Applied Acoustics 70 (2009) 1212–1218

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Technical Note

A hybrid compression method for head-related transfer functions

Lin Wang *, Fuliang Yin, Zhe Chen

School of Electronic and Information Engineering, Dalian University of Technology, Dalian, 116023, PR China

ARTICLE INFO

Article history:
Received 10 September 2008
Received in revised form 23 March 2009
Accepted 4 April 2009
Available online 10 May 2009

PACS:
43.60.Gk
43.66.Qp

Keywords:
Head-related transfer function
Data compression
Principal component analysis
Vector quantization
Curved surface fitting

ABSTRACT

The paper proposes a hybrid compression method to resolve the storage problem of a large number of head-related transfer functions (HRTFs). First, each HRTF is approximated by a minimum-phase HRTF and an all pass filter whose group delay equals the interaural time delay (ITD). Second, principal component analysis is applied to the entire HRTF set to derive several basis functions, with a weight vector set defining the contribution of the basis functions to each HRTF. Third, the weight set is vector quantized with the designed codebook. At last, the ITD is curved surface fitted with a cosine series bivariate polynomial. As a result, the HRTF can be reconstructed from the basis functions, codebook indexes, and ITD polynomial coefficients. Simulation results reveal that the proposed method may reduce the data size greatly with similar reconstruction precision comparing with the principal component analysis method.

1. Introduction

A set of transfer functions from different spatial locations to both ears, or head-related transfer functions (HRTFs), are primarily due to interactions of acoustic waves with a listener's head, torso and pinnae [1,2]. It is known that HRTFs convey important cues for human spatial hearing, such as interaural time difference (ITD), interaural level difference (ILD), and spectral cues, where ITD and ILD describe information in the horizon plane while spectral cues help resolving the cone of confusion by giving front–back and elevation perception. HRTFs serve a dominant role for implementation of binaural technology and virtual auditory reality, including 3D sound system. Spatial perception can be made through the convolution of the sound with the appropriate pair of HRTFs. There are two critical aspects to HRTFs: they vary according to sound source position and are unique for each individual. Theoretically, an ideal 3D sound field may be generated with individualized HRTFs, which are measured accurately for each subject separately. But in practice individualized HRTFs are difficult to be obtained because the measurement is time-consuming and requires specific equipments. In practice, generalized or non-individualized HRTFs are used, which may cause large localization errors which typically appear as up–down confusion, front–back confusion and in-head localization [3]. To solve the problems, some spectrum modification methods have been proposed to enhance the perception difference from different sound directions [4,5]. But these methods only improve the localization performance a little because of the diversity of individuals. Recently some methods have been proposed to select a most appropriate HRTF set from a large number of HRTF sets for the listener by subjective listening comparison [6,7]. The subjective selection method requires no extra measurement equipment and thus can resolve the localization error problem to a certain extent when individualized HRTFs are difficult to be acquired. But the HRTF sets of multi-subjects consume a lot of storage space, this impedes the application of the subjective selection method in binaural technology and virtual auditory reality system.

Much research work has been done to reduce the memory consumption of the HRTF. The research may be divided into three kinds. The first kind is to calculate the HRTF by accurately acoustic modelling for the given sound direction [8,9]. This method requires little memory, but its localization performance needs further examination because of the complexity of human body structure. The second kind is pole-zero modelling of the measured HRTF, which represents the HRTF as a low order pole-zero model [10–12]. It can reduce the data size and computation load greatly. The third kind is to represent the HRTFs as weighted combinations of a few basis functions by mathematical decomposition. As an example of the third method, Chen et al. [13] and Bai et al. [14] modeled the external ear as a multisensor broadband beamformer

* Corresponding author. Tel./fax: +86 411 84707049.
E-mail addresses: wanglin_2k@sina.com (L. Wang), flyin@dlut.edu.cn (F. Yin), eeyin@dlut.edu.cn (Z. Chen).

0003-682X/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.apacoust.2009.04.003

whose weight vector set represented the characteristics of HRTF. Kistler and Wightman applied principal component analysis (PCA) to the logarithms of the HRTF magnitudes and represented them as linear combinations of basis functions [15]. Chen et al. proposed a spatial feature extraction and regularization (SFER) model for the complex valued (magnitude and phase) HRTFs, which represents HRTF as weighted combinations of a set of complex valued eigen transfer functions [16]. It is believed that there are a lot of similarities existing in the HRTFs [15], therefore a HRTF can be expressed as a linear combination of a few basis functions. The compression performance of the third kind of method is the best among the three ones, but the total memory size of it is still huge when there are a large number of HRTFs.

To reduce the memory size of the HRTF database, especially the one composed of multi-subjects' HRTFs, a hybrid compression method for HRTF is proposed. The HRTF is firstly approximated by a minimum-phase HRTF and an all pass filter whose group delay equals the interaural time delay (ITD). Principal component analysis is applied to the minimum-phase HRTF magnitudes to derive a set of basis functions for the entire data set and a weight vector for each HRTF. The weight vector set is vector quantized with the designed vector codebook [17]. The ITD variation with azimuth and elevation is curved surface fitted with a cosine series bivariate polynomial model. Comparing with principal component analysis, the proposed method may reduce the data size greatly with similar reconstruction precision.

This paper is organized as follows. The hybrid compression method for HRTF is developed in Section 2 and described in three aspects: principal component analysis, vector quantization, and curved surface fitting. Experiment results and analysis are given in Section 3 to evaluate the performance of the proposed method. Finally, conclusions are drawn in Section 4.

2. The hybrid compression method for HRTF

The classical compression method, principal component analysis, is well known for its compression efficiency. It decomposes a group of HRTFs into principal components and the corresponding principal component weights, where principal components are shared by the whole group and principal component weights remain unique for each HRTF. Although principal component analysis may reduce the data size greatly, the weight vector of each HRTF still consume much memory. Based on principal component analysis, we employ vector quantization method for the weight vector and curved surface fitting method for the interaural time delay (ITD) data separately to get further compression.

First of all, each HRTF is minimum-phase reconstructed, which results in a minimum-phase HRTF cascaded with an all pass filter whose group delay equals the interaural time delay (ITD).

Second, principal component analysis is applied to the entire minimum-phase HRTF magnitude set. A small set of basis functions are derived from the principal component analysis. Each spectrum can be approximated with a weighted sum of the basis functions, and the weights define the relative contribution of each basis function to the spectrum.

Third, the entire weight set is vector quantized, where a weight vector can be represented as a vector index in the codebook, and thus the storage space is saved.

At last, the ITD variation with direction is curved surface fitted with a cosine series bivariate polynomial for each subject in the database separately. Thus only a few polynomial coefficients need to be stored instead of the ITD set.

The workflow of the proposed method is shown as in Fig. 1. The basis functions and the vector codebook are shared by the entire HRTF database, the ITD polynomial coefficients are calculated for each subject, respectively, and the weight vector index is unique for each HRTF direction. All these parameters have been computed offline and stored in the memory beforehand. When binaural synthesis, the HRTF can be reconstructed with its corresponding parameters read from the memory. Because principal component analysis, vector quantization and curved surface fitting are employed jointly to reduce the size of the HRTF data, we call the new method a hybrid compression method. And the proposed hybrid compression method is described in detail as follows.

2.1. Minimum-phase reconstruction and principal component analysis

Generally, HRTF $H(e^{j\omega})$ can be approximated by a minimum-phase sequence combined with a position-dependent interaural time delay [18]:

$$H(e^{j\omega}) = H_{min}(e^{j\omega})H_{ap}(e^{j\omega}) \quad (1)$$

where H_{min} is the minimum-phase function and H_{ap} is an all pass function with the group delay $\tau = \text{ITD}$, the interaural time difference.

$$H_{ap}(e^{j\omega}) = e^{j\omega\tau} \quad (2)$$

The phase of the minimum-phase function can be estimated from the magnitude function. Thus only the magnitude of the minimum-phase HRTF is processed by principal component analysis (PCA).

The key idea of principal component analysis is to reduce the dimensionality of a data set while retaining the primary variation in the data. It decomposes a set of magnitude spectra into weighted combinations of basis functions. Suppose d_k is the kth magnitude spectrum of the data set, it can be represented as a linear combination of basis functions by PCA. This is described as

$$d_k = \sum_{i=1}^{q} w_{ki}c_i \quad (3)$$

where c_i is the ith basis function, and w_{ki} is the ith weight for d_k, q is the total number of basis functions. The basis functions are shared by the whole spectrum set and is called principal component (PC); the weight define the relative contribution of each basis function to the spectrum and is called principal component weight (PC weight),

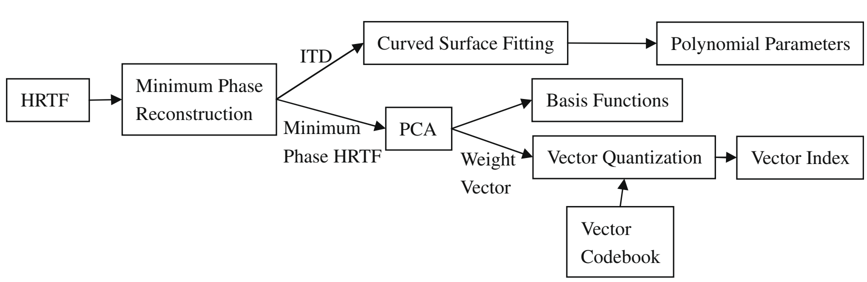

Fig. 1. Workflow diagram of the hybrid compression method for HRTF.