MODAL EXPANSION OF HRTFS: CONTINUOUS REPRESENTATION IN FREQUENCY-RANGE-ANGLE

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

Applied Signal Processing Group
Department of Information Engineering, RSISE
The Australian National University
Email:{Wen.Zhang, Thushara.Abhayapala,
Rodney.Kennedy}@anu.edu.au

Ramani Duraiswami

Perceptual Interfaces and
Reality Laboratory, UMIACS
University of Maryland, College Park
Email: ramani@umiacs.umd.edu

ABSTRACT

This paper proposes a continuous HRTF representation in both 3D spatial and frequency domains. The method is based on the acoustic reciprocity principle and a modal expansion of the wave equation solution to represent the HRTF variations with different variables in separate basis functions. The derived spatial basis *modes* can achieve HRTF near-field and far-field representation in one formulation. The HRTF frequency components are expanded using Fourier Spherical Bessel series for compact representation. The proposed model can be used to reconstruct HRTFs at any arbitrary position in space and at any frequency point from a finite number of measurements. Analytical simulated and measured HRTFs from a KEMAR are used to validate the model.

Index Terms— Modal Analysis, HRTF, Reconstruction

1. INTRODUCTION

The Head-Related Transfer Function (HRTF) is a description of the transformation of a given sound wave input (parameterized as frequency and source location) filtered by the diffraction and reflection properties of an individual's body. It is usually obtained from measurements on people (or dummy heads). Such data is naturally taken from systematic measurements over a discrete set of angles and at discrete frequencies or time samples. Whilst the data is discrete by necessity, it is understood that the underlying HRTF is fully continuous in space (both angle and range) and frequency.

Many techniques have been proposed to perform the interpolation [1] of the HRTF from the discrete measurements but the most appropriate interpolation can be still considered as an open question. An alternative method is to seek continuous representations of the HRTF in the 2D angle domain, like a weighted sum of spherical harmonics [2], so that the need for interpolation is removed. The work of [3] further applies a series of multi-poles (including spherical Hankel function and spherical harmonics) as basis to capture both angle and range variations.

Our approach is based on the acoustic reciprocity principle and modal expansion of the wave equation solution to develop a continuous HRTF representation in all frequency-range-angle domains. The contributions of our approach are as follows: i) separable basis functions to represent the HRTF dependence on each variable; ii) the normalized *spatial modes* to achieve near-field and far-field HRTFs representation in one formulation; iii) the decomposed model coefficients show that the HRTF spectrum has underlying pattern of spherical Bessel functions and we propose to use Fourier Spherical Bessel (FSB) series to expand the HRTF frequency components. The goal of this series representation is to generate coefficients more compact (few parameters) for HRTF representation with high accuracy. In summary, the proposed model can help to obtain the HRTF at any frequency for an arbitrary position from existing measurements conducted on a single sphere (source position)[1].

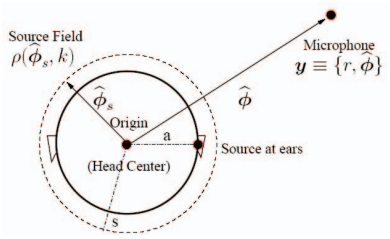


Fig. 1. Geometry of the HRTF measurement based on the reciprocity principle.

2. MODAL ANALYSIS OF HRTFS

HRTFs are usually obtained by emitting a signal from a loudspeaker at different positions in the space and recording it at a microphone in the listener's ear. At the physical level, the HRTF is characterized by the classical wave equation subject to boundary conditions. The general solution to the wave equation can be decomposed with respect to each variable (frequency, range and angle). In this section, we use the modal expansion of the wave equation solution to derive a general representation of the HRTF.

The wavefield between the microphone and the speaker location contains sources (the speaker), the scatterer (human head and body) and the receiver (listener's ear), which does not satisfy the condition of being source free for modal expansion analysis. We use the principle of reciprocity [4] as used in [3] to remove this difficulty. The reciprocity principle states that the acoustic signal received at a receiver is the same as measured at the source location when the source and receiver locations are interchanged and the same excitation is applied.

Here for HRTF analysis, we assume that the acoustic sources are located at the listener's ears and microphones are at some distance

[1] We only consider the case that no objects are within the space between the loudspeakers and human subjects.

away (Fig. 1). We analyze the wavefield at these microphone positions which is a typical exterior problem (all sources inside a spherical surface). From the Huygens-Fresnel principle [5], the scatterers can be thought as producing new wavefronts (or regarded as new sources). Hence, we consider all the scattering sources as the secondary level sources with the original sources at the listener's ears together constituting an *equivalent source field*. As shown in Fig. 1, this source field is encompassed in a sphere (origin is the head center) and as a function of angular position (source) and frequency. We write the source field at radius s (s is greater than the head radius a) as

$$\rho(\hat{\boldsymbol{\phi}}_s, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \alpha_n^m(k) Y_n^m(\hat{\boldsymbol{\phi}}_s), \tag{1}$$

where $\hat{\boldsymbol{\phi}}_s$ is a unit vector (2D angle) to the direction of the source and $Y_n^m(\cdot)$ are spherical harmonics. The $\alpha_n^m(k)$ are the spherical harmonics coefficients of the source field and obtained from

$$\alpha_n^m(k) = \int_{\mathbb{S}^2} \rho(\hat{\boldsymbol{\phi}}_s, k) \overline{Y_n^m(\hat{\boldsymbol{\phi}}_s)} d\hat{\boldsymbol{\phi}}_s \tag{2}$$

on the 2-sphere $\mathbb{S}^2$ [5] at wavenumber k, $k = 2\pi f/c$, where f is the frequency, c is the speed of sound propagation and $\overline{(\cdot)}$ stands for the complex conjugate. We can see that $\alpha_n^m(k)$ carry the information about the source and also the human scattering behavior.

Then the received signal at $\boldsymbol{y} \equiv \{r, \hat{\boldsymbol{\phi}}\}$ (the HRTF corresponding to that position) can be written in terms of the source field as

$$\widehat{H}(r, \hat{\boldsymbol{\phi}}, k) = \int_{\mathbb{S}^2} \rho(\hat{\boldsymbol{\phi}}_s, k) \frac{e^{ik\|s\hat{\boldsymbol{\phi}}_s - y\|}}{4\pi\|s\hat{\boldsymbol{\phi}}_s - y\|} d\hat{\boldsymbol{\phi}}_s \tag{3}$$

where r is the distance between the head center (origin, or source center) and receiver position and $\hat{\boldsymbol{\phi}}$ is the direction of the receiver. Note the integral is over the sphere to account for all of the sources.

Using the Jacobi-Anger expansion [5], we have

$$\frac{e^{ik\|s\hat{\boldsymbol{\phi}}_s - y\|}}{4\pi\|s\hat{\boldsymbol{\phi}}_s - y\|} = ik \sum_{n=0}^{\infty} \sum_{m=-n}^{n} j_n(ks) h_n^{(1)}(kr) \overline{Y_n^m(\hat{\boldsymbol{\phi}}_s)} Y_n^m(\hat{\boldsymbol{\phi}}), \tag{4}$$

where $j_n(\cdot)$ is the spherical Bessel function and $h_n^{(1)}(\cdot)$ is the spherical Hankel function of the first kind. By substituting (4) into (3), we can expand the 3D HRTF as

$$\widehat{H}(r, \hat{\boldsymbol{\phi}}, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \beta_n^m(k) h_n^{(1)}(kr) Y_n^m(\hat{\boldsymbol{\phi}}) \tag{5}$$

where

$$\beta_n^m(k) = 4\pi ik\, \alpha_n^m(k) j_n(ks). \tag{6}$$

In (5), HRTF dependence on each variable (frequency, range, and angle) is represented by separate basis functions (note the radial part is with variable *range measured in wavelengths* as it being a function of kr). The spatial modes $h_n^{(1)}(kr) Y_n^m(\hat{\boldsymbol{\phi}})$ account for the HRTF spatial variations as in [3] and $\beta_n^m(k)$ are the modal decomposed HRTF frequency components. In the next section we further develop this representation to i) link the near-field and far-field HRTFs directly and ii) parameterize the frequency components by a set of basis functions.

3. HRTF CONTINUOUS REPRESENTATION

In this section, we modify the basis functions in the general representation (5) to derive an efficient continuous representation of the HRTF in both spatial and frequency domains.

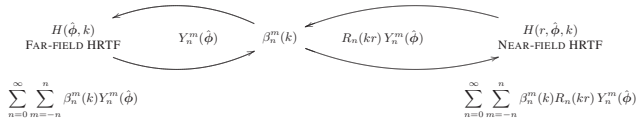


Fig. 2. Modal decomposition of the HRTF with radial invariant frequency components.

3.1. Normalized Modes for HRTF Spatial Representation

The spatial modes in (5) cannot represent the far-field HRTFs because the radial term tends to zero, viz.,

$$h_n^{(1)}(kr) \to (-i)^{(n+1)} \frac{e^{ikr}}{kr}, \text{ as } r \to \infty. \tag{7}$$

It is desired to normalize the spherical Hankel function, i.e.,

$$R_n(kr) \triangleq i^{(n+1)} kr e^{-ikr} h_n^{(1)}(kr), \tag{8}$$

so that we can achieve both near-field HRTF and far-field HRTF representation in one formulation. Further,

$$\lim_{r \to \infty} R_n(kr) = 1, \ \forall n. \tag{9}$$

The modified HRTF representation with normalized modes is in the following form

$$H(r, \hat{\boldsymbol{\phi}}, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \beta_n^m(k) R_n(kr) Y_n^m(\hat{\boldsymbol{\phi}}) \tag{10}$$

for the near-field HRTFs and when $r \to \infty$,

$$H(\hat{\boldsymbol{\phi}}, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \beta_n^m(k) Y_n^m(\hat{\boldsymbol{\phi}}). \tag{11}$$

Equations (10) and (11) show that the HRTF frequency components $\beta_n^m(k)$ are radially invariant and can be obtained from the spherical harmonic transform of the measurements on a single sphere, i.e.,

$$\beta_n^m(k) = \begin{cases} \frac{1}{R_n(kr)} \int_{\mathbb{S}^2} H(r, \hat{\boldsymbol{\phi}}, k) \overline{Y_n^m(\hat{\boldsymbol{\phi}})} d\hat{\boldsymbol{\phi}}, & \text{for near-field} \\ \int_{\mathbb{S}^2} H(r, \hat{\boldsymbol{\phi}}, k) \overline{Y_n^m(\hat{\boldsymbol{\phi}})} d\hat{\boldsymbol{\phi}}, & \text{for far-field} \end{cases} \tag{12}$$

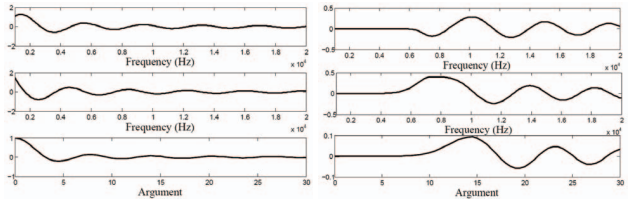
and later used for HRTF reconstruction at any spatial point, as shown in Fig. 2.

3.2. FSB Series for HRTF Frequency Representation

The goal of seeking an efficient continuous HRTF frequency representation is to determine the spectrum of HRTF with better spectral resolution and fewer parameters. The exhibiting characteristic of $\beta_n^m(k)$ is the underlying pattern of the spherical Bessel functions (implicitly shown in (6)), which means the signal can be more compactly represented by the spherical Bessel function rather than complex exponentials. Next, we use the spherical head model [6] as an example, in which the HRTFs are represented as

$$\varphi_H(r, \Theta, k) = \frac{-r}{ka^2} e^{-ikr} \sum_{n=0}^{\infty} (2n+1) P_n(\cos\Theta) \frac{h_n^{(1)}(kr)}{h_n'^{(1)}(ka)}, \ r > a \tag{13}$$

where Θ is the angle of incidence, $P_n(\cdot)$ is a Legendre polynomial of order n and $h_n^{(1)}(\cdot), h_n'^{(1)}(\cdot)$ are the spherical Hankel function of



(a) (b)

Fig. 3. Examples to demonstrate the structural similarities between the HRTF frequency components $\beta_n^m(k)$ and the spherical Bessel functions of the first kind. Top plots and middle plots are the real and imaginary parts of $\beta_n^m(k)$ with (a) $n = 0, m = 0$ and (b) $n = 12, m = 0$; and the bottom plots are the spherical Bessel functions $j_n(\cdot)$ at the corresponding order $n = 0$ and $n = 12$ with against arguments from 0 to 30.

the first kind and its derivative. Applying the addition theorem [5] to expand the above model in spherical harmonics $Y_n^m(\cdot)$, the HRTF frequency components are

$$\beta_n^m(k) = \frac{4\pi \overline{Y_n^m(\hat{\phi}_{\text{ear}})}}{i^{(n+2)}} \left(j_n(ka) - j_n^{'}(ka) \frac{h_n^{(1)}(ka)}{h_n^{'(1)}(ka)} \right). \quad (14)$$

The first component represents the incident wavefield and the second term is the scattered field. Both show the similar structures to the spherical Bessel functions; so we can observe the strong correlation between the HRTF frequency components and spherical Bessel functions in Fig. 3.

Thus, we apply the Fourier Spherical Bessel (FSB) series for HRTF frequency components representation. The FSB series [7] (derived from the Fourier Bessel series) are orthogonal basis functions on the interval $(0, 1)$

$$\int_0^1 x^2 j_n\left(xZ_q^{(n)}\right) j_n\left(xZ_r^{(n)}\right) dx = \frac{1}{2}\delta_{\ell,r}\left(j_{n+1}(Z_q^{(n)})\right)^2, \quad (15)$$

where $Z_q^{(n)}$ and $Z_r^{(n)}$ are the positive roots of the $j_n(\cdot)$. And the derived HRTF frequency components representation is

$$\beta_n^m(k) = \sum_{q=1}^{\infty} A_{n;q}^m \, j_n\left(\frac{Z_q^{(n)}}{k_{\max}} k\right), \quad (16)$$

where from (15)

$$A_{n;q}^m = \frac{2}{k_{\max}^3 J_{n+1}^2\left(Z_q^{(n)}\right)} \int_0^{k_{\max}} k^2 \beta_n^m(k) \, j_n\left(\frac{Z_q^{(n)}}{k_{\max}} k\right) dk. \quad (17)$$

$k_{\max}$ is the maximum wave number of an HRTF data set.

The above development leads to the HRTF functional model written as

$$H(r, \hat{\phi}, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \sum_{q=1}^{\infty} A_{n;q}^m j_n\left(\frac{Z_q^{(n)}}{k_{\max}} k\right) R_n(kr) Y_n^m(\hat{\phi}), \quad (18)$$

where the modal coefficients $A_{n;q}^m$ are solved from (12) and (17). Given discrete experimentally measured HRTFs, we use the left Riemann sum to approximate both integrals[2].

[2]The problem of no complete data set over the sphere in the HRTF measurements (low elevation measurements are missing) is solved by the extrapolation algorithm [8].

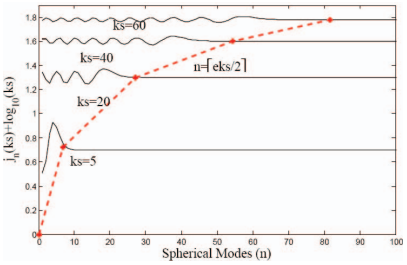


Fig. 4. Dependence of the spherical Bessel function $j_n(ks)$ in the order n at different ks shown on the vertically shifted curves.

4. IMPLEMENTATION DETAILS: CHOICE OF TRUNCATION NUMBER

For practical implementation, the key issue is the truncation of the spatial modes decomposition, through parameter N_0, and FSB series expansion, through parameter Q_0, in the proposed model (18) denoted as

$$H(r, \hat{\phi}, k) = \sum_{n=0}^{N_0-1} \sum_{m=-n}^{n} \sum_{q=1}^{Q_0} A_{n;q}^m j_n\left(\frac{Z_q^{(n)}}{k_{\max}} k\right) R_n(kr) Y_n^m(\hat{\phi}) \quad (19)$$

which transforms any HRTF data set to a coefficients set $\{A_{n;q}^m\}$ of the size $N_0^2 \times Q_0$.

4.1. Truncation of Spatial Modes Decomposition

The spherical Bessel function $j_n(ks)$ determines the truncation of the HRTF decomposition with spatial modes. Fig. 4 shows that when $n < \lceil eks/2 \rceil$, the spherical Bessel functions oscillate and there is no decay in the amplitude for growing n; however, when $n \geq \lceil eks/2 \rceil$ the functions monotonically decay to zero with growing n, where $\lceil \cdot \rceil$ is the integer ceiling function. So guided by [9], we choose the truncation order

$$N_0 = \lceil e k_{\max} s/2 \rceil + 1 \quad (20)$$

for a specified HRTF data set by the maximum wavenumber $k_{\max}$ and s taken to be the radius of the sphere which encloses all the scattering sources. Here we choose $s = 0.1$m to include all the head scattering (note larger values to add shoulder and torso reflection may further improve the reconstruction accuracy).

4.2. Truncation of FSB Series Expansion

Firstly, we should state that due to the structural similarities between FSB series and the HRTF frequency components, compared to other basis functions the FSB series provide the most accurate reproduction under the same truncation order. Further, Fig. 5 shows the energy of $A_{n;q}^m$ calculated from the spherical head mode at $r = 1.0$ m with $k_{\max} = 296$ (16 kHz) along the FSB series expansion order. We found 90% of the energy is kept in the first 10 components, i.e., $Q_0 = 10$ for the analytical simulated HRTFs. Section 5.2 shows for real data, this value may need to increase due to the more complicated waveforms caused by diffraction and reflection of the human (or dummy) head.

5. MODEL VALIDATION

The effectiveness of the proposed HRTF continuous representation is investigated by decomposing the experimentally measured (or

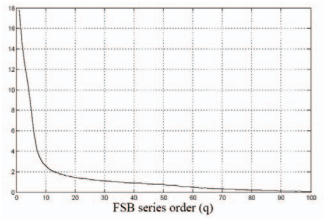
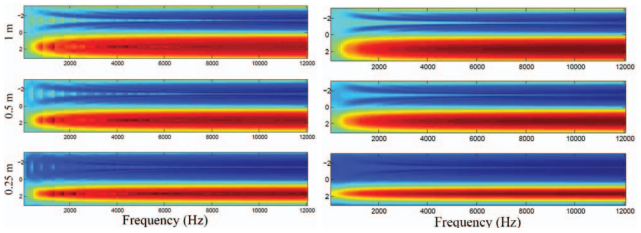


Fig. 5. Decomposed model coefficients energy of the HRTF from the spherical head model versus the FSB series expansion order q.



(a) (b)

Fig. 6. Analytically simulated HRTFs at $r = 1.0$ m (top plots) and the reconstructed HRTFs at $r = 0.5$ and $r = 0.25$ m (a) compared to the reference (b). The horizontal axis is frequency and the vertical axis is azimuth from $[-\pi, \pi]$.

analytically simulated) HRTFs on a single sphere and reconstruct HRTFs at any frequency for an arbitrary spatial location.

5.1. Analytical Solutions

We decompose the analytically simulated HRTFs [6] at 1.0 m with 961 samples on the sphere to obtain the $A_{n;q}^m$. The truncation number of N_0 is 32 for frequency up to 12 kHz according to (20) and $Q_0 = 10$ from analysis in Section 4.2. In Fig. 6, the plots on the right are the magnitudes of the analytical HRTF at different ranges at the equator, compared to the reconstructions from the proposed model on the left. We observe the reconstruction is perfect with average approximation error around 0.52%.

5.2. KEMAR Data

The MIT data is acquired using a KEMAR manikin [10] at a distance of 1.4 m with sampling rate of 44.1 kHz. We use this data to check the frequency reconstruction performance of the proposed model. Note that the data is not available on the whole sphere (710 samples with low elevation measurements missing). We use the extrapolation algorithm developed in [8] to estimate the missing HRTFs. The model decomposition is performed during the frequency range of $[0.2, 10]$ kHz so we choose $N_0 = 27$. Here we choose $Q_0 = 120$ (keeping 90% of the coefficients energy), which is larger than for the spherical head model due to complicated diffraction of the head. Fig. 7 shows the mean squared reconstruction error at each frequency bin. The maximum error is less than 1.4% and the error increases with frequency (because the truncation number N_0 is determined from the maximum wavenumber in (20), which can fit the low frequency data very well).

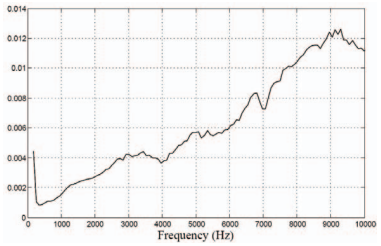


Fig. 7. The MIT left ear HRTFs reconstrunction error at each frequency bin.

6. CONCLUSION

A continuous functional model was developed for the HRTF representation in both spatial and frequency domains. The method is powerful for the computation of the HRTF at any arbitrary position in space and at any frequency point from a given set of measurements at a fixed distance. We observed good HRTF spatial and frequency components reconstruction and extrapolation results from both analytical solutions and KEMAR data.

7. REFERENCES

[1] K. Hartung, J. Braasch, and S. J. Sterbing, "Comparison of different methods for the interpolation of head-related transfer functions," in *Proc. of the 16th AES International Conference*, Rovaniemi, Finland, Apr. 1999, pp. 319–329.

[2] M. J. Evans, J. A. S. Angus, and A. I. Tew, "Analyzing head-related transfer function measurements using surface spherical harmonics," *J. Acoust. Soc. Am*, vol. 104, no. 4, pp. 2400–2411, Oct. 1998.

[3] R. Duraiswami, D. N. Zotkin, and N. A. Gumerov, "Interpolation and range extrapolation of HRTFs," in *Proc. IEEE ICASSP 2004,*, Montreal, Quebec, Canada, May 2004, vol. IV, pp. 45–48.

[4] P. M. Morse and K. U. Ingard, *Theoretical acoustics*, Princeton University Press, New Jersey, 1968.

[5] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*, Springer, New York, 1998.

[6] R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," *J. Acoust. Soc. Am.*, vol. 104, no. 5, pp. 3048–3058, Nov. 1998.

[7] T. D. Abhayapala, "Generalized framework for spherical microphone arrays: Spatial and frequency decomposition," in *Proc. IEEE ICASSP 2008*, Las Vegas, Nevada, Mar. 2008, pp. 5268–5271.

[8] W. Zhang, R. A. Kennedy, and T. D. Abhayapala, "Iterative extrapolation algorithm for data reconstruction over sphere," in *Proc. IEEE ICASS 2008*, Las Vegas, Nevada, Mar. 2008, pp. 3733–3736.

[9] T. D. Abhayapala, T. S. Pollock, and R. A. Kennedy, "Characterization of 3D spatial wireless channels," in *IEEE 58th Vehicular Technology Conference, VTC 2003-Fall*, Orlando, Florida, USA, Oct. 2003, vol. 1, pp. 123–127.

[10] W. G. Gardner and K. D. Martin, "HRTF measurements of a KEMAR," *J. Acoust. Soc. Am*, vol. 97, no. 6, pp. 3907–3908, Jun. 1995.