Applied Acoustics 87 (2015) 162–173

Contents lists available at ScienceDirect

Applied Acoustics

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

Individualization of head related transfer functions using principal component analysis

Kimberly J. Fink, Laura Ray *

Thayer School of Engineering, Dartmouth College, 14 Engineering Drive, Hanover NH 03755, United States

ARTICLE INFO

Article history:
Received 25 February 2014
Received in revised form 3 July 2014
Accepted 4 July 2014
Available online 25 July 2014

Keywords:
Head related transfer functions
Principal component analysis
Virtual auditory display

ABSTRACT

Prior research investigates virtual auditory displays (VADs) using models of HRTFs as a function of a finite number of principal components (PCs) and associated weights (PCWs). This paper studies the effect of PCWs on horizontal plane HRTFs derived from a database of HRIRs and provides a principled approach to PCW tuning. Tuning is first evaluated numerically to determine how variation of PCWs from an average PC model affects HRTF spectral characteristics. An average PC model at 50 azimuths in the horizontal plane is developed from a database of HRIRs of 34 subjects. HRIRs of nine additional subjects are used to test the validity of the average model and to conduct numerical optimization experiments, in which a cost function of spectral distortion is minimized by sequentially tuning PCWs. Sequential tuning mimics how a human would tune a VAD. Numerical results show that sequential tuning of a subset of PCWs reduces spectral distortion metrics when tuning an average HRTF to match an individual HRTF. These experiments show that tuning PCWs can change the shape and frequency location of the pinna notch. The numerical experiments also aid in developing a tuning method that is amenable to human tuning. Several variants of subject tuning experiments are conducted to verify that sequential tuning reduces listening errors. Results of a head steering task show an improvement of 30% in large heading errors when using a tuned VAD relative to an untuned VAD.

1. Introduction

Spatial localization of sound is known to depend on interaural time delay and level difference, and more generally on the interaction of the source with the head, torso, and pinna. Head-related transfer functions (HRTFs) model the acoustic transfer functions between the source and the listener's tympanic membrane from which listeners derive spatial information from binaural signals and have been used to create a virtual auditory display (VAD) through head phones [1,2]. A high quality VAD typically requires a large number of HRTFs, as well as a means of continuous representation or interpolation of HRTFs [3–5]. Non-individualized HRTFs are generally used to create a VAD owing to the difficulty of measuring individual HRTFs [6–8]. These, however, can create errors in sound localization such as front-back reversals, up-down confusions, and lateralization or inside of head localization [6–9]. This paper presents and evaluates a method that enables listeners to tune a VAD in the horizontal plane originating from an HRTF database. The method incorporates principal component analysis (PCA) to reduce dimensionality of head-related impulse responses (HRIRs) identified from HRIRS of 34 subjects in a public database and shows that principal component weights (PCWs) can be tuned sequentially in order to individualize the HRTFs.

Methods for HRTF customization are important in reducing listening errors. Principal component analysis (PCA) has been used by several authors to model HRIRs or HRTFs and to reduce the dimensionality of an HRIR or HRTF dataset. PCA reduces dimensionality by transforming a number of potentially correlated variables into a smaller number of uncorrelated principal components (PCs), where a small number of PCs recover a large percentage of the variability in the database [10–12]. Kistler and Wightman [12] use frequency-domain PCA to reduce dimensionality of 5300 HRTFs (10 subjects at 265 locations for both ears) measured at 11 elevations from -48° to 72° and 24 azimuths from -165° to 180° [2]. They found that the first five PCs retain ~90% of the variation in the original log-magnitude HRTF dataset. Middlebrooks and Green [13] also perform PCA in the frequency domain on the HRTFs of eight subjects at 360 locations for both ears. They also found that the first five PCs covered approximately 90% of the variation of the original HRTFs, and their PCs were similar to those of [12]. By splitting the subjects into two groups according to height and performing

* Corresponding author. Tel.: +1 603 646 1243.
E-mail address: lray@dartmouth.edu (L. Ray).

http://dx.doi.org/10.1016/j.apacoust.2014.07.005
0003-682X/

PCA twice [13], found a dependency of PCs on the physical size of subjects.

The CIPIC database is comprised of the measured HRIRs of 43 subjects (27 male and 16 female) and 2 KEMAR manikins at 1250 locations (25 azimuths and 50 elevations) [14]. It incorporates anthropometric measurements of many subjects. The HRIRs contain 200 samples at a sample frequency of 44.1 kHz. Zotkin et al. [3] use anthropometric database matching to generate semi-personalized HRTFs from the CIPIC database. Seven measurements of each pinna are taken with a camera. Using the CIPIC data for HRTFs and pinna measurements, the best matching set of HRTFs is found for the left and right ears. Although this method reduces localization errors, it requires taking a picture of a listener's pinna and automatically computing all the desired dimensions. Xu et al. [15] followed a similar approach finding that anthropometric data can improve HRTF personalization.

Refs. [16–21] develop customization methods that do not involve anthropometric data. They use PCA in the time domain, rather than in the frequency domain, on portions of the HRIRs in the CIPIC database in the median plane. Shin and Park [16] perform PCA on the first 0.2 ms of HRIRs of 45 subjects in the median plane at 45 elevations. Shin and Park [16] found that five PCs for each ear at each elevation recover more than 90% of the variation in the original 0.2 ms of the HRIRs. They also found that the PCs were similar for each ear, so ear symmetry was assumed. They allowed subjects to tune all five PCWs for the left ear HRIRs at each elevation, modifying the first 0.2 ms of the HRIR, and for the remainder, a KEMAR HRIR was used. Testing was done on four subjects with the customized and KEMAR HRIRs. The HRIRs of the four subjects were also measured and included in the subject testing. The front-back reversal rate is 28.1% for the KEMAR HRIRs, 13.1% for measured HRIRs, and 10.6% for tuned HRIRs.

Hwang et al. [18] perform PCA on the HRIRs of 45 subjects from the CIPIC database in the median plane. They, however, only perform a single analysis that includes HRIRs for 49 elevations in the median plane and every subject. The first 1.5 ms of the HRIRs are used, and once again, ear symmetry is assumed. With the first 12 PCs they reconstruct the original HRIRs with 4.8% modeling error and retain 90.2% of the variation. For customization, subjects tune the three weights with the highest standard deviations at each elevation; therefore, each elevation has its own set of PCWs being tuned. Customization of the HRIRs and subject testing is reported for three subjects. The average front-back reversal rate is 22.6% with KEMAR HRIRs, 0.37% with measured HRIRs, and 5.9% with customized HRIRs [18]. Refs. [17, 19–21] present similar methods and results.

While these studies provide evidence that PCW tuning gives customized HRIRs with improved front-back reversal rate, there are no studies that quantitatively relate tuning to changes in the HRTF or that relate tuning procedures amenable to human use to achievable variations in the HRTF through tuning. The objectives of this paper are (1) to investigate the effect of tuning PCWs on horizontal plane HTRFs derived a from PC model of the HRIR, and (2) to identify a procedure for individualizing an HRTF through tuning PCWs in a manner that is amenable to use by human subjects to reduce characteristic anomalies in average HRTFs, e.g., front-back reversals and lateralization. Specifically, we wish to determine how varying PCWs from an average PC model affects HRTF spectral characteristics, such as the pinna notch. Additionally, we identify a subset of PCWs that can be used to tune a VAD in order to individualize an HRTF. To do so, we construct a numerical nonlinear optimization problem in which PCWs are tuned sequentially, in a specified order, to minimize an objective function of spectral distortion. We use sequential tuning as a model of how a human would tune a VAD; while more complex nonlinear optimization results exist, it is too complex for a human to tune multiple parameters simultaneously.

We first develop a PC model of an average subject at each of 50 azimuths in the horizontal plane using HRIRs reported for 34 subjects in the CIPIC database. We identify a set of PCs that models the original, measured HRIRs with less than 5% error, averaged over 34 subjects. We hold out nine subjects from the CIPIC database in order to test the validity of the PC model derived from these 34 subjects, and we use HRIR data for these nine subjects with the optimization procedure to develop a tuning procedure using a small number of PCWs, in order to match the average PC model to the subject's HRTF derived from a measured HRIR. Two optimization experiments are reported. In the first, an average HRTF is constructed from the CIPIC database for a specified direction, and a subset of weights are tuned to match a specific subject's HRTF. In the second experiment, a subset of PCWs of the average HRTF for a specified direction are tuned to match the HRTF of a specific subject for that direction reflected to the back of the head. These numerical experiments thus investigate customization of an average HRTF for a given direction to a specific subject, and customization of an average HRTF to eliminate a front-back reversal. Following numerical experiments, results of subject testing, in which subjects tune PCWs sequentially to customize a VAD are presented. A head steering task is conducted with the tuned VAD to demonstrate effectiveness in a listening scenario.

2. Theory of principal component analysis of HRIRs

2.1. Principal component analysis (PCA)

PCA reduces the dimensionality of a dataset while keeping as much of the original variation as possible [10,11]. One PCA approach is based on singular value decomposition (SVD), in which the data are assembled into $N \times M$ matrix, X, where each row is an observation. The mean of X is calculated as

$$\bar{x}_m = \frac{1}{N}\sum_{n=1}^{N} X_{n,m} \tag{1}$$

where the subscript denotes the element. $\bar{x}_m$ is subtracted from X providing matrix $B = X - u\bar{x}$ where u is a $N \times 1$ vector of ones. SVD of B gives

$$B = USV^T \tag{2}$$

The original data X can be reconstructed by,

$$X = USV^T + u\bar{x} \tag{3}$$

V is an $M \times M$ matrix and the columns are the PCs. The PCWs are contained in $N \times M$ matrix US with each row corresponding to the PCWs for one observation from the original matrix X. If all the PCWs are used, the original data is reconstructed without error. If the number of PCs is truncated to reduce the size of the data, error is incurred.

2.2. Selection of the number of PCWs

PCA should provide a set of basis functions that can represent the HRIRs of a general population. While it is unknown whether the CIPIC or any other existing database is sufficiently large to provide a set of basis functions, we investigate this question by using a large number of subjects from the database and holding out subjects based on anthropometric data for validation, setting a maximum error threshold of 5% [17–21]. We perform PCA on the HRIRs of 34 subjects from the CIPIC database for all 50 azimuths in the horizontal plane. HRIRs of KEMAR manikins are removed