*INTERSPEECH 2019*
September 15–19, 2019, Graz, Austria

# Parameter-Transfer Learning for Low-Resource Individualization of Head-Related Transfer Functions

*Xiaoke Qi*[1], *Lu Wang*[2]

[1] School of Information Management for Law, China University of Political Science and Law, China
[2] College of Computer Science and Software Engineering, Shenzhen University, China

qixiaoke@cupl.edu.cn, wanglu@szu.edu.cn

## Abstract

Individualized head-related transfer functions (HRTFs) play an important role in accurate localization perception. However, it is a great challenge to efficiently measure continuous HRTFs for each person in full space. In this paper, we propose a parameter-transfer learning method termed PTL to obtain individualized HRTFs based on a small set of HRTF measurements. The key idea behind PTL is to transfer a HRTF generation model from other database to a target individual. To this end, PTL first pretrains a deep neural network (DNN)-based universal model on a large database of HRTFs with the assist of domain knowledge. Domain knowledge is used to generate the input features derived from the solution to sound wave propagation equation at the physical level, and to design the loss function based on the knowledge of objective evaluation criterion. Then, the universal model is transferred to a target individual by adapting the parameters of a hidden layer of DNN with a small set of HRTF measurements. The adaptation layer is determined by experimental verification. We also conduct the objective and subjective experiments, and the results show that the proposed method outperforms the state-of-the-arts methods in terms of LSD and localization accuracy.

**Index Terms**: head-related transfer functions, individualization, spatial hearing, transfer learning

## 1. Introduction

Currently, augmented reality (AR) and virtual reality (VR) technologies are becoming increasingly popular in our lives. Spatial hearing plays an important role not only in VR/AR to further improve the naturalness of the vision scene but also in other applications such as navigation for the blind [1], cochlear implants[2] and entertainment, eg., music and movie.

To generate more precise localization perception, the single-channel sound is required to pass through two filters that contain all the localization-related information of two ears, i.e., head-related transfer functions (HRTFs). HRTFs describe the propagation response of sound waves from the sound source to ear drums in the form of refraction, reflection and diffraction in free space, which is highly related to the anthropometric features of a human, such as head width, cavum concha height, and pinna height [3]. As each subject has different anthropometric data, HRTFs are highly individual-dependent. It has been shown that the use of non-individualized HRTFs is prone to cause up-down inversion, front-back confusion, and in-head localization [4]. Therefore, individualized HRTFs are essential for accurate localization perception in spatial hearing.

The most accurate method for HRTF acquisition is to directly measure the impulse responses from the sound source to the human ears over the full space [5]. However, it is greatly time consuming, expensive, and non-scalable. In light of this, several theoretical or numerical models have been proposed to approximate the complicated human anatomy, such as spherical head model [6], snowman model [7], structural models [8], boundary element method [9], finite-difference time-domain method [10]. However, they all require expensive acquisition hardware and are computationally intensive. Many methods have been proposed for a better balance between the performance and the computation complexity, such as using a representation of anthropometric features to individualize HRTFs [11], building a direct correlation between anthropometries and HRTFs using an artificial neural network[12]. Besides, there is another promising method, which simplifies the HRTF acquisition process by reducing the required number of HRTF measurements and interpolating over the full space. In [13], a spatial principal component analysis model was proposed to evaluate individual weights from a small set of measured data. As a result, HRTF magnitudes at 493 source directions can be recovered from 73 measured source directions with about 19 dB of average signal-to distortion ratio in the recovery. In [14], based on generic HRTFs, an optimal minimum mean-square error solution was obtained by minimizing the spatial aliasing error in the spherical harmonics (SH) representation of the HRTFs, and incorporating statistics calculated from sparse HRTF measurements. The experiments showed that using 25 individual HRTF measurements and generic HRTFs as a reference, the estimation error of less than -6dB up to 12kHz was achieved. However, it was shown in [15] that aliasing cancelation in [14] may affect some of the artifacts.

To generate a more accurate HRTF individualization model and reduce the number of HRTF measurements, we propose a Parameter-Transfer Learning method termed PTL. The key idea behind PTL is to transfer a HRTF generation model from a person to a target individual based on a small set of HRTF measurements. To this end, PTL first pretrains an universal generation model on a large database of HRTFs with the assist of domain knowledge. Then, the model is transferred to the target by adapting the parameters of the universal model using a small set of HRTF measurements. PTL-based model is inherently nonlinear, which has potential to acquire more information from an individual. As a result, less measurements lead to the same performance as those linear models.

## 2. Proposed Parameter-Transfer Learning-Based Individualization Model

### 2.1. System architecture

In this paper, we propose a parameter-transfer learning method termed PTL to obtain a more accurate HRTF individualization model, and reduce the number of HRTF measurements. The main idea behind the system is to first train an universal model



http://dx.doi.org/10.21437/Interspeech.2019-2558

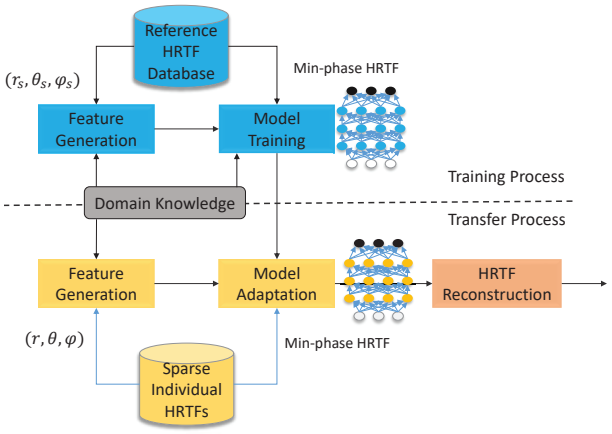


Figure 1: *The system architecture.*

from other subjects and fine-tune parameters of the model on a small set of HRTF measurements. Therefore, the model trained from others is transferred to a target individual.

The overall system architecture is shown in Fig. 1. The process of PTL can be divided into three key components: feature generation, universal model training and model adaptation. First, input features are generated by utilizing domain knowledge derived from the solution to sound wave propagation equation at the physical level, which depends on the position in three-dimensional space. Then, a deep neural network-based universal generation model is trained on a large database of HRTFs to make a mapping between the features and the HRTFs, where the loss function for model training is designed based on the knowledge of objective evaluation criterion, i.e., log-spectral distortion (LSD). To transfer the universal generation model to an individual, we utilize a model adaptation method to adjust the parameters of a hidden layer by using a small set of individual HRTF measurements. As a consequence, a HRTF individualization model is achieved and individualized HRTFs over full space will be constructed based on this model.

### 2.2. Feature generation

HRTFs can be characterized by the wave equation from the source to the ear canal. The source field can be represented by a specific set of orthogonal series, such as SH basis, spherical Fourier-Bessel (SFB) basis [16], which consists of spherical harmonics and spherical Bessel functions to represent the angular part and the radial part of HRTFs, respectively. Motivated by this domain knowledge, we generate the input features of the model based on SFB transform [17].

*The angular part* of SFB basis in our method exploits a real version of spherical harmonics by considering the property of the log-magnitude. Real spherical harmonics is a function of elevation $\phi$ and azimuth $\theta$, and can be expressed as [18] [19]

$$Y_n^m(\theta,\phi) = \sqrt{\frac{2n+1}{4\pi}\frac{(n-|m|)!}{(n+|m|)!}} P_n^{|m|}(\sin\theta) g(|m|\phi), \quad (1)$$

with

$$g(|m|\phi) = \begin{cases} \sin(|m|\phi), & m \le 0 \\ \cos(|m|\phi), & m > 0 \end{cases}, \quad (2)$$

where $n = 0, ...N$, and $|m| \le n$. $P_n^{|m|}(\cdot)$ is associated Legendre function of degree $n$ and order $m$.

The corresponding *radial part* on a solid sphere of radius $r$ uses normalized spherical Bessel function, which is defined as [20]

$$\Phi_{nl}(r) = \frac{1}{\sqrt{N_{nl}}} j_l(k_{nl} r), \quad (3)$$

where $j_l(x)$ is the spherical Bessel function of order $l$ and $j_l(x) = \sqrt{\pi/2x} J_{l+1/2}(x)$ with $J_{l'}(x)$ the Bessel function of order $l'$. Under the zero-value boundary condition, $k_{nl} = x_{nl}/a$ and $N_{nl} = a^3 j_{l+1}^2(x_{ln})/2$. $x_{ln}$ is the $n$th positive solution to $j_l(x) = 0$ in an ascent order, and $a$ is the maximum radius.

Finally, for each position $\mathbf{d} = (r, \theta, \phi)$, the set of the input features is generated by concatenating the angular part and the radial part as $\mathbf{F}(\mathbf{d}) = [Y_n^m(\theta,\phi), \Phi_{nl}(r)]$ with $n = 0, ...N$, $|m| \le n$, and $l = 1, ..., L$, which contains a total number of $N_t = [(N+1)^2 + NL]$ parameters.

As for the output label of the model, since human is not sensitive to the fine details of the phase spectrum of HRTFs in localization [21] and discrimination perception [22], the minimum phase HRTFs and interaural time delay can well approximate HRTFs [23] [24]. Moreover, considering that the phase part of the min-phase HRTFs can be obtained by Hilbert transform, we exploit the logarithmic magnitude of the min-phase HRTFs as the label.

### 2.3. Universal model training

Before the model is trained, the preprocessing is required to make the same variance for the training samples. For each position in HRTF database, a pair of training samples consists of input features and the corresponding HRTFs. For the input features, the preprocessing process is to normalize each dimension by the mean subtraction and then the standard variance division, which is calculated for the $j$-th feature $f_s(j)$ at the $s$-th position as

$$\tilde{f}_s(j) = \frac{f_s(j) - \mu_f(j)}{\sigma_f(j)}, j = 1, ..., N_t, \quad (4)$$

where $\mu_f(j)$ and $\sigma_f(j)$ denote the mean and standard variance of $j$-th item of the features for all the $S$ training positions.

For the corresponding HRTFs at the $s$-th position, the normalization is operated on each frequency bin, and is calculated as

$$\tilde{H}_s(i) = \frac{H_s(i) - \mu_h(i)}{\sigma_h(i)}, i = 1, ..., N_f, \quad (5)$$

where $\mu_h(i)$ and $\sigma_h(i)$ denote the mean and standard variance of HRTFs on $S$ training positions for the $i$-th frequency bin, respectively. $N_f$ is the number of the frequency bins.

After the preprocessing, pairs of features are fed into a DNN to train a model. The loss function is used for measuring the accuracy of the model. In order to design it properly, the knowledge of subjective perception should be considered. Since log-magnitude spectra preserves all of the perceptually-relevant information which is contained in a measured HRTF for a position [21], we design loss function derived from log-spectral distortion (LSD) which represents the difference between HRTFs on a logarithmic basis from human hearing, and has been widely used for objective evaluation of HRTFs models [25][26][27]. LSD expresses the distortion between the estimated and the measured HRTFs. Based on LSD, we define

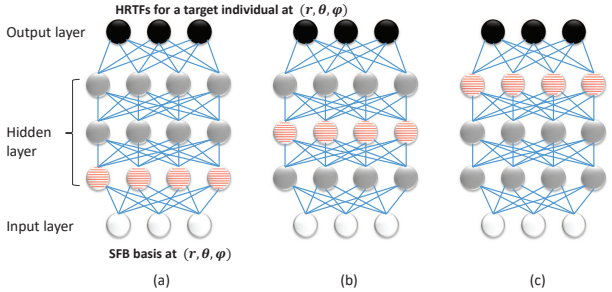


Figure 2: *The model adaptation structure. The white and black circles denote the nodes from the input layer and the output layer, respectively. The gray circles mean fixed network nodes during model adaptation, while the stripped circles mean that the parameters are fine-tuned on a small set of individual HRTF measurements. We respectively choose (a) the first hidden layer (termed PTL-F), (b) the middle hidden layer (termed PTL-M.) and (c) the last hidden layer as the adaptation layer (termed PTL-L), while keeping the parameter of other layers fixed*

a weighted mean square loss function, and is expressed as

$$L = \sqrt{\frac{1}{S \cdot N_f} \sum_{s=1}^{S} \sum_{k=k_1}^{k_2} \left[ \sigma_h(i) \left( \hat{H}_s(i) - \tilde{H}_s(i) \right) \right]^2}, \quad (6)$$

where $N_f$ is the number of the frequency bins from $k_1$ to $k_2$. $\hat{H}_s(i)$ denotes the estimated normalized HRTF for the $i$-th frequency bin on the $s$-th position. It is seen that we choose the standard variance as the weights for frequency bins to compensate the influence of preprocessing of HRTFs. By setting the loss function related to LSD, the model can maximize the objective performance by minimizing the loss function $L$.

### 2.4. Model adaptation

An important problem for PTL is to transfer the universal model to an individual. However, it is a great challenge for DNNs because a network often contains a large number of parameters, and we could not tune all the parameters in a deep network without enough data. A possible solution is to adapt the parameters of a specific layer based on sparse HRTF measurement, keeping the rest layers of the network unchanged. However, no conclusive study suggests on where exactly in the network this transformation should be implemented. In this paper, we choose the layer which provides the best empirical results.

As shown in Fig. 2, we respectively test three hidden layers as the adaptation layer while keeping the other nodes fixed, where the adaptation layer is (a) the first hidden layer (PTL-F), (b) the middle hidden layer (PTL-M), and (c) the last hidden layer (PTL-L). By analyzing the localization errors, we choose a proper layer as the adaptation layer. Then, with the loss function in Eq.( 6), the nodes at the adaptation layer are fine-tuned, and the network parameters are adapted to an individual based on its sparse HRTF measurements. Finally, we can synthesize HRTFs for a target subject over full space by generating features and then putting them forward to the PTL-trained model.

## 3. Performance Evaluation

In this section, we conduct experiments to evaluate the system performance. Two databases, PKU&IOA [28] and CIPIC [29], are used for universal model training and model adaptation, respectively. PKU&IOA database contains 793 locations for each distance and a total of 6344 HRTFs over the eight distances (20, 30, 40, 50, 75, 100, 130 and 160 cm) measured from the KEMAR mannequin. Each head-related impulse response (HRIR) has been windowed in about 15.625ms with the sampling rate of 65.536kHz. CIPIC database [29] contains HRIRs from the positions on 25 azimuths and 50 elevations at a distance of 1 m for 35 subjects. The length of each HRIR is approximately 4.5 ms with a sampling rate of 44.1 kHz.

Table 1: *Objective performance comparison of PTL-F, PTL-M, PTL-L, DNN model without adaptation(w/o adapt.), [13] and [14] in terms of LSD[dB] and RMSE.*

| Method | # of HRTFs | LSD | RMSE |
|---|---|---|---|
| [13] | 60 | 5.62 | 0.083 |
| [14] | 60 | 5.75 | 0.091 |
| DNN (w/o adapt.) | - | 7.19 | 0.11 |
| PTL-F | 60 | 5.48 | 0.082 |
| PTL-M | 60 | 5.51 | 0.081 |
| PTL-L | 60 | 4.95 | 0.072 |

First, the HRIRs for the two databases are resampled to the same sampling frequency, i.e., 44.1kHz. Then, HRIRs are converted to HRTFs using a 256-point FFT followed by constant-Q filtering, and the min-phase HRTFs are then obtained following by Hilbert transform. We evaluate the frequency bands between 200 Hz and 20 kHz. Together with ITD, there are 233 parameters for each direction, which are the outputs of DNN. For input features, we set $N = 10$, $L = 3$ and $a = 220$cm, as $N = 4$ is enough to preserve localization accuracy [24], and the inputs of DNN are computed by using Eq.(2) and (3), leading to $N_t = 130$ features. For DNN, we use the *Relu* activation function because of its nonlinearity and good performance in other tasks for hidden layers, and *linear* activation function for the output layer because of the large fluctuation for HRTFs in different frequency bands. The number of hidden layers is 3 with 512 nodes for each layer. The dropout fraction is set to 0.5, and the sparsity target is 0.2.

### 3.1. Objective evaluation

In this section, we conduct experiments to evaluate the objective performance of the proposed PTL method. LSD in frequency domain, and root mean square error (RMSE) in time domain are used as two metrics for objective evaluation, where RMSE is defined as the difference between the estimated and the measured HRIRs, which is expressed as

$$\text{RMSE} = \sqrt{\frac{1}{N_d N_t N_m} \sum_{m,t,d} (h_{m,t,d} - \hat{h}_{m,t,d})^2}, \quad (7)$$

where $h_{m,t,d}$ and $\hat{h}_{m,t,d}$ respectively denote the ground-truth HRIR and the estimated HRIR at $t$-th time sampling point of the $d$-th direction for subject $m$. $m = 1, ..., N_m$, $t = 1, ..., N_t$, $d = 1, ..., N_d$.

First, the influence of model adaptation on different layers is considered by comparing DNN model without adaptation with PTL-F, PTL-M and PTL-L under the same 60 HRTF measurements. The goal is to investigate the adaptation of the universal HRTF model trained on PKU&IOA to a target individual in CIPIC. The results are provided in Table 1. It is shown that PTL provides up to 2.24% relative improvement compared to the unadapted model. Especially, PTL-L

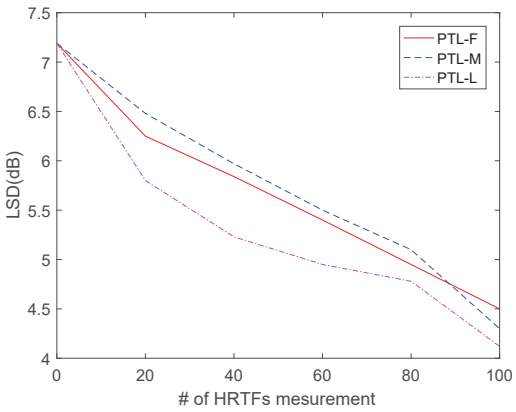


Figure 3: *The effect of the number of HRTF measurements for a target subject on the performance for PTL method.*

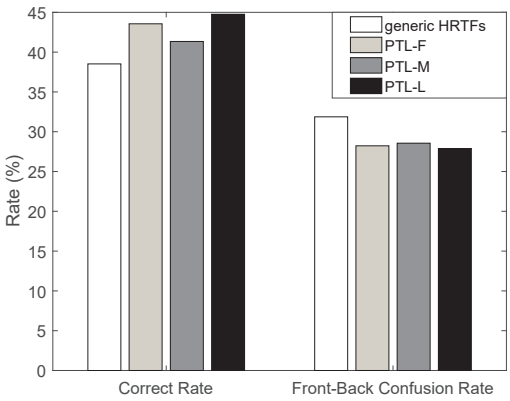


Figure 4: *Subjective performance comparison in terms of CR and FBR.*

achieves about 0.5dB LSD improvement over PTL-F and PTL-M methods. It indicates that a parametric transformation at the last hidden layer is most effective for adaptation of the universal model to the characteristics of a new subject. It is reasonable to perform the domain switch at the last hidden layer, because the input features of the model is related to the directions of measurements, and the output features are depend on individuals. Therefore, domain information of HRTFs is maximum in the last hidden layer of a HRTF generation model.

Furthermore, we also compare the performance of PTL with other two individualization methods [13] and [14] with the same number of HRTF measurements, as shown in Table 1. It is noticed that PTL-L achieves a best performance with up to 0.8dB of LSD improvement. This improvement mainly lies in two reasons. First, rather than a linear model in [13] and [14], PTL builds an inherently nonlinear model and has potential to make a more accurate mapping between localization features and individual HRTFs. Second, DNN is tolerent to noise, leading to a more robust PTL-based model .

Afterwards, we analyze the effect of the number of HRTFs for a target subject on the adaptation performance in terms of LSD, as shown in Fig. 3. The positions of HRTFs are randomly chosen for each subject from the CIPIC database. It is seen that the performance is better as the number of pre-measured HRTFs increases. When the number of HRTF measurements exceeds 60, LSD of PTL-L is reduced below 5dB.

### 3.2. Subjective evaluation

The subjective performance of the proposed PTL method is evaluated in terms of the perception correct rate (CR) and the front-back confusion rate (FBR) by comparing with generic HRTFs in PKU&IOA. 5 subjects without any hearing problem participated in experiments. Their small sets of HRTFs were measured as follows. First, 20 positions were chosen on a sphere at a distance of 1.2m from the head of a subject, and 15 of them were at the front and back because FBR is more sensitive to the individual. Then, a burst of 1s chirp signal with a bandwidth of 200Hz to 20kHz was generated and sent to each subject for 3 times at each position. After the sound signal was recorded at two ear canals, HRTFs could be calculated using a FFT-based synchronization method [30]. Finally, 20 HRTFs were obtained for each subject and used for mode adaptation. Prior to the experiments, the subjects performed the procedural training to reduce the influence of procedural factors on the results by playing binaural signals from 5 different directions with feedback, while in the test phase, no feedback was given. During the experiments, the subjects were allowed to repeatedly listen to spatial audio files and then were required to record their perceptive positions.

The perception results are illustrated in Fig. 4. It is shown that the proposed PTL achieves a better perception localization performance than the method using generic HRTFs. Especially, PTL-L provides best performance, which achieves 6.24% improvement in CR and and 3.98% reduction in FBR over generic HRTFs. The main reason is that the last hidden layer contains more domain information and the front-back localization correction is operated by measuring more front and back data.

To efficiently choose the positions of sparse HRTF measurements is a difficult task when we want to minimize the number of HRTFs while keeping an accurate localization perception. Our future work will focus on this problem, and conduct a more detailed subjective listening test.

## 4. Conclusions

In this paper, we propose a parameter-transfer learning method termed PTL to obtain accurate individualized HRTFs based on a small set of HRTF measurements. The key idea behind PTL is to transfer an HRTF generation model from other persons to a target individual. To this end, PTL first pretrains a universal neural network on a large database of HRTFs, and then transfers it to an individual by adapting the parameters of a layer of the neural network using a small set of HRTF measurements. The target layer for model adaptation is determined by experimental verification, and PTL-L method achieves best performance because domain information is maximum in the LAST hidden layer. Compared with other linear models, the experimental results show that the proposed PTL method achieves a better localization perception performance under the same HRTF measurements. Therefore, instead of measuring HRTFs over the full space for each subject, only few-shot samples of HRTFs are sufficient to help to build an effective individualization model.

## 5. Acknowledgements

This work is supported by the National Natural Science Foundation of China (NSFC) (No.61603390, 61872246), Scientific Research Foundation of CUPL (10819146), Guangdong Special Support Program, Shenzhen Science and Technology Foundation (JCYJ20170817095418831).

# 6. References


[1] J. M. Loomis, R. G. Golledge, and R. L. Klatzky, "Navigation system for the blind: Auditory display modes and guidance," *Navigation*, vol. 7, no. 2, 2006.

[2] R. Y. Litovsky, A. Parkinson, and J. Arcaroli, "Spatial hearing and speech intelligibility in bilateral cochlear implant users," *Ear and hearing*, vol. 30, no. 4, p. 419, 2009.

[3] 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.

[4] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer-functions," *Journal of the Acoustical Society of America*, vol. 94, no. 1, pp. 111–123, 1993.

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

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] X. Qi and J. Tao, "Sparsity-constrained weight mapping for head-related transfer functions individualization from anthropometric features," *Proc. Interspeech 2018*, pp. 841–845, 2018.

[12] 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.

[13] 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.

[14] D. L. Alon, Z. Ben-Hur, B. Rafaely, and R. Mehra, "Sparse head-related transfer function representation with spatial aliasing cancellation," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 6792–6796.

[15] Z. Ben-Hur, D. L. Alon, B. Rafaely, and R. Mehra, "Loudness stability of binaural sound with spherical harmonic representation of sparse head-related transfer functions," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2019, no. 1, p. 5, 2019.

[16] W. Zhang, T. D. Abhayapala, R. A. Kennedy, and R. Duraiswami, "Insights into head-related transfer function: Spatial dimensionality and continuous representation," *Journal of the Acoustical Society of America*, vol. 127, no. 4, pp. 2347–2357, 2010.

[17] Q. Wang, O. Ronneberger, and H. Burkhardt, "Fourier analysis in polar and spherical coordinates," *Albert-Ludwigs-Universität Freiburg, Institut für Informatik*, 2008.

[18] R. Duraiswami, Z. Li, D. N. Zotkin, E. Grassi, and N. A. Gumerov, "Plane-wave decomposition analysis for spherical microphone arrays," in *Applications of Signal Processing to Audio and Acoustics, 2005. IEEE Workshop on*. IEEE, 2005, pp. 150–153.

[19] M. Poletti, "Unified description of ambisonics using real and complex spherical harmonics," *Ambisonics Symp.*, vol. 1, no. 1, pp. 2–2, 2009.

[20] Q. Wang, O. Ronneberger, and H.Burkhardt, "Rotational invariance based on fourier analysis in polar and spherical coordinates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 9, pp. 1715–1722, 2009.

[21] D. J. Kistler and F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *The Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1637–1647, 1992.

[22] 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.

[23] B. Xie, *Head-related transfer function and virtual auditory display*. J. Ross Publishing, 2013.

[24] 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.

[25] 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.

[26] K. J. Fink and L. Ray, "Tuning principal component weights to individualize hrtfs," *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 389–392, 2012.

[27] 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.

[28] T. S. Qu, Z. Xiao, M. Gong, Y. Huang, X. D. Li, and X. H. Wu, "Distance-dependent head-related transfer functions measured with high spatial resolution using a spark gap," *IEEE Transactions on Audio Speech and Language Processing*, vol. 17, no. 6, pp. 1124–1132, 2009.

[29] 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.

[30] S. Boumard and A. Mammela, "Robust and accurate frequency and timing synchronization using chirp signals," *IEEE Transactions on Broadcasting*, vol. 55, no. 1, pp. 115–123, 2009.