Real-time rendering of moving sound source in a room based on principal components analysis

Dan Rao, Yu Liu

Acoustic Lab., Physics Dept., South China University of Technology, Guangzhou, 510641, P.R.China

Abstract

In free-field situation, convoluting sound signal with a pair of head-related transfer function can render a sound source virtually. In room, however, because there are many image sources generated by reflections from wall, the situation is more complex, especially when the source is moving. Rendering such scene requires 2*N (image source numbers) times of convolution computations so as to bring difficulty in the real-time implementation. In this paper, a method using principal components analysis (PCA) is proposed to reduce the rendering computation cost. In this method, HRIRs were expressed as weighted sum of a set of basics functions which can be derived from PCA. Therefore, the HRIRs of real source and image sources at different directions share a common set of basics function. In such way, computation reduced to 2*L(basics function number) times of convolution computations. We used Q=7 and 10 basics functions in our implementation, which were much less than the number of image sources (36 for considering second-order image sources in rectangular room). Objective analysis validated the proposed method. The performance and details of real-time implementation using VC++ of the proposed method on PC were also discussed in this paper.

Introduction

Virtual auditory display (VAD) is a technique to simulate a specific auditory scene by convoluting sound signal with head-related impulse responses (HRIRs, called as HRTFs in frequency domain) and replaying by headphones.. VAD has used in various fields, such as virtual reality, multimedia comunication and so on [1]. In order to simulate complex acoustical scenes(such as multi-source case or room-environment case), a large computation of convolution is needed. This led to great difficulty in real-time implementation. Therefore, reduction of computation is necessary.

Several methods of linear decomposition of HRIR were proposed to solve the problem of computation cost increasing with the increase of simulated sound source [2]. Among them, Principal component analysis (PCA) is an efficient method to reduce signal processing computation cost and compress data. Many research used PCA to decompose and analyze HRTF(or HRIR)[3][4]. However, application of PCA in rendering auditory scene in a room was seldom. In the present paper, a method to reduce computation cost of rendering auditory scene in a room based on PCA was proposed.

Methods

By using PCA, the HRIR of a given ear (left or right) at direction (θ_i, ϕ_i) can be expanded as a linear combination of basis functions:

$$h(\theta_i, \phi_i, t) = \sum_{q=0}^{Q} f_q(t)\, w_q(\theta_i, \phi_i) \tag{1}$$

Where the common basis functions (filters) $f_q(t)$ are direction-independent, $w_q(\theta, \phi)$ are corresponding weigh coefficients. The contribution of $f_q(t)$ and $w_q(\theta, \phi)$ to the expansion rank in decreasing order. The procedure of HRIRs expansion for two ears are the same.

In order to make PCA more efficient, the initial delay of HRIRs at different directions are aligned before performing PCA. The initial delay of each HRIR was calculated by onset-detecting method [4] and will be added back in real-time rendering implementation. Here the onset is defined as the instance at which the amplitude of impulse response reached 15% of the entire impulse response.

In a room, considering a rectangle room, the sound reflecting from walls can be equivalent to the sound emitting from image sources. The image source numbers increases exponentially as the increase of reflection order. So, often only a few order of image source was considered in VAD. However even several image order is considered, the source image number is very large. If the sound source is static, the HRIRs from all image source to two ears can be combined into a pair of binarual room impulse response. However, if the sound source is dynamic, each image source should be rendered separately in the traditional method, thus resulting in large computation cost. The traditional method given by the following equation:

$$E_l = \sum_{i=1}^{N} h(\theta_i, \phi_i, t) \otimes E_i \tag{2}$$

Where $\otimes$ denotes the linear convolution, E_i refers to signal of sound source or its' image sources. From this equation, it can be seen that the computation cost is linearly proportional to source numbers and therefore is exponentially proportional to reflection order.

Substituting Eq. (1) into Eq.(2), the method using PCA was obtained:

$$E_l = \sum_{q=0}^{Q} \left(\sum_{i=1}^{N} w_q(\theta_i, \phi_i) E_i\right) \otimes f_q(t) \tag{3}$$

From the above equation, it can be observed that before the convolution, weighted sum operations of signals were performed. In such way, the convolution computation is only related to the basic function number Q. Obviously, when (image) source numbers N is larger than Q, the PCA method will be more efficient. That is the main idea that using PCA to reduce the computation cost.

Figure.1 show the block diagram of real-time rendering by using PCA.

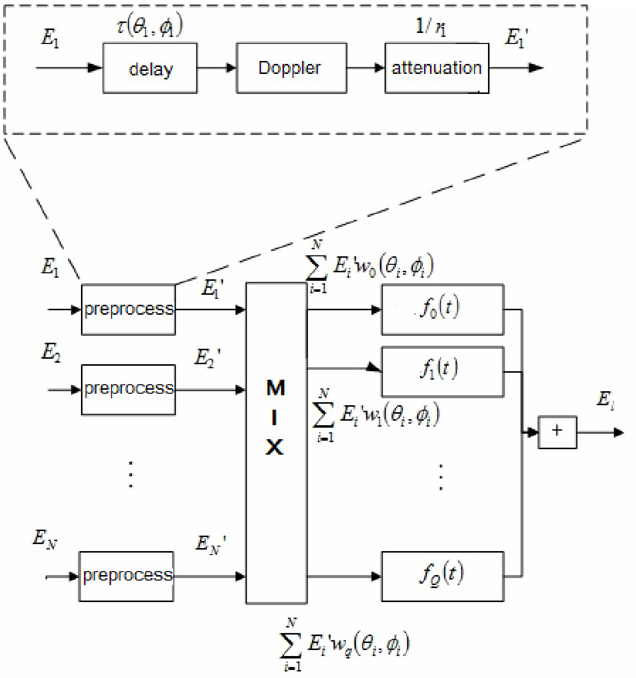

Figure1: Block diagram of real-time rendering by using PCA.

In the Figure 1, E_i denotes signals of sound sources (including real and image sources). The delay includes propagation delay and initial delay of HRIRs at different directions. The Doppler effect was simply achieved by linear interperation of the delay[5].

Performance analysis

To validate the performance of the proposed methods, the rendering of a sound source in a rectangle room according to the diagram of Figure 1 was implemented by using Virtual C++ under windows platform. The rectangle room is 15 m×8 m×5 m. For simplicity, visibility checking of image sources is not considered here. Simulation was conducted in a PC with Intel Core E6550 @3.15GHz CPU and 4Gbyte memory.

HRIRs used in this simulation are from MIT database [6], which are 512 samples in length. Two Q values were adopted in the simualtion:7,10. Their corresponding cumulative percentage variance were 92% and 96% respectively. We use the consuming time of processing 10s sound signal to evaluate the performance of rendering method. The performance results are shown in Table.1.

Table 1: The processing time for 10s signal

Image order(number)	Processing time(s)		
	Q=7	Q=10	Traditional method
1(6)	2.18	3.12	2.12
2(36)	2.58	3.54	11.31
3(216)	4.78	6.11	66.08

From Table 1, when Q < N (image order >1), the processing times of PCA-based method for Q=7, 10 are both less than that of raditional method. This is due to the fact that the most time-consuming computation is convolution and that for PCA method, the convolution computation is independent of source number while the traditional method is dependent of source number.

In addition, the slight increase in processing time with the increase of image order results from the increase of other computations, such as weighted sum computation (computation within the bracket of Eq.(3)), preprocessing computation.

Conclusions

From the above analysis and simualtion result, it can be seen that the proposed method using PCA is suitable for the real-time rendering of moving sound source, especailly for high image order. The future works will conduct psychoacoustic experiments to determining the best Q value and validate the performance of the proposed method.

Acknowledgements

The authors acknowledge the support State Key Lab of Subtropical Building Science, South China University of Technology.

Reference

[1]. Xie, B.S.: Head-Related Transfer Function and Virtual Auditory (in Chinese). National Defense Industry Press, Beijing, 2008.

[2]. Larcher, V., Jean-Marc, J., Guyard, J., et al. :Study and comparison of efficient methods for 3d audio spatialization based on linear decomposition of hrtf data [C]. AES 108th Convention, Paris, France, 2000, Reprint:5918. 123.

[3]. Kistler D.J, Wightman F.L.: A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. J. Acoust. Soc. Am. 91 (1992), 1637-1647.

[4]. Chen, J., Van Veen, B. D., and Hecox, K. E. :A spatial feature extraction and regularization model for the head-related transfer function, J. Acoust. Soc. Am. 97(1995), 439–452.

[5]. Saviojia, L., Huopaniemi, J., et.al: Creating interactive virtual acoustic environments, J. Audio. Eng. Soc., 47(1999), 675-705.

[6]. Gardner W. G., Martin K. D. : HRTF measurements of a KEMAR, J. Acoust. Soc. Am., 97 (1995): 3907-3908.