The Effectiveness of Chosen Partial Anthropometric Measurements in Individualizing Head-Related Transfer Functions on Median Plane

Hugeng, Wahidin Wahab & Dadang Gunawan

Department of Electrical Engineering
Faculty of Engineering, University of Indonesia
Depok, Indonesia
Email: hugeng@ui.ac.id

Abstract. Individualized head-related impulse responses (HRIRs) to perfectly suit a particular listener remains an open problem in the area of HRIRs modeling. We have modeled the whole range of magnitude of head-related transfer functions (HRTFs) in frequency domain via principal components analysis (PCA), where 37 persons were subjected to sound sources on median plane. We found that a linear combination of only 10 orthonormal basis functions was sufficient to satisfactorily model individual magnitude HRTFs. It was our goal to form multiple linear regressions (MLR) between weights of basis functions acquired from PCA and chosen partial anthropometric measurements in order to individualize a particular listener's HRTFs with his or her own anthropometries. We proposed a novel individualization method based on MLR of weights of basis functions by employing only 8 out of 27 anthropometric measurements. The experiments' results showed the proposed method, with mean error of 11.21%, outperformed our previous works on individualizing minimum phase HRIRs (mean error 22.50%) and magnitude HRTFs on horizontal plane (mean error 12.17%) as well as similar researches. The proposed individualization method showed that the individualized magnitude HRTFs could be well estimated as the original ones with a slight error. Thus the eight chosen anthropometric measurements showed their effectiveness in individualizing magnitude HRTFs particularly on median plane.

Keywords: *HRIR; HRTF Modeling; individualization; PCA; MLR.*

1 Introduction

With limited or no vision, one will naturally recognize the direction of sound sources using one's hearing. The main cues in pinpointing the direction of a sound are interaural time difference (ITD), interaural level difference (ILD), and spectral adjustment caused by pinnae, head, and torso. These main sound cues are encrypted in HRTF. While ITDs and ILDs have almost no effect on determination of sound directions on median plane, however, spectral variations in HRTFs play a main role to distinguish the sound directions [1]. HRTF is

Received October 22nd, 2011, Revised March 5th, 2011, Accepted for publication March 6th, 2011.

defined as the acoustic filter of the human auditory system, in frequency domain, from a sound source to the entrance of ear canal. Head-related impulse response (HRIR) is the counterpart of HRTF in the time domain. An important realization of binaural HRTFs is in the creation of Virtual Auditory Display (VAD) in virtual reality to filter monaural sound. The basis of this fact is that the human psychoacoustic characteristic tends to find a spatial sound to be convincing when it is attained using two channels. It is pointed out in [2] and [3] that HRTF varies with directions of sound sources and differs from subject to subject due to inter-individual disparity in anthropometric measurements.

The synthesis of accurate VAD systems requires a series of empirical measurements on individual HRTFs for each listener. These types of measurements tend to be impractical as it involves heavy and expensive equipments in addition to excessive measurement time.

The majority of commercial virtual auditory systems nowadays are synthesized using generic/non-individualized HRTFs that do not take inter-subject differences into account. These non-individualized HRTFs contain distortions such as in-head localization when using headphones, inaccurate lateralization, poor vertical effects, and weak front-back distinction caused by unsuitable HRTFs applied to a listener [3]. Therefore an individualization method to estimate proper HRTFs for a particular listener, that presents satisfactory sound cues without measurement of the individual HRTFs, is critically needed.

Currently the individualization of HRTF in frequency domain or HRIR in time domain is the subject of rigorous research. Numerous HRTF individualization methods have been elaborated, as in HRTF clustering and selection of a few most representative ones [4], HRTF scaling in frequency [5], a structural model of composition and decomposition of HRTFs [6], HRTF database matching [7], the boundary element method [8], HRIR subjective customization of pinna responses [9] and of pinna, head, and torso responses [10] in the median plane, and HRTF personalization based on multiple regression analysis (MRA) in the horizontal plane [11]. Shin and Park [9] proposed HRIR customization method based on subjective tuning of only pinna responses (0.2 ms out of entire HRIR) in the median plane using PCA of the CIPIC HRTF Database [12]. The customized pinna responses were attained by letting a subject tune several weights of corresponding basis functions. Hwang and Park [10] pursued a comparable method as [9], except that they fed PCA with the entire median HRIRs; each HRIR was 1.5 ms long (67 samples) since the arrival of direct pulse. This HRIR included the pinna, head, and torso responses. The weights of three dominant basis functions were tuned according to the three largest standard deviations at each elevation. Hu, et al. [11] personalized the estimated log-magnitude responses of HRTFs by MRA. Initially, the log-magnitude responses were estimated using PCA as linear combination of ten weighted basis functions. Subsequently the weights of the basis functions were estimated using some anthropometric measurements based on MRA.

Our individualization method was in line to the method in [11]. However we employed the magnitude responses of HRTFs in the PCA modeling, instead of the log-magnitude responses of HRTFs utilized by Hu, et al. Furthermore, our selection procedure of anthropometric measurements was notably different. The whole range of median magnitude HRTFs calculated from the original HRIRs in the CIPIC HRTF Database were included in a single analysis. Therefore all median magnitudes HRTFs for both ears shared the same set of basis functions, which covered the inter-individual variations as well as the inter-elevation variation. This paper presents an individualization method by developing the statistical PCA model of magnitude HRTFs and MLR between weights of basis functions and selected few anthropometric measurements. It is shown later that our results outperformed that of [11].

This paper is organized as follows. Section 2 discusses the proposed algorithm of individualization method, minimum phase analysis, PCA of magnitude HRTFs, minimum phase reconstruction and synthesis of HRIR models, individualization of magnitude HRTFs using MLR, and correlation analysis for the selection process of independent variables and dependent variables of MLR models. Section 3 elaborates the experiments' results, consisting of discussions of resulting basis functions and weights of basis functions from PCA, and the performance of the proposed individualization method. Section 4 concludes the paper.

2 Proposed Individualization Method

The aim of this research was developing a novel individualization method of HRTFs on median plane, by using multiple regression models between weights of PCA models of magnitude HRTFs and selected few anthropometric measurements. This method individualized magnitude HRTF models into suitable HRIRs for a particular listener, by utilizing only a little number of his or her own anthropometries. When a listener uses a spatial auditory application, suitable individualized binaural HRIRs are imperative. The selection of magnitude HRTFs to be modeled by PCA was based on the fact that modeling magnitude HRTFs provided best results among other data types of HRTFs in frequency domain, as shown in our previous work [13].

Figure 1 shows the schematic diagram of the proposed method. The database of HRIRs used in this research was provided by CIPIC Interface Laboratory of California University at Davis [2,12]. First, as seen in Figure 1, the entire original HRIRs on median plane of 37 subjects were attained from the database. There were 50 HRIRs for each ear of each subject. A total number of M=3700 HRIRs were used in modeling and individualizing HRIRs of a listener. Each HRIR was transformed into its corresponding complex HRTF using 256-points fast Fourier transform (FFT).

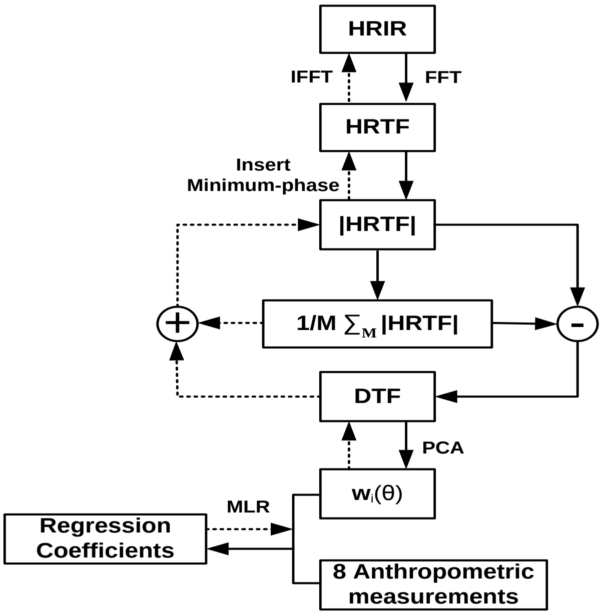

Figure 1 Proposed HRTFs individualization method.

Only 128 frequency components of magnitude of the complex HRTF were used as the object of HRTF modeling using PCA. At this step, the phase of the complex HRTF was disregarded. Next the mean of the overall magnitude HRTFs was computed. The resulting mean was subtracted from each magnitude HRTF to obtain its corresponding direct transfer function (DTF). This subtraction was performed in order to have centered data of magnitude HRTF, called DTF, which was critical in obtaining a PCA model with good results.

All DTFs were subsequently fed into PCA. The PCA delivered 128 ordered basis functions or principal components (PCs) and their weights (PCWs). The PCs were put in decreasing order according to their Eigen values. Note that each Eigen value determined the percentage variance of all DTFs as stated by its corresponding PC. The first PC that corresponds to the largest Eigen value indicated the largest percentage variance of the entire DTFs. To attain an individualized HRTFs of a new listener, we carried out multiple linear regression (MLR) between the PCWs resulted from PCA and a few anthropometric measurements of 37 subjects in the database. A thorough selection process of anthropometric measurements from a total of 27 measurements is explicated in a separated subsection below. The MLR method presented regression coefficients that correlated the PCWs and selected anthropometric measurements. Consequently these regression coefficients were applied to a set of anthropometric measurements of a new listener to obtain estimated PCWs for that listener. A linear combination of weighted PCs using these estimated PCWs resulted in an individualized DTF.

The dashed line in Figure 1 shows the reconstruction process of obtaining the desired individualized HRIRs. Each individualized DTF that was attained from the MLR method and PCA, was added to the mean of the previously calculated DTFs to generate its individualized magnitude HRTF. Minimum-phase was then inserted to the individualized magnitude HRTF in order to obtain an individualized complex HRTF. It is assumed that the phase of the HRTF can be estimated using minimum-phase [14]. Lastly, individualized HRIRs were obtained from the corresponding complex HRTFs by applying inverse fast Fourier transform (IFFT). The distance between a sound source in particular direction and each ear drum resulted in an initial left- and right-ear time delay, which were inserted correspondingly to the left- and right-ear HRIR.

The following subsections elaborate the minimum phase analysis, PCA of the magnitude HRTFs in the frequency domain, minimum phase reconstruction and the synthesis of HRIRs, MLR method, as well as the selection process of anthropometric measurements.

2.1 Minimum Phase Analysis

Each HRIR in the dababase was measured using a distance of one meter from the sound source to the center of subject's head. From the graph of HRIR versus time, it is noted that there exists a time delay due to this distance. This delay is the time taken by a sound wave to propagate from its source to the ear drum, before a maximum amplitude of HRIR occurs. To alleviate this time delay, HRIR can be reconstructed into a minimum-phase HRIR using Hilbert transform. In the minimum-phase HRIR, the phase is allowed to be arbitrary or it is set in such a way so that the magnitude response of HRIR is easier to attain. A linear time invariant filter, $H(z) = B(z)/A(z)$, is said to have minimum phase if all of its poles and zeros are inside the unit circle, $|z|=1$, in the z-plane. Equivalently, a filter, $H(z)$, has minimum phase if both itself and its inverse, $1/H(z)$, are stable. A minimum-phase filter is also causal since non causal terms in the transfer function correspond to poles at infinity. The simplest example of minimum-phase filter would be the unit-sample filter, $H(z) = z$, which consists of a zero at $z = 0$ and a pole at $z = \infty$. A filter has a minimum phase if both the numerator and denominator of its transfer function are minimum phase polynomials in z^{-1}, i.e. a polynomial of the form,

$$B(z) = b_0 + b_1 z^{-1} + b_2 z^{-2} + \ldots + b_M z^{-M}$$
$$= b_0(1-\zeta_1 z^{-1})(1-\zeta_2 z^{-1})\ldots(1-\zeta_M z^{-1}) \quad (1)$$

has a minimum phase if all of its roots, ζ_i, i=1,2,...,M, lie inside the unit circle, i.e. $|\zeta_i| < 1$. A common characteristic of minimum-phase impulse responses is that among all impulse responses, $h_i(n)$, having identical magnitude spectra, impulse responses with minimum phases undergo the fastest decay in the sense that,

$$\sum_{n=0}^{K} \left| h_{mp}(n) \right|^2 \geq \sum_{n=0}^{K} \left| h_i(n) \right|^2, \; n=0, 1, 2, ..., K, \quad (2)$$

where $h_{mp}(n)$ is a minimum-phase impulse response.

The equation above signifies that the energy in the first K + 1 samples of the minimum phase case is at least as large as any other causal impulse response having the same magnitude spectrum. Therefore, minimum-phase impulse responses are maximally concentrated toward time t=0 among the space of causal impulse responses for a particular magnitude spectrum. Due to this characteristic, minimum-phase impulse responses are often named minimum-delay impulse responses. It is known that in a minimum phase filter, $H(z) = e^{a(z)}e^{i\,b(z)}$, the relations, $b(z) = -\mathcal{H}\{a(z)\}$ and $a(z) = -\mathcal{H}\{b(z)\}$, are also valid, where $\mathcal{H}\{\}$ denoted the Hilbert transform. The logarithmic change of these relations is attained primarily through the calculation of real cepstrum.

Kulkarni, et al. [14] suggested that the phase of HRIR can be estimated by using minimum phase. A minimum phase system function, H(z), of an HRIR, h(n), comprises of poles and zeros which are located inside the unit circle $|z| = 1$ in the z-plane. The calculation of real cepstrum of an original HRIR, which has arbitrary phase, results in a minimum phase HRIR, $h_{mp}(n)$. It can be concluded that the minimum phase HRIR is the removed initial time delay version of the corresponding original HRIR. However both kinds of HRIR have the same magnitude spectrum in the frequency domain. The real cepstrum, v(n), of HRIR, h(n), is calculated as follow,

$$v(n) = \mathrm{Re}\{F_D^{-1}\{\ln|F_D\{h(n)\}|\}\} \quad (3)$$

where ln and Re{} denote respectively natural logarithm and the real part of a complex variable, $F_D\{\}$ and $F_D^{-1}\{\}$ are the discrete Fourier transform and its inverse respectively. This real cepstrum is subsequently weighted by the following window function,

$$w(n) = \begin{cases} 0 & \text{if } n < 0, \\ 1 & \text{if } n = 0, \\ 2 & \text{if } n > 0. \end{cases} \tag{4}$$

In case of a rational H(z), the window function can be observed as a complex conjugate inversion of the zeros outside the unit circle, so that a minimum phase HRIR is provided. Therefore the desired minimum phase HRIR, $h_{mp}(n)$, is resulted from:

$$h_{mp}(n) = \mathrm{Re}\{\exp(F_D\{w(n).v(n)\})\}. \tag{5}$$

2.2 PCA of Magnitude HRTFs in Frequency Domain

FFT was implemented to HRIRs of the database utilized to attain the complex HRTFs. The whole complex HRTFs were calculated from left-ear and right-ear HRIRs of 37 subjects on median plane. There are 50 HRIRs from different directions (50 elevations) on median plane for each ear of a subject, so that a total of 3700 complex HRTFs were formed by 256-points FFT. Only the magnitudes of all complex HRTFs were taken as the input of PCA modeling. Only 128 first frequency components of a magnitude HRTF were taken into analysis due to the symmetry characteristic of a magnitude spectrum.

A matrix composed of DTFs was needed by PCA. The original data matrix, $\mathbf{H}$ (NxM), was composed of magnitudes of HRTFs on median plane, in which, each column vector, $\mathbf{h}_i$ (i=1,2,…,M), represented a magnitude HRTF of an ear of a subject in a direction on median plane. The total number of magnitude HRTFs of each subject on median plane was 100 (2 ears x 50 elevations). Hence, the size of $\mathbf{H}$ was 128 x 3700 (N=128, M=3700). The empirical mean vector ($\boldsymbol{\mu}$: Nx1) of all magnitude HRTFs is given by,

$$\boldsymbol{\mu} = (1/M)\sum_{i=1}^{M} \mathbf{h}_i. \tag{6}$$

The DTFs matrix, $\mathbf{D}$, is the mean-subtracted matrix and given by,

$$\mathbf{D} = \mathbf{H} - \boldsymbol{\mu}.\mathbf{y}, \tag{7}$$

where $\mathbf{y}$ is a 1xM row vector of all 1's. The next step was to calculate a covariance matrix, $\mathbf{S}$, that is given by

$$\mathbf{S} = \mathbf{D}.\mathbf{D}^{*}/\,(M-1) \tag{8}$$

where * indicates the conjugate transpose operator. The basis functions or PCs, $\mathbf{v}_i$ (i=1,2,…,q), were the q Eigen vectors of the covariance matrix, $\mathbf{S}$, corresponding to q largest Eigen values. If q = N, then the DTFs could be fully reconstructed by a linear combination of the N PCs. However, q was set smaller than N because the goal of PCA is to reduce the dimension of dataset. An estimate of the original dataset was obtained in this research by only 10 PCs, which justified 92.49% variance in the original data **D**. By using only 10 PCs to model magnitude HRTFs, we expected to obtain satisfactory results. The PCs matrix, $\mathbf{V} = [\mathbf{v}_1\ \mathbf{v}_2\ \ldots\ \mathbf{v}_N]$ consisting of a complete set of PCs can be attained by solving the following Eigen equation,

$$\mathbf{S}\ \mathbf{V} = \mathbf{\Lambda} \tag{9}$$

where $\mathbf{\Lambda} = \text{diag}\{\lambda_1,\ldots,\lambda_{128}\}$, is a diagonal matrix formed by 128 Eigen values. Each Eigen value, λ_i, represents a sample variance of DTFs that is projected onto i-th Eigen vector or PC, $\mathbf{v}_i$.

Correspondingly, the weights of PCs (PCWs), **W**(10x3700), that relate to all DTFs, **D**, can be stated as,

$$\mathbf{W} = \mathbf{V^*.D}, \tag{10}$$

where the PCs matrix, **V,** had by now been reduced to $\mathbf{V} = [\mathbf{v}_1\ \mathbf{v}_2\ \ldots\ \mathbf{v}_{10}]$. PCWs represent the contribution of each PC to a DTF. They contain both the spatial features and the inter-individual difference of DTF. Thus, the matrix which consisted of models of magnitude HRTFs, $\hat{\mathbf{H}}$, can be stated as

$$\hat{\mathbf{H}} = \mathbf{V.W} + \boldsymbol{\mu}.\mathbf{y}. \tag{11}$$

Table 1 The percentage of variance related to basis functions.

PC	Eigen Value	Variance (%)	Cumulative Variance (%)
$\mathbf{v}_1$	11.61	37.96	37.96
$\mathbf{v}_2$	6.93	22.66	60.62
$\mathbf{v}_3$	2.92	9.56	70.18
$\mathbf{v}_4$	1.73	5.65	75.83
$\mathbf{v}_5$	1.28	4.17	80.00
$\mathbf{v}_6$	1.10	3.61	83.61
$\mathbf{v}_7$	0.91	2.96	86.56
$\mathbf{v}_8$	0.77	2.52	89.09
$\mathbf{v}_9$	0.55	1,79	90.88
$\mathbf{v}_{10}$	**0.49**	**1.61**	**92.49**
$\mathbf{v}_{11}$	0.43	1.41	93.90
$\mathbf{v}_{12}$	0.35	1.15	95.05
$\mathbf{v}_{13}$	0.33	1.08	96.13
$\mathbf{v}_{14}$	0.25	0.81	96.94
$\mathbf{v}_{15}$	0.21	0.67	97.61
$\mathbf{v}_{16}$	0.18	0.60	98.21
$\mathbf{v}_{17}$	0.17	0.55	98.76
$\mathbf{v}_{18}$	0.13	0.44	99.20
$\mathbf{v}_{19}$	0.12	0.40	99.60
$\mathbf{v}_{20}$	0.12	0.40	99.99

Table 1 shows the percentage variance and the cumulative percentage variance of all DTFs related to PC-1 to PC-20 ($\mathbf{v}_1$, $\mathbf{v}_2$, … , $\mathbf{v}_{20}$). The use of more PCs would reduce the modeling error between the magnitude HRTF of database and the model of magnitude HRTF, yet it would consume longer computation time and larger memory space. The PCs-matrix, **V**, that at first had 128x128 elements, had by now been reduced into a matrix of only 128x10 elements. We used only the first 10 PCs out of all 128 PCs. This ensures that we need only 10 PCWs to perform the model. It is obvious that the use of PCA is advantageous in terms of preserving memory space.

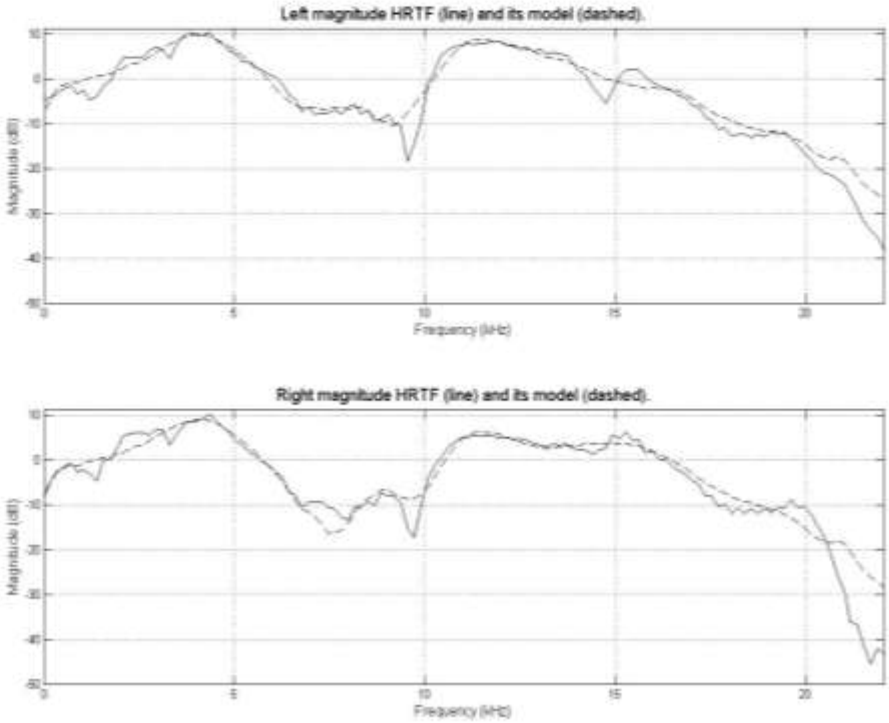

Figure 2 Comparison of original magnitude HRTFs and their corresponding PCA models.

Figure 2 shows a left magnitude HRTF of Subject 003 and its PCA model in the direction directly in front with azimuth 0^o and elevation 0^o (top panel). The bottom panel shows the right magnitude HRTF and its PCA model in the same direction. It is shown that the models approximate well the corresponding magnitude HRTFs.

2.3 Minimum Phase Reconstruction and Synthesis of HRIR Models

As elaborated in the previous subsection, we obtained PCs matrix, **V**, and PCWs matrix, **W**, using the PCA method. Both matrices together and an empirical mean vector, $\boldsymbol{\mu}$, were employed to yield the matrix of models of magnitude HRTFs, $\hat{\mathbf{H}}$, as suggested by Eq.(11). By now, we could calculate the models of magnitude HRTFs of both ears. In order to synthesize the models of complex HRTFs, the phase information of left- and right-ear model of magnitude HRTF should be inserted into those models. The reconstruction of the models of complex HRTFs was based on the approach taken by Kulkarni, et al. [14]. They assumed that the phase of a HRTF was minimum phase. The phase function for a given model of magnitude HRTF was computed using Hilbert transform of natural logarithm of the model of magnitude HRTF. The minimum phase, $\boldsymbol{\phi}_{mp}$, of a model of magnitude HRTF, $\hat{\mathbf{h}}_i$ (i=1,2,…,M), can be stated as,

$$\boldsymbol{\phi}_{mp} = \mathrm{Imag}\{\mathcal{H}\{-\ln(\hat{\mathbf{h}}_i)\}\}, \tag{12}$$

where Imag{} represents the imaginary part of a complex number and ln is the natural logarithm.

Thus, the model of minimum phase complex HRTF, $\hat{\mathbf{h}}_c$, can be calculated using,

$$\hat{\mathbf{h}}_c = \hat{\mathbf{h}}_i \,.\, \exp(j.\,\boldsymbol{\phi}_{mp}), \tag{13}$$

where exp() denotes the exponential function. The corresponding model of minimum phase HRIR, $\hat{\mathbf{h}}_{mp}(n)$, was given by the IFFT of its complex HRTF, $\hat{\mathbf{h}}_c$, from Eq. (13). Additionally, in reconstructing the model of left-ear minimum phase HRIR and the model of right-ear minimum phase HRIR for a particular direction of sound source into related model of left- and right-ear HRIR, respectively, we needed to insert respective time delay related to the propagation distance of a sound wave from the sound source to each ear drum of a subject, into each model of minimum phase HRIR. The time delays to be inserted were obtained from the means of time delays of respective directions on median plane from all subjects in the database utilized. The difference between left- and right-ear time delays is called interaural time difference (ITD), which is required by human to determine sound source direction.

The left panel of Figure 3 shows the original HRIRs of subject 003 due to direction with azimuth 0^o and elevation 0^o. The right panel shows related models of left and right HRIR. One can see the significant similarities between the original HRIRs and their models. These models resulted from the reconstructions of the PCA models of magnitude HRTFs into their corresponding HRIRs, as previously described. However, the models of magnitude HRTFs attained had not been individualized.

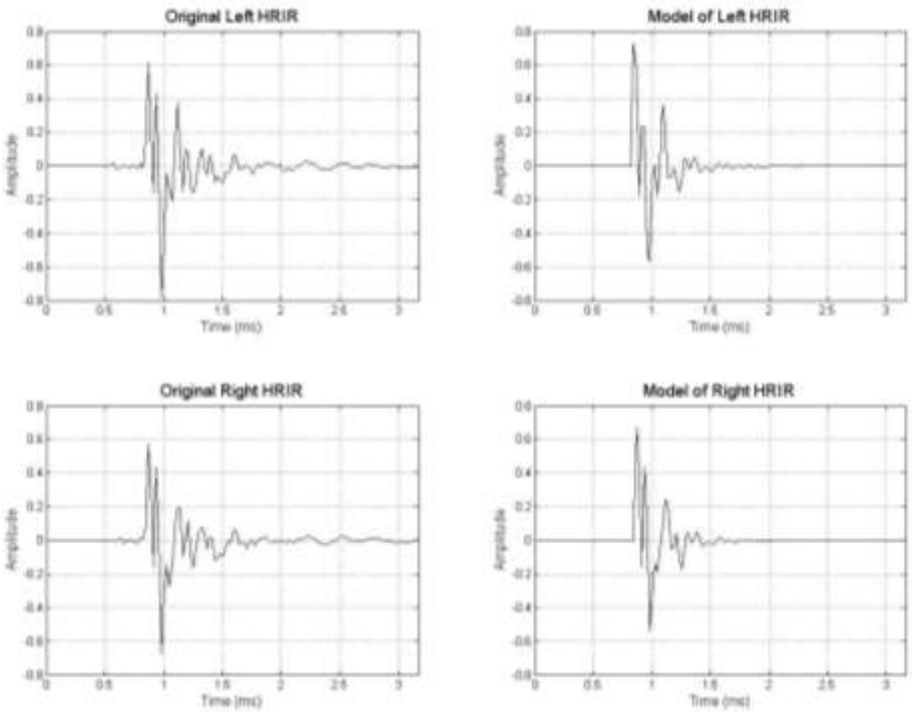

Figure 3 Comparison of original HRIRs and models of HRIRs obtained by reconstruction of models of magnitude HRTFs.

2.4 Individualization of Magnitude HRTFs Using MLR

As shown in Figure 1, the individualization of the models of magnitude HRTFs, which were resulted from PCA, were done through MLR of PCWs matrix, **W**, using anthropometric measurements of a listener. From the matrix **W** of Eq. (10), a weight vector, $\mathbf{w}_{i,ø}$ (37x1) could be extracted. This weight vector consisted of the i-th weights of the i-th PC, $\mathbf{v}_i$, of an ear of all subjects with elevation ø on median plane, where i=1,2,...,10. In this research, we utilized only 8 anthropometric measurements of a subject in the individualization process. The selection process of these 8 measurements will be elaborated in a separate subsection below. These selected measurements of all subjects being analyzed were then arranged in the columns of an anthropometric matrix, **X** (37x9), where the first column of **X** consists of all 1's.

The relation between the weights vector, $\mathbf{w}_{i,ø}$, and the anthropometric matrix, $\mathbf{X}$, is given as,

$$\mathbf{w}_{i,ø} = \mathbf{X} \, . \, \boldsymbol{\beta}_{i,ø} + \mathbf{E}_{i,ø} \, , \tag{14}$$

where $\boldsymbol{\beta}_{i,ø}(9x1)$ is the regression coefficients vector and $\mathbf{E}_{i,ø}(9x1)$ is the estimation errors vector. The regression coefficients were determined by implementing least-square estimation. This estimation was carried out by solving the optimization problem min{$\mathbf{E}_{i,ø}$}, where $\mathbf{E}_{i,ø}$ is the i-th dependent variable's estimation error. PCWs and anthropometric measurements are respectively the model's dependent and independent variables.

From Eq. (14), the regression coefficients due to i-th PCWs in elevation ø, $\mathbf{B}_{i,ø}$, can be estimated as,

$$\mathbf{B}_{i,ø} = (\mathbf{X}^T . \mathbf{X})^{-1} . \mathbf{X}^T . \, \mathbf{w}_{i,ø} . \tag{15}$$

As implied by Eq. (15), to improve the performance of the MLR method both dependent and independent variables must be carefully selected. The application of PCA on magnitude HRTFs, would significantly reduce the dimensions of independent variables as well as the complexity of the models. Numerous correlation analyses had been employed to select the independent variables in obtaining more accurate and simpler MLR method, as explained further in the subsection 2.5.

2.5 Correlation Analysis for Selection of Anthropometric Measurements

We utilized the CIPIC HRTF Database, which comprised both the measured HRIRs and some anthropometric measurements for 45 subjects, including the KEMAR mannequin with both small and large pinna. The detail definitions of the all 27 anthropometric measurements are given in [2, 12]. The modeling of a listener's own HRIRs via his or her own anthropometric measurements will directly affect the feasibility and complexity of the system. Implementing all measurements into the model would notably prove to be prohibitive. Some useful information will be obscured by the redundant measurements, which results in a worse regression model. Moreover, many measurements are very difficult to be measured accurately.

There are three parameters that are psychoacoustically important in the perception of natural sound, i.e. interaural time difference (ITD), interaural level difference (ILD) and pinna notch frequency, f_{pn}. ITD is defined as the time lapse between the arrival of first pulse of sound source from a particular direction on the left ear drum and that of the right ear drum. At the directions of sound source on median plane, ITDs are near zero, where for a perfect symmetric head, there is no ITD on that plane. Therefore ITD is a function of azimuth on planes with fixed elevation. ITD can be calculated from the time delay of maximum cross correlation of the left HRIR and right HRIR at a particular direction. On the other side, ILD is defined as level or magnitude difference (in dB) in frequency domain between the left magnitude HRTF and the right magnitude HRTF at a particular direction of sound source. For a particular direction, we obtained ILD from each frequency component in the range of 0 – 22050 Hz. ILDs are normally analyzed for a determined frequency component on the horizontal plane and on median plane. Another major psychoacoustic parameter is pinna notch frequency, f_{pn}. Pinna notch frequency is the notch frequency in the magnitude spectrum of HRTF caused by diffraction and reflection of sound wave on a pinna.

ITD and ILD are important for the perception of azimuth of sound source. Many variations of HRTF on the horizontal plane are affected by these parameters. On the contrary, ILD and f_{pn} are the key parameters in the perception of elevation of sound source and affect the variation of HRTF on median plane. It is generally problematical to characterize the range of HRTF disparity among subjects. However, maximum ITD, ITD_{max}, maximum ILD, ILD_{max}, and f_{pn} are simple and perceptually relevant parameters that characterize existing HRTF variation. Correlation analysis was applied to resolve which anthropometric measurements have strong correlations with ITD_{max}, ILD_{max}, and f_{pn}. From a few anthropometric measurements which are strongly correlated, four measurements were chosen from head and torso sizes, i.e. head width x_1 with $\rho = 0.736$, head depth x_3 with $\rho = 0.706$, neck width x_6 with $\rho = 0.726$, and shoulder width x_{12} with $\rho = 0,768$, where ρ indicates the correlation coefficient between the measurement and ITD_{max}. These four measurements were utilized in the individualization of magnitude HRTFs using MLR method. Correlation analysis between ILD_{max} and head and torso sizes provided weaker correlations but confirmed the chosen of x_1, x_6, and x_{12}.

The effects of pinna sizes are stronger with HRTFs on median plane than HRTFs on the horizontal plane [1]. However, the pinna sizes generally affect HRTFs in all directions. The correlation analysis between f_{pn} and anthropometric measurements provided weaker correlations than those of ITD_{max}. The four pinna sizes chosen were those with the strongest correlations with f_{pn}; i.e. cavum concha height, d_1, with $\rho = 0.435$, cavum concha width, d_3, with $\rho = 0.360$, pinna width, d_5, with $\rho = 0.204$, and pinna height, d_6, with $\rho = 0.280$. These selected sizes of pinna could be easily measured and represented the measures of height and width. Hence, eight anthropometric measurements, x_1, x_3, x_6, x_{12}, d_1, d_3, d_5, and d_6 were chosen and fed to the MLR method in order to generate regression coefficients. These eight anthropometric measurements are the same as the measurements that we used in our previous work [15, 16]. Subsequently, the regression coefficients were taken into account while estimating the PCWs of a DTF at each direction on median plane.

3 Experiments' Results and Discussion

This section elaborates the performance of the proposed individualization method from the objective simulation experiments between the original magnitude HRTFs of the database and the individualized models of magnitude HRTFs. The experiments were carried out by employing only the data on median plane of 37 subjects out of all 45 subjects in the database. This was due to the fact that the database had not included the complete set of anthropometric measurements of all subjects and the selected 8 anthropometric measurements were included only for 37 subjects.

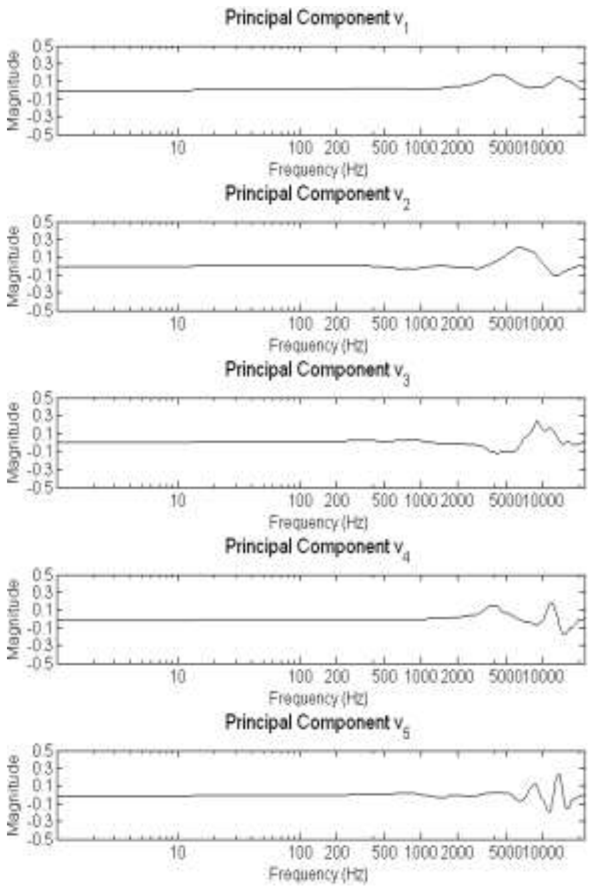

Figure 4 The first five basis functions or PCs extracted from PCA of 3700 DTFs from both ears of 37 subjects on median plane.

3.1 Basis Functions Resulted from PCA

The inputs of the PCA were 3700 DTFs processed from HRIRs on median plane of 37 subjects. By solving Eigen-equation, we attained 10 basis functions or PCs to model the given DTFs. Figure 4 shows the first five basis functions, $\mathbf{v}_1,...,\mathbf{v}_5$. Observing Figure 4, it is noted that all five basis functions are roughly constant and verge on zero at frequencies below 2 kHz. This implies that there is almost no direction-dependent variability in the DTFs in this frequency range. Regardless of the weights employed to the basis functions, the resulting weighted sum will be close to zero in this range.

Beyond approximately 2 kHz, all five basis functions have nonzero values. It can be observed that with the exception of the first basis function, the high-frequency dissimilarity in these basis functions signifies the direction-dependent high-frequency peaks and notches in the DTFs. The higher order basis function has more ripples and more details particularly for the frequencies above about 2 kHz. The trends explained above are similar for the sixth to tenth basis functions. In general, all basis functions seem to capture the high frequency spectral variability. They also reflect spectral differences between sources in front and sources behind the subject.

3.2 Weights of Basis Functions

Based on PCA, it was assumed that DTFs could be represented by a reasonably small number of basic spectral shapes or PCs. Therefore it was reasonable to expect that the amount each basic shape contributed to the DTF at a given source position would be simply related to source azimuth and elevation. In the case of source position on median plane, this amount or weight was related to elevation only.

The weights of PC-1 and those of PC-2 had much larger variability than the weights of other PCs on median plane. Figure 5 shows on the top panel and bottom panel, respectively, left ear PC-1 weights and left ear PC-2 weights for DTFs of subject 003, which were plotted as function of source elevation on median plane. It is shown that the PC-1 weights tend to increase in magnitude as the source moves from front below (elevation -45^o) to maximum at directly in front of the head (elevation 0^o). These weights tend to decrease for sources above (in front and rear). From Figure 5(b), the weights of PC-2 decrease from source -45^o to -11.25^o (front below) and then increase to a maximum at 56.25^o (front above). From this maximum, these weights tend to decrease for larger elevations.

The distributions of PC-1 weights and PC-2 weights were similar for all 37 subjects, which meant that they had low inter-subject variability. As can be observed in Figure 4, the first basis function has a nearly flat magnitude through all frequencies, implying that PC-1 weights are functioning as the amplification in HRTF modeling. The rest of the PC weights have smaller variability around the x-axis. Higher order PC has analogous flatter weights pattern for sources on median plane. It is observed that the patterns of PC weights are generally similar for all subjects and ears.

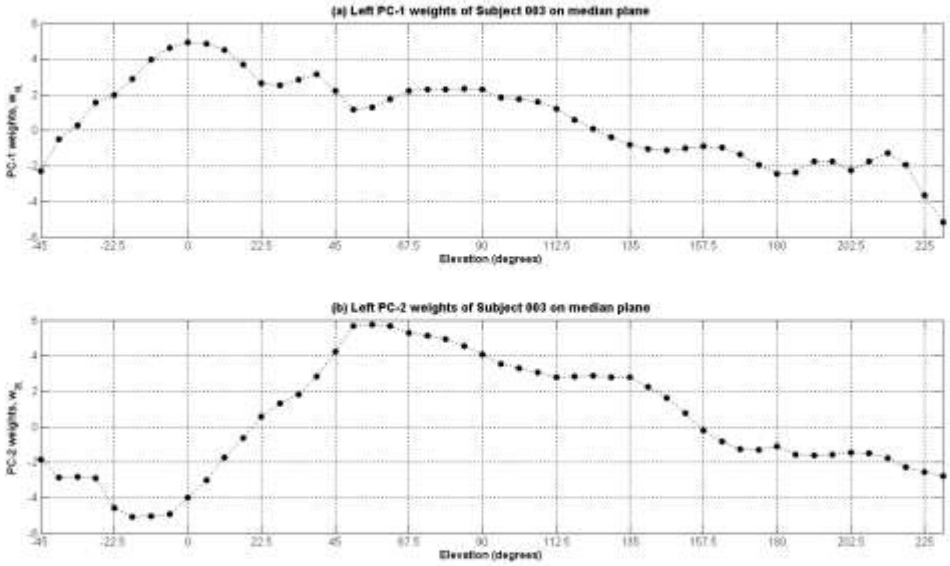

Figure 5 (a) Left ear PC-1 weights and (b) Left ear PC-2 weights for DTFs on median plane of subject 003.

3.3 Performance of the Proposed Individualization Method

The performances of the estimated magnitude HRTFs on median plane, resulting either from PCA or individualization, were evaluated by comparing the mean-square error of the disparity between the approximated magnitude HRTFs and the original magnitude HRTFs calculated from database, to the mean-square error of the original magnitude HRTFs in percentage, which is stated as

$$e_j(\o) = 100\ \%\ \text{x}\ \| \mathbf{h}_j(\o) - \hat{\mathbf{h}}_j(\o)\|^2 / \| \mathbf{h}_j(\o)\|^2 \qquad (16)$$

where $\mathbf{h}_j(\o)$ is the j-th original magnitude HRTF with elevation ø on median plane, $\hat{\mathbf{h}}_j(\o)$ is the corresponding approximated magnitude HRTF of $\mathbf{h}_j(\o)$. As the error increases, the performance of the estimated magnitude HRTF deteriorates. On the contrary better localization results will be achieved with small error, $e_j(\o)$.

Prior to individualizing magnitude HRTFs using MLR, mean error from PCA modeling of magnitude HRTFs was calculated for all data in the database. At first, PCA modeling was carried out for all data from all source directions of 45 subjects. This experiment yielded a mean error of 3.31% across all directions and subjects, with a mean error of 2.60% across directions on median plane. Subsequently, modeling was carried out using data at all directions of only 37 subjects, which yielded a mean error of 3.32% and a mean error of 2.56% on median plane. These two experiment results implied that the corresponding mean errors were mostly the same. In the next step, the data of both ears of 45 subjects at directions only on median plane were utilized, yielding a mean error of 2.53 %. Finally, PCA modeling was carried out using data of both ears of only 37 subjects at directions only on median plane. This experiment yielded a mean error of 2.52%. Again the same mean errors were attained from the last two experiments. In summary, the use of data of 45 subjects or 37 subjects would result in the same mean errors across related directions. Mean errors on median plane were the same either by using the data from all directions or only from directions on median plane. These mean errors are considerably smaller than the related mean errors obtained from our previous work on PCA modeling of minimum phase HRIRs [15].

In individualizing magnitude HRTFs, we used only the data of both ears of 37 subjects at directions on median plane, which meant that we used the results of fourth experiment mentioned above for individualization. In this research we attained a notably small mean error of PCA models of magnitude HRTFs on median, i.e. 2.52% compared to our previous work of 3.68% on the horizontal plane as in [16].

Consecutively, we individualized the PCA model of magnitude HRTFs on median plane using MLR with eight chosen measurements. The mean error of a subject was different from that of another subject in the database. It is observed that a good performance of the individualized left-ear magnitude HRTFs of a subject did not necessarily generate the same performance of the right-ear ones. The overall mean error was only 11.21%, which significantly outperformed the mean error of 22.50% found in [15] and also better than that of PCA model of magnitude HRTFs on horizontal plane, which was 12.17% as in [16].

The top panel of Figure 6 shows the left- and right-ear errors as a function of elevations on median plane of subject 003, while the bottom panel shows those of subject 163 after individualizing magnitude HRTFs. The mean error for left ear of subject 003 was 7.75% and that of right ear was 5.18%. These mean errors were better than those for both ears of subject 163. The mean errors for left ear and right ear of subject 163 were 11.38% and 12.09% respectively. The left-ear errors of subject 003 on median plane were approximately under 10%, except for some elevations at front below. However, the right-ear errors of subject 003 were overall smaller than the left-ear errors across elevations. The left-ear errors of subject 163 are observed to be almost the same as the right-ear errors, except for some elevations from -45^o to -16.875^o (front below of head). The general worse performance in these directions was due to the reflections of sound waves from the floor.

By comparing corresponding mean errors of subject 003 and subject 163 for directions on median plane stated above to those for directions on horizontal plane as in [16], it was confirmed that individualizing magnitude HRTFs on median plane in this research showed better performance than individualizing magnitude HRTFs on horizontal plane as in our previous work [16]. As elaborated in [16], that the performance of individualizing magnitude HRTFs on horizontal plane was much better than individualizing log-magnitude HRTFs on horizontal plane as in [11]. Hence, if Hu, et al. [11] had used the log-magnitude HRTFs on median plane, we believed that our work could outperform theirs.

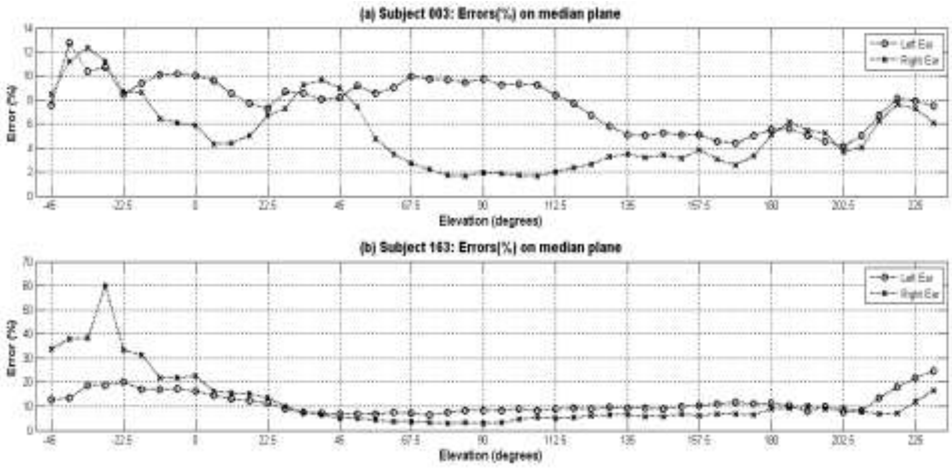

Figure 6 Left-ear and right-ear errors of (a) Subject 003 and (b) Subject 163 on median plane after individualization of magnitude HRTFs.

It was acknowledged that the proposed individualization method generated overall additional errors. These additional errors were due to the use of MLR. The disorderly behavior of weights of PCs across subjects and across directions had complicated the estimation of adequately accurate regression coefficients by the MLR. Moreover, we performed linear regression of anthropometric measurements to estimate the weights of PCs. Higher order regression might provide better estimates of these weights.

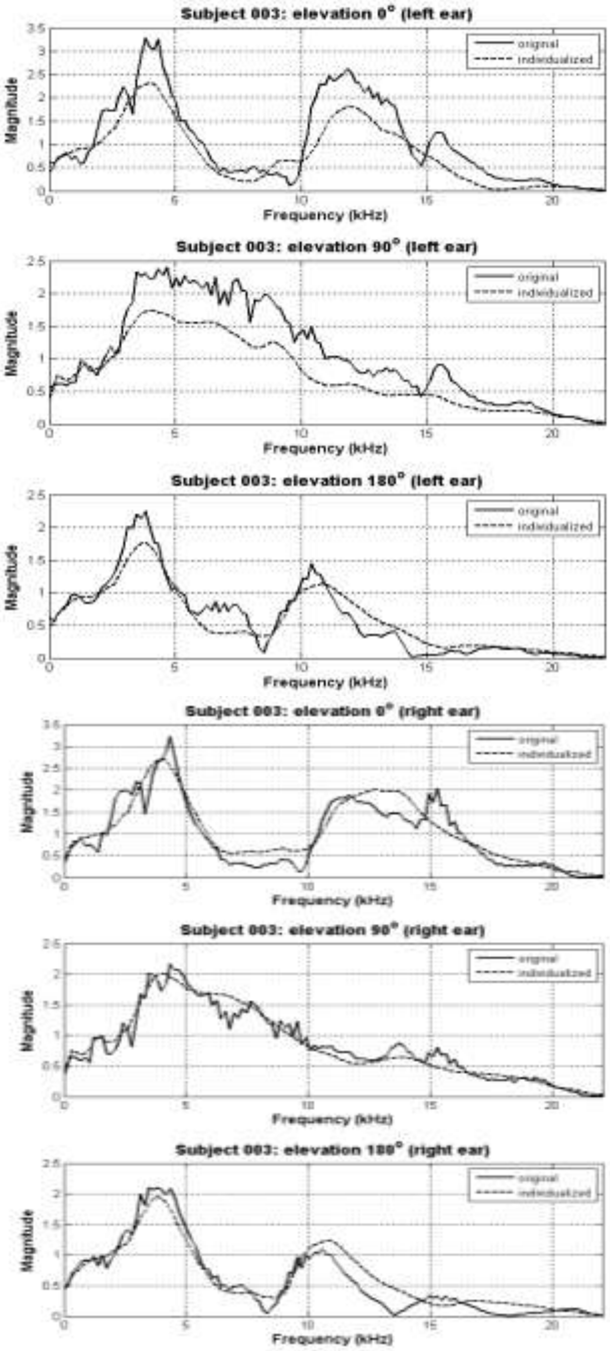

Figure 7 Magnitude responses of the original and individualized HRTFs of Subject 003 on median plane.

The original magnitude HRTFs of subject 003 could be well approximated by the corresponding individualized magnitude HRTFs. Figure 7 shows the individualized and original magnitude HRTFs for both the left and right ear in the extreme directions on median plane. The top, middle, and bottom panel corresponds to elevations 0°, 90°, and 180° respectively.

Informal listening tests underwent by five subjects had shown a good and natural perceived moving sound around the median plane by all subjects when the subjects' individualized reconstructed HRIRs, according to the sound source directions, were employed in the headphone simulation.

4 Conclusion

We proposed a novel individualization method of magnitude HRTFs for sources on median plane, based on principal components analysis and multiple linear regressions using chosen eight anthropometric measurements. The proposed method showed superior performance in the objective simulation experiments compared to that of similar researches and outperformed our previous works. The chosen anthropometric measurements had showed their effectiveness in individualizing magnitude HRTFs on median plane.

Acknowledgements

The authors wish to thank all staffs of CIPIC Interface Laboratory of California University at Davis, USA for providing the CIPIC HRTF Database. This work was supported in part by a grant from Electrical Engineering Department, Industrial Technology Faculty, Trisakti University, Jakarta, Indonesia.

References

[1] Blauert, J., *Spatial Hearing*, Revised Edition, Cambridge: MIT Press, 1997.

[2] Algazi, V.R., Duda, R.O., Thompson, D.P. & Avendano, C., *The CIPIC HRTF Database*, in Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'01), pp. 99-102, Oct. 2001.

[3] Wenzel, E.M., Arruda, M., Kistler, D.J. & Wightman, F.L., *Localization Using Nonindividualized Head-Related Transfer Functions*, J. Acoust. Soc. Am., **94**(1), 111-123, 1993.

[4] Shimada, S., Hayashi, M. & Hayashi, S., *A Clustering Method for Sound Localization Transfer Functions*, J. Audio Eng. Soc., **42**(7/8), 577-584, 1994.

[5] Middlebrooks, J.C., Virtual *Localization Improved by Scaling Non-individualized External-Ear Transfer Functions in Frequency*, J. Acoust. Soc. Am., **106**(3), 1493-1510, 1999.

[6] Algazi, V.R., Duda, R.O., Morrison, R.P. & Thompson, D.M., *Structural Composition and Decomposition of HRTF,* in Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'01), pp. 103-106, Oct. 2001.

[7] Zotkin, D.N., Hwang, J., Duraiswami, R., & Davis, L.S., *HRTF Personalization Using Anthropomentric Measurements*, in Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'03), pp. 157-160, 2003.

[8] Katz, B. F. G., *Measurement and Calculation of Individual Head-Related Transfer Functions Using a Boundary Element Model Including the Measurement and Effect of Skin and Hair Impedance*, PhD dissertation, Graduate Program in Acoustics, Pennsylvania State University, Pennsylvania, 1998.

[9] Shin, K. H. & Park, Y., *Enhanced Vertical Perception through Head-Related Impulse Response Customization Based on Pinna Response Tuning in the Median Plane*, IEICE Trans. Fundamentals, **E91-A**(1), 345-356, Jan. 2008.

[10] Hwang, S. & Park, Y., *HRIR Customization in the Median Plane via Principal Components Analysis*, in Proc. AES 31st Int. Conf. New Directions in High Resolution Audio, June 2007.

[11] Hu, H., Zhou, J., Ma, H. & Wu, Z., *Head Related Transfer Function Personalization Based on Multiple Regression Analysis*, in Proc. IEEE Int. Conf. on Computational Intelligence and Security, **2**, pp.1829 – 1832, 2006.

[12] CIPIC HRTF Database Files, Release 1.2, http://interface.cipic.ucdavis.edu/data/CIPIC_hrtf_database.zip, (23 September 2004).

[13] Hugeng, Wahab, W. & Gunawan, D., *Effective Preprocessing in Modeling Head-Related Impulse Responses Based on Principal Components Analysis*, Signal Processing: An International Journal (SPIJ), **4**(4), 201-212, 2010.

[14] Kulkarni, A., Isabelle, S.K. & Colburn, H.S., *On the Minimum-Phase Approximation of Head-Related Transfer Functions*, in Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'95), pp. 84-87, 1995.

[15] Hugeng, Wahab, W. & Gunawan, D., *Enhanced Individualization of Head-Related Impulse Response Model in Horizontal Plane Based on Multiple Regression Analysis*, in Proc. IEEE 2010 2nd Int. Conf. on Computer Engineering and Applications (ICCEA 2010), pp. 226-230, March 2010.

[16] Hugeng, Wahab, W. & Gunawan, D., *Improved Method for Individualization of Head-Related Transfer Functions on Horizontal Plane Using Reduced Number of Anthropometric Measurements*, Journal of Telecommunication, **2**(2), 31-41, 2010.