed as RMSE $= \sqrt{\frac{1}{N_m}\sum_{m,t,d}(h_{m,t,d}-\hat{h}_{m,t,d})^2}$. Furthermore, the relative RMSE (RRMSE) is also calculated as the normalized RMSE, i.e., RRMSE $= \sqrt{\frac{1}{N_m}\sum_{m,t,d}(h_{m,t,d}-\hat{h}_{m,t,d})^2/\sum_{m,t}||\hat{h}_{m,t,d}||^2}$, where $m = 1, ..., N_{s,t}, t = 1, ..., N_t, d = 1, ..., N_d$, $k = k_1, ..., k_2$ denotes the range of the considered frequency bins, and thus $N_b = k_2 - k_1 + 1$. N_t is the number of the sampling points for each HRIR. $N_f = N_s N_d N_b$, and $N_m = N_s N_d N_b$.

We compare our SWM with 4 other methods: L2 method without sparse constraint, i.e., $\mathbf{w}_h^* = \mathbf{A}_{tr}^\dagger \mathbf{A}_t$; the methods in [17] and [18]; and the optimal method calculated as $\mathbf{w}_h^{(opt)} = \mathbf{L}_{tr}^\dagger \mathbf{L}_t$, denoted as OptWT. OptWT achieves the theoretical lower bound for LSD. The frequency band between 200Hz and 20kHz is focused on. The results are shown as Table 1 in terms of LSD, RMSE and RRMSE. First, it is seen that the LSD performance of the proposed SWM dramatically outperforms other method except OptWT. While compared with OptWT, there only exits a 0.4059dB gap for SWM. It is also observed that the performance of $L2$ is much worse than other sparse methods, which agrees to the conclusion in [17]. Moreover, we observe that SWM obtains the best performance in terms

Table 2: *LSD (dB) comparison for different bands (kHz).*

Band	OptWT	SWM	[17]	[18]	[15]
0.2-1.0	0.8643	**1.0705**	2.0828	3.3380	1.2572
1.0-2.0	1.1555	**1.6788**	3.8203	2.4180	1.7882
2.0-4.0	2.2523	2.3885	4.9831	2.8034	**2.2142**
4.0-8.0	2.8938	**3.2106**	6.7398	3.6385	3.4673
8.0-15.0	4.1628	**4.9690**	9.4756	5.6147	5.8007
15.0-20.0	4.5525	**5.1561**	10.0142	5.8508	-

of RMSE and RRMSE in time domain. The results are not an exact match to those of LSD. The reason is complex because a point of an HRIR are recovered by processing all the frequency bands of an HRTF. It also infers that when using different metrics, the results might not be strictly identical each other.

We also shows the LSD for five frequency bands in Table 2 by comparing to [17], [18], [15] and OptWT. From the table, we can observe that LSD increases with frequency, and SWM performed better than other methods in most frequency bands except OptWT. Over high frequency bands, especially above 8kHz, SWM gains significantly lower LSD than [15].

4.2. Subjective evaluation

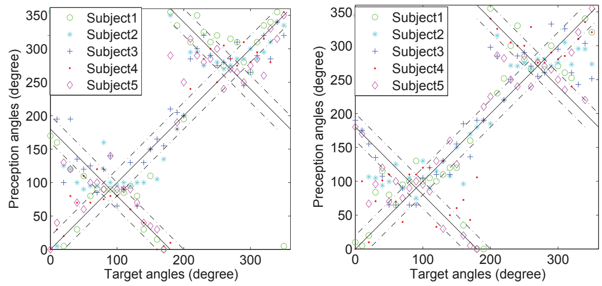
In this section, we conduct subjective experiments to evaluate the localization performance of proposed SWM. 5 listeners without any hearing problem participated in subjective experiments. Their anthropometric features are measured using the method in [24] via a camera.

Prior to the experiments, the subjects perform the procedural training to reduce the influence of procedural factors on the results by playing binaural signals from 5 different directions with the feedback, while in the test phase, no feedback is given. The localization is tested on a whole sphere at a distance of 1.2m. Furthermore, the test files are labeled by a random value from 1 to 1000 for the three kinds of HRTFs. During the experiments, the repeat is allowed. Finally, after listening to an audio, 5 subjects are required to record the corresponding perception direction.

The perception results of subjective localization experiments for 5 subjects are shown in Fig. 4 in terms of the target angles and the perception angles on the horizontal plane. In this figure, the results between the diagonal lines with positive slope are regarded as correct answers, with the interval of two neighbour lines of 20°. The results between two diagonal lines with negative slope are the front-back confusion judgements.

From Fig. 4, it can be observed that the correct rate (CR) using individual HRTFs is significantly higher than that using generic HRTFs and the individual HRTFs generated by our SWM. Among errors, the front-back confusion (FBR) happens quite frequently, and most of confusion errors are back-front confusion. Fig. 4 shows that there exists significant difference among subjects when considering the ability of localization perception. For example, the correct perception rate for subject 2 is only 40.11% and 42.89% using generic HRTFs and the individual HRTFs, respectively, while for subject 3, they are 41.42 and 54.81%, respectively. The noticeable perception difference is caused by several reasons. One possible reason is that some subjects are not familiar with the binaural audio even after procedural training. Another lies in different sensitivities among subjects to the gap between the estimated individual HRTFs by SWM and the ground truth.

Moreover, the statistical results tested on a sphere have been shown in Table 3 in terms of CR, FBR and up-down



(a) (b)

Figure 4: *Results of subjective localization experiments for 5 subjects. (a) generic HRTFs. (b) our SWM method.*

Table 3: *Subjective performance comparison in terms of correct rate (CR), front-back confusion rate (FBR) and up-down confusion rate (UDR) by using the proposed SWM, generic HRTFs and [18], respectively.*

HRIR data	CR (%)	FBR (%)	UDR (%)
Generic HRTFs	43.67	30.22	35.64
[18]	50.25	24.26	25.91
Proposed SWM	**53.33**	**17.78**	**21.59**

confusion rate (UDR) using the generic HRTFs, HRTFs from [18] and SWM. It can be seen that our proposed SWM method achieves best localization perception performance, and 9.66% and 3.08% gains of CRs are respectively achieved when compared to other methods. Furthermore, SWM generates a dramatic rate reduction of up to 12.44% and 14.05% in FBR and UDR, respectively. It infers that 1) the localization perception performance can be improved by the means of individualization, and 2) the improvement of combination weights is helpful for localization perception. Moreover, it is seen that the UDR is higher than FBR, because the main factor for up-down localization is the pinna, but they are not very sensitive to the variance of sound wave.

5. Conclusions

In this paper, an HRTF individualization method by exploring weight mapping based on anthropometric features is proposed. SWM first learns two sparse representations between the target subject and the training subjects in terms of anthropometric features and HRTFs, respectively. To this end, we use a non-negative sparsity-constrained model with considering the non-negative property of the anthropometric features. Next, a mapping between the two weight vectors is implemented by using a neural network based logistic regression, and an iterative data extension strategy is proposed in order to alleviate underfitting. The objective and subjective experimental results show that the proposed SWM method gains less LSD when compared with other methods, and achieves better localization perception performance. Our future work focuses on perception-based HRTF individualization, such as, setting different weights for different frequency bands based on perception sensitivity.

6. Acknowledgements

This work is supported by the National Natural Science Foundation of China (NSFC) (No.61603390), the the National Key Research & Development Plan of China (No. 2017YFB1002804).

7. References

[1] C. I. Cheng and G. H. Wakefield, "Introduction to head-related transfer functions (hrtfs): Representations of hrtfs in time, frequency, and space," *Journal of the Audio Engineering Society*, vol. 49, no. 4, pp. 231–249, 2001.

[2] H. Moller, "Fundamentals of binaural technology," *Applied Acoustics*, vol. 36, no. 3-4, pp. 171–218, 1992.

[3] B. S. Xie, "Recovery of individual head-related transfer functions from a small set of measurements," *Journal of the Acoustical Society of America*, vol. 132, no. 1, pp. 282–294, 2012.

[4] V. R. Algazi, C. Avendano, and R. O. Duda, "Estimation of a spherical-head model from anthropometry," *Journal of the Audio Engineering Society*, vol. 49, no. 6, pp. 472–479, 2001.

[5] V. R. Algazi, R. O. Duda, R. Duraiswami, N. A. Gumerov, and Z. H. Tang, "Approximating the head-related transfer function using simple geometric models of the head and torso," *Journal of the Acoustical Society of America*, vol. 112, no. 5, pp. 2053–2064, 2002.

[6] C. P. Brown and R. O. Duda, "A structural model for binaural sound synthesis," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 5, pp. 476–488, 1998.

[7] M. Otani and S. Ise, "Fast calculation system specialized for head-related transfer function based on boundary element method," *Journal of the Acoustical Society of America*, vol. 119, no. 5, pp. 2589–2598, 2006.

[8] H. Takemoto, P. Mokhtari, H. Kato, R. Nishimura, and K. Iida, "Mechanism for generating peaks and notches of head-related transfer functions in the median plane," *The Journal of the Acoustical Society of America*, vol. 132, no. 6, pp. 3832–3841, 2012.

[9] S. Hwang, Y. Park, and Y.-s. Park, "Modeling and customization of head-related impulse responses based on general basis functions in time domain," *Acta Acustica united with Acustica*, vol. 94, no. 6, pp. 965–980, 2008.

[10] K. J. Fink and L. Ray, "Individualization of head related transfer functions using principal component analysis," *Applied Acoustics*, vol. 87, pp. 162–173, 2015.

[11] K. Sunder, E.-L. Tan, and W.-S. Gan, "Individualization of binaural synthesis using frontal projection headphones," *Journal of the Audio Engineering Society*, vol. 61, no. 12, pp. 989–1000, 2013.

[12] H. M. Hu, L. Zhou, H. Ma, and Z. Y. Wu, "Hrtf personalization based on artificial neural network in individual virtual auditory space," *Applied Acoustics*, vol. 69, no. 2, pp. 163–172, 2008.

[13] F. Grijalva, L. Martini, S. Goldenstein, and D. Florencio, "Anthropometric-based customization of head-related transfer functions using isomap in the horizontal plane," *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (Icassp)*, 2014.

[14] X. Liu and X. Zhong, "An improved anthropometry-based customization method of individual head-related transfer functions," in *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 336–339.

[15] F. Grijalva, L. Martini, D. A. F. Florêncio, and S. Goldenstein, "A manifold learning approach for personalizing hrtfs from anthropometric features," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, pp. 559–570, 2016.

[16] Q. H. Huang and Q. L. Zhuang, "Hrir personalisation using support vector regression in independent feature space," *Electronics Letters*, vol. 45, no. 19, pp. 1002–U58, 2009.

[17] P. Bilinski, J. Ahrens, M. R. P. Thomas, I. J. Tashev, and J. C. Platt, "Hrtf magnitude synthesis via sparse representation of anthropometric features," *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (Icassp)*, 2014.

[18] J. He, W.-S. Gan, and E.-L. Tan, "On the preprocessing and post-processing of hrtf individualization based on sparse representation of anthropometric features," *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 639–643, 2015.

[19] A. Kulkarni, S. K. Isabelle, and H. S. Colburn, "Sensitivity of human subjects to head-related transfer-function phase spectra," *Journal of the Acoustical Society of America*, vol. 105, no. 5, pp. 2821–2840, 1999.

[20] G. D. Romigh, D. S. Brungart, R. M. Stern, and B. D. Simpson, "Efficient real spherical harmonic representation of head-related transfer functions," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 921–930, 2015.

[21] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale l_1-regularized least squares," *IEEE journal of selected topics in signal processing*, vol. 1, no. 4, pp. 606–617, 2007.

[22] T. J. Hastie, R. J. Tibshirani, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2011.

[23] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The cipic hrtf database," in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*. IEEE, 2001, pp. 99–102.

[24] D. N. Zotkin, R. Duraiswami, and L. S. Davis, "Rendering localized spatial audio in a virtual auditory space," *IEEE Transactions on Multimedia*, vol. 6, no. 4, pp. 553–564, 2004.