Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Psychological and Physiological Acoustics
Session 1pPPa: Binaural Hearing and Binaural Techniques I

1pPPa4. Towards optimal functional representation of head-related transfer functions in the horizontal plane

Wen Zhang*, Thushara D. Abhayapala, Rodney A. Kennedy and Mengqiu Zhang

***Corresponding author's address: Research School of Engineering, College of Engineering and Computer Science, Australian National University, Canberra, 0200, ACT, Australia, wen.zhang@anu.edu.au**

Head-related transfer function (HRTF) individualization using principle component analysis (PCA) modelling rely on the empirical data to reduce HRTF dimensionality for an optimal representation and to achieve HRTF personalization by tuning the model weights with the subject anthropometric parameters. However, for these representations, the basis is discrete and data dependent which can limit its usefulness in universal HRTF representation. This paper studies the optimal functional representation of magnitude HRTF of 45 subjects for sound sources in the horizontal plane. We firstly use circular harmonics to extract the subject-independent HRTF angular dependence. The remaining spectral components of 45 subjects are then modelled by PCA and two standard functions, i.e., Fourier series and Fourier Bessel series. The metric to evaluate the model efficiency is the expansion weights cumulative variance. We identify that individual magnitude HRTFs over 20 kHz range could be modelled adequately well by a linear combination of only 19 Fourier Bessel series; this is a near optimal representation in comparison with the statistical PCA model. Further analysis of the model weights with subjective anthropometric measurements will provide a promising method for HRTF individualization, especially considering the nature of data independent continuous basis functions employed in the proposed functional representation.

Published by the Acoustical Society of America through the American Institute of Physics

Received 21 Jan 2013; published 2 Jun 2013

INTRODUCTION

HRTFs are usually obtained from measurements on people (or a manikin) and so there is a rich variety of HRTFs corresponding to the different human subjects being measured. Statistical methods by means of principle component analysis (PCA) have been used for HRTF magnitude spectra decomposition to result in a small set of basis functions and position dependent weights [1, 2]. Studies [3, 4] show that only 10−20 principle components seem to be able to give satisfactory good results in terms of HRTF data modelling and its usage for binaural localization.

These optimal discrete methods rely on the empirical data and provide an optimal description for any given lower dimensional model. However, their dependence on the empirical data is a weakness as well as a strength. The clear strength is that the representation deals directly with the set of measurements and reality rather than an abstraction through a model which may be inaccurate. The weakness is that optimal discrete models only represent the given empirical data set and any changes to that data set change the model. That is, the principle components, which are basis vectors, will vary with any change to the data set (additions or omissions). Of course, if a data set is sufficiently large and rich, in the sense of capturing the true variance across the population, the basis functions may be fixed without much loss of optimality [5]. Further, because the data is discrete and taken from a measurement grid, the representation is not universal in the sense that HRTF measurements taken with differing measurement grids cannot be merged directly whereas the underlying continuous HRTFs should be consistent and is independent of how it is measured.

A method for HRTF continuous representation in the spatial domain is to express HRTFs as a weighted sum of Fourier series [6, 7] (for HRTF azimuth variation only) or spherical harmonics [8, 9] (for HRTF azimuth and elevation variations). The models have been verified for its accurate representation of the original measurements and interpolation results. The spectral components embedded in the expansion coefficients carry the information regarding the individuality. Orthogonal series based on bessel functions have been proposed for HRTF spectral representation with a satisfactory level of performance in terms of average mean-square-error (MSE) of each individual data set modelling [7, 9]. However, whether these orthogonal basis possesses the fastest convergence property to represent HRTF spectral variations has not been investigated.

In this work, we investigate optimal functional modelling of horizontal plane magnitude HRTFs based on the CIPIC database of 45 subjects [10]. As the minimum phase approximation of the HRTF works extremely well in practical spatial audio, implementation of modelling magnitude HRTFs have great utility and practical value. The purpose here is to seek continuous basis functions which can determine the HRTF spectra of different listening subjects with fewest parameters so that the most efficient HRTF data reduction can be achieved. The metric to evaluate the model efficiency is the cumulative variance of the expansion coefficients. We identify that individual magnitude HRTF spectral components over 20 kHz range could be represented by a linear combination of only 19 Fourier Bessel series; this is a near optimal representation in comparison with the statistical PCA model.

MAGNITUDE HRTF AZIMUTH MODELLING

Using a Fourier series expansion, magnitude HRTFs $H_{\text{mag}}(f,\phi)$ as a periodic function of the azimuth angle ϕ can be expanded in the following form

$$H_{\text{mag}}(f,\phi) = \sum_{m=-\infty}^{\infty} A_m(f) e^{im\phi}, \tag{1}$$

where $i = \sqrt{-1}$ and the m^{th} order Fourier series weight is given by

$$A_m(f) = \frac{1}{2\pi}\int_0^{2\pi} H_{\mathrm{mag}}(f,\phi)e^{-im\phi}\,d\phi. \tag{2}$$

The work of [6] shows that the horizontal plane HRTFs have a low pass character with limited spatial bandwidth. Most of energy in $A_m(f)$ is restricted in a region which is limited by $|m| \le 2\pi f a/c$, where c denotes the sound propagation velocity and a is the radius of the head (i.e., $c = 340$ m/s and $a = 0.09$ m). Thus, the infinite series (1) can be truncated to $|m| \le M$ to represent a band-limited magnitude HRTF function through

$$H^M_{\mathrm{mag}}(f,\phi) \approx \sum_{m=-M}^{M} A_m(f)e^{im\phi}. \tag{3}$$

As a guide, based on [6], a suitable M can be determined through

$$M = \lceil 2\pi a f_{\mathrm{max}}/c \rceil, \tag{4}$$

where $\lceil\cdot\rceil$ is the integer ceiling function, and f_{max} is the maximum frequency. For example, for the audible frequency range of 20 kHz, $M = 34$.

Given HRTF measurements of V_0 azimuthal angles, (3) can be represented in a matrix form, $\boldsymbol{H} = \boldsymbol{AE}$. Then, the coefficient matrix $\boldsymbol{A}$ can be estimated as a least square solution to this matrix equation, i.e., $\boldsymbol{A} = \boldsymbol{HE}^\dagger$, where $(\cdot)^\dagger$ represents the pseudo inverse operation. As the basis functions are continuous in the spatial azimuth domain, the model provides a natural method to obtain interpolation results for angles where no measurements are made.

MAGNITUDE HRTF SPECTRAL MODELLING

The goal is to find which closed form standard orthogonal functions match the experimentally measured HRTFs distribution most efficiently where efficiency is typically measured in terms of variance. We initially formulate a general orthogonal representation of the magnitude HRTF spectral components $A_m(f)$ for each m defined over the finite interval $(0, f_{\mathrm{max}})$. Unlike the expansion for the azimuthal variable there is no obviously compelling choice for the preferred orthogonal representation so this is a crucial issue in formulating our representation. We write

$$A_m(f) = \sum_{\ell=1}^{\infty} C_{m;\ell}\,\varphi_\ell(f), \tag{5}$$

where, $\varphi_\ell(f)$ is a suitable complete set of orthonormal functions defined on the finite interval $(0, f_{\mathrm{max}})$ and indexed by ℓ. Theoretically any square-integrable function can be represented by (5) [11], and arbitrarily well approximated by truncating to a finite number of terms L, i.e.,

$$A_m(f) \approxeq \sum_{\ell=1}^{L} C_{m;\ell}\varphi_\ell(f). \tag{6}$$

The order L is usually chosen as a tradeoff between accuracy and economy of representation, noting that the approximation can be made arbitrarily close by choosing L sufficiently large.

The expansion coefficients $C_{m;\ell}$ in (5) can be obtained from

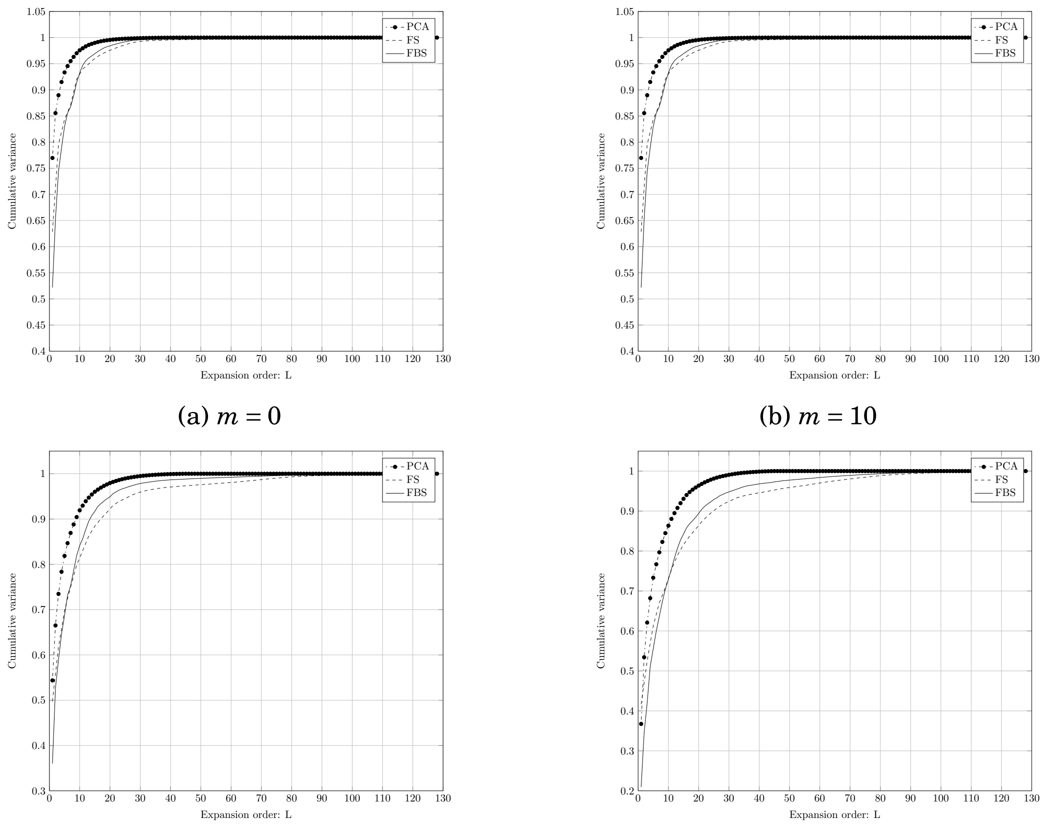
$$C_{m;\ell} = \int_0^{f_{\mathrm{max}}} A_m(f)\overline{\varphi_\ell(f)}W(f)df, \tag{7}$$

where $W(f)$ is the prescribed (non-negative) weighting function to make $\{\psi_\ell(f) = \sqrt{W(f)}\varphi_\ell(f)\}$ become orthonormal set in the specified region $(0, f_{\mathrm{max}})$.

TABLE 1: Candidate closed form orthogonal functions for magnitude HRTF spectral modelling.

Orthogonal Functions (general form)	Interval	Modified Basis $\varphi_\ell(f)$	Weight $W(f)$
Fourier Series $e^{i2\pi\ell x}$	$x \in [0,1]$	$\frac{1}{\sqrt{f_{\max}}} e^{i2\pi\ell f/f_{\max}}$	1
Fourier Bessel series of zero-th order $J_0(Z_\ell^{(0)}x)^1$	$x \in (0,1)$	$\frac{\sqrt{2}}{\sqrt{f_{\max}} J_1(Z_\ell^{(0)})} J_0(Z_\ell^{(0)} f/f_{\max})$	$\frac{f}{f_{\max}}$

$* \; Z_\ell^{(0)}$–ℓ-th zero of Bessel function of order 0.



(a) $m = 0$ (b) $m = 10$

(c) $m = 20$ (d) $m = 30$

FIGURE 1: The cumulative variance of the expansion coefficients.

Table 1 gives an example of two standard orthogonal bases with $W_\ell(f)$ defined for the HRTF spectral modelling purpose. Both Fourier series (FS) and Fourier Bessel series (FBS) have been proven sufficient in modelling complex HRTF spectra based on the average mean-square-error (MSE) performance in individual HRTF data set representation. However, which orthogonal basis possesses the fastest convergence property to represent the HRTF spectral variations has not been investigated. In this work, we treat each individual HRTF as a realization of a random process and use the cumulative variance of the expansion coefficients as the evaluation metric [12],

$$\Delta_{L;\varphi}^{(m)} = \frac{E\{\sum_{\ell=1}^{L} |C_{m;\ell}|^2\}}{E\{\sum_{\ell=1}^{L_0} |C_{m;\ell}|^2\}}.$$

$E\{\cdot\}$ represents the expectation operator performed on different persons HRTFs and L_0 denotes the maximum expansion order (the number of HRTF frequency samples). Notice that here we evaluate the efficiency for each expansion order m.

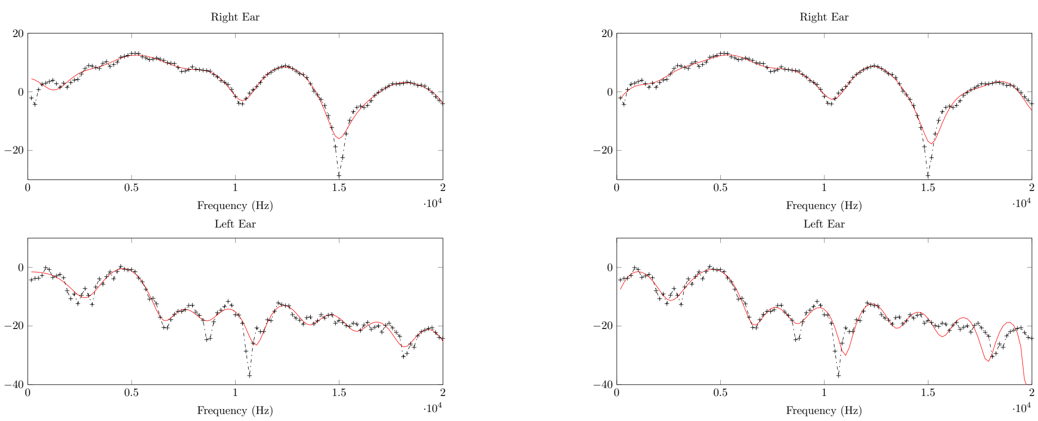
Figure 1 compares the efficiency of the candidate basis functions with the results from the PCA analysis performed on CIPIC database of 45 subjects (43 listeners and 2 manikin data). It can be seen that the cumulative variance of the decomposed coefficients from the Fourier Bessel series expansion becomes closer to that from the statistical PCA analysis results as the expansion order m gets larger. Even though our previous study shows that both Fourier series and Fourier Bessel series have comparable average MSE performance in modelling individual HRTF data set. The methodology adopted here proves the Fourier Bessel series are near-optimal continuous basis functions in terms of modelling horizontal plane HRTF magnitude spectra of different listeners. In the following, we will show that only 19 Fourier Bessel series (including more than 90% cumulative variance) can represent magnitude HRTF spectra within a bandwidth of 20 kHz.

MODAL VALIDATION

The relative mean square error (MSE) over audible frequency range of $[0.2, 20]$ kHz is used as the error metric

$$\varepsilon(\phi) = \frac{\sum_{j=1}^{N} |H_{\text{mag}}(f_j,\phi) - \widetilde{H}_{\text{mag}}(f_j,\phi)|^2}{\sum_{j=1}^{N} |H_{\text{mag}}(f_j,\phi)|^2}, \tag{8}$$

where $H_{\text{mag}}(f_j,\phi)$ and $\widetilde{H}_{\text{mag}}(f_j,\phi)$ are the original and reconstructed magnitude HRTFs, respectively.



(a) FBS model (b) FS model

FIGURE 2: Examples of measured HRTFs reconstructions using continuous orthogonal functions as bases. Original: + and reconstruction: line.

Figure 2 plots the original and reconstructed HRTF magnitude for CIPIC subject 003 ($\phi = 280°$). The reconstructions are obtained from 19 Fourier Bessel series or Fourier series ($L = -9, \ldots, 9$) to represent the whole magnitude HRTF spectra of range between $[0.2, 20]$ kHz. Both reconstructed results are smoothed form of HRTF magnitude spectra with capability to capture all the major peaks and troughs. Obviously, the model using Fourier Bessel series as bases performs better for representing contralateral sound (HRTFs at the head-shadowed side) at frequencies above 15 kHz.

The errors in fitting the HRTF measurements are further analyzed at each of the measured positions. The distribution of errors across all positions are presented in Fig. 3. Note that in

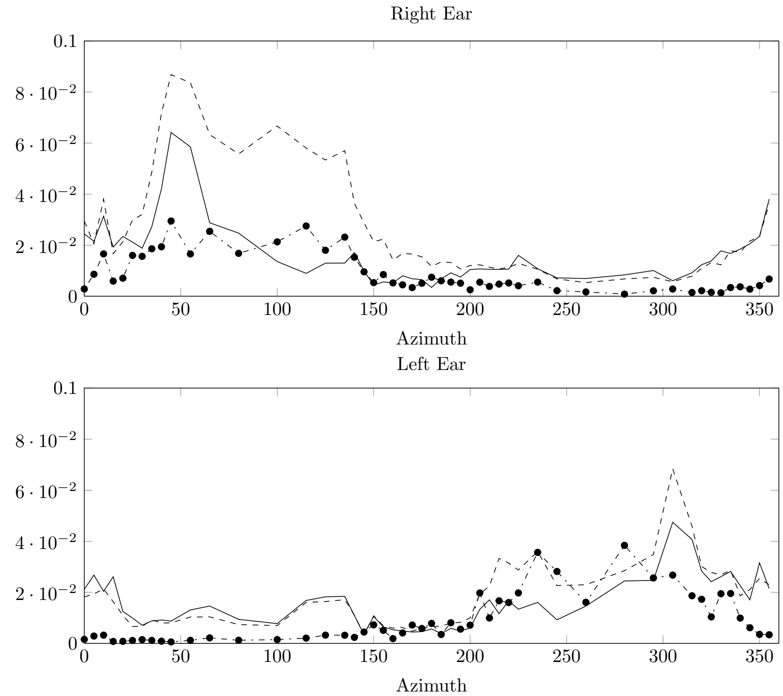


FIGURE 3: The reconstruction error distribution as a function of the source position (azimuthal angle ϕ). PCA: dotted line, FS: dashed line and FBS: solid line.

current data sets, a source located at 90° azimuth is directly across from the left ear (the right ear is the shadowed ear) and a source located at 270° is directly across from the right ear (the left ear is the shadowed ear). The calculated error in the model fitting leads to two observations. First, as expected, the statistical PCA model provides the best match to measurements while the continuous Fourier Bessel series model has slightly larger reconstruction errors. Models using Fourier series as basis functions performs the worst. With mean error less than 0.02, we expect the synthesis using the proposed continuous FBS model can provide satisfactory good results. Second, in all three models, the synthesis of HRTFs is better at the source-facing side of the head compared to the head's shadowed side. This is due to the fact of the comparative poorer signal-to-noise ratio (SNR) at the head shadowed side. The signal level at the head shadowed side is lower than that at the source facing side, resulting in the signal to noise ratio in the measurement is relatively poorer at the head shadowed locations. In addition, because of the diffraction around the head, the contralateral sounds have more variations. This results in the spectral shapes that are more complicated and more difficult to model.

CONCLUSION

This paper studies the optimal functional representation of magnitude HRTF of different listening subjects for sound sources in the horizontal plane. Fourier series were used to model the spatial azimuth dependence of the HRTF. The remaining spectral components are then modelled by two standard functions, i.e., Fourier series and Fourier Bessel series. The metric to evaluate the model efficiency is the expansion weights cumulative variance. By using PCA analysis results as the baseline, we identify the near optimal representation is to use a linear combination of 19 Fourier Bessel series to model magnitude HRTF spectral components over audible frequency range of 20 kHz. Further work will be analysis of the model weights with subjective anthropometric measurements to achieve HRTF individualization.

REFERENCES

[1] D. J. Kistler and F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction", Journal of the Acoustical Society of America **91**, 1637–1647 (1992).

[2] J. Breebaarta, "Effect of perceptually irrelevant variance in head-related transfer functions on principal component analysis", Journal of the Acoustical Society of America Express Letters **133**, EL1–EL6 (2013).

[3] J. Leung and C. Carlile, "PCA compression of HRTFs and localization performance", in *Proceedings of the International Workshop on the Principles and Applications of Spatial Hearing*, 3 pages (Sendai, Japan) (2009).

[4] Hugeng, W. Wahab, and D. Gunawan, "Improved method for individualization of head-related transfer functions on horizontal plane using reduced number of anthropometric measurements", Journal of Telecommunications **2**, 31–41 (2010).

[5] C. Jin, P. Leong, A. Corderoy, and S. Carlile, "Enabling individualized virtual auditory space using morphological measurements", in *Proceedings of the IEEE 2000 International Symposium on Multimedia Information Processing*, 235–238 (Sydney, Australia) (2000).

[6] T. Ajdler, C. Faller, L. Sbaiz, and M. Vetterli, "Sound field analysis along a circle and its applications to HRTF interpolation", Journal of the Audio Engineering Society **56**, 156–175 (2008).

[7] W. Zhang, R. Kennedy, and T. Abhayapala, "Efficient continuous HRTF model using data independent basis functions: Experimentally guided approach", IEEE Transactions on Acoustics Speech and Signal Processing **17**, 819–829 (2009).

[8] R. Duraiswami, D. N. Zotkin, and N. A. Gumerov, "Interpolation and range extrapolation of HRTFs", in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume IV, 45–48 (Montreal, Quebec, Canada) (2004).

[9] W. Zhang, T. Abhayapala, R. Kennedy, and R. Duraiswami, "Insights into head-related transfer function: Spatial dimensionality and continuous representations", Journal of the Acoustical Society of America **127**, 2347–2357 (2011).

[10] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database", in *Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 99–102 (New Paltz, New York) (2001).

[11] L. E. Franks, *Signal Theory* (Prentice Hall) (1969).

[12] M. Zhang, R. A. Kennedy, and T. D. Abhayapala, "Efficiency evaluation and orthogonal basis determination in functional HRTF modeling", in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 53–56 (Prague, Czech Republic) (2011).