INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

A Domain Knowledge-Assisted Nonlinear Model for Head-Related Transfer Functions Based on Bottleneck Deep Neural Network

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, Institute of Automation, Chinese Academy of Sciences, China
[3]School of Computer and Control Engineering, University of Chinese Academy of Sciences, China

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

Abstract

Many methods have been proposed for modeling head-related transfer functions (HRTFs) and yield a good performance level in terms of log-spectral distortion (LSD). However, most of them utilize linear weighting to reconstruct or interpolate HRTFs, but not consider the inherent nonlinearity relationship between the basis function and HRTFs. Motivated by this, a domain knowledge-assisted nonlinear modeling method is proposed based on bottleneck features. Domain knowledge is used in two aspects. One is to generate the input features derived from the solution to sound wave propagation equation at the physical level, and the other is to design the loss function for model training based on the knowledge of objective evaluation criterion, i.e., LSD. Furthermore, with utilizing the strong representation ability of the bottleneck features, the nonlinear model has the potential to achieve a more accurate mapping. The objective and subjective experimental results show that the proposed method gains less LSD when compared with linear model, and the interpolated HRTFs can generate a similar perception to those of the database.

Index Terms: head-related transfer functions, spherical Fourier-Bessel, bottleneck features, spatial hearing

1. Introduction

An HRTF is a response that characterizes how the sound wave propagates from the source to the ears of the listener, which is a complex function of the frequency, range, azimuth and elevation angle. Head-related transfer functions (HRTFs) have played a key role in spatial audio because they carry most of spatial information used in localization[1]. In order to give an immersive display in three-dimensional (3D) space, the high spatial resolution measurement is necessary to make discrete HRTFs continuous[2][3]. However, it is quite time-consuming and challenging because of uncontrollable factors during measurement process, such as the head movement of subjects, the measurement error, the position shifting, and so on. A promising solution is to model limited HRTFs in lower dimensional spaces[4], and then generate HRTFs by interpolating or extrapolating in full space. Therefore, the heart of the solution stands on more accurate modeling for HRTFs.

Many methods have been proposed for HRTFs linear modeling. One is based on principal components analysis (PCA) [5][6] or the spatial feature extraction method, such as spatial PCA [7]. The spatial variation is modeled by the combination of a small number of principal components. In order to interpolate in 3D space, [8] proposes a tensor modeling for distance-dependent HRTFs by adopting multilinear principal component analysis (MPCA), and then using a linear interpolation of two adjacent core tensors to interpolate HRTFs on a new distance. Another approach is surface spherical harmonics-based modeling (SHM) [9] [10]. Spherical harmonics (SH) are a complete set of continuous orthonormal basis functions on the sphere. By using SH, the model extracts the directional cues from HRTFs, and achieves an encouraging level in terms of log-spectral distortion (LSD). The main advantage of SHM is that the HRTFs can be modeled with a linear combination of relatively small set of SH expansion coefficients at the full space. Furthermore, a generative model of HRTFs in frequency-range-angle domains is proposed in [11], which combines spherical harmonics and spherical Hankle functions as the basis in 3D space. When used for interpolation, HRTFs are linearly generated by weighting the basis with the coefficients.

In order to exploit the inherent nonlinearity property of HRTFs, nonlinear techniques are used for dimensionality reduction. [12] proposes a interpolation method based on Isomap on a sphere. [13] uses locally linear embedding (LLE) to construct a manifold only on the median plane, and finds that their algorithm is not very stable when using all directions from one subject. However, these methods only generate model coefficients by using nonlinear methods, but still exploit linear weighting to synthesize HRTFs.

Motivated by this, we propose a domain knowledge-assisted nonlinear model to build the relationship between the features and HRTFs, and then generate HRTFs at any position of 3D space. Domain knowledge is utilized in two aspects. First, the features are generated based on sound wave propagation at the physical level, which is a function of range, azimuth and elevation angles. Second, the loss function for training bottleneck deep neural network (DNN) is designed based on the knowledge of subjective evaluation criterion to match the localization perception. The objective and subjective experiments suggest that the method can achieve a more accurate mapping between the features and the HRTFs.

2. Proposed domain knowledge-assisted nonlinear model

2.1. System architecture

The system architecture of the proposed domain knowledge-assisted modeling method is shown as Fig. 1. The main idea behind the system is to utilize the nonlinear property of neural network to improve the representation ability of the model to the

http://dx.doi.org/10.21437/Interspeech.2017-222

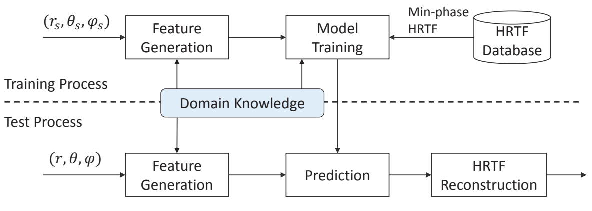

Figure 1: *The system architecture.*

HRTFs, and extract more correlated information between the features and the HRTFs with the assistant of domain knowledge.

The system contains two phases: the training phase and the test phase. In the training phase, the features are first generated as an input of the model by using domain knowledge. That is, HRTFs can be characterized by the wave equation from the source to the ear cannel, and the solution can be described by several variables, such as frequency, range, azimuth and elevation angles. Then, the nonlinear model is trained by using a bottleneck deep neural network to make a mapping between the features and the HRTFs. 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 [6] and discrimination perception [14], the minimum phase HRTFs and interaural time delay can well approximate HRTFs [15] [10]. 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. In the test phase, the features are obtained given any target position, and then the trained model is used for predicting log-magnitude HRTFs. Finally, by inverse Hilbert transform, the target HRTFs will be obtained.

In the next subsection, we will introduce two core parts of the system in detail: the domain knowledge-based feature generation and the model training process, which includes preprocessing, model structure and domain knowledge-based loss function design.

2.2. Feature generation

The source field can be represented by a specific set of orthogonal series, such as SH basis, spherical Fourier-Bessel (SFB) basis [11], 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 [16].

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 ϕ and azimuth θ expressed as [17] [18]

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

with

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

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

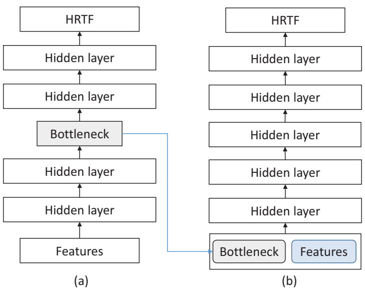

Figure 2: *The model structure.*

The corresponding *radial part* on a solid sphere of radius r uses normalized spherical Bessel function defined as [19]

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

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 nth 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 $n = 0, ...N$, $|m| \leq n$, and $l = 1, ..., L$, which contains a total of $N_t = [(N+1)^2 + NL]$ parameters.

2.3. Model training process

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 features, the preprocessing process is to normalize each dimension by the mean subtraction and then the standard variance division, which is calculated for the i-th item $f_s(i)$ of the feature at the s-th position as

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

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

For the corresponding HRTFs, the normalization is operated on each frequency bin, and for the HRTF $H_s(i)$ of i-th frequency bin at the s-th position $\mathbf{d}_s$ calculated as

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

where $\mu_h(i) = 1/S \sum_{s=1}^{S} H_s(i)$ 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 used in modeling.

After the preprocessing, pairs of samples are fed into bottleneck DNN to train a model, and the structure is shown in Fig. 2. First, bottleneck features, which have widely used in many fields to improve the accuracy performance, such as speech recognition [20], speech synthesis [21], and so on, are generated from a multi-layer perceptron. The structure for extracting bottleneck features is shown in Fig. 2(a), in which one of the internal layers has a small number of hidden units, making a more compact representation between both the features and the HRTFs. Then, as shown in Fig. 2(b), by concatenating the features generated from domain knowledge and the extracted bottleneck features, the model is trained to minimize a loss function.

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 [6], we design loss function derived from 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 [22][23][24]. Based on LSD, we define a weighted mean square loss function 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.

3. Performance evaluation

In this section, we conduct objective and subjective experiments to evaluate the performance of the proposed domain knowledge-assisted nonlinear model. PKU&IOA database is used for this purpose [25]. The 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 (1024 points) with the sampling rate of 65.536kHz. Prior to the nonlinear modeling, all the HRIRs are first converted to the HRTFs by using 1024-DFT, and the min-phase HRTFs are then obtained following by Hilbert transform. We evaluate the frequency bands between 200Hz and 20kHz. Therefore, for each position, there is a total number of 618 HRTF parameters for modeling double ears.

For the bottleneck 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 5 with 1024 nodes for each. The length of bottleneck features is set to 30.

3.1. Objective evaluation

First, objective performance is evaluated by comparing the proposed method (bottleneck DNN) with the spherical Fourier-Bessel (SFB) model in terms of LSD [11]. Because [11] models the complex HRTFs while the log-magnitude HRTFs are used in this paper, the complex basis is first modified to its real version.

Several experimental conditions are designed for interpolation and extrapolation performance evaluation. For the interpolation performance, we evaluate by respectively considering HRTFs on the sphere of $r = 100$cm, median planes ($\theta = 0^\circ$) and horizontal planes ($\phi = 0^\circ$) as test samples. Then, by setting HRTFs on the sphere of $r = 20, 130$cm and $\phi = -40^\circ$ as test samples, the performance of model extrapolation is evaluated. For a total of 6 experiments, the parameters are set to $N = 15$, $L = 3$ and $a = 220$cm.

Table 1: *LSD (dB) comparison for the proposed nonlinear model and the spherical Fourier-Bessel linear model when considering interpolation and extrapolation.*

Test conditions	Bottleneck DNN	NN	SFB
$r = 100$cm	3.234	3.281	4.861
$\theta = 0^\circ$	5.513	5.800	8.871
$\phi = 0^\circ$	4.040	4.693	4.004
$r = 130$cm	4.010	4.562	8.745
$r = 20$cm	3.717	3.759	5.399
$\phi = -40^\circ$	5.065	5.791	15.117

The LSD results are shown in Table 1 with the comparison of bottleneck DNN, NN without bottleneck features, and SFB. First, it can be seen that the LSD of SFB model is larger than other nonlinear methods in most cases, implying the efficiency of the nonlinear model. For the distance interpolation and extrapolation, bottleneck DNN achieves up to 4.735dB gain over SFB model. This gain comes from the stronger information extraction of the nonlinear model, which can make a more accurate mapping between the features and the HRTFs. Second, we observe that the objective performance of bottleneck DNN is slightly better than the NN. The reason is that the bottleneck layer can extract more high-dimensional features of the basis that may be lost in a single NN, and thus results in the performance improvement. Third, by comparing the performance of model interpolation and extrapolation, we can find that the SFB model is more sensitive to the interpolation or extrapolation, for which the former gives much better performance than the latter. Furthermore, notice that for the low elevation extrapolation of $\phi = -40^\circ$, SFB model hardly gives a reliable result. One possible reason comes from the dramatic inconsistency between the theory of wave propagation and HRTFs caused by knees, body, hair, and so on, and thus results in worse mapping between the basis and the the measured HRTFs. Another reason lies in little tolerance of SFB linear model to this inconsistency. Moreover, it is observed that worst LSD performance occurs on the median planes for bottleneck DNN, which implies that the generation of the up-down HRTFs is more complicated and does not exactly match with the wave propagation theory.

3.2. Subjective evaluation

The subjective performance is evaluated on the horizontal plane in terms of the perception correct rate, by comparing the measured HRTFs and interpolated HRTFs obtained via the proposed bottleneck DNN model and SFB model. 35 azimuth angles are chosen at a interval of 10° as the test directions.

The two experiments are designed as follows. For the first experiment, the stimulus signal is from a 5s human sound with the bandwidth of 200 $\sim$ 8kHz, and for the second experiment, the signal is a burst of 5s chirp noise with the bandwidth of 200 $\sim$ 20kHz. The sample ratio is 65.536kHz to match with that of the HRIRs. The binaural signal is produced by convolving the stimulus with the HRTFs of the desired target direction and then played via the earphones. 5 listeners without any hearing problem participated in these experiments. Before the experiments, the subjects perform the procedural training

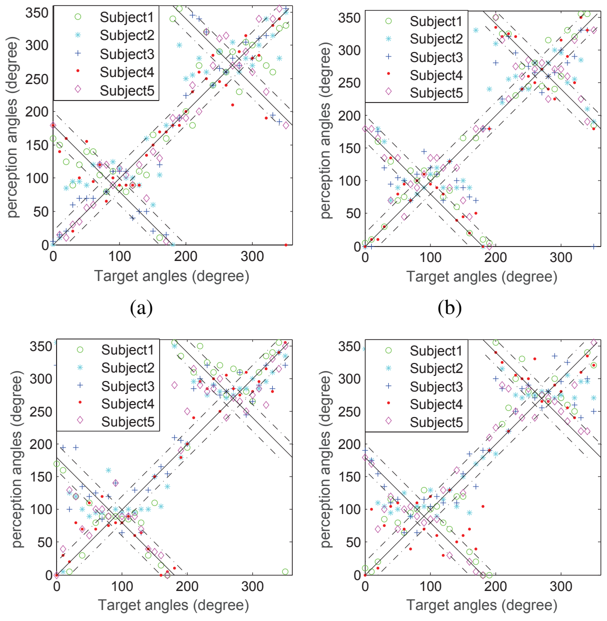

Figure 3: *Results of subjective localization experiments for 5 subjects. (a) Human sound with measured HRTFs. (b) Human sound with interpolated HRTFs by using our method. (c) Chirp noise with measured HRTFs. (d) Chirp noise with interpolated HRTFs by using our method.*

to reduce the influence of procedural factors to the results, by playing binaural signals from 5 different directions with the feedback, while in the test phase, no feedback is given. 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 a file, 5 subjects are required to record the corresponding perception angle.

The perception results of subjective localization experiments for 5 subjects are shown in Fig. 3 in terms of the target angles and the perception angles for measured HRTFs and the interpolated HRTFs via our method. In this figure, the results between the diagonal lines with positive slope are regarded as the 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. The average correct rates and front-back confusion rates are also calculated after all experiments are completed and listed in Table 2.

From Fig. 3, it can be intuitively observed that little difference exists between the measured and the interpolated HRTFs. The front-back confusion happens quite frequently for both sound and full-band chirp noise, and further the majority of errors are back-to-front confusion. The reason is that the discrimination of front and back mainly depends on the structure of one's own pinna, but the non-individual HRTFs used in the experiments could not provide this information. Furthermore, Fig. 3 shows that localization perception abilities exist significant difference among subjects. For example, the correct perception rate for subject 2 is only 36.11% and 38.89% for the measured HRTFs and the estimated HRTFs via our method, while for subject 3, the correct rate is 52.78% for two methods. The noticeable perception difference generates from several reasons. One possible reason is that some subjects are not familiar with the binaural audio even after procedural training. Another is that individual difference exists between subjects' anthropometric parameters and the KEMARs. Since an HRTF is a response of sound wave from a sound source to the ear canal via diffusing effects from head, torso, and pinna, HRTFs for subjects' own and KEMARs' are different. Therefore, when using KEMARs' HRTFs for all subjects, the localization perception is different among subjects.

Table 2: *The comparison of subjective localization results by using the measured HRTFs, SFB method and the proposed method, respectively.*

Source	**HRIR data**	**Correct rate (%)**	**Front-back confusion rate (%)**
Sound (200Hz ∼8kHz)	Measured	48.15	22.22
	SFB	45.82	24.36
	Proposed	47.22	22.92
Chirp (200Hz ∼20kHz)	Measured	44.84	26.59
	SFB	43.16	26.12
	Proposed	44.84	24.60

Furthermore, it is shown from Table 2 that the front-back confusion rate is much lower for the full frequency band noise than that for the low frequency sound. That is because the pinna also plays a key role in the localization for the high frequency bands, but that is lost in non-individual HRTFs. Moreover, by comparing the three methods, it indicates that the perception performance of our method is much closer to the measured one than SFB linear model. Especially for chirp noise, bottleneck DNN achieves the same correct rate, and slightly higher front-back confusion rate than the measured one. This result infers that from the perception perspective, the proposed model can achieve the similar performance with the measured HRTFs.

4. Conclusions

In this paper, a domain knowledge-assisted nonlinear modeling method is proposed based on bottleneck deep neural network. Domain knowledge is used for feature generation and loss function design. For feature generation, the spherical harmonics and spherical Bessel functions are exploited to represent the angular part and the radial part of HRTFs, respectively, based on the theory of sound wave propagation. When designing loss function, we utilize LSD to guide the choice of loss function for model training. Combining with the strong representation ability of the bottleneck features, our proposed nonlinear model has the potential to achieve a more accurate mapping between the features and the HRTFs. We also conduct objective and subjective experiments, and the results suggest that the proposed model achieves less LSD when compared with linear model, and generates a similar localization perception to the measured ones from the database. Future work will focus on the feature improvement and HRTFs individualization by exploiting the nonlinear methods.

5. Acknowledgements

This work is supported by the National Natural Science Foundation of China (No.61603390, No.61425017, No.61403386), the National High-Tech Research and Development Program of China (863 Program) (No.2015AA016305), the Major Program for the National Social Science Fund of China (13&ZD189), and the Strategic Priority Research Program of the CAS (Grant XDB02080006).

6. 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] W. Zhang, R. A. Kennedy, and T. D. Abhayapala, "Efficient continuous hrtf model using data independent basis functions: Experimentally guided approach," *IEEE Transactions on Audio Speech and Language Processing*, vol. 17, no. 4, pp. 819–829, 2009.

[3] D. N. Zotkin, R. Duraiswami, and N. A. Gumerov, "Regularized hrtf fitting using spherical harmonics," in *Applications of Signal Processing to Audio and Acoustics, 2009. WASPAA'09. IEEE Workshop on*. IEEE, 2009, pp. 257–260.

[4] S. Shekarchi, J. Christensen-Dalsgaard, and J. Hallam, "A spatial compression technique for head-related transfer function interpolation and complexity estimation," *Journal of the Acoustical Society of America*, vol. 137, no. 1, pp. 350–361, 2015.

[5] W. L. Martens, *Principal components analysis and resynthesis of spectral cues to perceived direction*. Ann Arbor, MI: MPublishing, University of Michigan Library, 1987.

[6] 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.

[7] 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.

[8] Q. Huang, K. Liu, and Y. Fang, "Tensor modeling and interpolation for distance-dependent head-related transfer function," in *2014 12th International Conference on Signal Processing (ICSP)*. IEEE, 2014, pp. 1330–1334.

[9] M. J. Evans, J. A. Angus, and A. I. Tew, "Analyzing head-related transfer function measurements using surface spherical harmonics," *The Journal of the Acoustical Society of America*, vol. 104, no. 4, pp. 2400–2411, 1998.

[10] 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.

[11] 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.

[12] F. Grijalva, L. C. Martini, D. Florencio, and S. Goldenstein, "Interpolation of head-related transfer functions using manifold learning," *IEEE Signal Processing Letters*, vol. 24, no. 2, pp. 221–225, 2017.

[13] R. Duraiswami and V. C. Raykar, "The manifolds of spatial hearing," *2005 IEEE International Conference on Acoustics, Speech, and Signal Processing, Vols 1-5*, pp. 285–288, 2005.

[14] 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.

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

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

[17] 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.

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

[19] 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.

[20] D. Yu and M. L. Seltzer, "Improved bottleneck features using pretrained deep neural networks." in *Interspeech*, vol. 237, 2011, p. 240.

[21] Z. Wu, C. Valentini-Botinhao, O. Watts, and S. King, "Deep neural networks employing multi-task learning and stacked bottleneck features for speech synthesis," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 4460–4464.

[22] 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.

[23] 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.

[24] 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.

[25] 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.