A REDUCED ORDER MODEL OF HEAD-RELATED IMPULSE RESPONSES BASED ON INDEPENDENT SPATIAL FEATURE EXTRACTION

Qinghua Huang, Kai Liu

School of Communication and Information Engineering,
Shanghai University, Shanghai, P.R. China
qinghua@shu.edu.cn

ABSTRACT

A reduced order model for large numbers of head-related impulse responses (HRIRs) is proposed for real-time three-dimensional (3D) sound rendering. Independent spatial features are firstly extracted from measured HRIRs using independent component analysis (ICA). These spatial feature vectors are not only mutually statistical independent but independent from all the measured azimuths. Therefore filtering sound sources with numerous HRIRs is transformed into filtering them using the extracted lower-dimensional feature vectors. Furthermore balanced model truncation (BMT) method in a state space is adopted to reduce the order of each independent feature vector. Simulation results demonstrate that our proposed algorithm not only acquires better approximated accuracy but has significantly lower computational complexity.

Index Terms—Three-dimensional sound, head-related impulse response, binaural synthesis, independent component analysis, reduced order modeling

1. INTRODUCTION

3D sound reproduction has been an important research area in audio signal processing. It seeks to create a more spatial and immersive auditory scene for a listener via either conventional headphones or loudspeakers. Binaural synthesis is a technique to reproduce a virtual sound image at one position in the 3D space from one monaural sound source using only a pair of headphones. It has many applications such as virtual reality, entertainment and training simulation et al. HRIR or its frequency domain representation called head-related transfer function (HRTF) describes the acoustic filtering of a plane wave from a spatial sound source to the listener's eardrum. Binaural synthesis of a sound source at a desirable location in 3D space is realized by filtering the sound with the left and right ear HRIRs corresponding to the sound source position [1].

Many typical acoustic scenes contain multiple sound sources, source and listener motion. A framework filtering source signals with numerous HRIRs simultaneously is often adopted. Therefore the number of HRIRs required for these acoustic scenes is often large. It is desirable to develop some reduced order HRIR models to save storage and appropriate for real-time processing. Mackenzie et al. proposed a low-order model for HRTFs using BMT method [2]. Adams et al. adopted state-space system to describe the HRTF filter array and employed Hankel operator for state order reduction [3]. Principal component analysis (PCA) has been applied to reduce dimension of HRTF array. It decorrelates only second-order statistics [4]. A 3D array beamformer was used to synthesize HRTFs by non-uniform sampling techniques and singular value decomposition (SVD) [5]. It can also save memory storage for HRTFs. However the accuracy of matching array coefficients with measured HRTFs has an effect on the location performance.

In this paper, we propose to adopt ICA to extract independent spatial features for HRIRs. Combined with BMT method, the orders of the independent spatial features can be reduced. Therefore the computational cost can greatly decrease, and simultaneously the good approximate performance can be achieved. The simulation results can demonstrate the validity of the proposed method.

2. INDEPENDENT SPATIAL FEATURE EXTRACTION FOR HRIRS

Canonical subspace projection techniques such as PCA and ICA have been widely used in feature extraction and data compression. They project observed data from a high-dimensional space to a lower-dimensional space. The appropriate lower-dimensional space is spanned by a set of basis vectors which have some desirable properties, such as sparse, uncorrelated or independent.

In binaural synthesis for virtual auditory space, let $h^L(\theta_i,\phi_j,n)$ and $h^R(\theta_i,\phi_j,n)$ be the HRIRs for the left and right ears from the direction which azimuth is θ_i, elevation is ϕ_j and discrete time index is n. The HRIR filter array

matrix for a specific elevation ϕ_j and all observed azimuths ($i = 1, 2, \cdots, D$) is expressed as

$$H = \begin{bmatrix} h^L(\theta_1,1) & h^L(\theta_1,2) & \cdots & h^L(\theta_1,N) \\ h^R(\theta_1,1) & h^R(\theta_1,2) & \cdots & h^R(\theta_1,N) \\ \cdots & \cdots & \cdots & \cdots \\ h^L(\theta_D,1) & h^L(\theta_D,2) & \cdots & h^L(\theta_D,N) \\ h^R(\theta_D,1) & h^R(\theta_D,2) & \cdots & h^R(\theta_D,N) \end{bmatrix}$$

For convenience, ϕ_j is omitted. H is a $2D \times N$ real matrix. Each row data of H denotes a HRIR for a specific direction, and the length of each HRIR is N. Generally the dimension of the array is high and the HRIRs from D directions are correlated. Moreover each HRIR is a multivariate function. If we can separate these variables by finding a linear combination of a series of basis functions and reduce the dimension of HRIR array, the problem will be simplified. PCA and ICA are two appropriate techniques. The former gets M principal components by decorrelating the measured array only using second-order moments. The latter obtains M independent basis vectors in the time-domain. The M independent basis vectors represent M independent spatial features of HRIRs from different directions. 'Independent' has two manifold meanings, one is that the features are mutually independent and the other is that the features are independent from all the measured azimuths. Therefore the independent spatial feature extraction for HRIRs provides a more efficient representation and can reduce computational cost. The aim is to extract M independent spatial features from H using the following linear transformation

$$H = WS = \begin{bmatrix} w^L(\theta_1,1) & \cdots & w^L(\theta_1,M) \\ w^R(\theta_1,1) & \cdots & w^R(\theta_1,M) \\ \cdots & \cdots & \cdots \\ w^L(\theta_D,1) & \cdots & w^L(\theta_D,M) \\ w^R(\theta_D,1) & \cdots & w^R(\theta_D,M) \end{bmatrix} \begin{bmatrix} s(1,1) & \cdots & s(1,N) \\ s(2,1) & \cdots & s(2,N) \\ \cdots & \cdots & \cdots \\ s(M,1) & \cdots & s(M,N) \end{bmatrix} \quad (1)$$

where $W \in \mathbb{R}^{2D \times M}$ represents weight coefficient matrix which is correlated with spatial directions and $S \in \mathbb{R}^{M \times N}$ consists of the basis vectors which are uncorrelated with spatial directions. One random chosen left ear HRIR from the horizontal plane can be represented using these M independent basis vectors. This is shown in Fig.1 in which azimuth variable θ_i is omitted from the weight coefficient $w^L(\theta_i, j)$ and the length of the HRIR is 128.

Independent basis vectors can be acquired using the second-order blind identification (SOBI) algorithm which is based on a joint diagonalization of a set of spatial covariance matrices [6]. The first step is whitening that is achieved by applying a whitening matrix U to the HRIR array H

$$G = UH \quad (2)$$

where U is a $M \times 2D$ matrix and G is a whitened data array. Given the whitened data $G \in \mathbb{R}^{M \times N}$, M independent basis vectors are obtained by a linear transformation as follows

$$S = VG \quad (3)$$

where $V \in \mathbb{R}^{M \times M}$ represents the transformed matrix and $S \in \mathbb{R}^{M \times N}$ is the independent basis vectors. For a fixed set of time lags $\{\tau_j \mid j = 1, 2, \cdots, K\}$, the corresponding sample covariance matrices $R(\tau_j)$ of the whitened data are computed. Then the transformed matrix V can be obtained as joint diagonalizer of the set $\{R(\tau_j) \mid j = 1, 2, \cdots, K\}$. Therefore the weight matrix W is

$$W = (VU)^{\#} \quad (4)$$

where the superscript $\#$ denotes the Moore-Penrose pseudo-inverse operator. After getting the weight matrix and independent basis vectors, each HRIR can be expressed as a linear combination of these M independent basis vectors as follows

$$\hat{h}^L(\theta_i, n) = \sum_{j=1}^{M} w^L(\theta_i, j) s(j, n) \quad (5)$$

$$\hat{h}^R(\theta_i, n) = \sum_{j=1}^{M} w^R(\theta_i, j) s(j, n) \quad (6)$$

Binaural synthesis $o^L(\theta_i, n)$ and $o^R(\theta_i, n)$ of one monaural source $e(n)$ are as follows

$$o^L(\theta_i, n) = e(n) * \hat{h}^L(\theta_i, n) = \sum_{j=1}^{M} w^L(\theta_i, j)\left[e(n) * s(j, n)\right] \quad (7)$$

$$o^R(\theta_i, n) = e(n) * \hat{h}^R(\theta_i, n) = \sum_{j=1}^{M} w^R(\theta_i, j)\left[e(n) * s(j, n)\right] \quad (8)$$

Therefore when $2D$ HRIRs filter sound sources simultaneously, the process can be substituted by only M independent basis vectors filtering the sound sources. The dimension of filter array greatly decreases.

3. REDUCED ORDER MODELING FOR INDEPENDENT SPATIAL FEATURES

Although independent basis vectors can be extracted from original high-dimensional HRIR array, the filter implementation of each independent basis vector is a high-order finite impulse response (FIR) structure. Therefore in this section a reduced order modeling for independent basis vectors is performed.

3.1. Reducing order modeling

The reduced order model of independent basis vectors can be achieved by constructing a multiple input and single output (MISO) system of FIR filters. The state-space form of the MISO system can be written as

$$\vec{x}(n) = A\,\vec{x}(n-1) + B\vec{u}(n) \tag{9}$$

$$y(n) = C\vec{x}(n) \tag{10}$$

where $\vec{x}(n)$ is the MN-dimensional state vector, $\vec{u}(n)$ is the M-dimensional input vector and $y(n)$ is the scalar output. Let (A,B,C) represent the state-space system. The impulse response of the system is

$$\vec{s}(n) = \begin{bmatrix} s(1,n) & \cdots & s(M,n) \end{bmatrix} = CA^{n-1}B \qquad n > 0 \tag{11}$$

A reduced K order model (A_K, B_K, C_K) of the N order system (A,B,C) is obtained using BMT method [3]. BMT method adopts the K largest Hankel singular values and applies a balancing similarity transform to (A,B,C). The full description of the BMT technique can be found in Ref.[2]. Therefore the impulse response of the approximated reduced order system (A_K, B_K, C_K) is

$$\hat{\vec{s}}(n) = C_K A_K^{n-1} B_K \tag{12}$$

It is the reduced order realization of independent basis vectors.

3.2. Computational cost

We employ the number of multiplication operations required per sample period to define the computational cost C [3]. An FIR filter array of order N with D inputs and one output requires $C = D(N+1)$ multiplication operations per sample period. After independent spatial feature extraction and reduced order modeling, an approximate state-space system of order K with M inputs and one output demands $C = K^2 + (M+1)K$ multiplication operations. For example, there is 37 inputs and one output FIR filter array that models 37 HRIRs with order $N = 511$. The computational cost of the array is $C = 18944$. After independent spatial feature extraction, the cost of new FIR filter array of order $N = 511$ with $M = 18$ inputs is 9216. After reducing the order of state-space model for independent spatial features, the cost of a 64 order state-space system with 18 inputs and one output is 5312.

4. SIMULATION RESULTS AND EVALUATION

The HRIRs in the simulation are obtained from KEMAR database using a sampling rate of 44.1 kHz [7]. We chose five different elevations ($\phi_j = 0^\circ, \pm 10^\circ, \pm 20^\circ$). The azimuth is from 0° to 180° in a 5° step. Therefore HRIRs for the left and right ears are respectively from 37 different azimuths. PCA for basis vectors extraction is in comparison to our proposed algorithm. Moreover the number of the independent spatial feature vectors is equal to that of the principal basis vectors. It is chosen to be 18. In Fig.1 and Fig.2, the approximations using different methods of a pair of HRTFs at 50 degree azimuth on the horizontal plane are shown. These basis vectors capture more than 99% energy of the HRIR data.

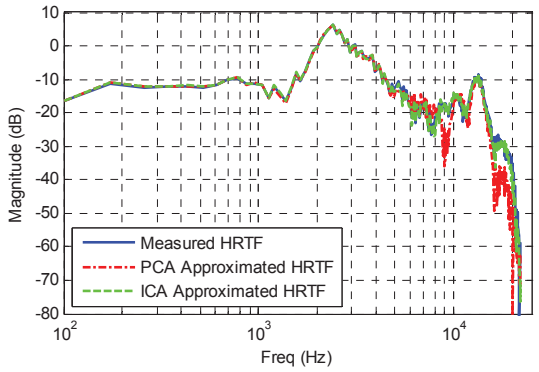

Fig. 1. Approximated left ear HRTF at 50 degree on the horizontal plane

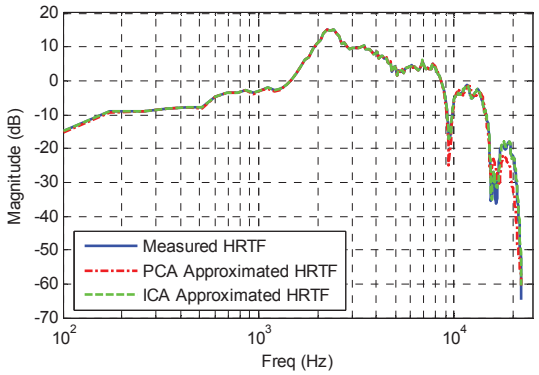

Fig. 2. Approximated right ear HRTF at 50 degree on the horizontal plane

The average signal distortion ratio (ASDR) acts as an objective measure to evaluate the approximate performance. It is computed using eq.(13) in decibel (dB). PCA, ICA and these two methods combined with BMT were carried out and compared. Table 1 gives the approximated performance for HRIRs from five different elevations. From the table, we can see that the approximated accuracy of the horizontal plane is worse than that of other elevations. With the increase of elevation, the approximated results using the same principal components and independent components have been improved.

$$ASDR_j = \frac{1}{2}10\lg\frac{\sum_{i=1}^{D}\sum_{n=1}^{N}\left|h^L(\theta_i,\phi_j,n)\right|^2}{\sum_{i=1}^{D}\sum_{n=1}^{N}\left|h^L(\theta_i,\phi_j,n)-\hat{h}^L(\theta_i,\phi_j,n)\right|^2} + \frac{1}{2}10\lg\frac{\sum_{i=1}^{D}\sum_{n=1}^{N}\left|h^R(\theta_i,\phi_j,n)\right|^2}{\sum_{i=1}^{D}\sum_{n=1}^{N}\left|h^R(\theta_i,\phi_j,n)-\hat{h}^R(\theta_i,\phi_j,n)\right|^2} \tag{13}$$

Simple subject tests using headphone listening were conducted. Ten test subjects participated in the listening experiments and four test stimuli were chosen. The stimuli are pink noise samples of duration 1 second with 50

millisecond onset and offset time [8]. They were rendered at four different positions on the horizontal plane. The subjects were asked to grade the localization similarity of the test stimuli rendering using the approximated HRIR models compared to those rendering using the corresponding KEMAR HRIRs on a continuous 1.0 to 5.0 scale in Table 2 [9]. Fig.3 shows the results of the subjective tests using ICA_BMT model.

Table 1. ASDR for different methods and different elevations

ϕ_j	-20°	-10°	0°	10°	20°
PCA	30.05	28.60	26.61	27.15	29.75
ICA	**32.74**	**31.41**	**29.91**	**30.73**	**33.19**
PCA_BMT	25.82	24.75	24.89	25.94	25.82
ICA_BMT	27.10	26.46	26.60	26.90	27.10

Table 2. Localization impairment scale

Grade	Localization similarity
1	Very different
2	Slightly different
3	Slightly similar
4	Very similar
5	No difference

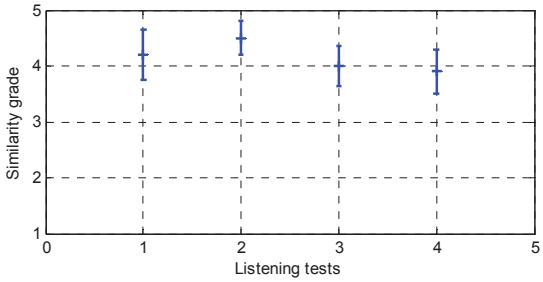

Fig. 3. Listening test results using ICA_BMT model

5. CONCLUSIONS

Large numbers of HRIRs and high-order FIR structure greatly affect the real-time immersive rendering of multiple sound sources or moving sound source. We proposed a reduced order model of HRIRs based on ICA. Independent basis features are firstly extracted from HRIR array, and then they are reduced order in a MISO state space using BMT method. Therefore the computational cost significantly decreases, at the same time better approximated performance can be achieved. Experimental results have shown the good performance in comparison to other methods. However, the best numbers of the extracted features and the reduced order of BMT model are not known in advance. In our future work, we will further study how to automatically choose the two optimal numbers to achieve better performance.

Acknowledgement: This work was supported by Shanghai Natural Science Foundation of China (No. 08ZR1408300).

6. REFERENCES

[1] R. Rabenstein, and S. Spors, "Sound field reproduction," *Springer Handbook of Speech Processing*, Part I, pp. 1095-1114, 2008.

[2] J. Mackenzie, J. Huopaniemi, V. Valimaki, and I. Kale, "Low-order modeling of head-related transfer functions using balanced model truncation," *IEEE Transaction on Signal Processing Letters*, Vol. 4, No. 2, pp. 39-41, February 1997.

[3] N. H. Adams, and G. H. Gregory, "State-space synthesis of virtual auditory space," *IEEE Transaction on Audio, Speech, and Language Processing*, Vol. 16, No. 5, pp. 881-890, July 2008.

[4] P. S. Chanda, S. Park, and T. I. Kang, "A binaural synthesis with multiple sound sources based on spatial features of head-related transfer functions," *Proc. Neural Networks*, pp. 1726-1730, July 2006.

[5] R. B. Mingsian, and O. Kwuen-Yieng, "Head-related transfer function (HRTF) synthesis based on a three-dimensional array model and singular value decomposition," *Journal of Sound and Vibration*, Vol. 281, pp. 1093-1115, 2005.

[6] A. Belouchrani, K. A. Meraim, J. F. Cardoso, and E. Moulines, "A blind source separation technique based on second order statistics," *IEEE Trans. on Signal Processing*, Vol. 45 No. 2, pp. 434-444, 1997.

[7] MIT Media Lab Machine Listening Group, "HRTF measurements of a KEMAR dummy-head microphone," *http:// sound.media.mit.edu/KEMAR.html*.

[8] P. S. Chanda, S. Park and T.I. Kang, "A binaural synthesis with multiple sound sources based on spatial features of head-related transfer function," International Joint Conference on Neural Networks, Vancouver, Canada, July 16-21, pp. 1726-1730, 2006.

[9] J. Huopaniemi, N. Zacharov, and M. Karjalainen, "Objective and Subjective Evaluation of Head Related Transfer Function Filter Design," In 105th Audio Engineering Society Convention, August 1998.