Horizontal plane HRTF reproduction using continuous Fourier-Bessel functions

Wen Zhang[1,2], Thushara D. Abhayapala[1,2], Rodney A. Kennedy[1]

[1]*Department of Information Engineering, Research School of Information Sciences and Engineering, The Australian National University, Canberra ACT 0200, Australia*

[2]*Wireless Signal Processing Program, Canberra Research Laboratory, NICTA, Canberra ACT 2612, Australia*

Correspondence should be addressed to Wen Zhang (wen@rsise.anu.edu.au)

ABSTRACT

This paper proposes a method to reproduce the Head Related Transfer Function (HRTF) in the horizontal auditory scene. The method is based on a separable representation which consists of a Fourier Bessel series expansion for the spectral components and a conventional Fourier series expansion for the spatial dependence. The proposed representation can be used to predict HRTFs at any azimuth position and at any frequency sampling point from a finite number of measurements. Implementation details are demonstrated in the paper. Measured HRTFs from a KEMAR manikin and analytically simulated HRTFs were used to validate the fidelity and predictive capabilities of the method. The average mean square error for model reconstruction is less than two percent.

1. INTRODUCTION

The ability of human beings to localize the sound in three dimensions depends on the way the sound waves from the same source differ from each other as they reach the left and right ears. The head, torso, shoulders and the outer ears modify the sound arriving at a person's ears. This modification can be described by a complex response function, the Head Related Transfer Function (HRTF), which depicts how a given sound wave input (parameterized as frequency and source location) is filtered by the diffraction and reflection properties of the individual body shape before the sound reaches the listener's eardrum. Theoretically, HRTFs can be used to generate binaural sound, a "virtual acoustic environment", as it contains all the information about the sound source's location (its direction and distance from the listener).

To synthesize the auditory scene, a straightforward way is to filter the original monaural sound through a proper set of measured HRTFs for each individual. However, to represent the entire auditory scene, one common approach is to have functional representations of the HRTF, such as filter bank models [1, 2] and transfer decompositions (principal component analysis [3, 4] and spherical harmonics [5]). As the weights of these models are available only for either the measured directions or sampled frequency points, the HRTF must be interpolated between two discrete measurement positions or discrete frequencies. Many techniques have been proposed now to perform the interpolation of the HRTF, such as the bilinear method [6], pole-zero approximation models [7] and spherical spline-based methods [8]. Nevertheless, the most appropriate interpolation is still considered as an open question. A more efficient approach is to have continuous functional representation of HRTFs.

This paper proposes a method to reproduce HRTFs at all possible positions in the horizontal plane based on a novel HRTF Fourier Bessel functional representation. The Fourier Bessel functional model uses the Fourier series to separate the spatial and spectral dependence of the HRTF. According to the strong correlation between the measured spectral structure of the HRTF and the family of Bessel functions of the first kind, we use the Fourier-Bessel series to represent the Fourier series weights (spectral components of HRTFs). Both Fourier series and Fourier-Bessel series are well-studied complete orthogonal functions and can be readily applied to accurately model the HRTF. Further, by applying these two sets of continuous functions, each individualized HRTF is transformed to a coefficient matrix, which could be very easily saved and processed to predict HRTFs at any arbitrary position in two dimensions. Given the

characteristics of the Fourier Bessel functional model, the reproduction implementation details are elaborated in the paper. Finally, the reproduction results of the proposed method are validated by comparing a KEMAR manikin measured [9] and analytically simulated HRTFs [10] with the corresponding reproduced HRTFs.

2. HRTF FOURIER BESSEL FUNCTIONAL MODEL

A continuous functional representation is a mathematical model or equation that represents the HRTF as a function of continuous variables (source positions and frequency points). The horizontal plane HRTF is denoted by $H(f,\phi)$ as a function of frequency f and azimuthal angle ϕ. In this section we develop spatial and spectral continuous representations of the HRTF.

2.1. HRTF spatial components modelling

One of the noticeable characteristics of the HRTF is that the function is periodic with period 2π in the azimuthal variable. A periodic function is most naturally expanded using a Fourier series, which makes use of the orthogonality between the complex exponentials to break up an arbitrary periodic function into a series of simple terms which converge in the mean. When truncated this infinite series can provide very good approximation to the original function given the HRTF has a low pass character in the azimuthal variable. Thus, the continuous Fourier series can be used as basis to extract the spatial dependence of HRTFs written as

$$H(f,\phi) = \sum_{m=-\infty}^{\infty} A_m(f) e^{im\phi}, \qquad (1)$$

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

$$A_m(f) = \frac{1}{2\pi} \int_0^{2\pi} H(f,\phi) e^{-im\phi} d\phi. \qquad (2)$$

Equations (1) and (2) allow the calculation of the Fourier series weights $A_m(f)$ given a continuous HRTF function. In a practical context we need to reconsider equations (1) and (2). Two important modifications are required. Firstly (1) should be modified by truncating to a certain order M depending on what accuracy is desired. Secondly (2) should be modified by replacing the integral with a finite summation to calculate coefficients $A_m(f)$ at discrete frequencies. As the solved Fourier series weights $A_m(f)$ are only available at some discrete frequencies, the problem of modelling $A_m(f)$ as a function of frequency is addressed below to achieve the goal of a continuous representation.

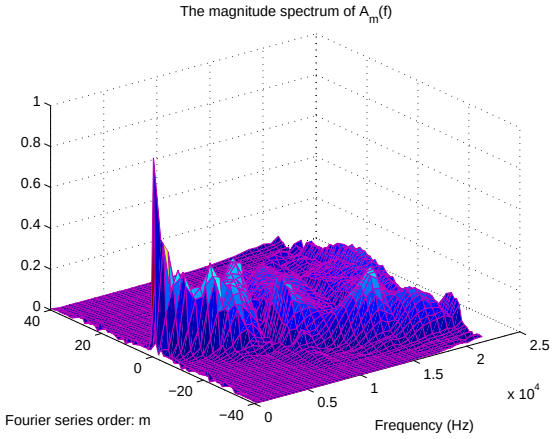

Fig. 1: The magnitude spectrum of $A_m(f)$ calculated from the MIT data [9].

2.2. HRTF spectral components modelling

Because the physical understanding of the HRTF variance with frequency remains an open issue, there is no obvious choice for the representation of the spectral components. A practically effective strategy is to use a representation consisting of a complete set of orthogonal functions which can be served as basis functions to any order of $A_m(f)$. The concept can be expressed as

$$A_m(f) = \sum_{k=1}^{\infty} C_{mk} \varphi_k(f), \qquad (3)$$

where, $\varphi_k(f)$ is a suitable orthogonal set of functions defined on the interval $f \in (0, f_{\max})$ ($f_{\max}$ are the maximum measurement frequency).

Theoretically any general function can be perfectly approximated by Eq. (3) without error; while for practical implementation, truncation of the equation to a finite number of terms, K, will bring in inaccuracies in the representation. In principle any orthogonal set of functions on a finite interval can be adapted for the purpose and provide an exact representation and in that sense are equivalent. However, under truncation to a certain number of terms some orthogonal sets will perform better

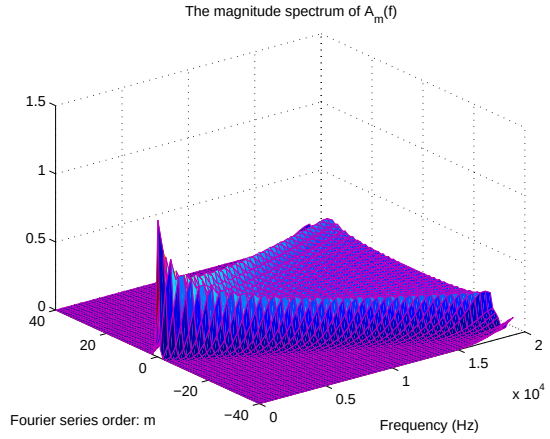

Fig. 2: The magnitude spectrum of $A_m(f)$ calculated from an analytical model [10].

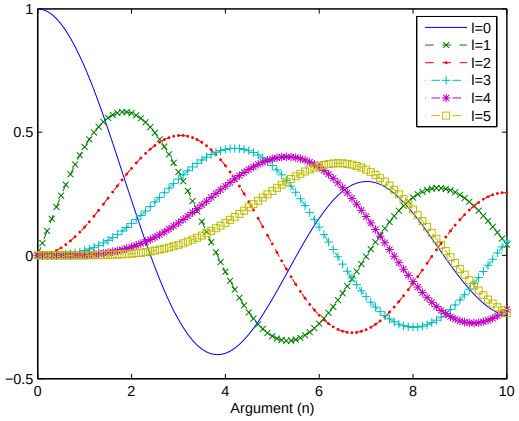

Fig. 3: Plot of the Bessel functions of the first kind $J_\ell(n)$ for various orders ℓ.

than others. That is, under truncation, different orthogonal sets of functions are not equivalent and there will exist preferred choices.

The aim of this work is to find the most efficient orthogonal sets under truncation based on the pattern of $A_m(f)$ (the most resemble one should give the lowest truncation number). Fig. 1 and 2 depict a mesh plot of the magnitude spectrum of $A_m(f)$ calculated from the MIT data [9] and a simple analytical model [10]. The figures reveal that the energy of the HRTF frequency components does not uniformly spread over the Fourier series expansion order m. Instead, the pattern of $A_m(f)$ from both data sets exhibits similarities to the family of Bessel functions of the first kind (Fig. 3), such as the finite value of $A_0(f)$ at the origin $f = 0$, the prominence of other $A_m(f)$s after the origin and proportionally decaying oscillation like sine or cosine functions. This resemblance demonstrates that there is a strong correlation between the spectral components of HRTFs and the family of Bessel functions of the first kind. Thus, in this paper, a complete orthogonal set, the Fourier Bessel series, is proposed to model the HRTFs frequency components $A_m(f)$ whose dependence on m is explained next.

The Fourier Bessel series make use of the orthogonality between Bessel functions of the first kind for a specific order ℓ on the interval (0,1) to expand a general function [11]. The Fourier Bessel series expansion of the frequency component of HRTFs is given as

$$A_m(f) = \sum_{k=1}^{\infty} C_{mk} J_\ell(\beta_k^{(\ell)} \frac{f}{f_{\max}}), \tag{4}$$

where $\beta_1^{(\ell)}, \beta_2^{(\ell)}, ..., \beta_k^{(\ell)}$ are the positive roots of $J_\ell(x) = 0$ and ℓ is the specific order of the Bessel function of the first kind. C_{mk} are complex coefficients whose dependence on the choice of ℓ is suppressed.

The coefficients of the Fourier Bessel series expansion can thereby be resolved from [11] as

$$C_{mk} = \frac{2}{[J_{\ell+1}(\beta_k^{(\ell)})]^2} \int_0^{f_{\max}} f A_m(f) J_\ell\left(\beta_k^{(\ell)} \frac{f}{f_{\max}}\right) df. \tag{5}$$

As the Fourier Bessel series has infinite number of sets, we need to decide which specific order ℓ to use. Firstly, the zero-th order $J_0(\cdot)$ should be included due to the dominance of $A_0(f)$; among all orders only $J_0(0) = 1$ could give very efficient approximation to $A_0(f)$ in the limit as $f \to 0$. However, using only one term $J_0(\cdot)$ is not the optimum representation for the high order $A_m(f)$ when $m \neq 0$. Noticing that high order Fourier Bessel series can match high order frequency components, we believe the representation with ℓ depending on m is more effective. In this paper, we have not mathematically derived a specific formula and a simple linear relationship $\ell = |m|$ is applied. The real measurements reconstruction performance in section 4 proves that this choice works well,

such as high approximation accuracy and significantly reduced number of parameters (the verification of this simple formula will be left as on open problem).

Thus, the HRTF functional model in the frequency domain is obtained as

$$H(f,\phi) = \sum_{m=-\infty}^{\infty} \sum_{k=1}^{\infty} C_{mk} J_{|m|}\left(\beta_k^{(|m|)} \frac{f}{f_{\max}}\right) e^{im\phi}, \quad (6)$$

where

$$C_{mk} = \frac{1}{\pi [J_{|m+1|}(\beta_k^{(|m|)})]^2} \int_0^{f_{\max}} \int_{-\pi}^{\pi} fH(f,\phi) J_{|m|}\left(\beta_k^{(|m|)} \frac{f}{f_{\max}}\right) e^{-im\phi} d\phi df. \quad (7)$$

Eq. (7) illustrates how to calculate the model parameters from a continuous HRTF function. While for experimentally measured HRTFs, the model coefficients C_{mk} are calculated using the left Riemann sum to approximate the integral. The implementation about appropriate truncation of (6) for efficient representation is illustrated in section 3.

Now, the HRTF representation (6) are functions of continuous variables representing the spectral and spatial cues. On the one hand, the proposed model can achieve HRTF reconstruction at any frequency point for an arbitrary azimuth. On the other hand, given the basis functions remain same for all listeners, for each individualized HRTF measurement, only a coefficient matrix C_{mk} needs to be saved. Note this coefficient matrix is much smaller in size compared to original measurements; hence the goal of data compression is achieved.

3. HRTF REPRODUCTION IMPLEMENTATION

In this section, issues about how to implement the proposed HRTF model (6) are elaborated including the limitation on experimental measurement.

3.1. Angular Sampling

The all-important implementation issue is to record the plane sound field with enough samples at the desired fidelity. Ajdler in [12] proves that the appropriate way to sample HRTFs is through

$$w_t = \pm \frac{c\ell_\theta}{0.09}, \quad (8)$$

where c is the speed of sound propagation, ℓ_θ and w_t are the angular pulsation and temporal sampling frequency. ℓ_θ as defined above depends on the angular sampling interval. For example, under uniformly spaced sampling (the most widely used sampling method for HRTF measurement), $\ell_\theta = 2\pi/\triangle\theta$ and $\triangle\theta$ is the angular spacing between two consecutive HRTF measurement positions.

Further, according to (8), in order to avoid spatial aliasing in HRTF reproduction, there is a maximum azimuthal angular spacing $\triangle\theta_{\max}$ beyond which fidelity is lost. The $\triangle\theta_{\max}$ is related to the temporal sampling frequency w_t. When the sampling frequency is 44.1 kHz (as used in MIT data set), the maximum angular sample interval is approximate 5 degrees.

In summary, the proposed HRTF reproduction method requires the sound field (HRTF) to be sampled uniformly on the horizontal plane and the angular sample interval no more than $\triangle\theta_{\max}$ for a specific temporal sampling frequency following (8).

3.2. Choice of Truncation Number

The primary parameters that influence the model fitting are the truncation numbers M and K in (6). The general rule is to monitor the distribution of model parameters C_{mk} over the Fourier series order m and the Fourier Bessel series order k for each specific data set, like Fig. 4 from analytical model and Fig. 5 from MIT data. The structure of the coefficients has a clear downward trend in the average contribution made by Fourier series harmonics and Fourier Bessel series of increasing degree. Hence, a higher truncation number results in better approximation; but too large value may lead to over fitting. In addition, make sure the value of C_{mk} is solved over large orders. For example, the value of C_{mk} calculated from MIT data is nearly zero when $M > 20$ and $K > 100$, which proves that our calculation order is adequate. Then, the criteria is that at least 90% of the total "energy" of C_{mk} is contained in the approximation. In this paper, for both analytical solutions and MIT measurements, the lowest truncation numbers are chosen as $M_L = 16$ and $K_L = 87$ accordingly.

Once the angular sampling interval and truncation numbers are determined, we could estimate the model parameters C_{mk} using (7) from the experimental measurements and further reproduce HRTFs in the whole horizontal auditory scene.

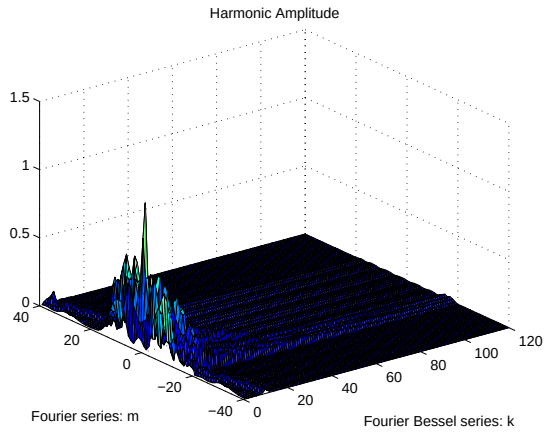

Fig. 4: The amplitude of model parameters C_{mk} (solved from analytical solutions) over Fourier series m and Fourier Bessel series k.

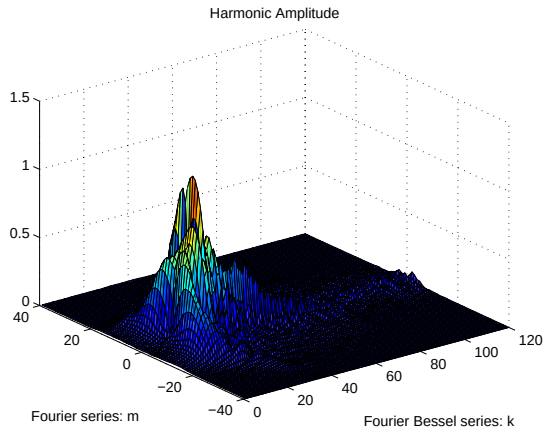

Fig. 5: The amplitude of model parameters C_{mk} (solved from MIT data) over Fourier series m and Fourier Bessel series k.

4. PRACTICAL RESULTS AND PERFORMANCE ASSESSMENT

The fidelity and predictive capabilities of our proposed horizontal planar HRTF reproduction method is validated by comparing the measured (or analytically simulated) and model reproduced data. Two sets of data are employed in evaluation: 1) the MIT data acquired using a KEMAR manikin [9], and 2) some analytical solutions [10].

The error metric is defined as the percent mean square error in the magnitude and phase spectrum at each azimuthal location

$$e_i = \frac{\sum_{n=1}^{N} \|H(f_n,\phi_i) - \hat{H}(f_n,\phi_i)\|^2}{\sum_{n=1}^{N} \|H(f_n,\phi_i)\|^2} \times 100\%, \quad (9)$$

where for each azimuth, HRTFs are measured (or simulated) at N frequency points. $H(f_n,\phi_i)$ is the original HRTFs at the n^{th} frequency point and i^{th} azimuth; and $\hat{H}(f_n,\phi_i)$ is the reproduced HRTF.

The error performance is investigated for two different classes of directions; those for which measurements (analytical simulations) were conducted and contributed toward the derivation of the model parameters (Reconstruction) and directions for which had no experimental reference (Interpolation).

4.1. Real Data

The proposed method is tested on the MIT measurements first. The MIT measurements were made in an anechoic chamber; and both the "small" and "large" pinna models were tested on the KEMAR DB 4004. The measurements are the head related impulse response in the time domain at the 44.1 kHz sampling rate and each response is 512 samples long. In the horizontal plane, a full 360 degree of azimuth was sampled in equal sized increments (5 degrees approximately).

In order to use one set of measurement to examine two kinds of error performance, 72 MIT measured HRTFs in the horizontal plane are divided into two groups with each having 36 data sets spaced by 10°. The first group is used to determine the model parameters C_{mk}; and the reconstruction performance is evaluated by comparing the measured and model reconstructed HRTFs at these locations. The second group data is used to demonstrate the predictive power of the model (interpolation performance) as these locations are not used to determine C_{mk}. Note that from (8), we can only interpolate the HRTFs

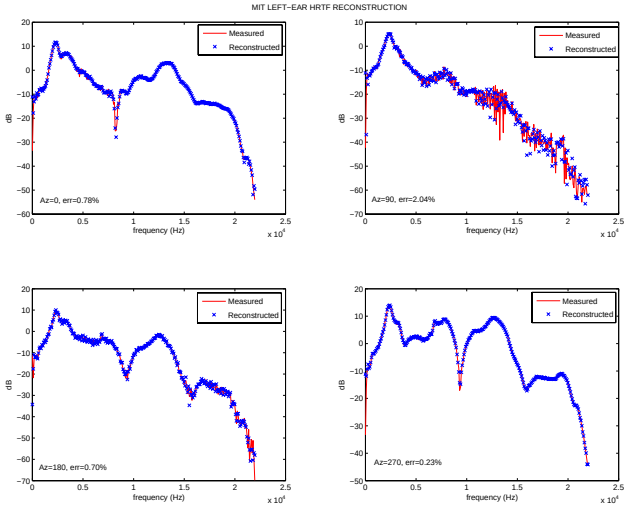

Fig. 6: Example MIT (left ear) measured and model reconstructed HRTFs using frequency domain model at $\phi = 0°, 90°, 180°$, and $270°$.

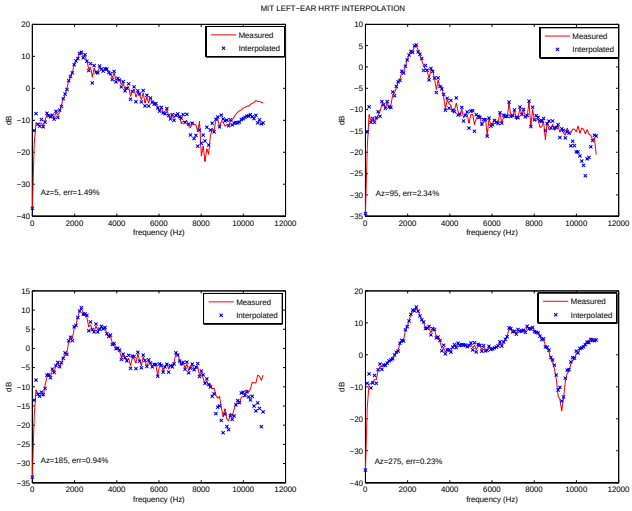

Fig. 7: Example of model interpolated MIT left ear HRTFs at $\phi = 5°, 95°, 185°, 275°$ with original measurements overlaid.

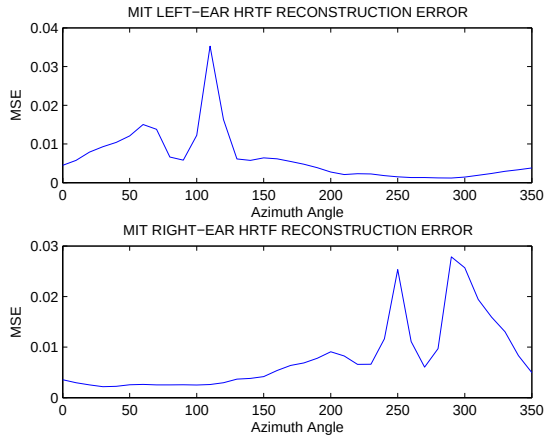

Fig. 8: Reconstruction error distributions as a function of the source position azimuthal angle ϕ for the MIT HRTFs. The error reaches its maximum for source locations where the ear is maximally shadowed.

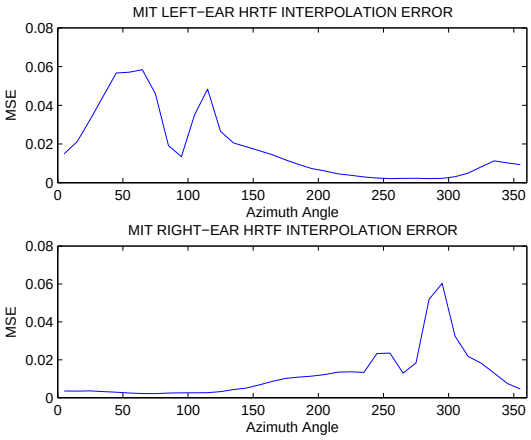

Fig. 9: Interpolation error distributions as a function of the source position azimuthal angle ϕ for the MIT data.

up to 10.8 kHz using the angular spacing of 10°. In addition, the second group locations are at the midpoints of those used to determine the modal parameters and thus represent the locations of maximum interpolation error.

Fig. 6 compares measured and model reconstructed HRTF magnitude spectrum at four typical directions in the horizontal plane. The responses qualitatively show that the model reproduces the experimental measured responses with very high accuracy. Fig. 7 illustrates the predictive power of the model by comparing measured and model interpolated HRTF magnitude spectrum. The precise interpolation results prove that the proposed continuous functional model can achieve HRTF estimation at any point in the horizontal plane.

The distribution of errors across all positions is also presented in Fig. 8 and 9 for the MIT measurements. The worst reconstruction error is around 3% at an azimuth of $100° - 130°$ for the left ear and $280° - 320°$ for the right ear. In MIT measurements, a source located at 90° azimuth is directly across from the right ear (the left ear is the shadowed ear) and a source located at 270° is directly across from the left ear (the right ear is the shadowed ear). Figures reveal that the reproduction of HRTFs is usually better at the source-facing side of the head than at the head's shadowed side. Observe that the interpolation errors are slightly larger than previous reconstructional errors because these directions are actually not used to determine the model parameters.

4.2. Analytical Solutions

The proposed method is also tested on synthetic data. We used the well known analytical sphere HRTF model presented in [10] and a circle of 72 points on the horizontal plane (note these 72 points are uniformly sampled according to section 3.1). The HRTFs are computed at each point at a distance of 1.2m from the sphere (head). Then the same validation procedure is employed. The results are very similar to those from the MIT database described above. As an example, the error distribution from fitting the model to the analytic solution is shown in Fig. 10 and 11 (the left and right ears have symmetrical error performance as the head is modelled as a sphere in the analytical model). Note that, as with the MIT data, the reconstruction error reaches maximum (nearly 1.4%) at the head's shadowed area.

4.3. Discussion of the Model

The reproduction of HRTFs is usually better at the source-facing side of the head than at the head's shadowed side. Two factors contribute to the relative large errors at the head shadowed side. The most significant is the relatively lower signal-to-noise ratio (SNR) at the head shadowed side. The energy in the contralateral HRTFs is less than that in the ipsilateral HRTFs. Hence, the head-shadowed HRTFs contribute relatively little to

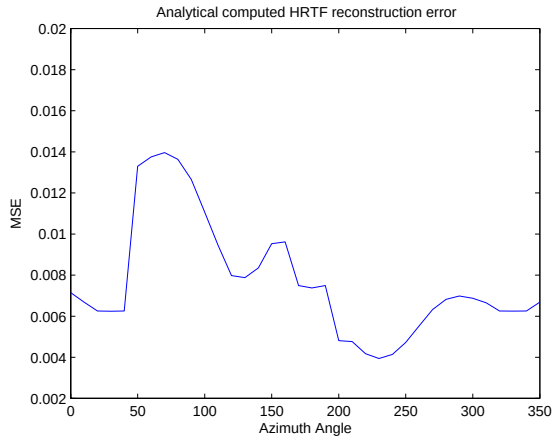

Fig. 10: Reconstruction error distributions as a function of the source position azimuthal angle ϕ for the analytical model.

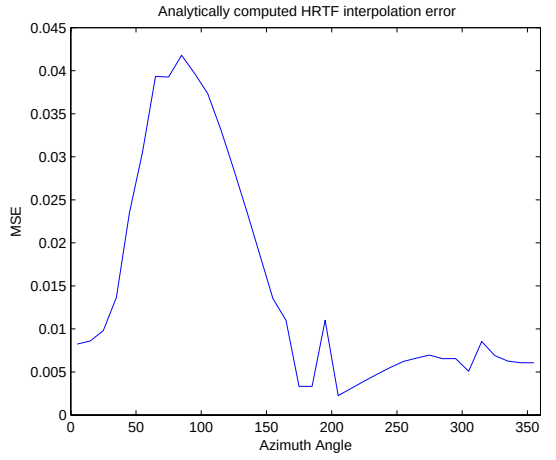

Fig. 11: Interpolation error distributions as a function of the source position azimuthal angle ϕ for the analytical model.

estimate the model parameters C_{mk}. A second factor is that the contralateral sounds may produce more variations because of the diffraction around the head. This results in the spectral shapes that are more complicated and more difficult to model. Therefore, it is obviously noted from the error distributions that the error reaches its maximum at the head shadowed side.

In view of the above two factors, we believe the synthesis error at the head shadowed side can be reduced by increasing the truncation number M and K (including more basis functions) in the proposed model. As stated in section 2, truncation of the infinite series model (6) may cause the inaccuracy of the representation. While the coefficients C_{mk} show a clear downward trend in the average contribution to approximate HRTFs; the high order parameters correspond to the fine detail in the response. Hence, a higher truncation number results in more accurate representation.

5. CONCLUSION

In this paper, a method was developed for HRTFs reproduction in the horizontal plane. The method is based on a continuous HRTF functional model which uses Fourier series and Fourier Bessel series to separate the spatial and spectral dependence of the HRTF. Measured HRTFs from the KEMAR manikin and analytical simulated HRTFs were used to investigate the model performance. The average MSE between measured and model reconstructed HRTFs is less than two percent; and the average MSE between measured and model interpolated HRTFs is less than four percent. The implementation results show that the method could give indistinguishable reproduction compared with the original measurements.

6. REFERENCES

[1] F. L. Wightman and D. J. Kistler, “Headphone simulation of free-field listening. II: Psychophysical validation,” *J. Acoust. Soc. Am.*, 85, pp. 858-867, 1989.

[2] A. Kulkarnj and H. S. Colburn, “Infinite-impulse-response models of the head-related transfer function,” *J. Acoust. Soc. Am.*, 115, pp. 1714-1728, 2003.

[3] W. L. Martens, “Principal component analysis and resynthesis of spectral cues to perceived direction,”

in *The International Computer Music Conference*, edited by J. Beauchamp, San Francisco, CA, pp. 274-281, 1987.

[4] D. J. Kistler and F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *J. Acoust. Soc. Am.*, 91, pp. 1637-1647, 1992.

[5] J. E. Michael, A. S. A. James and I. T. Anthony, "Analyzing head-related transfer function measurements using surface spherical harmonics," *J. Acoust. Soc. Am.*, 104, pp. 2400-2411, 1998.

[6] L. Savioja, J. Huopaniemi, T. Lokki and R. V*ää*n*ä*nen, "Creating interactive virtual acoustic environments," *J. Audio Eng. Soc.*, 47(9), pp. 675-705, 1999.

[7] M. A. Blommer and G. H. Wakefield, "Pole-zero approximations for head-related transfer functions using a logarithmic error criterion," *IEEE Trans. Speech Audio Processing*, 5, pp. 278-287, 1997.

[8] S. Carlile, C. Jin and V. V. Raad, "Continuous virtual auditory space using HRTF interpolation: Acoustic and psychophysical errors," in *2000 International Symposium on Multimedia Information Processing*, Sydney, Australia, pp. 220-223, 2000.

[9] G. G. William and D. M. Keith, "HRTF measurements of a KEMAR," *J. Acoust. Soc. Am.*, 97, pp. 3907-3908, 1995.

[10] R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model", *J. Acoust. Soc. Am.*, 104, pp. 3048-3058, 1998.

[11] W. Kaplan, "Fourier Bessel Series", in *Advanced Calculus (4th ed.)*, MA: Addision-Wesley, pp. 512-518, 1992.

[12] T. Ajdler, L. Sbaiz and M. Vetterli, "Plenacoustic function on the circle with application to HRTF interpolation," in *Proc. 2005 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Philadelphia, PA, USA, pp. 273-276, 2005.