ANTHROPOMETRIC-BASED CUSTOMIZATION OF HEAD-RELATED TRANSFER FUNCTIONS USING ISOMAP IN THE HORIZONTAL PLANE

Felipe Grijalva[1] *Luiz Martini*[1] *Siome Goldenstein*[2] *Dinei Florencio*[3]

[1]School of Electrical and Computer Eng., University of Campinas. Campinas, Brazil
[2]Institute of Computing, University of Campinas. Campinas, Brazil
[3]Multimedia, Interaction, Communication Group, Microsoft Research. Redmond, WA, USA

ABSTRACT

In this paper, we introduce a new anthropometric-based method for customizing of Head-Related Transfer Functions (HRTF) in the horizontal plane. The method uses Isomap, artificial neural networks (ANN), and a neighborhood-based reconstruction procedure. We first modify Isomap's graph construction step to emphasize the individuality of HRTFs and perform a customized nonlinear dimensionality reduction of the HTRFs. We then use an ANN to model the nonlinear relationship between anthropometric features and our low-dimensional HRTFs. Finally, we use a neighborhood-based reconstruction approach to reconstruct the HRTF from the estimated low-dimensional version. Simulations show that our approach performs better than PCA and confirm that Isomap is capable of discovering the underlying nonlinear relationships of sound perception.

Index Terms— HRTF, Manifold, Isomap, Auditory Augmented Reality, Virtual Auditory Display

1. INTRODUCTION

Head Related Transfer Function (HRTF) is the spectral filtering of a sound source caused by the head, pinna and torso before it reaches the eardrum. HRTFs are complex-valued functions that contain various types of localization cues, such as Interaural Time Difference (ITD), Interaural Level Difference (ILD) and spectral coloring. These static cues, in conjunction with dynamic cues (e.g. head movements), define our three dimensional perception of audio [1].

As auditory augmented reality applications become more important [2], there is increasing research effort in the customization of HRTFs. A significant problem for the implementation of 3D sound systems is the fact that spectral features of HRTFs differ among individuals [1]. Various studies show a decrease in localization accuracy due to nonindividualized HRTFs [3, 4]. Thus, it is necessary to personalize HRTFs to guarantee high quality 3D sound perception. However, custom HRTF measurement is a complex, time consuming, and not scalable procedure [5]. To avoid HRTF measurements, several theoretical models (spherical head model [6], the snowman model [7]) and numerical methods (boundary element method [8]) have been proposed. Nevertheless, theoretical models are approximations of complicated anatomy and numerical methods are computationally intensive.

On the other hand, since HRTFs are closely related to certain anthropometric parameters, they can therefore be customized from anthropometric measurements [9]. Anthropometric regression methods predict the individualized HRTFs of a new subject using a model derived from a baseline database. Usually, some dimensionality reduction is applied to the HRTFs prior to customization.

2. PRIOR WORK

Nishino et al. [10] performed Principal Component Analysis (PCA) on the log magnitude HRTFs in the horizontal plane for each direction and ear separately. Then, linear regression analysis for each direction and ear is applied on a baseline database, using 9 anthropometric parameters as inputs and 5 PCA weights as outputs. For a new subject outside the training database, the PCA weights are predicted from the linear models and then used to reconstruct the log magnitude of HRTFs. Finally, minimum-phase reconstruction [11] estimates the final complex-valued HRTFs.

Due to the inability of linear methods (such as PCA) to represent the complex relationship between HRTF and multiple variables (i.e direction, frequency and individual), Grindlay et al. [12] introduced a multilinear tensor framework representation for HRTF decomposition. The tensor has 3 modes: frequency mode, direction mode and subject mode. A single linear regression model is used for mapping anthropometric features to a 5 dimension vector representing the subject mode in the tensor. Li et al. [13] employ a similar approach for dimensionality reduction but instead of linear regression, they use an artificial neural network (ANN).

Moreover, nonlinear techniques have been applied to both

This work was supported by Microsoft-FAPESP grant 2012/50468-6, CNPq 307018/2010-5, and CAPES, and it is part of a project that was approved by Unicamp's IRB *CAAE 15641313.7.0000.5404.*

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

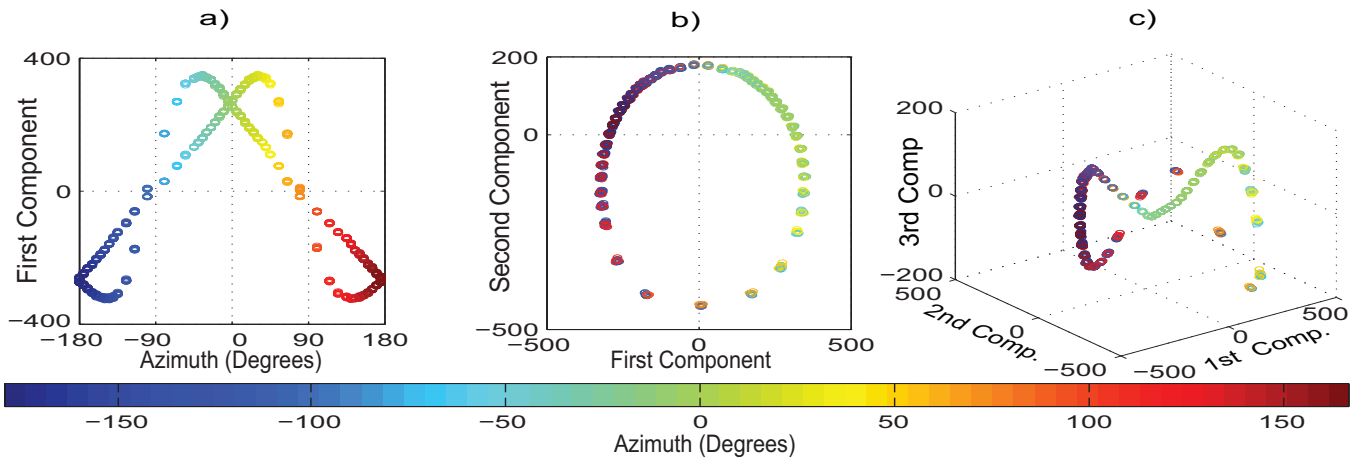


Fig. 1. Isomap Manifold for K=61 neighbors a) One-dimensional manifold as a function of azimuth b) Manifold embedded in two dimensions c) Manifold embedded in three dimensions

dimensionality reduction of HRTFs (e.g Isomap, Locally Linear Embedding) and to regression of HRTFs based on anthropometric features (e.g Support Vector Regression [14], ANNs [15, 13]). In [16], Duraiswami et al. present an exploratory study on learning the nonlinear manifold structure in vertical plane HRTFs using Locally Linear Embedding (LLE). They also propose a new method for HRTF interpolation and a new distance metric between two HRTFs based on the geodesic distance on the learned manifold.

Kapralos et al. [17, 18] conducted a comparative study from a quantitative point of view between PCA, Isomap and LLE for HRTF dimensionality reduction, finding that Isomap and LLE perform better than PCA in subjective experiments.

As in [15, 13], we employ an ANN for regression to predict the HRTFs for a new subject based on his anthropometric parameters. Unlike this prior work, we use nonlinear reduction technique, Isomap, to construct a manifold structure in horizontal plane HRTFs..

Our work is inspired on the successful results by Duraiswami et al [16] and Kapralos et al [17, 18] using LLE and Isomap for HRTF interpolation and dimensionality reduction. Their findings support the idea suggested by Seung et al [19] that nonlinear manifold techniques are crucial for understanding how perception arises from the dynamics of neural networks in the brain. However, neither of them addresses the customization of HRTFs as we do.

As in previous work [10, 15], we use the minimum phase approximations for HRTFs, a minimum-phase function cascaded with a pure delay [11]. In practice, the pure delay is the ITD and it is commonly cascaded in either the left or right HRTF of each left-right HRTF pair [10]. Calculation of ITD is beyond the scope of this paper. Several studies address the ITD calculation based on anthropometric parameters, notably in [20]. Here, we focus only on the spectral features of HRTFs magnitude and, unless otherwise stated, when we refer to HRTF we are referring to its magnitude.

3. HRTF CUSTOMIZATION

In this section we describe our HRTF personalization method. First, we reduce the HRTF dimensionality using Isomap. Then, we train an ANN with anthropometric parameters as inputs and the low-dimensional HRTFs as output – for each new subject with known anthropometric features, the ANN model predicts the low-dimensional HRTF representation. Finally, we use neighbor reconstruction mapping to recover the high-dimensional HRTFs from the low-dimensional space.

3.1. Dimensionality Reduction using Isomap.

In general, dimensionality reduction algorithms provide a method for taking a dataset represented in a $D \times N$ matrix $\mathbf{X}$ consisting of N sample vectors $\mathbf{x}_i$, *i.e.* $\mathbf{X} = \{\mathbf{x}_1, ..., \mathbf{x}_N\} \subset \mathbb{R}^D$ and calculating a corresponding low-dimensional representation in a $d \times N$ matrix $\mathbf{Y} = \{\mathbf{y}_1, ..., \mathbf{y}_N\} \subset \mathbb{R}^d$, where $d < D$. Here, consider all HRTFs in the horizontal plane as points in the D high-dimensional space.

Isomap is a nonlinear dimensionality reduction algorithm, first introduced in [21]. The first step in the Isomap algorithm is to construct a graph $G(V, E)$ on the input data set $\mathbf{X}$. Each sample $\mathbf{x}_i \in \mathbf{X}$ is represented by a node $v_i \in V$, and two nodes $v_i$and v_j are connected by an edge $(v_i, v_j) \in E$ with length $d_{\mathbf{X}}(\mathbf{x}_i, \mathbf{x}_j)$ if $\mathbf{x}_i$ is one of the K nearest neighbor of $\mathbf{x}_j$. The edge length $d_{\mathbf{X}}(\mathbf{x}_i, \mathbf{x}_j)$ is given by the Euclidean distance between $\mathbf{x}_i$ and $\mathbf{x}_j$ [21, 22].

The second step in Isomap involves computation of the shortest paths between all nodes in G. Distances are stored pairwise in a matrix $\mathbf{D}_G$. The distance matrix $\mathbf{D}_G$ represents geodesic distances between all samples on the manifold [22]. Because these distances are Euclidean, Isomap makes the same assumption of local linearity as LLE [22].

The third and final step is to construct the d-dimensional embedding calculating the eigenvectors of $\tau(\mathbf{D}_G)$, where $\tau(\mathbf{D}) = -\mathbf{HSH}/2$ and $S_{ij} = D_{ij}^2$ ($\mathbf{S}$ is the matrix of squared distances) and $H_{ij} = \delta_{ij} - 1/N$. Recall that N is

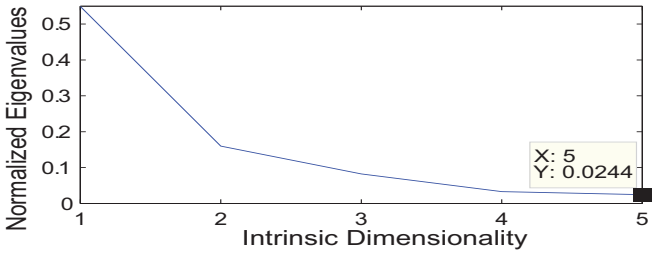


Fig. 2. Intrinsic Dimensionality Estimation.

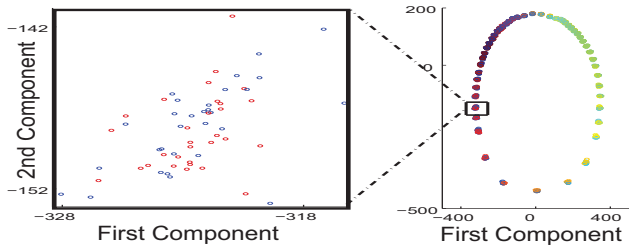


Fig. 3. Variability inside clusters due to inter-subject differences. Red and blue points represent HRTFs at symmetric azimuths of left and right ears respectively

the number of sample points and δ is the Kronecker delta function. Finally, let λ_p be the p^{th} eigenvalue (in decreasing order) of the matrix $\tau(\mathbf{D}_G)$, and v_p^i be the i^{th} component of the $\mathbf{p}^{th}$ eigenvector. Then set the p^{th} component of the d-dimensional coordinate vector $\mathbf{y}_i$ equal to $\sqrt{\lambda}v_p^i$ [21].

Isomap first step is the construction of a graph. The simplest approach is to select, for each data point, a fixed number of nearest neighbors, K, as measured by Euclidean distance. Other criteria, however, can also be used to choose neighbors, and in general, neighborhood selection in Isomap presents an opportunity to incorporate a priori knowledge [23].

We know that some correlation exists due to left-right symmetry of HRTFs at frequencies below 5.5 KHz [24]. Moreover, to emphasize the individuality of HRTFs across directions, Nishino et al. [10] perform PCA reduction separately for each direction and ear. Here, instead of applying Isomap separately for each direction and ear, we propose construct the graph taking into account this knowledge.

One of our contributions is our graph G construction procedure. Consider again the high-dimensional dataset in a $D \times N$ matrix $\mathbf{X} = \{\mathbf{x}_1, ..., \mathbf{x}_N\} \subset \mathbb{R}^D$ formed by N HRTFs of two ears of P subjects at M azimuths in the horizontal plane (i.e. $N = 2 \cdot P \cdot M$).

We connect each datapoint $\mathbf{x}_i$ to $K = 2P + 1$ neighbors and we set its edge lengths to $s_{ij} d_{\mathbf{X}}(\mathbf{x}_i, \mathbf{x}_j)$, where s_{ij} is a scale factor, according to the following rules: 1) If $\mathbf{x}_i$ and $\mathbf{x}_j$ represent HRTFs of the same azimuth and ear but different subject, then connect them and set $s_{ij} = 1/100$ in order to emphasize the individuality of HRTFs across directions. 2) Let θ_i and θ_j be azimuths of HRTFs represented by $\mathbf{x}_i$ and $\mathbf{x}_j$ respectively. Regardless of the subject, if $\mathbf{x}_i$ and $\mathbf{x}_j$ represent HRTFs of opposite ears and θ_j is the mirror horizontal azimuth of θ_i (i.e. $\theta_j = 360 - \theta_i$), then connect them and set $s_{ij} = 1/100$ in order to take advantage of left-right symmetry. 3) Let θ_i and θ_j be azimuths of HRTFs of the same subject represented by $\mathbf{x}_i$ and $\mathbf{x}_j$ respectively. If θ_j is the nearest azimuth greater than θ_i or if θ_j is the nearest azimuth less than θ_i , then connect $\mathbf{x}_i$ and $\mathbf{x}_j$ and set $s_{ij} = 1$.

Before applying Isomap, we first need to select the number of neighbors, K, and the intrinsic dimensionality, d. Due to our proposed graph construction explained above, the number of neighbors is set to $K = 2P+1$, where P is the number of subjects on the dataset $\mathbf{X}$. The intrinsic dimensionality was estimated analyzing the residual variance. Figure 2 shows the normalized eigenvalues (in decreasing order) calculated over the complete dataset $\mathbf{X}$. Since eigenvalues give the variance in each dimension, when they are lower than a threshold, little is gained from adding a dimension [22]. Using 0.025 for the threshold we find the intrinsic dimensionality $d = 4$ [25].

Unlike previous works [16], we apply Isomap only once, over the entire dataset – a single procedure for the HRTFs of all subjects, ears and directions taking into account our proposed neighborhood selection. Figure 1 shows the Isomap manifold calculated for all directions and ears of 30 individuals (i.e. $P = 30$, so $K = 2P+1 = 61$ neighbors) from CIPIC database [9] in the horizontal plane, where the color represent the azimuth angle. In Figure 1a, we plot the first embedded component of Isomap as a function of azimuth in order to highlight the symmetric properties of HRTFs. In Figure 1b and 1c, the manifold embedded in two and three dimensions show the variability of HRTFs across directions. Note that for each direction there are small clusters of reduced HRTFs. The variability inside these clusters is due to inter-subject differences (see Figure 3). Figure 1b illustrates that clusters are not uniformly distributed – the large gaps between some clusters is due to the HRTF non-uniform sampling in CIPIC database.

3.2. Regression using an Artificial Neural Network

ANN is a system inspired by human brain capable of approximating nonlinear functions of their inputs. Since the relationship between HRTFs and anthropometric parameters is very complex, it is difficult to express them with linear functions. Here, we apply a back propagation ANN with sigmoid activation function in the hidden layer and a linear activation function in the output layer. The inputs are s anthropometric parameters, the azimuth angle in the horizontal plane and the ear (Left=1, Right=-1). The outputs are the coordinates of the HRTFs in the low-dimensional space obtained in Section 3.1. In order to determine the number of hidden nodes, we varied it from 5 to 30 and selected 20 hidden nodes that produced the lowest mean squared error. Note that our approach requires training only one ANN for all directions and ears. After the regression model is learned, the individual HRTF on the low-dimensional space for a new subject can be predicted by his anthropometric parameter measurements.

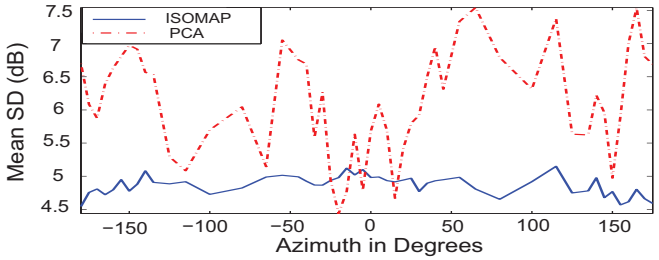
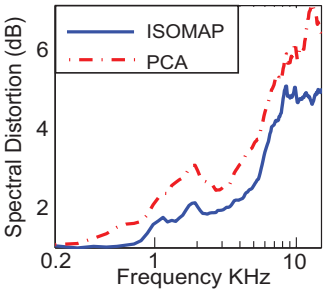


Fig. 4. Mean Spectral Distortion as a function of azimuth.



Azimuth	Mean SD Isomap	Mean SD PCA
-180	4.5±1.1	6.7±1.4
-100	4.7±0.7	5.7±3.2
0	5±1	5.7±3
100	4.9±1	6.3±3.7
145	4.7±1.1	6±2.3

(a) (b)

Fig. 5. a)Spectral Distortion b) Confidence Interval

3.3. Neighborhood Reconstruction Mapping

Unlike PCA and similar linear reduction methods, Isomap produce a low-dimensional embedding

$$\mathbf{Y}_{d\times N} = \{\mathbf{y}_1, ..., \mathbf{y}_N\} \in \mathbb{R}^d$$

from the samples in $\mathbf{X}$ without generating an explicit map [22]. As we are interested in reconstructing an HRTF in the high-dimensional space from the low-dimensional HRTF predicted by the ANN, we need to project a low-dimensional point $\mathbf{y}$ back into the original space. Since Isomap assumes that a sample and its neighbors are locally linear, we can perform the mapping using a linear combination of a sample's K neighbors [22], and the reconstructed HRTF, $\widehat{H}$,

$$\widehat{H} = \sum_{i}^{K} w_i \mathbf{x}_i \qquad (1)$$

to calculate the weights w_i, we follow Brown et al. [22], and choose w_i to be the inverse Euclidean distance between the sample and the neighbor i in the low-dimensional space.

4. SIMULATIONS

We use the publicly available CIPIC database [9] which contains head related impulse responses (HRIRs) measured for 45 subjects at 1250 directions (25 azimuths and 50 elevations in interaural coordinate system). We employ 50 azimuth directions per subject and ear corresponding to horizontal plane. Each HRIR is 200 samples long (roughly 4.5 ms at 44.1 KHz sampling rate and 16 bit resolution). Each HRIR was transformed into an HRTF by a 512-point FFT. To reduce the effects of error due to nonlinearity introduced by equipments used to measure HRIRs, HRTFs were filtered to preserve frequencies between 200 Hz and 15 kHz, leaving 172 frequencies in each HRTF magnitude. We use only subjects that has the complete anthropometric parameters (i.e. 35 subjects). Performance was evaluated using a K-fold cross-validation approach. We split the HRTF dataset into 7 folds of 5 subjects each (6 folds for training and 1 fold for testing). Because the number of subjects for training each fold is $P = 30$, then according to our neighborhood selection proposed, the number of neighbors for Isomap is set to $K = 2P + 1 = 61$

The CIPIC database also contains anthropometric measurements. We selected 8 anthropometric parameters for regression in accordance to [26]: head width, head depth, neck width, shoulder width, *cavum concha* height, *cavum concha* width, pinna height, and pinna width. As explained in Section 3.2, the azimuth angle, the L/R ear, and the eight anthropometric parameters are the inputs for the ANN and the outputs are the low-dimensional HRTFs reduced using Isomap. We used Matlab Neural Network Toolbox 8.0.

We implemented a PCA-based customization, for comparison, with seven principal components (90% of variance). We used a similar ANN structure for the regression model and K-fold cross-validations for testing. We used Matlab Dimensionality Reduction Toolbox [25] for both PCA and Isomap.

We choose the mean spectral distortion as an error metric,

$$SD_M = \sqrt{\frac{1}{N_f}\sum_{f_k}\left(20log_{10}\frac{|H(f_k)|}{|\widehat{H}(f_k)|}\right)^2} \qquad (2)$$

where H and $\widehat{H}$ represent the measured and reconstructed HRTF respectively and N_f is the number of frequency points. The reconstructed HRTF, $\widehat{H}$, was calculated using Equation 1.

As can be seen in Figure 4, our approach performs better than PCA. The confidence interval ($\pm 2\sigma, 95\%$) shows that our method has less variability than PCA (see Figure 5b). Moreover, our approach achieves better performance even with less dimensions than PCA. As in other studies [10], error increases at high frequencies due to complex scattering caused by pinna (Figure 5a) but in our approach it stays roughly below 5dB.

5. CONCLUSIONS

In this paper, we have introduced a new method for customizing HRTFs in the horizontal plane. Unlike previous works, we perform dimensionality reduction once on the entire HRTF dataset for all subjects, directions and ears in the horizontal plane. Besides using Isomap as a nonlinear dimensionality reduction technique, we introduce a brand-new graph construction technique that incorporates important prior information about the HRTFs. The results show that incorporating prior knowledge in the neighborhood selection in Isomap can lead to a better manifold representation, and we can conclude that Isomap is a promising reduction technique for HRTFs analysis and synthesis. As future work, we plan to extend our approach to estimate HRTFs beyond just the horizontal plane.

6. REFERENCES

[1] Durand R. Begault, *3D Sound for Virtual Reality and Multimedia*, AP Professional, 1994.

[2] Aki Härmä, Julia Jakka, Miikka Tikander, Matti Karjalainen, Tapio Lokki, Jarmo Hiipakka, and Gaëtan Lorho, "Augmented Reality Audio for Mobile and Wearable Appliances," *J. Audio Eng. Soc.*, 2004.

[3] Elizabeth M. Wenzel, "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Am.*, 1993.

[4] HG Fisher and SJ Freedman, "The role of the pinna in auditory localization.," *J. Aud. Res.*, 1968.

[5] Henrik Mø ller, "Fundamentals of binaural technology," *Appl. Acoust.*, 1992.

[6] PMC Morse and KU Ingard, *Theoretical acoustics*, McGraw-Hill, 1986.

[7] V. Ralph Algazi, Richard O. Duda, Ramani Duraiswami, Nail A. Gumerov, and Zhihui Tang, "Approximating the head-related transfer function using simple geometric models of the head and torso," *J. Acoust. Soc. Am.*, 2002.

[8] Makoto Otani and Shiro Ise, "Fast calculation system specialized for head-related transfer function based on boundary element method," *J. Acoust. Soc. Am.*, 2006.

[9] V Algazi, R Duda, D Thompson, and C Avendano, "The cipic hrtf database," in *Work. Appl. Signal Process. to Audio Acoust.* 2001, IEEE.

[10] Takanori Nishino, Kazuhiro Iida, Naoya Inoue, Kazuya Takeda, and Fumitada Itakura, "Estimation of HRTFs on the horizontal plane using physical features," *Appl. Acoust.*, 2007.

[11] D J Kistler and F L Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction.," *J. Acoust. Soc. Am.*, 1992.

[12] Graham Grindlay and M. Alex O. Vasilescu, "A Multilinear (Tensor) Framework for HRTF Analysis and Synthesis," in *ICASSP.* 2007, IEEE.

[13] Lin Li and Qinghua Huang, "HRTF personalization modeling based on RBF neural network," in *ICASSP*. 2013, IEEE.

[14] Q Huang and Y Fang, "Modeling personalized head-related impulse response using support vector regression," *J. Shanghai Univ.*, 2009.

[15] Hongmei Hu, Lin Zhou, Hao Ma, and Zhenyang Wu, "HRTF personalization based on artificial neural network in individual virtual auditory space," *Appl. Acoust.*, 2008.

[16] R. Raykar, VC Duraiswami, "The Manifolds of Spatial Hearing," in *ICASSP*. 2005, IEEE.

[17] Bill Kapralos and Nathan Mekuz, "Application of dimensionality reduction techniques to HRTFs for interactive virtual environments," in *Int. Conf. Adv. Comput. Entertain. Technol.* 2007, ACM.

[18] Bill Kapralos, Nathan Mekuz, Agnieszka Kopinska, and Saad Khattak, "Dimensionality reduced HRTFs: a comparative study," in *Int. Conf. Adv. Comput. Entertain. Technol.* 2008, ACM.

[19] HS Seung and DD Lee, "The manifold ways of perception," *Science*, 2000.

[20] V. Ralph Algazi, Carlos Avendano, and Richard O. Duda, "Estimation of a Spherical-Head Model from Anthropometry," *J. Audio Eng. Soc.*, 2001.

[21] J B Tenenbaum, V de Silva, and J C Langford, "A global geometric framework for nonlinear dimensionality reduction.," *Science*, 2000.

[22] W Michael Brown, Shawn Martin, Sara N Pollock, Evangelos A Coutsias, and Jean-Paul Watson, "Algorithmic dimensionality reduction for molecular structure analysis.," *J. Chem. Phys.*, 2008.

[23] K Saul Lawrence and T Roweis Sam, "Think globally, fit locally: Unsupervised learning of nonlinear manifolds," *J. Mach. Learn. Res.*, 2002.

[24] BoSun Xie, XiaoLi Zhong, Dan Rao, and ZhiQiang Liang, "Head-related transfer function database and its analyses," *Sci. China Ser. G Physics, Mech. Astron.*, 2007.

[25] Laurens van der Maaten, Eric Postma, and Jaap van den Herik, "Dimensionality reduction: A comparative review," *J. Mach. Learn. Res.*, 2009.

[26] Wahidin Wahab and Dadang Gunawan, "Enhanced Individualization of Head-Related Impulse Response Model in Horizontal Plane Based on Multiple Regression Analysis," in *Int. Conf. Comput. Eng. Appl.* 2010, IEEE.