INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

A Sparse Spherical Harmonic-Based Model in Subbands for Head-Related Transfer Functions

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

[1]National Laboratory of Pattern Recognition (NLPR),
[2]CAS Center for Excellence in Brain Science and Intelligence Technology,
Institute of Automation, Chinese Academy of Sciences, Beijing, China, 100190

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

Abstract

Several functional models for head-related transfer function (HRTF) have been proposed based on spherical harmonic (SH) orthogonal functions, which yield an encouraging performance level in terms of log-spectral distortion (LSD). However, since the properties of subbands are quite different and highly subject-dependent, the degree of SH expansion should be adapted to the subband and the subject, which is quite challenging. In this paper, a sparse spherical harmonic-based model termed SSHM is proposed in order to achieve an intelligent frequency truncation. Different from SH-based model (SHM) which assigns the degree for each subband, SSHM constrains the number of SH coefficients by using an l_1 penalty, and automatically preserves the significant coefficients in each subband. As a result, SSHM requires less coefficients at the same SD level than other truncation methods to reconstruct HRTFs . Furthermore, when used for interpolation, SSHM gives a better fitting precision since it naturally reduces the influence of the fluctuation caused by the movement of the subject and the processing error. The experiments show that even using about 40% less coefficients, SSHM has a slightly lower LSD than SHM. Therefore, SSHM can achieve a better tradeoff between efficiency and accuracy.

Index Terms: head-related transfer functions, spherical harmonic, sparse representation, spatial hearing

1. Introduction

Virtual auditory displays (VAD) have attracted more and more attention since virtual reality has been making a great revolution in society. VAD gives us quite vivid auditory perception, and pretends that we are surrounded by three-dimension (3D) sound. In order to make VAD, head-related transfer function (HRTF), which contains all of information in localization and gives a spatial perception, is necessary for naturalness and localization accuracy. HRTF describes the propagation from the sound source to ear drums in free space[1], which is highly individual-dependent. In order to generate the spatial audio for everybody, theoretically it should be measured throughout the space for each subject. Therefore, a great challenge for VAD is to efficiently store HRTF database from dense measurements because of the dependent property of HRTFs. One promising solution is to model HRTFs in lower dimensional spaces. Then, the model can be used to compress[2], interpolate to make discrete HRTFs continuous[3][4], fast convolve to generate 3D audio[5], and guide to sample during measurement[6][7] and so on.

Many methods have been proposed for HRTF modeling. One approach is based on principal components analysis (PCA) [8][9] or the spatial feature extraction method, such as spatial PCA [10]. The spatial variation is modeled by the combination of a small number of principal components. However, besides the principal components coefficients, the basis matrix of these methods should be saved since it is changed with the subject, resulting in less efficiency. Moreover, it is not flexible to interpolate HRTF for movement or head rotation in VAD. Another approach is surface spherical harmonics-based modeling (SHM) [11]. 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 spectral distortion (SD). 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, its basis is universal for all subjects, and thus only the SH coefficients are required to store.

In practice, SHs are truncated using a degree, and then the SH coefficients are estimated by solving linear equations of SHs and HRTFs. The degree plays an important role in modeling. [11] showed that the magnitude and phase of HRTFs can be separately modeled by using the degree of 7. [12] studied the impact of reconstructed HRTFs with reduced-degree SH expansions on the perception of virtual sounds, and found that the accurate localization performance is retained with the degree of 4. However, two problems exists in these traditional methods. First, it is not reasonable to assign the same degree for all subbands, because the importance for each subband to human perception is different. Second, SHM with a lower degree will suffer from underfitting and result in a large interpolation SD with a high probability. As a solution, [13] proposed two truncation methods respectively for spherical head and human subjects. By developing it, [14] assigned an upper SH degree, which is higher at the low frequencies than [13], and bounded by 30 at the high frequencies. However, these methods have no reason to select a degree varying with subbands, and cannot make a subject-adaptive truncation.

Motivated by this, we present a sparse spherical harmonic-based model termed SSHM. SSHM avoids to directly set a degree, since it is a great challenge to design a systematical truncation method where the degree adapts with the subband and the subject. Instead, by constraining the sparsity of the model, SSHM automatically preserves the significant coefficients in each subband for different subjects, and discards the relatively insignificant ones. Its advantage lies in adaptively controlling the frequency truncation without fixing a degree.

The remainder of this paper is organized as follows. Section

Copyright © 2016 ISCA

http://dx.doi.org/10.21437/Interspeech.2016-987

2 presents an overview of spherical harmonic-based model, and the related methods for frequency truncation. Section 3 describes the proposed sparse spherical harmonic-based model. The performance evaluation results are shown in Section 4. Finally, Section 5 gives the conclusions.

2. Overview of SHM

First, the spherical harmonic-based model (SHM) is introduced. Spherical harmonic is a function of elevation θ and azimuth ϕ [15] [16], which can be expressed as

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

where $n = 0, 1, 2, \ldots$, and $|m| \leq n$. $P_n^{|m|}(\cdot)$ is associated Legendre function of degree n and order m.

At the direction (θ, ϕ), SHM models the HRTFs as

$$H(\theta, \phi, f) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} C_n^m(f) Y_n^m(\theta, \phi), \quad (2)$$

where f is the frequency bin. $C_n^m(k)$ are the complex coefficients. In practice, this representation of (2) is truncated by using a degree of $N(f)$ for the frequency bin f. Thus, $C_n^m(k)$ can be approximated by using a limited number of samples over the space S as

$$C_n^m(f) = \sum_{s=1}^{S} H(\theta_s, \phi_s, f) Y_n^{m*}(\theta_s, \phi_s) \sin\theta_s, \quad (3)$$

where $(\cdot)^*$ denotes the conjugate operator.

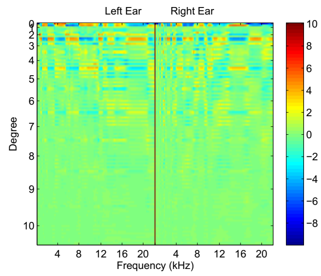
There are two popular approaches to determine $N(f)$. One is to choose a constant degree for any f, i.e., $N(f) = N$, while the other is to set the degree changing with frequency. [13] sets the frequency-dependent degree as $N(f) = \lceil \frac{e\pi f s_1}{c} \rceil = \lceil \frac{eks_1}{2} \rceil$, where $e = \exp(1)$. $k = 2\pi f/c$ denotes wavenumber and c is the speed of sound propagation in air. s_1 is the typical head radius, with 0.2m for $f \leq$ 3kHz and 0.09m otherwise. For the frequency of 20kHz, the degree can reach to 49. For another example, $N(k)$ is chosen larger at low frequencies and bounded by 30 at high frequencies[14], which is expressed as $N(k) = \min\left\{30, \lceil \frac{e\pi f s_2}{c} \rceil\right\}$, with $s_2 = 0.5$m. Though the method makes the degree varying with the frequency, it did not consider the difference between subjects, as well as the non-monotonicity with the subbands.

3. Proposed SSHM

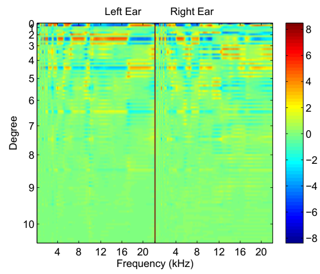
In this section, the proposed SSHM will be described in detail. Motivated by the drawbacks in the existing SHM methods, SSHM is proposed to achieve an adaptive degree truncation in subbands. Prior to SSHM, the preprocessing of HRTF database is discussed, and the parameters used for modeling are determined. Then, the optimization problem for SSHM is derived to obtain the sparse coefficients. As a result, SSHM allows to faster generate continuous HRTFs over the whole space, which is prerequisite for VAD.

3.1. Problem formulation

In the SHM method, frequency truncation plays an important role in the tradeoff between the efficiency and the accuracy. An optimal degree setting will definitely improve the performance of the model. However, it is not clear that which frequency is dominant in human localization perception, and which subject requires more or less coefficients to achieve a better balance. Therefore, it is quite a challenge to make a degree to adapt with the frequency as well as the subject.



(a)



(b)

Figure 1: *The normalized spherical harmonic coefficients at each frequency from CIPIC database. The degree is set to 10. (a) Subject 003, (b) Subject 008.*

To show this problem, the values of the normalized coefficients with the degree of 10 for each frequency bin are respectively drawn in Fig. 1 for subject 003 and 008 from CIPIC database [17]. In this figure, we use the normalized coefficients to reduce the fluctuation range, which is calculated as

$$\hat{C}_n^m(f_i) = \frac{C_n^m(f_i) - \mu(f_i)}{\sigma(f_i)}, \quad (4)$$

where $\mu(f_i)$ and $\sigma^2(f_i)$ respectively denote the mean and variance of the coefficients at the frequency f_i. From Fig. 1, it can be observed that: 1) the number of the significant coefficients non-monotonically varies with subbands, and thus the methods in Section 2 cannot be expected to achieve a best performance; 2) the position of the significant coefficients is different for each subband, which infers that the traditional truncation to the degree will result in redundance of the coefficients or underfitting; 3) the number of the significant coefficients varies with subjects, and thus an optimal model should follow this property. Based on these observations, it is concluded that it is quite challenging to adaptively preserve the coefficients for frequency bins from different subjects.

A heuristic method is used for SHM to set the coefficients with small values to zero, and keep the remaining coefficients unchanged. However, the method will definitely increase the log-spectral distortion (LSD). In this paper, we present SSHM, an adaptive degree modeling method for HRTFs. Instead of assigning the degree for each frequency bin of the subject, SSHM uses sparse representation to adaptively select the significant coefficients according to HRTFs for different subjects and frequency bins. Since SSHM automatically generates sparse coefficients based on database, it inherently satisfies the three observations, and gives a more efficient representation of HRTFs.

3.2. HRTF preprocessing

Human is not sensitive to the fine details of the phase spectrum of HRTFs in localization [9] and discrimination perception [18]. Therefore, the minimum phase HRTFs and interaural time delay (ITD) can well approximate HRTFs [19]. Besides, since the phase part of the min-phase HRTFs can be obtained by Hilbert transform of the its magnitude as

$$|H_{\min}(\theta,\phi,f)| = |H(\theta,\phi,f)|, \tag{5}$$

$$\varphi_{\min}(\theta,\phi,f) = -\frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{\ln|H_{\min}(\theta,\phi,f)|}{f-\xi}d\xi, \tag{6}$$

the magnitude of the min-phase HRTFs and ITD are sufficient to model HRTFs.

In SSHM, we use the logarithmic magnitude of HRTFs because it is more approaching human's auditory perception, which is experimentally verified in [12] by comparing with the complex HRTFs, and HRTF magnitudes. Prior to SSHM, the average log-magnitude spectrum across all locations is calculated and subtracted from each sample for each frequency bin to create directional spectra, which is expressed as

$$H_{avg}(f_i) = \sum\nolimits_{j=1}^{S} 20\log_{10}|H_{\min}(d_j,f_i)|, \tag{7}$$

$$H_p(d_s,f_i) = 20\log_{10}|H_{\min}(d_s,f_i)| - H_{avg}(f_i), \tag{8}$$

where $d_s = (\theta_s,\phi_s) \in \mathbb{D}$ denotes the s-th sampling point of the direction set $\mathbb{D}$. S is the total number of the directions.

Since the averages include spatial features shared by all HRTFs, the resulting log-magnitudes represent primarily frequency-dependent spatial effects. Along with ITD, they are used to model HRTFs by the proposed sparse spherical harmonic-based method.

3.3. Sparse representation for SHM

SSHM aims to adaptively select the coefficients by a sparse coefficient matrix. By discarding the insignificant coefficients, SSHM also reduces the influence of fluctuation caused by the movement of the subject and the processing error.

SSHM models the log-magnitude at each frequency bin for left and right ears and ITD over the space by using an l_1 penalty to achieve a sparse solution. Let $H_p^L(d_s,f_i)$ and $H_p^R(d_s,f_i)$ respectively denote the log-magnitude spectrum for the left and right ears after subtracting the average at the frequency f_i and the direction d_s. The problem for SSHM is formulated as

$$\min_{C_n^i}\left(\sum_{s=1}^{S}\left(R_s^i - \sum_{n=1}^{N_m} C_n^i Y_n(d_s)\right)^2 + \lambda_i\sum_{n=1}^{N_m}|C_n^i|\right), \tag{9}$$

$$\text{s.t.}, R_s^i = \begin{cases} T(d_s), & \text{for } i=0 \\ H_p^L(d_s,f_i), & \text{for } i\in[1,L] \\ H_p^R(d_s,f_{i-L}), & \text{for } i\in[L+1,2L] \end{cases}, \tag{10}$$

where L is the unique number of the log-magnitudes of HRTFs. Since the HRIRs are real, HRTFs have conjugate symmetry. By considering the log-magnitude, $L = Q/2 + 1$ coefficients are unique for each ear with the order of discrete fourier transform (DFT) Q. $N_m = (N_0+1)^2$ and N_0 is the allowable maximum degree. The N_m basis functions are $\mathbf{Y}(d_s) = [Y_1(d_s),\cdots,Y_{N_m}(d_s)]^T = [Y_0^0(d_s),Y_1^{-1}(d_s),Y_1^0(d_s),Y_1^1(d_s),\cdots,Y_{N_0}^{N_0}(d_s)]^T$. C_n^i are the sparse coefficients. The shrinking parameter λ_i controls the sparsity level of the model.

Eq.(9) is a convex optimization problem. We choose least absolute shrinkage and selection operator (LASSO) to solve it, 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[20][21]. To prevent overfitting, the K-fold cross-validation approach is used to train λ_i [22]. Here, we use tenfold, and λ_i is chosen with the minimum cross-validation error. As a result, the coefficients are obtained with nonzero terms at the position set of $\mathbb{P}_i$ for ITD and frequency bins of each ear.

3.4. HRTF reconstruction and continuous interpolation

Given any arbitrary direction, HRTFs can be generated by using the SH extension coefficients. Because of the sparsity of SSHM, the reconstruction or interpolation can be processed faster.

There are three steps to obtain the HRTFs for two ears given a direction $d_s = (\theta_s,\phi_s)$. First, the spherical harmonic basis functions are calculated by using (1), which can be prepared outline in advance. Then, the log-magnitude of the min-phase HRTFs (for $i\in[1,2L]$) and ITD (for $i=0$) can be estimated by

$$\hat{R}_s^i = \sum\nolimits_{l\in\mathbb{P}_i} C_l^i Y_l(d_s),\ i\in[0,2L], \tag{11}$$

which will be used to obtain the min-phase HRTFs $\hat{H}_{\min}(d_s,f_i)$ by using (5)-(8) and (10). Finally, the HRTFs for the two ears are approximated as

$$\hat{H}^L(d_s,f_i) = \hat{H}_{\min}(d_s,f_i)e^{-j2\pi f_i(T_0+\hat{T}(d_s))}, \tag{12}$$

$$\hat{H}^R(d_s,f_i) = \hat{H}_{\min}(d_s,f_{i+L})e^{-j2\pi f_i T_0}, \tag{13}$$

where $i = 1,...,L$, and T_0 is the propagation delay from the sound source to the right ear, which can be estimated by r/c with the path distance of r.

4. Performance evaluation

In this section, the performance of the proposed SSHM is evaluated. CIPIC database is used for this purpose [17]. The database is obtained from 43 subjects. Each ear of each subject is measured with 25 azimuths and 50 elevations at a distance of 1m. Each head-related impulse response (HRIR) has been windowed in about 4.5ms (200 points). Prior to SSHM, all the HRIRs first are converted to the HRTFs by using 256-DFT, and the min-phase HRTFs after subtracting the average are then calculated by (5). There are the total of 259 parameters required to be modeled for each direction.

For objective evaluation, the log-spectral distortion (LSD) between the estimated and the measured HRTFs is applied. Let $\mathbb{D}$ denote the set with S directions, and LSD is defined as

$$\text{LSD} = \sqrt{\frac{1}{SN_f}\sum_{d\in\mathbb{D}}\sum_{k=k_1}^{k_2}\left(20\log_{10}\frac{|H(d,f_k)|}{|\hat{H}(d,f_k)|}\right)^2}, \tag{14}$$

where k_1 and k_2 respectively denote the beginning and the end of the considered frequency bins, and thus $N_f = k_2 - k_1 + 1$.

In order to evaluate the relative efficiency of SSHM to SHM, a relative reduction of coefficients is defined as

$$P_r = 1 - \frac{\sum_{i=0}^{2L} M_i}{N_0(2L+1)}, \tag{15}$$

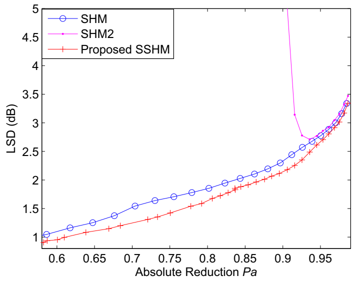


Figure 2: *The performance of SSHM between efficiency and accuracy compared with SHM and SHM2.*

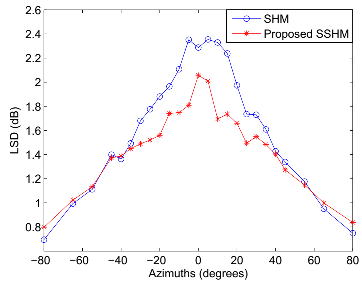


Figure 3: *The interpolation performance of SSHM compared with SHM with the degree of 20 in the horizontal plane.*

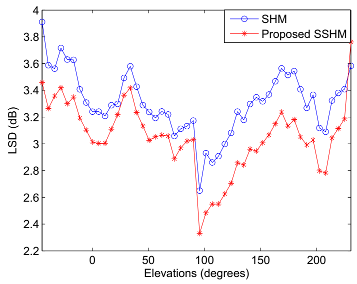


Figure 4: *The interpolation performance of SSHM compared with SHM with the degree of 20 in the median plane.*

where M_i denotes the number in the position set $\mathbb{P}_i$. Furthermore, an absolute reduction of coefficients P_a is used to evaluate the efficiency of different methods, which is defined as the ratio of the reduced number of coefficients to the number stored in CIPIC database.

4.1. Reconstruction performance

First, the reconstruction performance of SSHM is evaluated by comparing with other three methods: constant degree, and frequency-dependent truncation methods shown in [13] and [14] in terms of LSD, P_a and P_r. Two frequency bands are considered: audible band between 50Hz and 8kHz, and full band between 20Hz and 22.05kHz. The results are shown in Table 1 where SSHM-X denotes that the allowable maximum degree is X. First, for the audible band, it can be seen that on the same level of efficiency, SSHM achieves 0.09dB, 0.077dB gains over constant degree and [14], respectively. The LSD performance of [13] is dramatically worse than others, because of its too few coefficients at the low frequencies. For the full band, SSHM requires about 7% less number of the coefficients than the constant method at the same LSD of about 0.83dB. Besides, SSHM is significantly better than [13].

Table 1: *Reconstructed LSD comparison of SSHM with constant degree, [13] and [14] for subject 008 from CIPIC database.*

Band (Hz)	Methods	LSD (dB)	P_a	P_r
audible band (50-8000)	$N = 28$	0.401	0.848	–
	[13]	1.336	0.963	–
	[14]	0.388	0.849	–
	SSHM-40	0.311	0.848	0.499
full band (20-22050)	$N = 28$	0.831	0.568	–
	[13]	0.939	0.448	–
	[14]	0.719	0.537	–
	SSHM-40	0.866	0.636	0.579
	SSHM-50	0.830	0.637	0.728

Moreover, the tradeoff performance of SSHM between efficiency and accuracy is evaluated with a heuristic sparse method, which reduces the number of the non-zero coefficients on SHM by setting the small values to zero, such that the number of the remaining non-zero coefficients is equal to that of SSHM. It is used to evaluate the performance of SSHM under the same sparse level with SHM, and we refer it as SHM2. The results are shown in Fig. 2, where N_0 from the right to left is 4 to 34 with a step of 1. In this figure, it is clear that SSHM performs better than SHM and SHM2 in all conditions. Besides, SHM2 shows a worst LSD performance and is even diverged for a larger degree, because it discards much information. The experiments infer that SSHM can obtain a better tradeoff between the efficiency and the accuracy.

4.2. Interpolation performance

Then, the performance of SSHM on interpolation is evaluated in the horizontal plane and the median plane by comparing with SHM. The results are shown in Fig. 3 and Fig. 4, respectively.

Fig. 3 shows that SSHM achieves a significant improvement near the front of the horizontal plane. The average interpolation LSD of SHM over the 25 directions of the horizontal plane is 3.44dB. By using 38.17% less coefficients compared with SHM, the average LSD of SSHM is only 3.18dB. Furthermore, from Fig. 4, it can be seen that by reducing the number of coefficients by 39.32%, the average LSD of SSHM is 1.53dB over the median plane, compared with SHM of 1.75dB. Thus, though using less coefficients, SSHM can achieve better interpolation performance than SHM. The reason behind this is its great robustness to fluctuation. Furthermore, as we can see, the interpolation performance in the horizontal plane is much better than that in the median plane. One posssible reason is that the quite large fluctuation, especially above the ears and below the knees, takes more errors during the HRTF measurements, and thus the interpolated HRTFs are not completely matched to the measured ones.

5. Conclusions

In this paper, a sparse spherical harmonic-based model termed SSHM is proposed. SSHM uses a sparse linear combination to adaptively preserve the significant coefficients in each subband independent of the subject, resulting in more efficient representation. The experiments show that SSHM requires less coefficients at the same level of LSD compared with other frequency truncation methods. Furthermore, the interpolation performance shows that even using about 40% less coefficients, SSHM has a slightly lower LSD than SHM. SSHM can achieve a better tradeoff between the efficiency and the accuracy.

6. Acknowledgements

This work is supported by the National High-Tech Research and Development Program of China (863 Program) (No.2015AA016305), the National Natural Science Foundation of China (NSFC) (No.61425017, No.61403386, No.61305003), the Major Program for the National Social Science Fund of China (13&ZD189), and the Strategic Priority Research Program of the CAS (Grant XDB02080006).

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] 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.

[3] 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.

[4] 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.

[5] L. S. Davis, R. Duraiswami, E. Grassi, N. A. Gumerov, Z. Li, and D. N. Zotkin, “High order spatial audio capture and its binaural head-tracked playback over headphones with hrtf cues,” in *Audio Engineering Society Convention 119*. Audio Engineering Society, 2005.

[6] W. Zhang, M. Q. Zhang, R. A. Kennedy, and T. D. Abhayapala, “On high-resolution head-related transfer function measurements: An efficient sampling scheme,” *IEEE Transactions on Audio Speech and Language Processing*, vol. 20, no. 2, pp. 575–584, 2012.

[7] A. P. Bates, Z. Khalid, and R. A. Kennedy, “Novel sampling scheme on the sphere for head-related transfer function measurements,” *IEEE-ACM Transactions on Audio Speech and Language Processing*, vol. 23, no. 6, pp. 1068–1081, 2015.

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

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] J. Zhang and D. X. Mao, “Dependence of binaural loudness summation on interaural level difference and frequency for pure tones,” *Science China-Physics Mechanics & Astronomy*, vol. 53, no. 5, pp. 834–841, 2010.

[14] M. Q. Zhang, R. A. Kennedy, and T. D. Abhayapala, “Empirical determination of frequency representation in spherical harmonics-based hrtf functional modeling,” *IEEE-ACM Transactions on Audio Speech and Language Processing*, vol. 23, no. 2, pp. 351–360, 2015.

[15] 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.

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

[17] 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.

[18] 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.

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

[20] M. R. Osborne, B. Presnell, and B. A. Turlach, “On the lasso and its dual,” *Journal of Computational and Graphical statistics*, vol. 9, no. 2, pp. 319–337, 2000.

[21] J. Friedman, T. Hastie, and R. Tibshirani, “Regularization paths for generalized linear models via coordinate descent,” *Journal of statistical software*, vol. 33, no. 1, p. 1, 2010.

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