EFFICIENCY EVALUATION AND ORTHOGONAL BASIS DETERMINATION IN FUNCTIONAL HRTF MODELING

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

Applied Signal Processing Group, CECS, The Australian National University, Australia
Email: karan.zhang@cecs.anu.edu.au, \{rodney.kennedy, thushara.abhayapala\}@anu.edu.au

ABSTRACT

This paper considers the problem of how to evaluate the efficiency of a 3D continuous functional HRTF model in representing measured data. The proposed method is based on Karhunen-Loève theorem. After investigating the optimization problem of representing the process with an finite linear combination of orthogonal functions in \( L^2 \) space by means of the least squared method, the variance of random variables involved in the model is found the key metric in efficiency evaluation of a given functional model. Then, the coefficients of the 3D continuous HRTF model are analyzed. The results show that the efficiency of the model in spatial component expansion is around 70% and the best choice in frequency component expansion is the spherical Bessel function.

Index Terms — HRTF Model, Functional Model, Karhunen-Loève Theorem, Least Squared Method

1. INTRODUCTION

Many applications, such as virtual sound synthesis, binaural noise reduction, sound localization, sound separation, require individualized Head-Related Transfer Functions (HRTFs), which can be measured in an anechoic room [1]. However, such a database can only provide HRTFs for a finite number of source positions. In the attempt to represent individualized HRTFs in the entire auditory scene, it is useful to consider a continuous representation [2].

A number of studies have established filter-band models and geometric models for HRTFs, such as BEM-based model [3] and the snowman model [4]. Although the sound synthesized by these models are very close to that synthesized by the original HRTF measurements, the coefficients in the models are still complicated functions of angle and frequency, which limits their wide application.

The use of surface spherical harmonics [5] extracts the directional cues from HRTFs as continuous functions while the frequency cues are still embedded in the coefficients. Recently, a fast multipole accelerated boundary element method [6] for computation of HRTFs was developed. Further, a similar continuous functional HRTF model [2] separated the frequency components from the spatial components, which is easy to use and can reconstruct HRTFs at any arbitrary position in space and at any frequency point as well. But, 1) the efficiency of the model and 2) whether alternative representations are superior are key questions.

The aim of this paper is to answer these two questions. We firstly give a brief introduction of the 3D continuous HRTF model [2], which is under the evaluation. Then, based on Karhunen-Loève theorem [7], we identify that the sum of the variances of the random variables involved in the model is the key metric in efficiency evaluation and guides the choice of the orthogonal basis used in the functional modeling. Finally, by using the proposed method, the evaluation is separately conducted on spatial and frequency component expansion of the continuous HRTF model.

2. 3D FAR-FIELD CONTINUOUS HRTF MODEL

The 3D Far-Field continuous HRTF model [2] can be used to reconstruct HRTFs at any arbitrary position in space and at any frequency point from a finite number of measurements. In this model, the HRTF spatial components are expanded using spherical harmonics, i.e.,

\[
\hat{H}(\theta_s, \phi_s, k) = \sum_{n=0}^{N(k)} \sum_{m=-n}^{n} \gamma_n^m(k) Y_n^m(\theta_s, \phi_s),
\]

where \( \theta_s \) and \( \phi_s \) define the source position, \( k = 2\pi f / c \) is the wavenumber, \( Y_n^m(\theta_s, \phi_s) \) are the spherical harmonics characterized by degree \( n \) and order \( m \), \( \gamma_n^m(k) \) are the spherical harmonic coefficients of the equivalent source field at wavenumber \( k \), and the maximum degree can be effectively truncated to \( N(k) = e c k s / 2 \), in which \( c \approx 2.7183 \) and \( s = 9 \) cm [2].

Given a set of HRTF measurements, we can compute \( \gamma_n^m(k) \) by

\[
\gamma_n^m(k) = \int_{S^2} H(\theta_s, \phi_s, k) Y_n^m(\theta_s, \phi_s) \, d\theta_s d\phi_s,
\]

where \( H(\theta_s, \phi_s, k) \) are HRTFs measured on the 2-sphere \( S^2 \) at \( k \). Since \( \gamma_n^m(k) \) are still functions of frequency, further
decomposition is necessary (Details in [2]). It is given by

$$\gamma_n^m(k) = \sum_{q=1}^{Q} A_{n,q}^m \sqrt{\frac{2}{j_n(Z_q^{(n)})}} k_{\text{max}} j_n(\frac{Z_q^{(n)}}{k_{\text{max}} - k}),$$

(3)

where $A_{n,q}^m$ are random variables, termed as Fourier spherical Bessel coefficients, $j_n(\cdot)$ is the spherical Bessel function of the degree $n$, $Z_q^{(n)}$ are the positive roots of $j_n(\cdot)$.

3. EFFICIENCY EVALUATION OF FUNCTIONAL MODEL

We assume a random process can be expressed as a linear combination on infinite deterministic orthonormal functions weighted by random variables. We represent/approximate this process using a finite number of the orthonormal functions. Further, the functions are orthonormal with respect to an inner product $\langle \cdot, \cdot \rangle$. Then, the expected value of the squared norm of error of the representation is

$$\Delta = E\{\langle \mathbf{X}, \mathbf{X} \rangle \} - E\{\langle \hat{\mathbf{X}}, \hat{\mathbf{X}} \rangle\}$$

(4)

where $\mathbf{X}$ indicates the random process and $\hat{\mathbf{X}}$ indicates the finite dimensional representation of the random process.

Since the random process $\mathbf{X}$ is independent of the representation $\hat{\mathbf{X}}$, to minimize $\Delta$ is equivalent to maximize $E\{\langle \hat{\mathbf{X}}, \hat{\mathbf{X}} \rangle\}$, which means the most efficient representation must have the largest expected value of the squared norm.

3.1. Key parameter identification

Suppose $x(t)$ is a random process defined on a finite interval $|t| \leq T$ with zero-mean and finite energy. Given a set of orthonormal basis $\{\psi_n(t)\}$, $x(t)$ can be expanded as

$$x(t) = \sum_{n=1}^{\infty} \alpha_n \psi_n(t), \quad |t| \leq T,$$

(5)

where $\alpha_n$ are random variables, $\{\psi_n(t)\}$ are deterministic functions which are orthonormal complete basis in $L^2(T)$ with respect to the inner product. Then the approximation by means of the truncated $N$-dimensional representation, where $n = N$, can be made arbitrarily close by choosing

$$\alpha_n = \langle x(t), \psi_n \rangle = \int_{-T}^{T} x(t) \psi_n^*(t) \, dt,$$

(6)

and making $N$ sufficiently large [7, 8].

For the $N$-dimensional representation, the squared $L^2(T)$ norm of error for a particular realization of the random process is $\|\epsilon_i\|^2 = \langle \epsilon_i, \epsilon_i \rangle$ where

$$\epsilon_i = x_i(t) - \sum_{n=1}^{N} \alpha_{i,n} \psi_n(t),$$

(7)

and $i$ is the index of the realizations.

Let $\Delta_{\psi}$ denote the expected value of the squared norm of error given by the selected $N$-basis functions $\{\psi_n\}$. Then we have

$$\Delta_{\psi} = E\{\langle \epsilon, \epsilon \rangle\} = E\{\langle \mathbf{X}, \mathbf{X} \rangle\} - E\{\sum_{n=1}^{N} |\alpha_n|^2\}.$$  

(8)

As said before, to minimize $\Delta_{\psi}$, we need maximize the second term in (8), the expected value of the $N$ terms energy of $\alpha$, denoted by $A_{\psi}$ when $\alpha_n$ is calculated from $\{\psi_n(t)\}$. Hence, using (6), we have

$$A_{\psi} = E\{\sum_{n=1}^{N} \int_{-T}^{T} x(t) \psi_n^*(t) \psi_n(s) \, dt \, ds\}$$

$$= \sum_{n=1}^{N} \int_{-T}^{T} k_{xx}(t, s) \psi_n^*(t) \psi_n(s) \, dt \, ds,$$

(9)

where $k_{xx}(t, s) \triangleq E\{x(t) x^*(s)\}$ is the autocorrelation function of the random process. Notice that $A_{\psi}$ depends on the choice of the $\{\psi_n\}$.

Let $\xi_n(t)$ be the eigenfunctions of the following integral equation with kernel $k_{xx}(t, s)$,

$$\int_{-T}^{T} k_{xx}(t, s) \xi_n(s) \, ds = \lambda_n \xi_n(t).$$

(10)

Applying (10) into (9), we can calculate $A_{\xi} = \sum_{n=1}^{N} \lambda_n$. Then the mean squared error of the $N$-dimensional representation when the orthonormal basis is chosen to be the eigenfunctions corresponding to the $N$ largest eigenvalues is

$$\Delta_{\xi} = \sum_{n=1}^{N} \int_{-T}^{T} k_{xx}(t, t) \, dt - A_{\xi} = \sum_{n=N+1}^{\infty} \lambda_n.$$

(11)

It is the minimum [8] since for any other orthonormal set, $\{\psi_n(t)\}$,

$$A_{\psi} = \sum_{n=1}^{N} \int_{-T}^{T} k_{xx}(t, s) \psi_n^*(t) \psi_n(s) \, dt \, ds \geq A_{\xi}.$$  

(12)

Clearly, $A_{\psi}$ is maximized by the optimal expansion, also known as Karhunen-Loève expansion, with the basis of $N$ eigenfunctions of the integral equation (10) whose kernel is the autocorrelation function of the random process.

Since $\mathbf{X}$ is a zero-mean process, $\{\alpha_n\}$ are also zero-mean,

$$E\{\alpha_n\} = E\{\int_{-T}^{T} x(t) \psi_n^*(t) \, dt\} = \int_{-T}^{T} E\{x(t)\} \psi_n^*(t) \, dt = 0.$$  

(13)

Then, $A_{\psi}$ is the sum of the $N$ largest variance of $\alpha_n$ and it will not larger than that given by the KL expansion according
to (12). Thus, the sum of the \( N \) largest variances among the \( \alpha_n \) is the key parameter to judge how close of the given orthonormal basis to the optimal basis even though the random process is unknowable.

3.2. Evaluation Method

As long as we know the autocorrelation functions of the process, the optimal functional model can be derived in principle from the Equation (10). However, quite often, the distribution properties in the random process under analysis are too complicated to be described simply or in closed form. As a result, it is impossible to find the KL expansion for the random process. In such a situation, two circumstances are under consideration.

1) Since the representation is not optimal expansion, redundancy must exists in the model. Then, the question is how to measure the redundancy, or equivalent, the efficiency.

As we know more terms in the model imply less error in the representation. Then there is a tradeoff between efficiency and accuracy. Once we have a \( N \)-terms expansion, we can calculate the variances for \( N \) coefficients. It can be found that the variances of some coefficients are very small, which means the corresponding terms can be removed from the model without enlarging the error significantly. Then only \( N' \) (\( N' < N \)) terms are left in the model while achieving similar accuracy. Hence, we define the efficiency of the model \( \eta \) as

\[
\eta = \frac{N'}{N} \times 100\%.
\] (14)

2) Although the eigenfunctions \( \{\xi_n\} \) are not known or not determinable, arbitrary orthonormal basis \( \{\psi_n\} \) are usable. But there are many choices and there is the question of how close of each orthonormal basis to the optimal basis.

Different orthonormal bases give different sets of coefficients whose variances depend on the choice of the basis. For the same terms, the sums of variance of different sets of coefficients are different. Combining the previous conclusion, we can determine that the basis providing the largest sum of variance is the best basis.

4. APPLICATION OF EFFICIENCY EVALUATION IN HRTF MODELING

In this section, we apply the propose method to evaluate the efficiency of the 3D far-field continuous HRTF model.

4.1. Efficiency in Spatial Component Expansion

Equation (1) shows that the value of \( N(k) \) depends on the maximum wavenumber \( k_{\text{max}} \) and the radius of the sphere \( s \) [2]. Thereupon, \( M(k) \),

\[
M(k) = (N(k) + 1)^2
\] (15)
terms of spherical harmonics are required to represent HRTFs at a particular wavenumber \( k \) corresponding to all directions.

To examine if significant redundancy exists in this expansion, we compute the variance of the spherical harmonic coefficients along each frequency and sort them in descending order. The variance of \( M(k) = 256 \) spherical harmonic coefficients for the frequency at 4kHz are plotted in Figure 1 (a). We find that, for the frequency of 4 kHz, the cumulative energy of 189 largest variances cover 99.9\% of the total energy of 256 variances. The stars in Figure 1 (b) indicate the number of spherical harmonic coefficients with 99.9\% energy of the total energy of \( M(k) \) variances, denoted by \( M'(k) \), and the triangles show the number calculated from (15).

Figure 2 presents the comparison among the measured left ear HRTF at azimuth of \(-45^\circ\) and elevation of \(0^\circ\) and the reconstructed HRTFs using the calculated \( M(k) \) and the further truncated \( M'(k) \) number of spherical harmonics (\( M = 256, M' = 189 \) @ 4kHz).

From the above analysis, the differences of the number shown in Figure 1 (b) reveal the redundancy, or equivalently, the efficiency. According to 14, the efficiency is

\[
\eta = \frac{\sum_{k=0}^{k_{\text{max}} \rightarrow 0} M'(k)}{\sum_{k=0}^{k_{\text{max}}} M(k)} \times 100\%.
\] (16)

Then, the efficiency of the continuous HRTF model given by
Table 1. Properties of HRTF Frequency Component Models Using Four set of Orthogonal Basis with Argument Normalized as \( k' = k/k_{\text{max}} \) where \( k \in [0, k_{\text{max}}] \)

| Orthogonal Functions | Argument | Coefficient \((A_{n,q}^m)^2\) | \( E\left\{ \sum_{q=1}^{Q} |A_{n,q}^m|^2 \right\} \) |
|----------------------|----------|--------------------------|---------------------------------|
| Complex Exponentials  | \( e^{i\cdot} \) | \( k' \) | \( \frac{1}{k_{\text{max}}} \int_0^{k_{\text{max}}} \gamma_n^m(k)e^{i\cdot k} dk \) | 104.1 |
| Legendre Polynomials  | \( P_q(\cdot) \) | \( 2k'-1 \) | \( \frac{\sqrt{2(2q+1)}}{k_{\text{max}}} \int_0^{k_{\text{max}}} \gamma_n^m(k)P_q\left(2\frac{k}{k_{\text{max}}}-1\right) dk \) | 11.1 |
| Chebyshev Polynomials | \( U_q(\cdot) \) | \( 2k'-1 \) | \( \frac{\sqrt{n}}{k_{\text{max}}^{j,n}(Z_q^{j,n})} \int_0^{k_{\text{max}}} k\gamma_n^m(k)j_n\left(\frac{Z_q^{j,n}}{k_{\text{max}}}k\right) dk \) | 100.8 |
| Spherical Bessel Functions of order \( n \) \( j_n(\cdot) \) | \( k' \) | \( \gamma_n^m(k) \) | 104.7 |

(1) is around 70% when presenting CIPIC HRTF data up to 4 kHz.

4.2. Orthonormal Basis Selection in Frequency Component Expansion

A rich set of the orthogonal sets other than spherical Bessel functions is available to expand the frequency component, including Complex Exponentials, Legendre Polynomials, and Chebyshev Polynomials. From the KL theory, all of them are equivalent in expanding the spherical harmonic coefficients with infinite terms. However, under truncation, different set of orthogonal functions exhibit different properties.

Table 1 summarizes the expansions for \( \gamma_n^m(k) \) using four families of these orthogonal functions along with the sum of \( Q \) largest variance of the coefficients. The value of \( Q \) is set to 16 that the cumulative energy of 16 largest variance cover 93%, 79%, 64% and 95% of the total energy of variance of complex exponentials, Legendre polynomials, Chebyshev polynomials, and spherical Bessel functions coefficients, respectively. The sum corresponding to spherical Bessel functions is quite close to that of complex exponentials. But, we need to notice that 16 terms of spherical harmonics account for 95% of the variance in the coefficients while the same terms of complex exponentials achieve only 93% of the variance, which indicates that the spherical Bessel functions are more efficient than complex exponentials.

After comparing four sets of orthonormal basis, we conclude that the orthonormal basis — spherical Bessel functions — is the best choice for representing the frequency component of HRTFs.

5. CONCLUSION

In this paper, we successfully evaluate the efficiency of the 3D continuous HRTF model. The results show that the efficiency of the model in spatial component expansion is around 70% and the best choice in frequency component expansion is the spherical Bessel function.

6. REFERENCES


