Efficiency Evaluation and Orthogonal Basis Determination in Functional HRTF Modeling

M. Zhang, R. A. Kennedy, and T. D. Abhayapala

karan.zhang@cecs.anu.edu.au

Research School of Engineering
ANU College of Engineering and Computer Science
The Australian National University
Canberra ACT 0200 Australia

25 May 2011, ICASSP2011, Prague, Czech Republic

Outline

Introduction

Head Related Transfer Function (HRTF) — acoustic transfer function describing the filtering effects of the pinna, head, and torso of a listener

- **It is a function of position (θ, ϕ);**
- **It is strongly individual.**

Introduction

Head Related Transfer Function (HRTF) — acoustic transfer function describing the filtering effects of the pinna, head, and torso of a listener

- It is a function of position (θ, ϕ);
- It is strongly individual.

HRTF Measurement — the direct way

- **accurate;**
- **time-consuming.**

Introduction

Head Related Transfer Function (HRTF) — acoustic transfer function describing the filtering effects of the pinna, head, and torso of a listener

- It is a function of position (θ, ϕ);
- It is strongly individual.

HRTF Measurement — the direct way

- accurate;
- time-consuming.

HRTF Modeling — the efficient way

- **a continuous model;**
- **an individualized model.**

Introduction

3D Far-Field Continuous HRTF Model [1]

$$\hat{H}(\theta_s, \phi_s, k) = \sum_{n=0}^{N(k)} \sum_{m=-n}^{n} \sum_{q=1}^{Q} A_{n,q}^{m} \frac{\sqrt{2}}{\dot{j}_{n+1}(Z_q^{(n)})} \frac{k}{k_{\max}} j_n(\frac{Z_q^{(n)}}{k_{\max}} k) Y_n^m(\theta_s, \phi_s) \quad (1)$$

Merits:

- Separates the frequency components from the spatial components;
- It can reconstruct HRTFs at any arbitrary position (θ_s, ϕ_s) in space and at any frequency point $k = 2\pi c/f$.

[1] W. Zhang, T. D. Abhayapala, R. A. Kennedy, and R. Duraiswami, Insights into head related transfer function: Spatial dimensionality and continuous representation, J. Acoust. Soc. Am., vol. 127, no. 4, pp. 2347 - 2357, April 2010.

Introduction

3D Far-Field Continuous HRTF Model [1]

$$\hat{H}(\theta_s, \phi_s, k) = \sum_{n=0}^{N(k)} \sum_{m=-n}^{n} \sum_{q=1}^{Q} A_{n,q}^{m} \frac{\sqrt{2}}{j_{n+1}(Z_q^{(n)})} \frac{k}{k_{\max}} j_n\left(\frac{Z_q^{(n)}}{k_{\max}} k\right) Y_n^m(\theta_s, \phi_s) \tag{1}$$

Questions:

- How efficient is this model?
- Are there alternative representations which are superior?

[1] W. Zhang, T. D. Abhayapala, R. A. Kennedy, and R. Duraiswami, Insights into head related transfer function: Spatial dimensionality and continuous representation, J. Acoust. Soc. Am., vol. 127, no. 4, pp. 2347 - 2357, April 2010.

Outline

Efficiency Evaluation of Functional Model

We **assume** a random process can be expressed as a linear combination on **infinite** number of deterministic orthonormal functions, which are orthonormal with respect to an inner product$\langle \cdot, \cdot \rangle$, weighted by random variables.

Efficiency Evaluation of Functional Model

We assume a random process can be expressed as a linear combination on infinite number of deterministic orthonormal functions, which are orthonormal with respect to an inner product$\langle \cdot, \cdot \rangle$, weighted by random variables.

We **represent/approximate** this process using a **finite** number of the orthonormal functions.

Efficiency Evaluation of Functional Model

We assume a random process can be expressed as a linear combination on infinite number of deterministic orthonormal functions, which are orthonormal with respect to an inner product$\langle \cdot, \cdot \rangle$, weighted by random variables.

We represent/approximate this process using a finite number of the orthonormal functions.

The expected value of the squared norm of error of the finite representation is

$$\triangle = E\{\langle \mathbf{X}, \mathbf{X} \rangle\} - E\{\langle \hat{\mathbf{X}}, \hat{\mathbf{X}} \rangle\} \quad (2)$$

$\mathbf{X}$ — the random process
$\hat{\mathbf{X}}$ — the finite dimensional representation of the random process

Outline

Key Metric Identification

Now we focus on a simpler case for illustration.

Key Metric Identification

Now we focus on a simpler case for illustration.

$x(t)$ — **a random process defined on a finite interval** $|t| \leq T$ **with zero-mean and finite energy**

Key Metric Identification

Now we focus on a simpler case for illustration.

$x(t)$ — a random process defined on a finite interval $|t| \leq T$ with zero-mean and finite energy

$\{\psi_n(t)\}$ — **deterministic functions which are any orthonormal complete basis in $L^2(T)$ with respect to the inner product**

$$\langle \psi_n, \psi_m \rangle = \int_{-T}^{T} \psi_n(t)\psi_m^*(t)dt = \delta_{nm}, \tag{3}$$

where $*$ represents complex conjugation

Key Metric Identification

Now we focus on a simpler case for illustration.

$x(t)$ — a random process defined on a finite interval $|t| \leq T$ with zero-mean and finite energy

$\{\psi_n(t)\}$ — deterministic functions which are any orthonormal complete basis in $L^2(T)$ with respect to the inner product

$$\langle \psi_n, \psi_m \rangle = \int_{-T}^{T} \psi_n(t)\psi_m^*(t)dt = \delta_{nm}, \tag{3}$$

where $*$ represents complex conjugation

Then, the random process can be expressed as

$$x(t) = \sum_{n=1}^{\infty} \alpha_n \psi_n(t), \quad |t| \leq T \tag{4}$$

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

Key Metric Identification

For the N-dimensional representation, the error for a particular realization of the random process is

$$\epsilon_i = x_i(t) - \sum_{n=1}^{N} \alpha_{i,n}\psi_n(t) \tag{5}$$

i — the index of the realization

The expected value of the squared norm of error when using N-basis functions $\{\psi_n\}$ is

$$\triangle_\psi = E\{\langle\epsilon,\epsilon\rangle\} = E\{\langle\mathbf{X},\mathbf{X}\rangle\} - E\{\sum_{n=1}^{N}|\alpha_n|^2\} \tag{6}$$

Key Metric Identification

Evaluating the second term in (6)

$$\mathscr{A}_\psi = E\{\sum_{n=1}^{N} |\alpha_n|^2\} = E\{\sum_{n=1}^{N} \iint_{-T}^{T} \mathbf{x}(t)\psi_n^*(t)\mathbf{x}^*(s)\psi_n(s)\, dtds\}$$
$$= \sum_{n=1}^{N} \iint_{-T}^{T} E[\mathbf{x}(t)\mathbf{x}^*(s)]\psi_n^*(t)\psi_n(s)\, \mathrm{d}t\mathrm{d}s$$
$$= \sum_{n=1}^{N} \iint_{-T}^{T} k_{xx}(t,s)\psi_n^*(t)\psi_n(s)\, dtds, \qquad (7)$$

$k_{xx}(t,s) \triangleq E\{\mathbf{x}(t)\mathbf{x}^*(s)\}$ —the autocorrelation function of the random process

Key Metric Identification

Evaluating the second term in (6)

$$\mathscr{A}_\psi = E\{\sum_{n=1}^{N} |\alpha_n|^2\} = E\{\sum_{n=1}^{N} \iint_{-T}^{T} \mathbf{x}(t)\psi_n^*(t)\mathbf{x}^*(s)\psi_n(s)\, dtds\}$$
$$= \sum_{n=1}^{N} \iint_{-T}^{T} E[\mathbf{x}(t)\mathbf{x}^*(s)]\psi_n^*(t)\psi_n(s)\, \mathrm{d}t\mathrm{d}s$$
$$= \sum_{n=1}^{N} \iint_{-T}^{T} k_{xx}(t,s)\psi_n^*(t)\psi_n(s)\, dtds, \qquad (7)$$

$k_{xx}(t,s) \triangleq E\{\mathbf{x}(t)\mathbf{x}^*(s)\}$ —the autocorrelation function of the random process

$\mathscr{A}_\psi$ depends on the choice of the $\{\psi_n\}$.

Key Metric Identification

Any complete orthonormal basis $\{\psi_n\}$ can be used, but which is the best choice?

Let $\{\xi_n(t)\}$ be the eigenfunctions of the following integral equation with kernel, the autocorrelation function of the random process

$$\int_{-T}^{T} k_{xx}(t,s)\xi_n(s)\,ds = \lambda_n \xi_n(t) \tag{8}$$

Let $\{\xi_n(t)\}$ be the basis functions to expand the random process

$$x(t) \cong \sum_{n=1}^{N} \beta_n \xi_n(t) \tag{9}$$

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

Key Metric Identification

$$\mathscr{A}_\xi = E\{\sum_{n=1}^{N} |\beta_n|^2\} = \sum_{n=1}^{N} \iint_{-T}^{T} k_{xx}(t,s)\xi_n^*(t)\xi_n(s)\,dtds$$
$$= \sum_{n=1}^{N} \int_{-T}^{T} \lambda_n \xi_n(t)\xi_n^*(t)\,dt$$
$$= \sum_{n=1}^{N} \lambda_n \geq \mathscr{A}_\psi = E\{\sum_{n=1}^{N} |\alpha_n|^2\} \qquad (10)$$

Key Metric Identification

Optimal Expansion (Karhunen-Loève Expansion)

$$x(t) \cong \sum_{n=1}^{N} \beta_n \xi_n(t)$$

Quite often, the distribution properties in the random process are either unknown or too complicated to be described simply or in closed form.

As a result, it is extremely difficult to find the KL expansion for the random process. (The eigenfunctions would not be closed form functions in general.)

Key Metric Identification

Given a set of realizations of the random process $\{x(t)\}$, and an arbitrary orthonormal basis $\{\psi_n\}$, we calculate $\{\alpha_n\}$ for each realization.

We evaluate $\{\alpha_n\}$ and use sample variance to determine the terms $E\{|\alpha_n|^2\}$.

Then, the sum of N largest sample variances of $\{\alpha_n\}$ is $\mathscr{A}_\psi$ and it will not larger than $\mathscr{A}_\xi$, which is given by KL expansion.

Key Metric Identification

Given a set of realizations of the random process $\{x(t)\}$, and an arbitrary orthonormal basis $\{\psi_n\}$, we calculate $\{\alpha_n\}$ for each realization.

We evaluate $\{\alpha_n\}$ and use sample variance to determine the terms $E\{|\alpha_n|^2\}$.

Then, the sum of N largest sample variances of $\{\alpha_n\}$ is $\mathscr{A}_\psi$ and it will not larger than $\mathscr{A}_\xi$, which is given by KL expansion.

Key metric — Sum of the N largest sample variances among random variables involved in the model

Outline

Evaluation Method

1. Since the representation is not optimal expansion, redundancy must exists in the model. Then, how do we measure the redundancy, or equivalent, the efficiency?

For the N-terms expansion

- Calculate the variances for N coefficients;
- Choose $N'(N' < N)$ coefficients which show significant contribution to the model;
- Calculate the efficiency of the model η,

$$\eta = \frac{N'}{N} \times 100\%. \tag{11}$$

Evaluation Method

2. Although the eigenfunctions $\{\xi_n\}$ are unknown or not determinable, arbitrary orthonormal basis are usable. But there are many choices. Then, which one is the best?

For different orthonormal basis,

- Calculate coefficients $\{\alpha\}$;
- Calculate $\mathscr{A}$;
- Compare the value of $\mathscr{A}$, the larger the better.

Outline

Outline

Efficiency in Spatial Component Expansion

Spatial Component Expansion — Spherical Harmonics

$$\hat{H}(\theta_s, \phi_s, k) = \sum_{n=0}^{N(k)} \sum_{m=-n}^{n} \gamma_n^m(k) Y_n^m(\theta_s, \phi_s) \qquad (12)$$

$(\theta_s,\ \phi_s)$ — the sound source position
$k = 2\pi f/c$ — the wavenumber
$Y_n^m(\theta_s, \phi_s)$ — the spherical harmonics characterized by degree n and order m
$\gamma_n^m(k)$ — the spherical harmonic coefficients of the equivalent source field at wavenumber k
$N(k) = eks/2$ — the maximum degree, in which $e \approx 2.7183$ and $s = 9$ cm.

Efficiency in Spatial Component Expansion

$$M(k) = (N(k) + 1)^2 \tag{13}$$

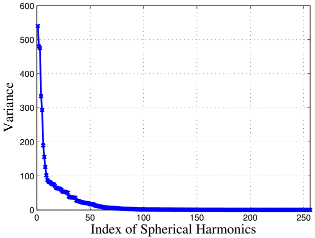
$M(k)$ terms of spherical harmonics are used to represent HRTFs at a particular wavenumber k corresponding to all directions.

$M'(k)$ indicate the number of spherical harmonic coefficients with 99.9% energy of the total energy of $M(k)$ variances.

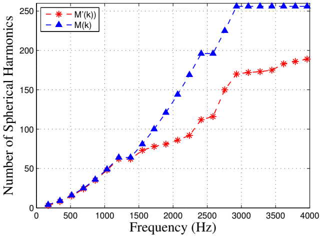
According to (11), the efficiency is

$$\eta = \frac{\sum_{k=0}^{k_{\max}} M'(k)}{\sum_{k=0}^{k_{\max}} M(k)} \times 100\%. \tag{14}$$

Efficiency in Spatial Component Expansion



(a)



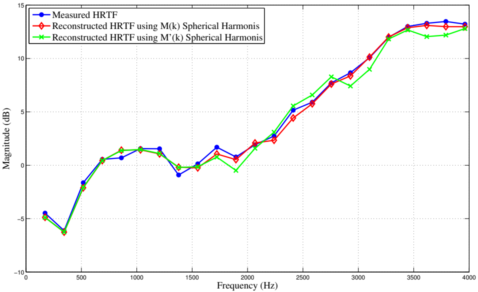
(b)

(a) The variance of all spherical harmonic coefficients for the frequency at 4kHz.

(b) The difference between $M(k)$ and $M'(k)$.

The efficiency η is around 70%.

Efficiency in Spatial Component Expansion



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

Outline

Ortho. Basis Selection in Freq. Component Expansion

Frequency Component Expansion — Spherical Bessel Function

$$\gamma_n^m(k) = \sum_{q=1}^{Q} A_{n,q}^m \frac{\sqrt{2}}{j_{n+1}(Z_q^{(n)})} \frac{k}{k_{\max}} j_n\left(\frac{Z_q^{(n)}}{k_{\max}} k\right) \qquad (15)$$

$A_{n,q}^m$ — the spherical Bessel coefficients

$$\frac{\sqrt{2}}{k_{\max}^2 j_{n+1}(Z_q^{(n)})} \int_0^{k_{\max}} k \gamma_n^m(k) j_n\left(\frac{Z_q^{(n)}}{k_{\max}} k\right) dk \qquad (16)$$

$j_n(\cdot)$ — the spherical Bessel function of the degree n

$Z_q^{(n)}$ — the positive roots of $j_n(\cdot)$

Ortho. Basis Selection in Freq. Component Expansion

- Complex Exponentials
- Legendre Polynomials
- Chebyshev Polynomials

Ortho. Basis Selection in Freq. Component Expansion

- Complex Exponentials

$$\frac{1}{k_{\max}} \int_0^{k_{\max}} \gamma_n^m(k) e^{j\pi qk/k_{\max}} \, dk \tag{17}$$

- Legendre Polynomials

$$\frac{\sqrt{2(2q+1)}}{k_{\max}} \int_0^{kmax} \gamma_n^m(k) P_q(2\frac{k}{k_{\max}} - 1) \, dk \tag{18}$$

- Chebyshev Polynomials

$$\frac{2\sqrt{2}}{k_{\max}\sqrt{\pi}} \int_0^{k_{\max}} [1-(2\frac{k}{k_{\max}}-1)^2]^{\frac{1}{4}} \gamma_n^m(k) U_q(2\frac{k}{k_{\max}}-1) \, dk \tag{19}$$

Ortho. Basis Selection in Freq. Component Expansion

Properties of HRTF frequency component models using four sets of orthogonal basis

Orthogonal Functions	$E\{\sum_{q=1}^{Q} \lvert A_{n,q}^{m}\rvert^2\}$
Complex Exponentials	104.1
Legendre Polynomials	11.1
Chebyshev Polynomials	100.8
Spherical Bessel Functions	104.7

Q is set to 16 that the cumulative energy of the 16 largest variance cover 93%, 79%, 64% and 95% of the total energy of variance of the respective coefficients.

Outline

Conclusion

- **Problem:** How do we evaluate the efficiency of a given functional HRTF model?
- **Key Metric:** The sum of the variance of random variables involved in the model
- **Results:**
  - The efficiency of the model in spatial component expansion is 70%;
  - The best choice in frequency component expansion is the spherical Bessel function.