Insights into head-related transfer function: Spatial dimensionality and continuous representation

Wen Zhang,[a] Thushara D. Abhayapala, and Rodney A. Kennedy
Department of Information Engineering, Research School of Information Sciences and Engineering, College of Engineering and Computer Science, The Australian National University, Canberra ACT 0200, Australia

Ramani Duraiswami
Perceptual Interfaces and Reality Laboratory, Institute for Advanced Computer Studies, University of Maryland, College Park, Maryland 20742

(Received 12 October 2009; revised 27 January 2010; accepted 1 February 2010)

This paper studies head-related transfer function (HRTF) sampling and synthesis in a three-dimensional auditory scene based on a general modal decomposition of the HRTF in all frequency-range-angle domains. The main finding is that the HRTF decomposition with the derived spatial basis function modes can be well approximated by a finite number, which is defined as the spatial dimensionality of the HRTF. The dimensionality determines the minimum number of parameters to represent the HRTF corresponding to all directions and also the required spatial resolution in HRTF measurement. The general model is further developed to a continuous HRTF representation, in which the normalized spatial modes can achieve HRTF near-field and far-field representations in one formulation. The remaining HRTF spectral components are compactly represented using a Fourier spherical Bessel series, where the aim is to generate the HRTF with much higher spectral resolution in fewer parameters from typical measurements, which usually have limited spectral resolution constrained by sampling conditions. A low-computation algorithm is developed to obtain the model coefficients from the existing measurements. The HRTF synthesis using the proposed model is validated by three sets of data: (i) synthetic HRTFs from the spherical head model, (ii) the MIT KEMAR (Knowles Electronics Mannequin for Acoustics Research) data, and (iii) 45-subject CIPIC HRTF measurements.
 [DOI: 10.1121/1.3336399]

PACS number(s): 43.60.Ac, 43.60.Uv, 43.66.Pn [EJS] Pages: 2347–2357

I. INTRODUCTION

A. Motivation and background

People hear sound in three dimensions and the perception of the spatial aspects of sound has been essential to people's lives. Multiple cues are involved for the spatial localization[1] including the amplitude and the time arrival of the sound at each ear and, most importantly, the spectrum of the sound, which is modified by the interaction between the sound wave and a person's body (the torso, head, and external pinna shape). The head-related transfer function (HRTF),[2] an acoustic transfer function from the sound source to a listener's eardrums, contains all the listening cues used by the hearing mechanism for decoding spatial information encoded in binaural signals. The HRTF changes with direction from which sound arrives to the listener, and any sound source can be realistically located by filtering sound with the HRTF corresponding to the desired location and presenting the resulting binaural signals to the subject using two playback channels achieved typically by a pair of headphones.[3]

Nowadays, in practice, hundreds of measured HRTFs from all directions surrounding a subject (person or dummy head) are fully recorded and have always been directly applied to study the transformation characteristics of the external ear and to synthesize virtual reality over headphones.[2,4,5] Two major problems with the direct use of measured HRTFs are that first it is impossible to simulate every conceivable direction and create source movement (panning) smoothly through the space given the HRTF measurements are discrete by necessity, and second there is no standard HRTF spatial sampling theory to make HRTF measurement practical for commercial applications.

One common approach toward the goal to study the HRTF is to model the HRTF or head-related impulse response (HRIR) by a reduced number of parameters and to make the processing more effective by operating in this parametric domain. In the case of discrete data and sets of measurements corresponding to different human subjects, many techniques have been proposed for HRTF modeling. The filter bank models[3,6] could achieve accurate reconstruction of the original HRTF measurements, but the expansion weights in the model are coupled with both angle and frequency variables, which limits the usefulness of the model for HRTF analysis. Statistical methods have been used to analyze the HRTF in an effort to reduce the redundancy (correlation) of a data set. One important study is principal component analysis (PCA).[7,8] However, the facts are that this PCA represen-

[a]Author to whom correspondence should be addressed. Electronic mail: wen.zhang@anu.edu.au

tation is not continuous and the basis vectors may change for each individual. Both filter bank models and statistical models such as PCA only allow the synthesis of the measured HRTF samples. Interpolation is still required[9–11] between the discrete measurement positions.

HRTFs have also been represented as a weighted sum of spherical harmonics in three dimensions,[12] and as a series of multipoles based on the reciprocity principle.[13] The spherical harmonics, a complete orthogonal basis function on the two-sphere, provide a natural continuous representation in the angular domain. Therefore, it leads to a straightforward solution to the problem of HRTF interpolation in elevation and azimuth. In both models, the expansion weights are functions of frequency; analyzing these components can provide a new means to study the scattering behavior of the human body.

B. Contributions and organization

In the previous work,[14] we studied horizontal plane HRTF representation. Here we extended our work to three-dimensional auditory scene. Three main contributions of this paper are summarized below.

In Sec. II, we use the acoustic reciprocity principle and modal expansion of the wave equation solution to develop a general HRTF representation in all frequency-range-angle domains. We show that the HRTF decomposition with the derived spatial basis function *modes* can be truncated to a finite number and still with relative high accuracy. This means that the HRTF is essentially a mode-limited function; a finite number of spatial modes (named the *dimensionality*) can represent the HRTF corresponding to all directions. The value of dimensionality also determines the required spatial resolution in HRTF measurement.

In Sec. III, we further develop the general HRTF model to a continuous representation. We apply *normalized spatial modes* to achieve near-field and far-field HRTF representations in one formulation, which provides a way to obtain the range dependence of the HRTF from measurements conducted at only a single range. We study the radially invariant HRTF spectral components and find that the HRTF spectrum has an underlying pattern similar to the spherical Bessel functions. We use an orthogonal property of the Bessel functions to form frequency basis functions, Fourier spherical Bessel (FSB) series, to model the HRTF spectral components. Besides achieving much higher spectral resolution, this series representation can has far few parameters compared to the measurements for a more efficient HRTF representation.

The practical model implementation issues are discussed in Sec. IV. A low-computation algorithm is proposed to calculate the model coefficients from discrete measurements. The proposed method separates the HRTF azimuth and elevation sampling effects, from which we have the following observations: (i) The HRTF measurements that are coarsely sampled in elevation can still be reconstructed with reasonable accuracy and (ii) as for the azimuth, we need finer azimuthal sampling on the elevations closer to the equator but less azimuthal sampling points closer to the pole.

Section V validates the developed HRTF sampling theory and the proposed HRTF continuous representation by decomposing the experimentally measured[15,16] (or analytically simulated[17]) HRTFs on a single sphere and synthesizing HRTFs at any frequency for an arbitrary spatial location to check both reconstruction and interpolation performances.

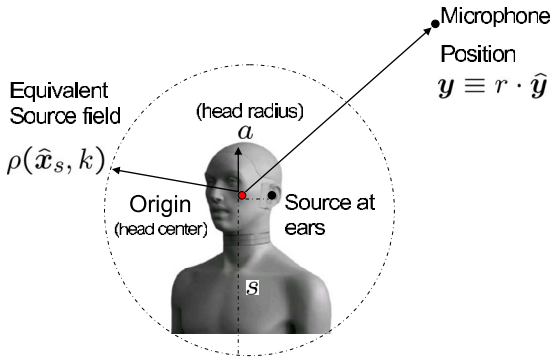

FIG. 1. (Color online) Geometry of HRTF measurement based on the reciprocity principle.

II. MODAL ANALYSIS OF HRTF

HRTFs are usually obtained by emitting a signal from a loudspeaker at different positions in 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 obtained by separation of variables (frequency, range, azimuth, and elevation angles). Thus, in principle, we can use the wave equation solution to expand the HRTF with separable basis functions.

A. Theoretical development

When sound propagates from the source to the listener, the received sound at the listener's ear is transformed by the structure and shape of the listener's body. We seek a representation of the sound pressure at the listener's ear (left or right), where two sources should be taken into account: one is the original acoustic source from the speaker and the other is the secondary source due to the scattering of human body. It is a complicated problem to apply the wave equation in this configuration because the receiver, the listener's ear, is within the scatterer region of human body. The *principle of reciprocity*[18] can be used to remove this difficulty and to develop a general representation of the HRTF.[13]

To apply the principle of reciprocity to the HRTF analysis, we assume that the original acoustic source is located at the listener's ear and microphones are some distance away (Fig. 1). Here, we consider all the scattering sources of human body as the secondary level sources with the original sources at the listener's ear together constituting the source field. From Huygens' principle,[19] the sum of the waves from all the sources (including both original and secondary sources) to any point beyond the scatterers (the human body) can be calculated by integration or numerical modeling. To exactly model the effect of the source field, we develop an *equivalent source field* on a sphere of radius s with origin at the head center, as shown in Fig. 1, where the sphere should

be large enough to enclose all the sources. Note that the reason we choose the sphere to include all the sources is because we can use a specific set of orthogonal series, spherical harmonics, to represent the source field; for example, we write the equivalent source field as a function of angular position and wavenumber; i.e.,

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

where $\hat{\boldsymbol{x}}_s$ is a unit vector (or a set of 2D angles, elevation and azimuth $\{\theta_s,\phi_s\}$) pointing into the equivalent source direction and $\boldsymbol{x}_s\equiv s\cdot\hat{\boldsymbol{x}}_s$ defines the equivalent source position. The wavenumber is defined as $k=2\pi f/c$, where f is frequency and c is the speed of sound propagation. $Y_n^m(\hat{\boldsymbol{x}}_s)$ are the spherical harmonics characterized by two indices, degree n and order m,

$$Y_n^m(\hat{\boldsymbol{x}}_s)\triangleq\sqrt{\frac{2n+1}{4\pi}\frac{(n-|m|)!}{(n+|m|)!}}P_n^{|m|}(\cos\,\theta_s)e^{im\phi_s}. \tag{2}$$

$\alpha_n^m(k)$ are the spherical harmonic coefficients of the equivalent source field at wavenumber k and obtained from

$$\alpha_n^m(k)=\int_{\mathbb{S}^2}\rho(\hat{\boldsymbol{x}}_s,k)\overline{Y_n^m(\hat{\boldsymbol{x}}_s)}d\varrho(\hat{\boldsymbol{x}}_s) \tag{3}$$

on the two-sphere $\mathbb{S}^2$, where $\overline{(\cdot)}$ stands for the complex conjugate and $\int_{\mathbb{S}^2}d\varrho(\hat{\boldsymbol{x}}_s)=\int_0^{2\pi}\int_0^{\pi}\sin\,\theta_s d\theta_s d\phi_s$. We can see that the $\alpha_n^m(k)$ carry information about the original source and also the human body scattering behavior. Then the received signal at $\boldsymbol{y}\equiv r\cdot\hat{\boldsymbol{y}}$ (the HRTF corresponding to that position) can be written in terms of the equivalent source field as

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

where r is the distance between the head center (origin or source center) and the receiver position and $\hat{\boldsymbol{y}}$ is the direction of the receiver. The integral is over the sphere to account for all sources. Using the Jacobi–Anger expansion,[19] we have

$$\frac{e^{ik\|\boldsymbol{x}_s-\boldsymbol{y}\|}}{4\pi\|\boldsymbol{x}_s-\boldsymbol{y}\|}=ik\sum_{n=0}^{\infty}\sum_{m=-n}^{n}j_n(ks)h_n^{(1)}(kr)\overline{Y_n^m(\hat{\boldsymbol{x}}_s)}Y_n^m(\hat{\boldsymbol{y}}),\quad r>s, \tag{5}$$

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 Eq. (5) into Eq. (4), we can expand the HRTF at position $\boldsymbol{y}$ as

$$\hat{H}(r,\hat{\boldsymbol{y}},k)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\hat{\beta}_n^m(k)h_n^{(1)}(kr)Y_n^m(\hat{\boldsymbol{y}}), \tag{6}$$

where

$$\hat{\beta}_n^m(k)=4\pi ik\alpha_n^m(k)j_n(ks). \tag{7}$$

In Eq. (6), the HRTF dependence on each variable (frequency, range, and 2D angle) is represented by separable basis functions. The spatial modes, i.e.,

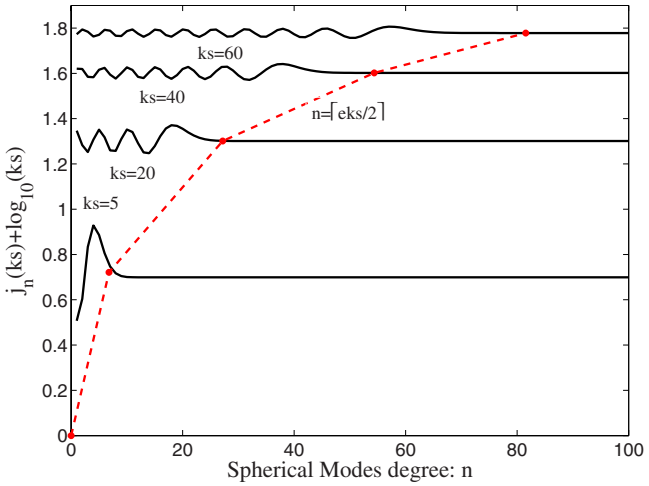

FIG. 2. (Color online) Dependence of the spherical Bessel function $j_n(ks)$ vs degree n at different ks shown on the vertically shifted curves.

$$h_n^{(1)}(kr)Y_n^m(\hat{\boldsymbol{y}}), \tag{8}$$

account for the HRTF spatial variations and $\hat{\beta}_n^m(k)$ are the modal decomposed HRTF spectral components.

B. Dimensionality of HRTF as a mode-limited function

In this section, we show that the HRTF decomposition in Eq. (6) can be well approximated by choosing a sufficiently large truncation order N, viz.,

$$\hat{H}(r,\hat{\boldsymbol{y}},k)\cong\sum_{n=0}^{N}\sum_{m=-n}^{n}\hat{\beta}_n^m(k)h_n^{(1)}(kr)Y_n^m(\hat{\boldsymbol{y}}), \tag{9}$$

which indicates that the HRTF is essentially a mode-limited function.[20] The required number $(N+1)^2$ of spatial modes (8) to represent the HRTF spatial variations should be determined by a typical size of human head/torso and by bounds on the spherical Bessel function $j_n(ks)$, which decides the upper limit of $\hat{\beta}_n^m(k)$ in Eq. (6) [because first, both the source field coefficients $\alpha_n^m(k)$ and the spherical harmonics $Y_n^m(\cdot)$ are bounded functions; second, the spherical Hankel function $h_n^{(1)}(kr)$ has a weaker impact than the same order of the spherical Bessel function $j_n(ks)$ (Ref. 21)]. We define this number of spatial modes as the spatial dimensionality of the HRTF.

Figure 2 illustrates typical dependence of the spherical Bessel function on the degree n for various values of ks. It is clearly seen that there are two distinct regions separated by value[22,23]

$$N=\lceil eks/2\rceil. \tag{10}$$

For $n<\lceil eks/2\rceil$, 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, and the decay is very fast. Therefore, we only need to include all spatial modes lower than the order of $N=\lceil eks/2\rceil$ for HRTF spatial representation. This yields the spatial dimensionality of the HRTF, or the required number of weights $\{\hat{\beta}_n^m(k)\}$ that can represent HRTFs corresponding to all directions; i.e.,

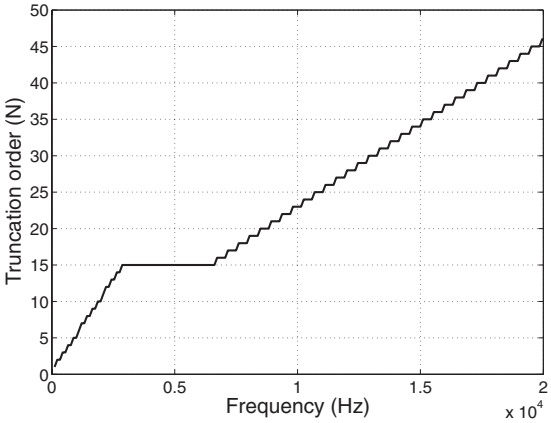

FIG. 3. Calculated the required truncation order N for the HRTF representation as a function of frequency.

$$\mathrm{DIM}(H) = (N+1)^2 = (\lceil eks/2 \rceil + 1)^2. \tag{11}$$

In order to obtain the required number of weights, Eq. (11) also defines the least number of the HRTF samples in the space. The dimensionality depends on the wavenumber k and the radius of the equivalent source field s, and we have the following comments.

(1) The dimensionality increases with wavenumber/frequency. This is because for a fixed size region of sphere, the low frequency HRTF requires fewer spatial modes since the waves are spatially varying more slowly; for increasing frequency, we need more spatial modes as the smaller wavelength indicates faster changes in the space.

(2) The value of s relates to the typical size of human head/torso. For example, for the spherical head, the value of s is simply the radius of the head (0.09 m). While for the Knowles Electronics Mannequin for Acoustics Research (KEMAR) or human subjects, we need to enlarge the radius of the equivalent source field to include the main torso effect, i.e., the shoulder reflection. However, the torso only contributes to the HRTF at frequencies below 3 kHz. For frequencies above 3 kHz, it is the pinna effect that allows the perception of elevation effects.[24,25] So we propose to set two separate values of the equivalent source field radius for two ranges of frequency, that is,

$$s = \begin{cases} 0.20 \text{ m} & \text{for } f \le 3 \text{ kHz} \\ 0.09 \text{ m} & \text{for } f > 3 \text{ kHz}. \end{cases} \tag{12}$$

(3) Figure 3 plots the calculated truncation order N required for HRTF representation (9) as a function of frequency (note that an interpolation function on the truncation order derived from the two equivalent source field radii is applied; the value of N in the frequency range of [3, 6] kHz is decided by the maximum value at $f=3$ kHz after which the source field radius reduces to 0.09 m). In the case for a given frequency range, such as the audible frequency range (200 Hz—20 kHz), the maximum number of the discrete frequency points included in the frequency range determines the least number of measurements. For example, 20 kHz bandwidth has the highest truncation order $N=46$ and requires at least 2209 HRTF measurements in the space.

III. HRTF CONTINUOUS REPRESENTATION

In this section, we further develop general representation (9) into a continuous HRTF model, which can (i) link near-field and far-field HRTFs directly, and (ii) parametrize the spectral components by a set of basis functions.

A. Normalized modes for HRTF spatial representation

The spatial modes in Eq. (6) cannot directly represent far-field HRTFs because the radial term tends to zero, viz.,

$$h_n^{(1)}(kr) \sim (-i)^{(n+1)} \frac{e^{ikr}}{kr} \to 0 \quad \text{as} \quad r \to \infty. \tag{13}$$

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

$$R_n(kr) \triangleq i^{(n+1)} kr e^{-ikr} h_n^{(1)}(kr), \tag{14}$$

so that we can achieve both near-field HRTF and far-field HRTF representations in one formulation. As demonstrated later in this section, we will show that this definition is consistent with the analytical spherical HRTF model.

Referring to Eq. (9), the modified HRTF representation with the normalization is then

$$H(r,\hat{\mathbf{y}},k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \beta_n^m(k) R_n(kr) Y_n^m(\hat{\mathbf{y}}), \tag{15}$$

noting that $\lim_{r\to\infty} R_n(kr)=1, \quad \forall n$, when $r\to\infty$, we have the normalized far-field representation

$$H(\hat{\mathbf{y}},k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \beta_n^m(k) Y_n^m(\hat{\mathbf{y}}). \tag{16}$$

Equations (15) and (16) show that the HRTF spectral components $\beta_n^m(k)$ are radially invariant and can be obtained from the spherical harmonic transform of the measurements at a single radius; i.e.,

$$\beta_n^m(k) = \begin{cases} \dfrac{1}{R_n(kr)} \displaystyle\int_{S^2} H(r,\hat{\mathbf{y}},k) \overline{Y_n^m(\hat{\mathbf{y}})} d\varrho(\hat{\mathbf{y}}) & \text{near-field} \\ \displaystyle\int_{S^2} H(\hat{\mathbf{y}},k) \overline{Y_n^m(\hat{\mathbf{y}})} d\varrho(\hat{\mathbf{y}}) & \text{far-field}, \end{cases} \tag{17}$$

and later used for HRTF reconstruction at any spatial point. In addition, from Eq. (7), we have

$$\beta_n^m(k) = \frac{\hat{\beta}_n^m(k)}{i^{(n+1)} kr e^{-ikr}} = \frac{4\pi \alpha_n^m(k) j_n(ks)}{i^n r e^{-ikr}}. \tag{18}$$

Example of spherical head model. We use the spherical head model[17] as an example to solve the HRTF spectral components, in which the HRTFs are represented as

$$\varphi_H(r,\Theta,k) = \frac{-r}{ka^2} e^{-ikr} \sum_{n=0}^{\infty} (2n+1) P_n(\cos\Theta) \frac{h_n^{(1)}(kr)}{h_n'^{(1)}(ka)}, \quad r > a, \tag{19}$$

where a is the spherical head radius, Θ is the angle of incidence (the angle between the ray from the center of the sphere to the source, $\hat{\boldsymbol{y}}$, and the ray to the measurement point on the surface of the sphere, $\hat{\boldsymbol{\vartheta}}_{\text{ear}}$), $P_n(\cdot)$ is the Legendre function of degree n, and $h_n^{(1)}(\cdot)$ and $h_n'^{(1)}(\cdot)$ are the spherical Hankel function of the first kind and its derivative. Applying the addition theorem,[19] we have

$$P_n(\cos\Theta) = \frac{4\pi}{2n+1} \sum_{m=-n}^{n} Y_n^m(\hat{\boldsymbol{y}}) \overline{Y_n^m(\hat{\boldsymbol{\vartheta}}_{\text{ear}})}. \tag{20}$$

Then we can expand the spherical head model HRTF with the normalized modes, where the spectral components are

$$\beta_n^m(k) = \frac{4\pi \overline{Y_n^m(\hat{\boldsymbol{\vartheta}}_{\text{ear}})}}{i^n} \left(j_n(ka) - j_n'(ka) \frac{h_n^{(1)}(ka)}{h_n'^{(1)}(ka)} \right). \tag{21}$$

B. Fourier spherical Bessel series for HRTF spectral representation

The goal of seeking an efficient continuous HRTF spectral representation is to determine the spectrum of the HRTF with higher spectral resolution and fewer parameters from a finite number of measurements, which usually have limited spectral resolution constrained by the sampling rate and number of samples (or the record time).

The $\beta_n^m(k)$ exhibit an underlying pattern similar to the spherical Bessel functions [implicitly shown in Eq. (18)].[26] An example is the spectral components of the spherical head HRTF, Eq. (21), in which the first component represents the incident wavefield and the second term is the scattered field. Both terms show the similar structures to the spherical Bessel functions, so we can observe the strong correlation between the HRTF spectral components and spherical Bessel functions in Fig. 4.

Figure 5 shows the energy spread of the HRTF spectrum over the spatial modes (n,m) and wavenumber k, which has a significant triangular null region and has been described as the butterfly shape of the HRTF spectrum[14,27] for the horizontal plane HRTF. The explanation for this special shape of the spectrum is because the HRTF dimensionality increases linearly with frequency, as shown in Sec. II B. At low frequencies, only low order spatial modes are signification and the high spatial modes have very small contributions; at higher frequencies, the higher order spatial modes become significant. Therefore, most of $\beta_n^m(k)$ energy is present in a triangular shaped region and, outside this region, the energy is greatly reduced.

In Fig. 4, the resemblance between the patterns of $\beta_n^m(k)$ and the spherical Bessel functions of the same degree indicates that the HRTF spectrum can be compactly represented by the spherical Bessel functions. Here, we apply the FSB series for the representation of the HRTF spectral components. The FSB series[28] (derived from the Fourier Bessel series used for the horizontal plane HRTF spectral representation[14]) are orthogonal basis functions on the interval (0,1) as follows:

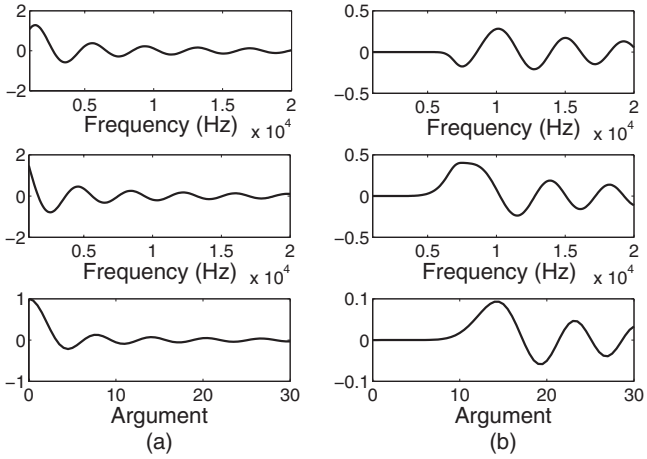

FIG. 4. Examples to demonstrate the structural similarities between the HRTF spectral components $\beta_n^m(k)$ and the spherical Bessel functions of the first kind. Top plots and middle plots are the real and imaginary parts of $\beta_n^m(k)$ with (a) $n=0,m=0$ and (b) $n=12,m=0$; and the bottom plots are the spherical Bessel functions $j_n(\cdot)$ at the corresponding degrees $n=0$ and $n=12$ against arguments from 0 to 30.

$$\int_0^1 x^2 j_n(xZ_\ell^{(n)}) j_n(xZ_h^{(n)}) dx = \frac{1}{2} \delta_{\ell,h} (j_{n+1}(Z_\ell^{(n)}))^2, \tag{22}$$

where $Z_\ell^{(n)}$ and $Z_h^{(n)}$ are the positive roots of the $j_n(\cdot)$, and $\delta_{\ell,h}$ is the Dirac delta function. The derived HRTF spectral component representation is

$$\beta_n^m(k) = \sum_{\ell=1}^{\infty} A_{n;\ell}^m j_n\left(\frac{Z_\ell^{(n)}}{k_{\max}} k \right), \tag{23}$$

where from Eq. (22)

$$A_{n;\ell}^m = \frac{2}{k_{\max}^3 j_{n+1}^2(Z_\ell^{(n)})} \int_0^{k_{\max}} k^2 \beta_n^m(k) j_n\left(\frac{Z_\ell^{(n)}}{k_{\max}} k \right) dk. \tag{24}$$

$k_{\max}$ is the maximum wavenumber of a HRTF data set being modeled. In Eq. (23), the HRTF spectral components are decomposed as a linear combination of FSB series. Given

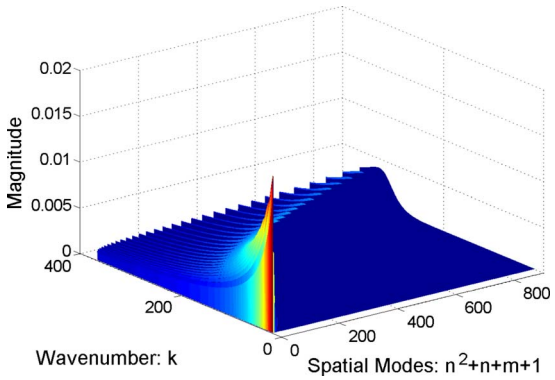

FIG. 5. (Color online) Magnitude of the HRTF spectral components over spatial modes and wavenumber for the spherical head case. The spatial mode of degree n and order m corresponds to number of $n^2+n+m+1$ on x-axis.

the FSB series expansion is convergent, Eq. (23) can be truncated as

$$\beta_n^m(k) \cong \sum_{\ell=1}^{L} A_{n;\ell}^m j_n\left(\frac{Z_\ell^{(n)}}{k_{\max}}k\right), \tag{25}$$

where by choosing L sufficiently large the contribution of the neglected higher order FSB terms can be made sufficiently small. Section IV gives a practical way to determine L.

Besides compact representation, the continuous FSB series can achieve HRTF spectral reconstruction at any frequency value (not necessity of measured frequencies) and therefore provide a way for generating HRTFs at higher spectral resolutions than the measurements.

C. Proposed continuous HRTF model

In summary, the above development leads to the HRTF functional model written as

$$H(r,\hat{\mathbf{y}},k) = \sum_{n=0}^{N}\sum_{m=-n}^{n}\sum_{\ell=1}^{L} A_{n;\ell}^m j_n\left(\frac{Z_\ell^{(n)}}{k_{\max}}k\right) R_n(kr) Y_n^m(\hat{\mathbf{y}}), \tag{26}$$

which can transform any HRTF data set to a set of coefficients $\{A_{n;\ell}^m\}$ of cardinality $(N+1)^2 \times L$. This HRTF representation exhibits three significant advantages.

- First, the representation has well studied closed form orthogonal basis functions, which can make the HRTF approximation easily implemented and model parameters $A_{n;\ell}^m$ simply computed using Eqs. (17) and (24). A low-computation algorithm is developed in Sec. IV given finite discrete measured HRTFs.
- Second, using continuous basis functions, the proposed model is powerful for the computation of the HRTF at any frequency point for an arbitrary direction from a given set of measurements at a fixed radius.
- Third, the basis functions are independent of the data. As the basis is same for all subjects, the model coefficients $A_{n;\ell}^m$ carry all information about the individuality. Thus, the model has capability to represent the individualized HRTF by assigning a subject specific set of parameters to the model.

IV. IMPLEMENTATION ANALYSIS

In this section, we investigate the modal decomposition of the discrete measured HRTFs using the proposed functional model (26). A practical method to solve integral equations (17) and (24) given the typical HRTF measurement setup is introduced in the following.

A. Typical HRTF measurement setup

Typically HRTFs are measured from humans or mannequins for both left and right ears at a fixed radius from the head center. Thus, the source location is specified by a 2D angle, elevation θ, and azimuth ϕ (denoted as a unit vector $\hat{\mathbf{y}}$ in our previous analysis). The elevation angle θ from top to bottom is defined as changing from 0° to 180°; and the azimuth ϕ is counterclockwise rotating from 0° to 360°, where 0° and 180° are the direct front and back directions and 90° and 270° are defined as the left and right sides (note that this definition is in accordance with the right hand coordinate system and may be different from others).

It is commonly believed that the HRTF should be sampled uniformly on the sphere; however, arranging points evenly on the sphere is a complicated mathematical problem. Two most used strategies for HRTF measurement are equidistance in the azimuth arc[15] and equiangular.[16] In the former one, the sampling points are distributed equally in the azimuth arc at all elevations, resulting in a decline of azimuth resolution toward the pole of the sphere. While the latter one applies the equal angular interval along both elevation and azimuth and samples the sphere with very high spatial resolution. These two sampling arrangements are compared in Sec. V using the proposed modal decomposition method, which helps us to thoroughly investigate the azimuth and the elevation sampling effects.

B. Practical modal decomposition method

The proposed modal decomposition method is a two-step procedure corresponding to approximating the two integral equations (17) and (24) given the discrete measured HRTFs.

1. Estimating HRTF spectral components

The HRTF spectral components are obtained from the spherical harmonic transform of the measurements on a single sphere, as shown in Eq. (17). We rewrite the spherical harmonic transform in elevation and azimuth, given $\hat{\mathbf{y}} = \{\theta, \phi\}$, as

$$\beta_n^m(k) = \int_{-\pi}^{\pi}\int_{0}^{\pi} H(\theta,\phi,k)\overline{Y_n^m(\theta,\phi)}\sin\theta d\theta d\phi. \tag{27}$$

Note here we only consider extracting $\beta_n^m(k)$ from far-field HRTFs; same procedure can be applied to near-field measurements with an additional step of dividing the normalized spherical Hankel function.

Let $H(\theta_q,\phi_v,k)$ be the HRTFs measured at several elevations θ_q, $q=1,\ldots,Q$, and several different/same azimuths ϕ_v, $v=1,\ldots,V_q$, at each elevation. We write the far-field measured HRTF decomposition with spherical harmonics (16) as

$$H(\theta_q,\phi_v,k) = \sum_{n=0}^{N}\sum_{m=-n}^{n} \beta_n^m(k) Y_n^m(\theta_p,\phi_v), \tag{28}$$

where the spherical harmonics $Y_n^m(\cdot)$ are defined in terms of the associated Legendre function $P_n^{|m|}(\cdot)$ and the exponential function, as shown in Eq. (2). We use Eq. (2) to express Eq. (28) in terms of the normalized Legendre function $\mathcal{P}_n^{|m|}(\cdot)$ and the normalized exponential function $E_m(\cdot)$ as

$$H(\theta_q,\phi_v,k) = \sum_{n=0}^{N}\sum_{m=-n}^{n} \beta_n^m(k)\mathcal{P}_n^{|m|}(\cos\theta_q)E_m(\phi_v), \tag{29}$$

where $E_m(\phi_v) \triangleq (1/\sqrt{2\pi})e^{im\phi_v}$ and

$$\mathcal{P}_n^{|m|}(\cos\theta_q) \triangleq \sqrt{\frac{2n+1}{2}\frac{(n-|m|)!}{(n+|m|)!}}P_n^{|m|}(\cos\theta_q). \tag{30}$$

Azimuth harmonics. At each elevation with the use of the orthogonality of the exponential functions over circle, we get an *azimuth harmonics*

$$a_m(\theta_q,k)=\sum_{n=|m|}^{N}\beta_n^m(k)\mathcal{P}_n^{|m|}(\cos\theta_q); \tag{31}$$

while given V_q azimuth samplings at each elevation θ_q, we have

$$a_m(\theta_q,k)\approx\Delta\phi_v\sum_{v=1}^{V_q}H(\theta_q,\phi_v,k)E_{-m}(\phi_v),$$

$$|m|\le\lfloor(V_q-1)/2\rfloor, \tag{32}$$

where $\Delta\phi_v$ is the azimuth sampling interval in radians. The azimuthal HRTFs of V_q samples contain at most V_q expansion components, which means we can estimate $a_m(\theta_q,k)$ for $|m|\le\lfloor(V_q-1)/2\rfloor$. However, if the sampling is non-uniform, we should emphasize that approximation (32) is determined by the maximum sampling interval; only the coefficients of $|m|\le\lfloor(2\pi/\Delta\phi_v^{\max}-1)/2\rfloor$ can be accurately solved.

Least-squares fitting. By writing Eq. (31) for a specific order of m for all measured elevations, we can now form a system of simultaneous equations given by

$$\mathbf{P}_m\mathbf{b}_m=\mathbf{a}_m,\quad m=-N,\ldots,N, \tag{33}$$

where the matrix $\mathbf{P}_m$ and the vector $\mathbf{b}_m$ are in the following forms:

$$\mathbf{P}_m=\begin{bmatrix}\mathcal{P}_{|m|}^{|m|}(\cos\theta_1) & \cdots & \mathcal{P}_N^{|m|}(\cos\theta_1)\\ \vdots & \ddots & \vdots\\ \mathcal{P}_{|m|}^{|m|}(\cos\theta_Q) & \cdots & \mathcal{P}_N^{|m|}(\cos\theta_Q)\end{bmatrix}, \tag{34}$$

$$\mathbf{b}_m=[\beta_{|m|}^m(k),\beta_{|m|+1}^m(k),\ldots,\beta_N^m(k)]^T, \tag{35}$$

and

$$\mathbf{a}_m=[a_m(\theta_1,k),a_m(\theta_2,k),\ldots,a_m(\theta_Q,k)]^T. \tag{36}$$

The HRTF spectral components $\beta_n^m(k)$ can be calculated by solving these linear equations described by Eq. (33) for each order m. Since there will be noise in the HRTF measurement, it is necessary to solve Eq. (33) in the least-squares sense by minimizing the mean squared error $\|\mathbf{P}_m\mathbf{b}_m-\mathbf{a}_m\|^2$. Another issue in the HRTF measurement is that no samplings are made for lower elevations (i.e., $\theta\ge140°$) because of the strong distortions from the ground and measurement apparatus. To avoid the enlargement of the unmeasured HRTFs, we need to regularize the solution (the power in the $\|\mathbf{b}_m\|^2$ may be included as a constraint). The minimum norm least-squares solution is denoted by

$$\mathbf{b}_m^+=\mathbf{P}_m^+\mathbf{a}_m, \tag{37}$$

where $\mathbf{P}_m^+$ is the general inverse of $\mathbf{P}_m$.[29] Given the size of $\mathbf{P}_m$ is $Q\times(N-|m|+1)$, there are two cases of interest and the Tikhonov regularized solutions are given explicitly by

$$\mathbf{P}_m^+=[\mathbf{P}_m^T\mathbf{P}_m+\lambda\mathbf{I}]^{-1}\mathbf{P}_m^T,\quad Q\ge(N-|m|+1), \tag{38}$$

$$\mathbf{P}_m^+=\mathbf{P}_m^T[\mathbf{P}_m\mathbf{P}_m^T+\lambda\mathbf{I}]^{-1},\quad Q<(N-|m|+1), \tag{39}$$

where λ is the regularization control parameter and $\mathbf{I}$ is the identity matrix. A systematic approach to evaluate λ for a meaningful result is given in the work.[30] In our experiment, we set a small value of $\lambda=10^{-5}$, which was seen to achieve reasonable reconstruction and interpolation quality.

Insights into spatial sampling. The main contribution of this low-computation algorithm is based on factorization of the spherical harmonics, which helps to separate the azimuth and the elevation sampling effects. We have the following comments regarding on the HRTF spatial sampling.

(1) In theory, it is necessary to set Eq. (33) as an overdetermined system; i.e., the number of elevation samples should be greater than $(N+1)$, so that the least-squares solutions are valid. However, with the use of the regularization technique, we can loosen this condition. Our experiment results show the HRTF measurements that are coarsely sampled in elevation (given the total number of samples greater than the dimensionality) can still be reconstructed with reasonable accuracy.
(2) For smaller elevations (θ toward the pole), the associated Legendre functions $P_n^{|m|}(\cos\theta)$ have values close to zero for higher m; i.e., $P_n^{|m|}(\cos 0)=0$, for $m\ne0$.[31] This means as elevation increases from the pole toward the equator, higher order m of coefficients begin to appear. Thus, in principle, we need less dense azimuth sampling closer to the pole and more azimuth sampling points on the elevations closer to the equator. This shows that the sampling of equidistance in the azimuth arc is appropriate for the HRTF measurement, which we will further corroborate using the real data validation.

2. Calculating model coefficients

From the estimated HRTF spectral components $\beta_n^m(k)$, the model coefficient $A_{n;\ell}^m$ is obtained by using the left Riemann sum to approximate integral (24). The most important issue is to determine the truncation order L. We define the relative power of the ℓth order FSB series term against the total power as

$$\eta_\ell=\frac{|A_{n;\ell}^m|^2}{\sum_{\ell=1}^{\infty}|A_{n;\ell}^m|^2},\quad \ell=1,2,\ldots. \tag{40}$$

In Eq. (40), the denominator is an infinite sum over model coefficients. Since only HRTFs at discrete frequencies are obtained by measurements, we evaluate the contribution of the FSB series over the maximum order L_f (the number of HRTF frequency samples); and the relative power ratio is defined as

$$\eta=\frac{\sum_{\ell=1}^{L}|A_{n;\ell}^m|^2}{\sum_{\ell=1}^{L_f}|A_{n;\ell}^m|^2}. \tag{41}$$

TABLE I. MIT KEMAR data measurement steps (angles in degrees).

Elevation (θ)	Azimuth resolution (ϕ)	No. of azimuthal measurements
70–110	5.00	72
60 and 120	6.00	60
50 and 130	6.43	56
40	8.00	45
30	10.00	36
20	15.00	24
10	30.00	12
0	⋯	1

Then for each HRTF spectral component $\beta_n^m(k)$, calculate η for $L=1,2,\ldots,L_f$ and when η reaches a power threshold (such as 0.9), L is chosen as the truncation order above which the contribution of higher order FSB series is negligible.

V. SIMULATION RESULTS

A. HRTF database

Three sets of HRTF database are used.

(1) *Analytically simulated HRTF from the spherical head model.*[17] The HRTF for an ideal rigid sphere is defined as the pressure on the sphere at the defined ear position divided by the pressure that would exist at the sphere center in the absence of the sphere. The synthetic data are without noise influence and provide reliable reference to check the proposed sampling theory and the continuous model performance.

(2) *The HRTF database for KEMAR from the MIT media laboratory.*[15] KEMAR is designed according to the mean anatomical size of the population; thus results from KEMAR HRTF represent the mean performance. The measurements are the head-related impulse responses in the time domain at 44.1 kHz sampling rate and each response is 512 samples long, from which a 512-tap HRTF can be obtained by the discrete-time Fourier transform. The speakers were at a distance of 1.4 m away from the head center. The HRIRs (or HRTFs) were sampled in the equidistance in the azimuth arc, where the measurements are available for elevation steps of 10° ranging from 0° (north pole) to 130° (40° underneath the horizontal plane) and for full azimuth cover but have a decline of azimuthal resolution toward the pole, as shown in Table I.

(3) *HRTF database of human subjects from CIPIC interface laboratory.*[16] The HRIR measurements performed at CIPIC include 45 subjects. A 200 samples long pseudo-random signal generated by the snapshot system (sampling frequency is 44.1 kHz) is used as the test signal. For each subject, the HRTFs are measured at 1250 points on the sphere of 1 m away from the listener. The elevation varies uniformly from 0° to 135° in the step of 5.625°; and there are 50 azimuth samples at each elevation but not uniformly sampled; i.e., $\phi=[0:5:45, 55,65,80,100,115,125,135:5:225,235,245,260,280, 295,305,315:5:355]°$. The samplings are more dense near the median plane but very coarse near the ear where the azimuth varies in the step of 20°.

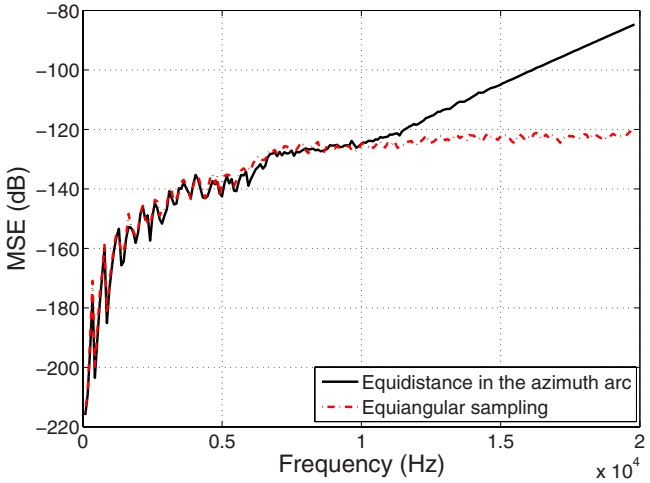

FIG. 6. (Color online) Synthetic HRTF reconstruction error performance for the audible frequency range of [0.2, 20] kHz.

B. Results for dimensionality and analysis

Simulations are run on some audible frequency range for each HRTF database, where the total number of spatial samples determines the maximum frequency point that can be reconstructed with high accuracy. The relative mean square error (MSE) over all M angles (including both azimuth and elevation) at each frequency is used as the error metric

$$\varepsilon(f)=\frac{\sum_{j=1}^{M}|H(f,\hat{\mathbf{y}}_j)-\tilde{H}(f,\hat{\mathbf{y}}_j)|^2}{\sum_{j=1}^{M}|H(f,\hat{\mathbf{y}}_j)|^2}, \tag{42}$$

where $H(f,\hat{\mathbf{y}}_j)$ and $\tilde{H}(f,\hat{\mathbf{y}}_j)$ are the original and the reconstructed HRTFs, respectively.

Figure 6 plots the synthetic HRTF reconstruction performance for the whole audible frequency range up to 20 kHz. The analytically simulated HRTFs are generated at 1.0 m away from the head center on a sphere according to the equidistance in the azimuth arc sampling (2640 samples) and the equiangular sampling (4371 samples); both satisfy the required dimensionality (2209 samples) for reconstruction up to 20 kHz. We can see that the maximum reconstruction error is at the highest frequency. This shows that more than dimensionality large number of measurements can fit the low frequency data very well. Both equiangular and equidistance samplings have very small reconstruction errors; but the equiangular method needs much more samples and its biggest failure is that the sampling points near the pole are dense, small, and can be very distorted when measurements in this region are contaminated by noise.

We next investigate the dimensionality results of the HRTF measurements on KEMAR mannequin and human subjects. Figure 7 shows that both data set reconstruction

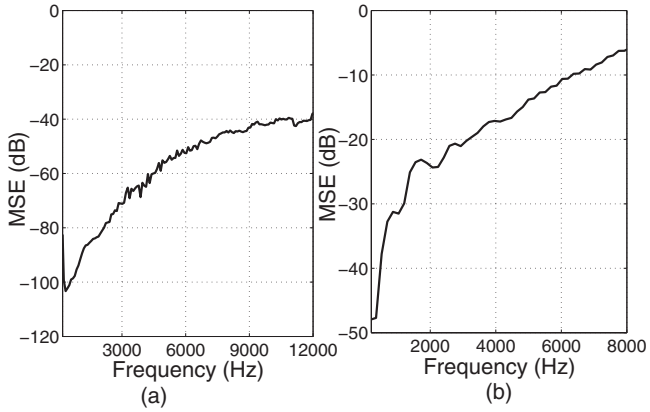

FIG. 7. HRTF measurement reconstruction error performances. (a) MIT KEMAR mannequin measurements of frequency range [0.2, 12] kHz. (b) HRTFs of CIPIC subject 3 of frequency range [0.2, 8] kHz.

errors are larger than that of the theoretical model due to the possibility of noise contamination at some measurements. This is especially more likely to occur for human subjects where the movement of the subject in the measurements can lead to inconsistency in the measured response. Thus, we can see that the reconstruction of KEMAR HRTFs is more accurate than that of the human subject data. In addition, both data sets have very similar error pattern.

As introduced in Sec. V A, the MIT KEMAR measurements are equidistance sampled, which has 72 azimuth samples on the horizontal plane and less azimuth samplings for the elevations toward the pole. In total there are 710 spatial samples on the sphere, which means we can solve the spatial mode decomposition up to $N=25$ corresponding to the frequency about 12 kHz according to Eq. (11). In Fig. 7(a), the MIT data reconstruction shows a reasonable match to the original data in the frequency range of [0.2, 12] kHz with maximum error less than −40 dB.

The CIPIC data have finer elevation samplings but are not uniformly sampled in azimuth. Even though the CIPIC measurements are sampled at a much higher spatial resolution (1250 samples on sphere), it has even larger errors [Fig. 7(b)] compared to MIT measurements. This is due to the possible large measurement variations and the coarse azimuthal sampling (50 not uniformly azimuths at each elevation). Very large azimuthal interval of 20° at both ear sides determines that the model coefficients can only be solved accurately for low order N. This is corroborated in Fig. 7(b), where the CIPIC data reconstruction errors are less than −17 dB for $f \leq 4$ kHz and increase to large values for higher frequencies.

In summary, the simulation results prove that the proposed dimensionality (11) determines the required number of spatial samples in the HRTF measurement. Only when the number of measurements is larger than the required dimensionality for a given frequency range (or a particular frequency point), reasonable reconstruction with high accuracy can be achieved. As for the HRTF measurement, equidistance in the azimuth arc is appropriate; with the use of the regularization technique, the spatial sampling for elevations can be coarse while the sampling along the azimuth should be finer (especially for the measurements close to the equator).

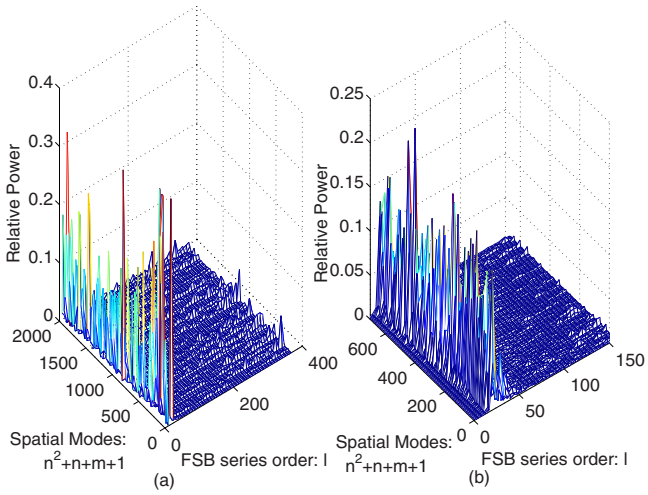

FIG. 8. (Color online) The relative power distribution of the FSB series components for the HRTF spectral representation. (a) Analytically simulated HRTFs from the spherical head model and (b) MIT KEMAR mannequin left ear HRTFs. For all spatial modes, the relative contribution of lower order FSB series is significant.

C. Continuous model performance

1. HRTF reconstruction results

Figure 8 shows how the relative power distribution of each FSB series coefficient varies with the spatial modes. We can clearly see that for all spatial modes, the relative contribution of lower order FSB series is significant, which corresponds to the smooth HRTF spectral variations. It also proves that using the relative power ratio as the criterion to choose the truncation order of the FSB series expansion is appropriate. We suggest the truncation order of the FSB series expansion based on the power criterion of 0.9. Table II summarizes the number of FSB series and the number of spatial modes (i.e., the dimensionality results given in Sec. V B) for the three sets of HRTF database representation for a given frequency range. Note that we only validate the CIPIC data at low frequencies here because its spectral components are accurately solved up to 4 kHz, as stated in Sec. V B. It can be seen that the number of FSB series for the HRTF spectral representation increases with frequency; in addition, the human subjects' HRTF needs more basis functions to emulate.

TABLE II. Summary of the number of spatial modes and the number of FSB series for the three sets of HRTF database.

	Given frequency range (kHz)	No. of spatial modes	No. of FSB series	Average MSE (dB)
Synthetic HRTF	[0.2, 20]	$47^2=2209$	85	−78.7
KEMAR HRTF	[0.2, 12]	$26^2=676$	67	−28.6
Subject HRTF	[0.2, 4]	$16^2=256$	16	−8.6

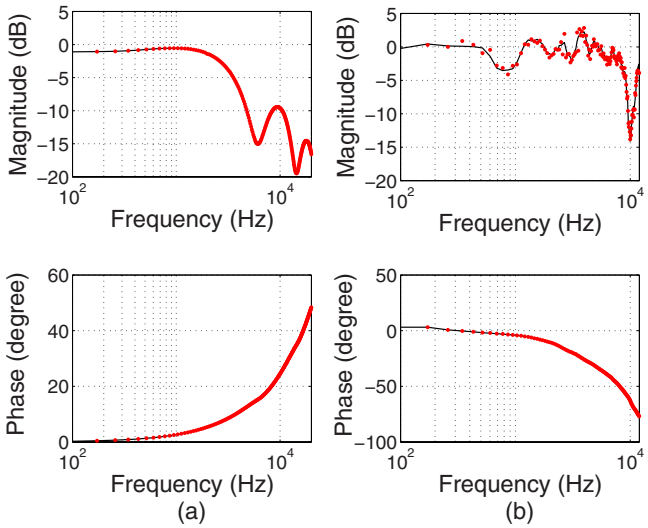

FIG. 9. (Color online) Examples of analytical simulated and measured HRTF reconstruction using the proposed continuous model. (a) Analytical simulated HRTFs at elevation 90° and azimuth of 80° (−85.3 dB error) and (b) left ear MIT KEMAR data at elevation 60° and azimuth of 0° (−41.5 dB error). Original: dotted line (·) and reconstruction: solid line (–).

Figure 9 plots the original and reconstructed HRTF magnitude and phase for synthetic HRTF (θ=90°, ϕ=80°) and KEMAR left ear measurements (θ=60°, ϕ=0°). The reconstruction errors for both data sets are shown in Table II. It is clear that the reconstructed responses closely match the synthetic and the KEMAR responses in both cases. We also use CIPIC subject measurements to check the model performance. The emulation error of subjects' HRTF tends to be a larger value (average MSE around −8.6 dB), which demonstrates that human subjects are harder to model than the spherical head and the KEMAR mannequin.

2. HRTF interpolation and range extrapolation

We further investigate the HRTF interpolation and range extrapolation performances using the proposed continuous model. The MIT KEMAR data are measured at the sampling frequency of 44.1 kHz with 512 samples for each measurement. Figure 10 plots the polar response magnitudes for data

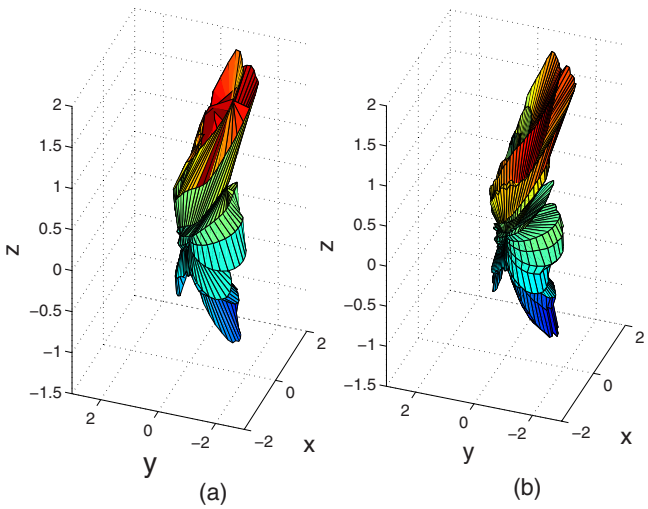

FIG. 10. (Color online) MIT left ear HRTF polar response at 8 kHz: (a) original data and (b) synthesized response over the sphere.

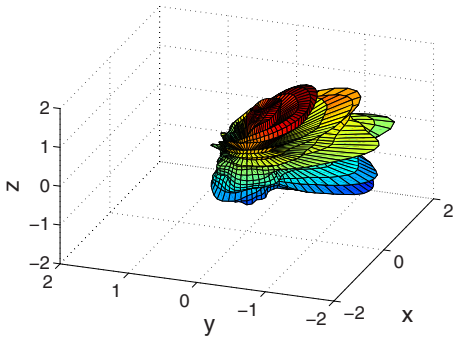

FIG. 11. (Color online) MIT left ear HRTF polar response at 4.15 kHz, which is not measured but interpolated by applying the modal decomposed coefficients to the continuous spectral modeling basis functions, FBS series.

at 8 kHz, where the synthesized polar responses (generated at much higher spatial resolutions of $\Delta\theta$=5° and $\Delta\phi$=5° at each elevation) are smooth forms of the original data and the match is reasonably accurate. In addition, the polar response at not measured frequency (f=4.15 kHz) is interpolated by applying the decomposed model coefficients to the continuous HRTF model, as shown in Fig. 11. We can see that the proposed continuous FSB series can achieve reasonable HRTF spectrum interpolation.

In Fig. 12, the plots on the left are the magnitudes of the analytical HRTFs at different ranges on the horizontal plane, compared to the range extrapolation results from the proposed model on the right. We observe that the reconstruction is perfect with average approximation error around 0.52% (−45 dB).

3. Discussion

We summarize the performance of the proposed continuous model in the following three aspects. First, the proposed continuous functional HRTF model provides accurate reconstruction to the experimental measurements. The interpolated results are also reasonable emulations. Second, as given in Table II, each of the individualized HRTF data set is trans-

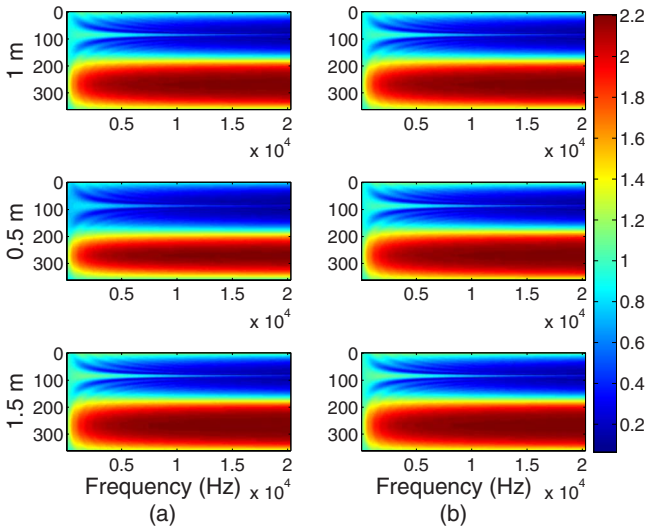

FIG. 12. (Color online) Analytically simulated HRTFs at r=1.0 m (top plots) and the extrapolated HRTFs at r=0.5 and r=1.5 m (b) compared to the reference (a). The horizontal axis is frequency and the vertical axis is azimuth from [0°, 360°).

formed to a set of coefficients. This coefficient set is much smaller in size compared to the original HRTF database. For example, for frequency range of [0.2, 12] kHz, original MIT databases have 81 920 sample points (710 directions and 160 frequency samples for each position); now the transformed coefficient just has 45 292 entries ($N=25$, $\ell=67$). The data that need to be saved have been reduced by nearly 45%. Compared to the statistical PCA model,[8] which is truly the optimal low-dimensional description for the HRTF data set, the disadvantage of our model using more basis functions is countered by the universality (data independent and measurement grid independence of the basis) and the continuous nature of the basis functions (eliminating the need for interpolation). Third, the proposed model can be regarded as noise discriminated as the basis functions we choose have structural similarities to the HRTF being analyzed. Thus, the unwanted components (noise or distortion) will not be represented with the same accuracy as the signal interested. For example, the noise components of high spatial bandwidth ($n>N$) are removed and the noise with frequency components outside the triangular shaped region will be significantly reduced.

VI. CONCLUSION

A general HRTF representation in all frequency-range-angle domains was developed in this paper. The HRTF spatial dimensionality is defined as the required number of spatial modes to represent HRTFs corresponding to all directions. A continuous functional model can represent the HRTF in both spatial and spectral domains. The model 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. A practical method was developed to obtain the model coefficients. We observed good HRTF spatial and spectral components' reconstruction and interpolation results from both analytical solutions and measurement data. We also need to state that the current approach is dealing with the representation of empirical measurements at the technical level. Psychoacoustic validation has to be performed in the future to confirm the error bounds and the truncation orders given in the paper.

[1] L. Rayleigh, "On our perception of sound direction," Philos. Mag. **13**, 214–232 (1907).
[2] J. Blauert, *Spatial Hearing-Revised Edition: The Psychophysics of Human Sound Localization* (MIT, Cambridge, MA, 1996).
[3] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening I: Stimulus synthesis II: Psychophysical validation," J. Acoust. Soc. Am. **85**, 858–878 (1989).
[4] J. Braasch and K. Hartung, "Localization of distracted sound sources: Determining the role of binaural cues using unilaterally attenuated and interaurally uncorrelated signals (A)," J. Acoust. Soc. Am. **105**, 1151 (1999).
[5] S. Carlile, "The auditory periphery of the ferret. I: Directional response properties and the pattern of interaural level differences," J. Acoust. Soc. Am. **88**, 2180–2195 (1990).
[6] A. Kulkarni and H. S. Colburn, "Infinite-impulse-response models of the head-related transfer function," J. Acoust. Soc. Am. **115**, 1714–1728 (2004).
[7] W. L. Martens, "Principal components analysis and resynthesis of spectral cues to perceived direction," in Proceedings of the 1987 International Computer Music Conference (1987), pp. 274–281.
[8] 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**, 1637–1647 (1992).
[9] K. Hartung, J. Braasch, and S. J. Sterbing, "Comparison of different methods for the interpolation of head-related transfer functions," in Proceedings of the 16th Audio Engineering Society International Conference: Spatial Sound Reproduction (1999), pp. 319–329.
[10] S. Carlile, C. Jin, and V. V. Raad, "Continuous virtual auditory space using HRTF interpolation: Acoustic and psychophysical errors," in Proceedings of the IEEE 2000 International Symposium on Multimedia Information Processing (2000), pp. 220–223.
[11] M. Matsumoto, S. Yamanaka, M. Tohyama, and H. Nomura, "Effect of arrival time correction on the accuracy of binaural impulse response interpolation," J. Audio Eng. Soc. **52**, 56–61 (2004).
[12] 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. **104**, 2400–2411 (1998).
[13] 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, ICASSP 2004 (2004), Vol. **IV**, pp. 45–48.
[14] W. Zhang, R. A. Kennedy, and T. D. Abhayapala, "Efficient continuous HRTF model using data independent basis functions: Experimentally guided approach," IEEE Trans. Audio, Speech, Lang. Process. **17**, 819–829 (2009).
[15] W. G. Gardner and K. D. Martin, "HRTF measurements of a KEMAR," J. Acoust. Soc. Am. **97**, 3907–3908 (1995).
[16] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in Proceedings of the 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (2001), pp. 99–102.
[17] R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," J. Acoust. Soc. Am. **104**, 3048–3058 (1998).
[18] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, NJ, 1987).
[19] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory* (Springer, New York, 1998).
[20] F. J. Simons, F. A. Dahlen, and M. A. Wieczorek, "Spatiospectral concentration on a sphere," SIAM Rev. **48**, 504–536 (2006).
[21] N. A. Gumerov and R. Duraiswami, *Fast Multipole Methods for the Helmholtz Equation in Three Dimensions* (Elsevier, Oxford, 2005).
[22] R. A. Kennedy, P. Sadeghi, T. D. Abhayapala, and H. M. Jones, "Intrinsic limits of dimensionality and richness in random multipath fields," IEEE Trans. Signal Process. **55**, 2542–2556 (2007).
[23] T. D. Abhayapala, T. S. Pollock, and R. A. Kennedy, "Characterization of 3D spatial wireless channels," in Proceedings of the IEEE 58th Vehicular Technology Conference, VTC 2003-Fall (2003), Vol. **1**, pp. 123–127.
[24] V. R. Algazi, C. Avendano, and R. O. Duda, "Elevation localization and head-related transfer function analysis at low frequencies," J. Acoust. Soc. Am. **109**, 1110–1122 (2001).
[25] E. A. G. Shaw, "Acoustical features of the human external ear," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Mahwah, NJ, 1997), pp. 25–47.
[26] W. Zhang, T. D. Abhayapala, R. A. Kennedy, and R. Duraiswami, "Modal expansion of HRTFs: Continuous representation in frequency-range-angle," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2009 (2009), pp. 285–288.
[27] T. Ajdler, C. Faller, L. Sbaiz, and M. Vetterli, "Sound field analysis along a circle and its applications to HRTFs interpolation," J. Audio Eng. Soc. **56**, 156–175 (2008).
[28] T. D. Abhayapala, "Generalized framework for spherical microphone arrays: Spatial and frequency decomposition," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2008 (2008), pp. 5268–5271.
[29] C. R. Rao and S. K. Mitra, *Generalized Inverse of Matrices and Applications* (Wiley, New York, 1972).
[30] D. N. Zotkin, R. Duraiswami, and N. A. Gumerov, "Regularized HRTF fitting using spherical harmonics," in Proceedings of the 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (2009), pp. 257–260.
[31] T. D. Abhayapala and A. Gupta, "Alternatives to spherical microphone arrays: Hybrid geometries," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2009 (2009), pp. 81–84.