Effective Preprocessing in Modeling Head-Related Impulse Responses Based on Principal Components Analysis

Hugeng hugeng@ui.ac.id
Department of Electrical Engineering
University of Indonesia
Depok, 16424, Indonesia

Wahidin Wahab wahidin.wahab@ui.ac.id
Department of Electrical Engineering
University of Indonesia
Depok, 16424, Indonesia

Dadang Gunawan guna@eng.ui.ac.id
Department of Electrical Engineering
University of Indonesia
Depok, 16424, Indonesia

Abstract

It was found in previous works in modeling head-related impulse responses (HRIRs) using principal components analysis (PCA), both in frequency and time domain, that different sets of measured HRIRs were used, which were obtained from measurements by various institutions involving different kinds of subjects of human being, anesthetized live cat and acoustic manikin. Groups of researchers also applied different number of basis functions resulted from PCA, i.e. 4 – 10 basis functions. Then, the performance of the models was tested using different parameters, i.e. spectral distortion score and mean square error (MSE). Since there were varied factors mentioned above, a fair comparison among these models is difficult to achieve. Using PCA, we modeled the original HRIRs, minimum-phase HRIRs, direct-pulse HRIRs, normalized HRIRs in the time domain. However, in frequency domain, the models of magnitude head-related transfer functions (HRTFs), log-magnitude HRTFs, standardized log-magnitude HRTFs were performed. We performed a comprehensive comparison of various types of preprocessing of the previous data types in modeling HRIRs based on PCA using ten basis functions, CIPIC HRTF Database, and MSE. Our results showed that models of magnitude HRTFs had overall smallest average MSE. On the other hand, the best models in time domain were achieved from minimum-phase HRIRs.

Keywords: HRIR Model, HRTF Model, Principal Components Analysis.

1. INTRODUCTION

Human being can employ two ears to distinguish directions of various sound sources even by visually impaired people. There are three primary cues in perceiving the direction of a sound source, i.e. interaural time difference (ITD), interaural level difference (ILD), and spectral modification caused by pinnae, head, and torso. These primary sound cues are encrypted in binaural Head-Related Transfer Functions (HRTFs). HRTF is defined as the transfer function of the acoustic filter of human auditory system, in frequency domain, from a sound source to the entrance of ear canal. The counterpart of HRTF in time domain is known as head-related impulse response (HRIR). By convolving binaural HRIRs with a sound signal, a listener can localize the direction of the sound. Spatial sound synthesis has been widely utilized in various fields because of the great supports of digital signal processing in headphone system and multichannel speaker system. To implement binaural HRTFs in the creation of virtual auditory space, control of ITD, ILD, and spectral modification is the most essential part to give information of sound direction. On the horizontal plane, ITD and ILD play the main role in perceiving the direction of the sound source [1]. Although these two cues are remaining almost the same on the median plane, perception of sound direction is much affected by spectral modification mainly due to reflection and diffraction of pinnae folds [2].

HRIRs are resulted by reflections and diffractions of sound wave by human body, therefore, HRIRs vary significantly from subject to subject [3-4]. Poor vertical effect and front-back reversal were observed when non-individualized HRIRs were used to generate spatial sound image [3]. Application of individual HRIRs will be the most appropriate approach to create precise localization cues. However, measuring individual HRTFs of a listener is a time consuming process and economically expensive. Recently researchers have studied many other approaches to overcome the individual variability of HRIRs without directly measuring individual HRIRs. One effective and simple approach is modeling the HRIRs which includes several parameters that can be adapted by anthropometric measurements of a listener to produce his / her individualized HRIRs.

Several authors had proposed functional models of HRTFs [5-7]. They sought a mathematical equation that represented the HRTF as a function of frequency and direction, such that the models could provide explicit mathematical relationship between the HRTF and source location. Functional approaches would reduce the storage requirement and represent the HRTF at an arbitrary direction. However, attempts to describe the HRTF in a simple mathematical equation have been of only limited success. Following the approach as in [6], Algazi et al. proposed a structural model of HRTF, which attempted to relate functional HRTF to anthropometric measurements [8].

In this research, we concerned on the low-dimensional and orthogonal models for a set of HRIRs which have been generated by using the Karhunen-Loeve expansion or Principal Components Analysis (PCA). Kistler and Wightman [4] proposed a model based on PCA and minimum-phase reconstruction. The set of HRIRs was preprocessed to result in a set of logarithms of HRTF magnitudes. Then, PCA was employed to the logarithms of the HRTF magnitudes after the removal of direction-independent and subject-dependent spectral features. Middlebrooks and Green [9] and Hu et al. [10-11], also applied the same preprocessing to the HRIRs data as in [4]. Another research was the spatial feature extraction and regularization model for the HRTFs proposed by Chen et al.[12]. They preprocessed the set of HRIRs to become a set of complex-valued HRTFs. Using PCA, the models of HRTFs were expressed as weighted combinations of a set of complex-valued eigen transfer functions. The sample weights were determined by projecting all measured complex-valued HRTFs onto the eigen transfer functions. A functional representation for weights was attained by applying a thin plate generalized spline smoothing model to regularize the sample weights. This approach maintained the phase of the spectral components and model accuracy at a whole upper 3/4 sphere but it dealt with large amounts of complex-valued computation in matrix-vector products and two-dimensional splines.

Another group of researchers, Inoue et al. [13], attempted to estimate HRTFs using anthropometric measurements. They modeled the magnitude HRTFs using PCA and the weights were estimated by a regression model using anthropometric measurements. The estimated HRTF of the main response was reconstructed as a minimum-phase response. Then, ITDs were modeled using multiple regression analysis from anthropometric measurements. The complete models of HRTFs for both ears were reconstructed by combining the magnitudes model and ITDs model. They evaluated the estimation performance using a spectral distortion (SD) score. Our previous work [14] has similar approach as [13] in preprocessing the set of HRIRs. However, we evaluated the model performance using mean square error (MSE). Xu et al. [15] also proposed the modeling of HRIRs in the frequency domain. They preprocessed the database of HRIRs into a set of standardized log-magnitude HRTFs. Because of the variations of HRTFs, it is necessary to standardize them before performing PCA. The log-magnitude HRTFs were standardized by subtracting the mean and dividing by the standard deviation. This preprocessing was different from that of [4] only at the dividing by standard deviation. They developed a method to individualize HRTFs using anthropometric measurements. They compared the existing [4],[13] and their improved methods by a paired-sample *t*-test. The performance of the HRTFs models was also measured using SD score. Using PCA, Sodnik and Tomazic [16] modeled the linear amplitude spectra of HRTF, called magnitude HRTF by other researchers, thus excluding the phase spectrum and ITD. They concentrated on the azimuth perception on the horizontal plane. The performance of the HRTF models was measured by subjective localization tests from 15 subjects.

Similar work of modeling HRTF using PCA in the time domain was performed by Wu et al. [17]. Each HRIR of the set of all HRIRs used in their model was normalized by its level gain. They defined the square root of the energy of HRIR as its level gain. The normalized HRIRs, having same energy and onset time, would have different spectral characteristics. The variances of HRIRs are decreased by these normalizing procedures. The object for measuring HRIRS was an anesthetized live cat. They resulted in that the modeled HRIRs were nearly identical to the measured HRIRs. Other works were proposed in [18-19]. Before PCA, they preprocessed the median HRIRs to remove the initial time delay and to extract the early response that lasted for certain time interval since the arrival of the direct pulse. Shin and Park [18] extracted only the response of pinna with length of about 0.23 ms (10 samples with sampling frequency, f_s, of 44,100 Hz). In spite of this, Hwang and Park [19] included also the response of head and torso. They employed 1.5 ms of HRIR (67 samples with same f_s). They individualized the resulted HRIR model subjectively and then tested the individualization method by subjective listening tests. We also proposed an individualization method for HRIR model using PCA based on multiple regression analysis [20]. We preprocessed the measured HRIRs from CIPIC HRTF Database into minimum-phase HRIRs using cepstrum analysis. We tested our individualization method both objectively and subjectively. The performance of our HRIRs models was tested objectively using MSE as used in [4],[17].

The above mentioned previous works in modeling HRIRs, both in frequency and time domain, used different sets of HRIRs, which were obtained from measurements by various institutions using subjects of human being, anesthetized live cat and acoustic manikin. In modeling HRIRs using PCA, they also used different number of basis functions, 4-10 basis functions. Also the performance of the models was tested using different parameters, i.e. SD score and MSE. Due to the varied subjects, sets of HRIRs, number of basis functions used in PCA and performance parameters used, a fair comparison among these models is difficult to obtain.

2. MODELING HRIRS BASED ON PCA

This paper presents a comprehensive comparison of various preprocessings of a set of HRIRs in modeling HRIRs based on PCA using ten basis functions and CIPIC HRTF Database [21]. The goal of our work is to obtain an effective preprocessing in modeling HRIRs based on PCA by analyzing this comparison.

Types of Preprocessing Used

In this research, we used the following data types in modeling HRIRs, in the time domain, i.e. original HRIRs (directly from database, without preprocessing), direct-pulse HRIRs or HRIRs with initial time delay removed [18-19], minimum-phase HRIRs [20], and normalized HRIRs [17]. In spite, in the frequency domain, the data types used in modeling HRIRs were magnitude HRTFs [13-14],[16], log-magnitude HRTFs [4],[9-11], and standardized log-magnitude HRTFs [15]. We defined the above terms of data types to avoid ambiguity of possibly different terms used among groups of researchers. Each preprocessing involved in achieving related data type is explained in more detail in the given references.

As proposed in [19], direct-pulse HRIRs were obtained before PCA by removing the initial time delay and extracting the early samples that lasted for 1.5 ms or 67 samples with f_s = 44,100 Hz since the arrival of direct pulse. The initial time delay of HRIR indicates the propagation time of sound from sound source to eardrum. If it is required later, it can be re-inserted afterwards. The response of 1.5 ms includes the effects of pinna, head and torso. As explained in [14], the minimum-phase HRIR can be obtained through the calculation of real cepstrum of its original HRIR, which has arbitrary phase. It can be said that the minimum-phase HRIR is the removed initial time delay version of HRIR, similar with the direct-pulse HRIR. Original HRIR and its correspond minimum-phase HRIR have the same magnitude spectrum in the frequency domain. We took only 67 first samples out of 200 samples in the minimum-phase HRIR. Thus, in both types of preprocessing, the size of dataset to be analyzed in PCA was reduced without loss of meaningful information by the preprocessing.

Due to the head shadow effect, each HRIR has different energy which equals to the sum-square of samples of that HRIR. The square root of the energy was defined in [17] as the level gain of the HRIR. The normalized HRIRs were obtained by normalizing all original HRIRs by their correspond level gains. The normalized HRIRs, having the same energy and onset time, would have different spectral characteristics. The normalizing procedures decrease the variances of HRIRs. We took only 120 samples of the original HRIRs and the normalized HRIRs for modeling in PCA because these responses already include responses of pinna, head, and torso, on other side, the level of rest samples is very small and not significant.

Complex HRTFs were attained by implementing 256-points fast Fourier transform (FFT) to all original HRIRs from the database used. The magnitude HRTFs are then the absolute values of these complex HRTFs. Only 128 first frequency components of a magnitude HRTF, |HRTF|, were taken into analysis because of the symmetry property of a magnitude spectrum. Then, we defined a log-magnitude HRTF as twenty times the base-10 logarithm of its magnitude HRTF ($20 \log_{10} |HRTF|$). Finally, a standardized log-magnitude HRTF was obtained by standardizing or normalizing a log-magnitude HRTF with its standard deviation.

Principal Components Analysis

We explain below only the modeling of magnitude HRTFs using PCA with the same number of basis functions, subjects, and sound directions used as in the modeling of other data. We took also only 128 first frequency components of log-magnitude HRTF and standardized log-magnitude HRTF for analysis in PCA because of the same reason as before. The variable N below is equal 120 in modeling original HRIRs and normalized HRIRs and is replaced by 67 when modeling minimum-phase HRIRs and direct-pulse HRIRs. In modeling other data than magnitude HRTFs, the data of magnitude HRTFs is simply replaced by the desired data.

The entire magnitude HRTFs were computed from left-ear and right-ear HRIRs of 45 subjects from all sound sources with 1250 directions in sphere. There are 25 azimuths and 50 elevations of sound sound directions for each ear of a subject, so that a total of 112,500 magnitude HRTFs were produced. A matrix composed of DTFs is needed by PCA. The original data matrix, $\mathbf{H}$ (NxM), is composed of magnitudes of HRTFs, in which, each column vector, $\mathbf{h}_i$ (i=1,2,…,M), represents a magnitude HRTF of an ear of a subject in a direction in sphere. The number of magnitude HRTFs of each subject is 2500 (2 ears x 1250 directions). Hence, the size of **H** is 128 x 112,500 (N=128, M=112,500). The empirical mean vector ($\boldsymbol{\mu}$: Nx1) of all magnitude HRTFs is given by,

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

The DTFs matrix, **D**, is the mean-subtracted matrix and is given by,

$$\mathbf{D} = \mathbf{H} - \boldsymbol{\mu}.\mathbf{y}, \tag{2}$$

where **y** is a 1xM row vector of all 1′s. The next step is to compute a covariance matrix, **S**, that is given by

$$\mathbf{S} = \mathbf{D}.\mathbf{D}^{*} / (M-1) \tag{3}$$

where * indicates the conjugate transpose operator. The basis functions or principal components (PCs), $\mathbf{v}_i$ (i=1,2,…,q), are the q eigenvectors of the covariance matrix, **S**, corresponding to q largest eigenvalues. If q = N, then the DTFs can be fully reconstructed by a linear combination of the N PCs. However, q is set smaller than N because the goal of PCA is to reduce the dimension of dataset. An estimate of the original dataset is obtained here by only 10 PCs, which account for more than 90% (exactly 94.30%) variance in the original data **D**. By using only 10 PCs to model magnitude HRTFs, we expected to obtain satisfactory good results. The PCs matrix, $\mathbf{V} = [\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_N]$, that consisted of complete set of PCs can be obtained by solving the following eigen equation,

$$\mathbf{S}\ \mathbf{V} = \boldsymbol{\Lambda}\ \mathbf{V} \tag{4}$$

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

Then, the weights of PCs (PCWs), **W**(10x112,500), that correspond to all DTFs, **D**, can be obtained as,

$$\mathbf{W} = \mathbf{V}^{*}.\mathbf{D}, \tag{5}$$

where PCs matrix now was 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 consisted of models of magnitude HRTFs, $\hat{\mathbf{H}}$, is given by,

$$\hat{\mathbf{H}} = \mathbf{V}.\mathbf{W} + \boldsymbol{\mu}.\mathbf{y}. \tag{6}$$

The performance of the models of magnitude HRTFs, resulting from PCA, was 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

$$e_j(\theta,\phi) = 100\ \%\ \text{x}\ \|\ \mathbf{h}_j(\theta,\phi) - \hat{\mathbf{h}}_j(\theta,\phi)\|^2 \ /\ \|\ \mathbf{h}_j(\theta,\phi)\|^2 \tag{7}$$

where $\mathbf{h}_j(\theta,\phi)$ is the j-th original magnitude HRTF in the direction with azimuth, θ, and elevation, ϕ, $\hat{\mathbf{h}}_j(\theta,\phi)$ is the corresponding approximated or model of magnitude HRTF, $\mathbf{h}_j(\theta,\phi)$. As the error increases, the performance of the model of magnitude HRTF deteriorates. On the contrary better localization results will be achieved with small error, $e_j(\theta,\phi)$, which is called MSE by some researchers [4],[17].

3. EXPERIMENTS' RESULTS AND ANALYSES

Analysis of Percentage Cumulative Variance and Average MSE

Tabel 1 shows the percentage cumulative variance of DTFs or direct impulse responses (DIRs) of each data type in the database explained by PC-1 ($\mathbf{v}_1$) until PC-2, PC-5, PC-10, and PC-13 ($\mathbf{v}_2$, $\mathbf{v}_5$, $\mathbf{v}_{10}$, $\mathbf{v}_{13}$) respectively. Percentage cumulative variance is obtained from the percentage of cumulative sum of first largest eigen values that correspond to first PC until a particular PC, compared to the total cumulative sum of all PCs. As we can see, some cumulative variances of PCs from several data types do not exceed 90% for first 10 PCs, especially worse from original HRIRs and normalized HRIRs. This fact influences the performance of models of those data types. The average MSE across sound directions and subjects of the models of original HRIRs was only 48.5% using 10 PCs, similarly that of normalized HRIRs was only 32.0%. These bad results due to the time delay and many details included in those data types. It is quite difficult for PCA to estimate the time delay and details in those HRIRs.

By using PCA with 10 PCs, the best result for modeling HRIRs in time domain was achieved from minimum-phase HRIRs. The cumulative variance and average MSE were 90.40% and 7.26% respectively. In the frequency domain, we attained best result from magnitude HRTFs with 94.30% of cumulative variance and only 3.30% of average MSE. Much smaller average MSEs are observed from models in frequency domain since they have smoother spectra than impulse responses in time domain.

By using the same setup for PCA, the same database, and the same performance parameter as in [18], our models of minimum-phase HRIRs in median plane outperformed the models of direct-pulse HRIRs in median plane as proposed by Hwang and Park [18]. Using 10 PCs, their models had average MSE of about 6.67% [18] compared to the average MSE of our models of 5.31%. We had also performed the individualization of magnitude HRTFs in horizontal plane using multiple linear regressing (MLR) [14]. Comparable work was done by Hu etal. [10]. However, they performed the individualization of log-magnitude HRTFs in the horizontal plane. As we explained in [14], by comparing the performance shown in [10], our individualization of magnitude HRTFs was much better than that of their individualization of log-magnitude HRTFs. We believed that the selection of preprocessing used in our work supported the better results.

The application of more PCs would reduce the modeling error between each data type of HRTF in database and its model, however, it costed more computing time and larger memory space. The PCs-matrix, **V**, that at first has 128x128 elements was reduced into a matrix of only 128x10 elements in the cases of HRTFs modeling since we used only the first 10 PCs out of all 128 PCs. Thus, we needed only 10 PCWs to perform the model. One can see obviously the advantage of PCA in reducing significantly the memory space needed.

Data Type	Cumulative Variance (%)				Average MSE (%)			
	2PCs	5PCs	10PCs	13PCs	2PCs	5PCs	10PCs	13PCs
Original HRIRs	27.98	50.90	70.89	79.16	83.41	66.25	48.50	41.49
Direct pulse HRIRs	47.02	73.07	89.15	92.91	42.02	20.97	8.91	6.24
Minimum phase HRIRs	56.98	77.89	**90.40**	93.57	28.97	18.44	**7.26**	4.82
Normalized HRIRs	25.97	46.41	67.26	75.18	72.37	52.38	32.00	24.26
Magnitude HRTFs	72.61	86.94	**94.30**	96.13	11.35	6.95	**3.30**	2.35
Log-magnitude HRTFs	76.46	85.91	92.38	94.24	17.81	10.17	5.17	3.70
Standardized log-magnitude HRTFs	73.20	81.48	88.69	91.16	18.61	12.46	7.43	5.71

TABLE 1: Percentage Cumulative Variance Explained and Average MSE across Directions and Subjects.

Because of lack of space in this paper, the following results' analyses were performed only for models of minimum-phase HRIRs and of magnitude HRTFs using only 10 PCs and the corresponding MSEs in the sphere from Subject-003 of CIPIC HRTF Database.

Analysis of Basis Functions of PCA in Modeling Minimum Phase HRIRs

The modeling method, PCA, of minimum-phase HRIRs was performed as explained in the following. At the beginning, all original HRIRs in the database (with total number of 112,500 HRIRs) were converted into a set of minimum-phase HRIRs using the method explained in [19]. Mean of these HRIRs was then calculated. This mean was subtracted from each minimum-phase HRIR to give corresponding minimum-phase direct impulse response (DIR). All minimum-phase DIRs were applied to compute samples covariance matrix, **S**, as shown in Eq. 3. By solving the eigen equation (Eq. 4) that involving **S**, we obtained matrix of basis functions, **V**, that consisted of 67 complete basis functions, i.e. $\mathbf{V} = [\mathbf{v}_1 \;\; \mathbf{v}_2 \;\; \ldots \;\; \mathbf{v}_{67}]$. A perfect modeling of minimum-phase DIRs was achieved by employing complete basis functions. But we used only 10 basis functions (PCs) which explained 90.4% variance of all minimum-phase DIRs.

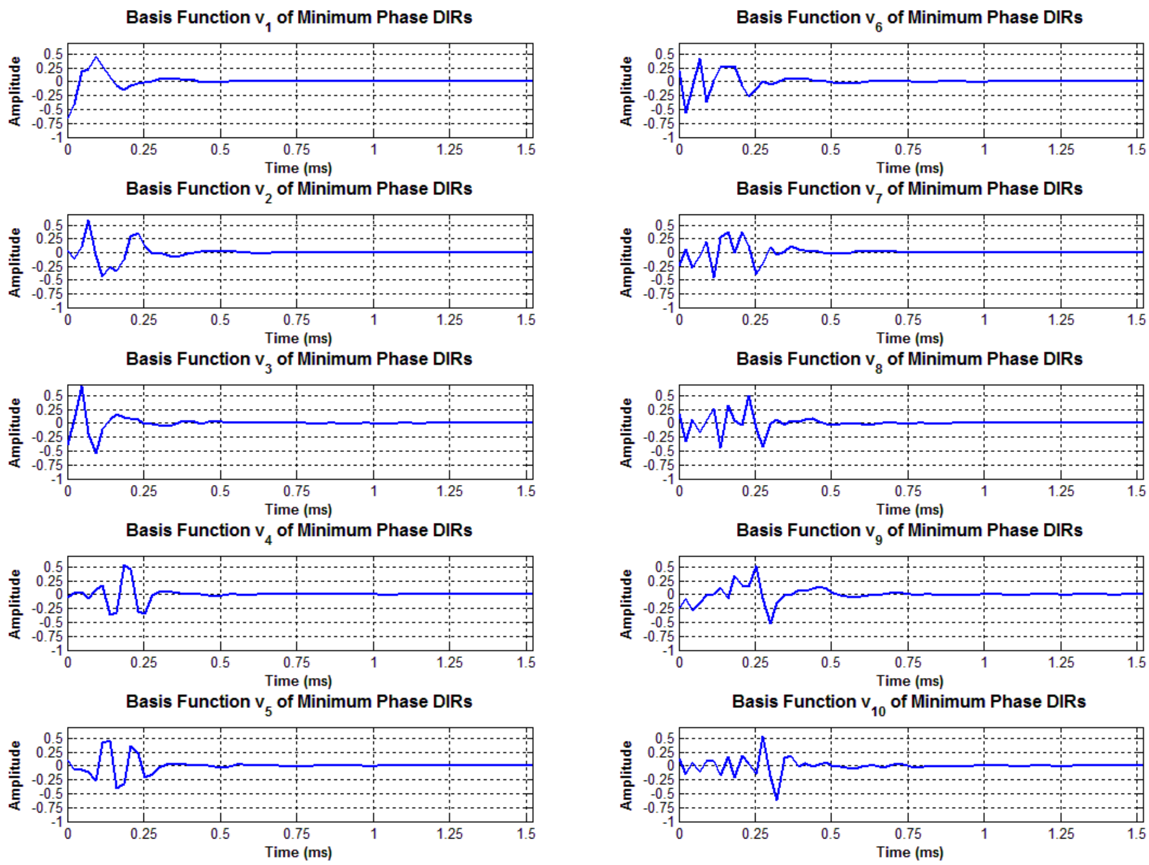

FIGURE 1: The First Ten Basis Functions Extracted from PCA in Modeling Minimum Phase HRIRs

Fig. 1 shows basis functions $\mathbf{v}_1$ to $\mathbf{v}_{10}$ that were applied in modeling minimum-phase DIRs. Basis functions $\mathbf{v}_1$ to $\mathbf{v}_{10}$ correspond to 10 consecutive largest eigen values, from the largest to the smallest one ($\lambda_1 > \lambda_2 > \ldots > \lambda_{67}$ where λ_i is the i-th eigen value resulted from PCA). As shown in the figure, $\mathbf{v}_1$ has the simplest form. When the eigen value become smaller, the correspond basis function has a form with more details. Basis functions $\mathbf{v}_1$ to $\mathbf{v}_5$ have non zero amplitudes from 0.3 ms, but $\mathbf{v}_6$ to $\mathbf{v}_{10}$ have non zero amplitudes from 0.6 ms. The intervals of amplitude levels were in general similar for these basis functions.

Analysis of MSEs of Models of Minimum Phase HRIRs in Sphere from Subject-003

From the observation of experiments' results, there are similar errors that occur in the models of both ears in the particular regions of interest. Hence, we discuss here only the errors of left ear models of minimum-phase HRIRs. Fig. 2 shows MSEs of the models of minimum-phase HRIRs from left ear of Subject-003 in sphere of all sound directions. As seen, small MSEs occur in center regions around the head, i.e. region near median plane with azimuth 0^0 and plane above head with elevation 90^0 of sound directions. Left ear average MSE across directions on the median plane is 4.23% and that of the plane above head is 2.03%. Across directions on horizontal plane (elevation 0^0), the left ear average MSE is 5.15%. Directions in the ipsilateral side provided smallest MSEs. Ipsilateral side of left ear is region with azimuth $-90^0 < \theta < 0^0$. Left ear average MSE across directions with azimuth -80^0 is 2.65%.

Less fine left ear models of minimum-phase HRIRs occur at the regions below subject, i.e. planes with elevation -45^0 (front below) and 230.6^0 (rear below), that have average MSEs 13.18% and 12.62% respectively. Worst left ear modeling occurs for contralateral plane with azimuth 65^0, that has average MSE of 15.33^0. However, average MSE on the contralateral plane with azimuth 80^0 is only 6.86%. Across all directions in sphere, left ear average MSE was 5.07% and that of right ear was 5.39%.

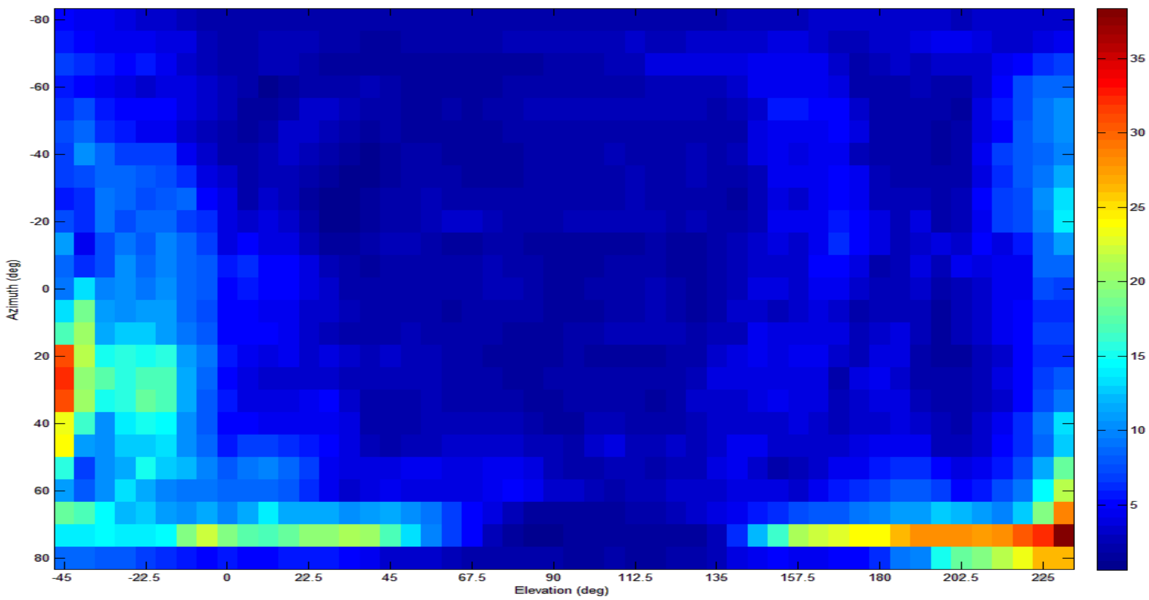

FIGURE 2: Percentage Mean Square Errors among Original and Models of Minimum Phase HRIRs

Analysis of Basis Functions of PCA in Modeling Minimum Phase HRIRs

It is shown from Fig. 3 that all of the first 10 basis functions can be said to be constant near zero at frequencies below 2 – 3 kHz because there are no dependencies between sound directions and variations of magnitude HRTFs in this frequency region. The sum of linear combination of these ten basis functions will be near zero or in other words that it is independent from the sound directions. Above ± 3 kHz, all basis functions posses non-zero magnitudes. Higher frequency variations in these basis functions (except in first basis function) represent higher frequency peaks and notches that depend on sound directions of all magnitude HRTFs. As happened in modeling minimum-phase HRIRs, first basis functions in this modeling has also the simplest form of magnitude spectrum. One can said this basis function as the amplification in the region from about 3 kHz to 22.05 kHz. More detailed magnitude spectrum is found at the basis function which corresponds to smaller eigen value. In general, the magnitudes of the spectrum can be seen to be in similar levels interval.

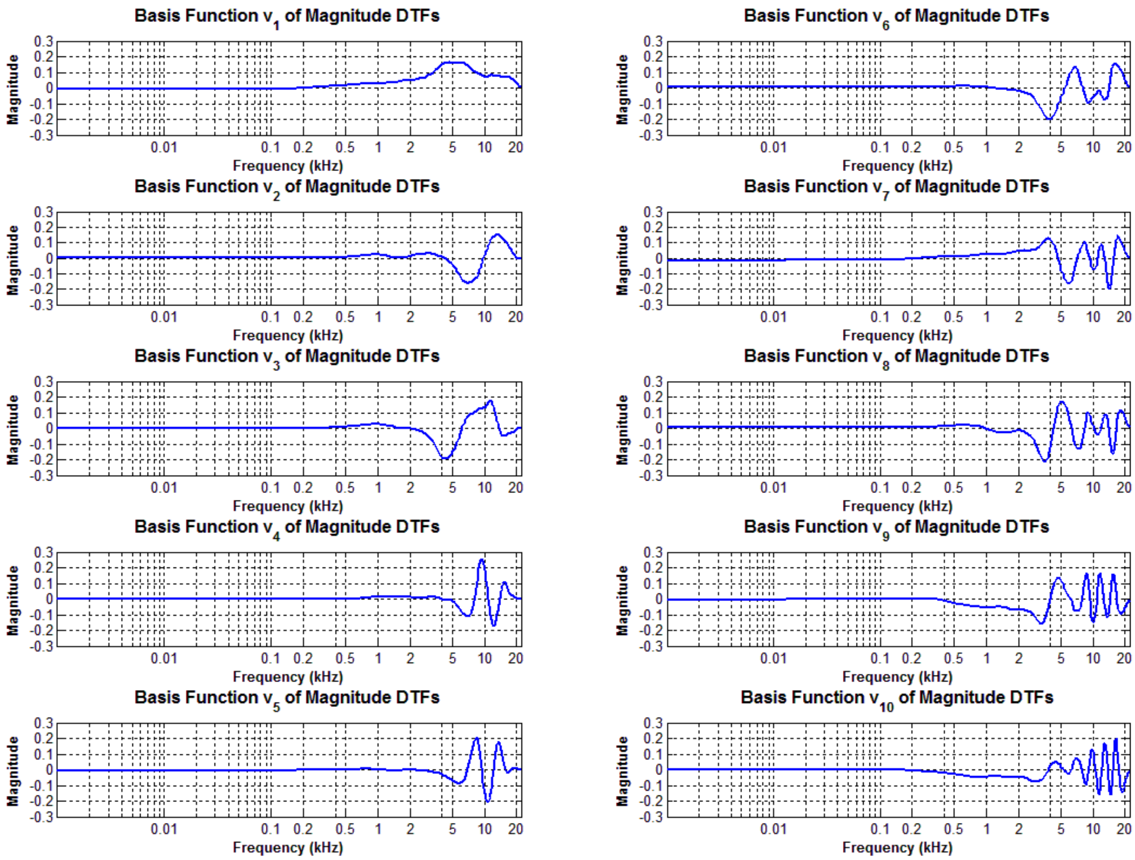

FIGURE 3: The First Ten Basis Functions Extracted from PCA in Modeling Magnitude HRTFs

Analysis of MSEs of Models of Magnitude HRTFs in Sphere from Subject-003

We can see in Fig. 4, a very good performance of models of magnitude HRTFs from Subject-003 across all directions in sphere with MSE about 5%. Fig. 4 shows MSEs of the models of magnitude HRTFs from left ear of Subject-003 of sound sources in sphere. Region with smallest MSEs is the region in the center of sphere, i.e. region around median plane. Average MSE across directions on median plane is 1.95%, while that on horizontal plane is 2.39%. The average MSE of directions on nearest ipsilateral plane with azimuth -80^0 is 1.47%, while that of directions on farthest contralateral plane with azimuth 80^0 is 2.62%. Smallest average MSE was achieved from the directions on the plane above head with elevation 90^0, i.e. 1.15%. On front below plane with elevation -45^0 and rear below plane with elevation 230.6^0, the average MSEs are worse, i.e. 6.30% and 5.05% respectively. The overall left ear average MSE in sphere is 2.44 and that of right ear is 2.53%.

For models of both minimum-phase HRIRs and magnitude HRTFs, fine performance of models in the region near the hearing ear causes from the fact that the HRIRs has generally larger energies compared to those of directions far from the hearing ear. One advantage of minimum-phase HRIRs is that they have removed time delays. On the other hand, magnitude HRTFs have quite smooth spectrum. PCA produced a set of PCs that could approximate well respective data types.

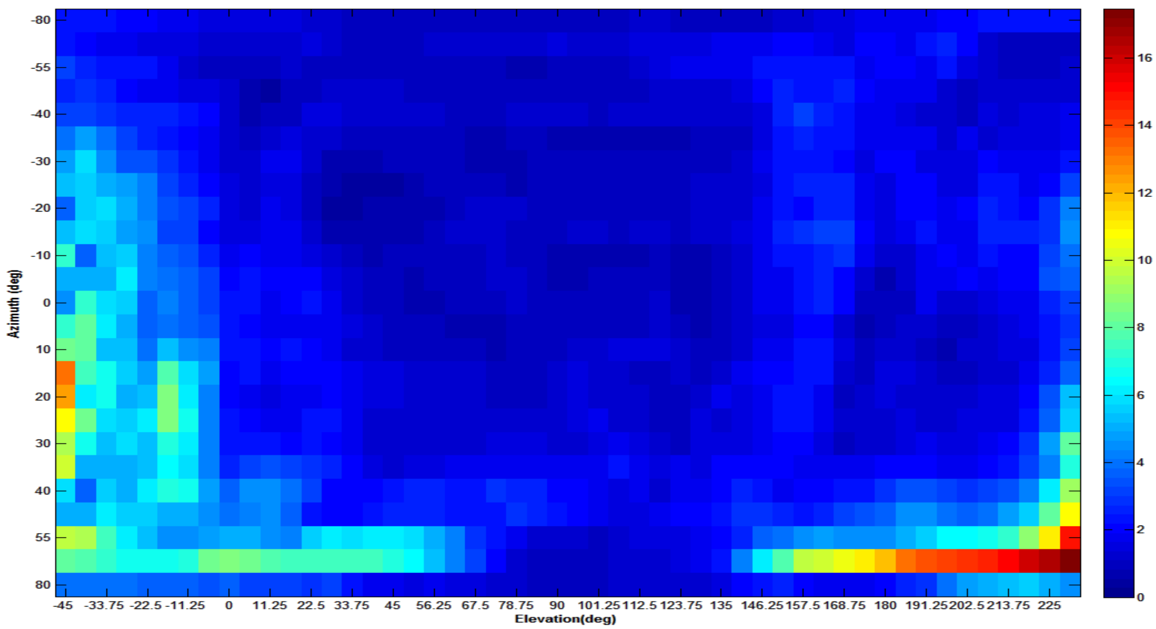

FIGURE 4: Percentage Mean Square Errors among Original and Models of Magnitude HRTFs

4. CONSLUSION

We proposed, using PCA, the modeling of magnitude HRTFs in frequency domain and the modeling of minimum-phase HRIRs in time domain. Using PCA with 10 basis functions from CIPIC HRTF Database, we compared the performances of models of 7 data types, i.e. original HRIRs, minimum-phase HRIRs, direct-pulse HRIRs, and normalized HRIRs, in time domain; also magnitude HRTFs, log-magnitude HRTFs, and standardized log-magnitude HRTFs, in frequency domain. Magnitude HRTFs showed the best performance with smallest average MSE across all HRIRs in database. On the other hand, the best models in time domain were achieved from minimum-phase HRTFs.

5. ACKNOWLEDGMENT

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 by a grant from Electrical Engineering Department, Industrial Technology Faculty, Trisakti University, Jakarta, Indonesia.

6. REFERENCES

1. D. R. Begault. ”*3-D sound for virtual reality and multimedia*”, Moffett Field, CA: Ames Research Center, 2000

2. J. Hebrank and D. Wright. “*Spectral cues used in the localization of sound sources on the median plane*”. J. Acoust. Soc. Amer., 56:1829-1834, 1974

3. E. M. Wenzel, M. Arruda, D. J. Kistler and F. L. Wightman. “*Localization using nonindividualized head-related transfer functions*”. J. Acoust. Soc. Amer., 94:111-123, 1993

4. 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. Amer., 91(3):1637-1647, 1992

5. D. W. Batteau. *"The role of the pinna in human localization"*. In Proceedings of Royal Society, London, 168(series B):158-180, 1967

6. K. Genuit. *„A description of the human outer ear transfer function by elements of communication theory“*. In Proceedings of 12th Int. Congress on Acoustics. Toronto, ADSTR, B6-8, 1986

7. J. Chen, B. D. Van Veen and K. E. Hecox. *"External ear transfer function modeling: A beamforming approach"*. J. Acoust. Soc. Amer., 92:1933-1945, 1992

8. V. R. Algazi, R. O. Duda, R. P. Morrison and D. M. Thompson. *"Structural composition and decomposition of HRTF"*. In Proceedings of IEEE WASPAA2001. New Paltz, NY, pp. 103-106, 2001

9. J. C. Middlebrooks and D. M. Green, *"Observations on a principal components analysis of head-related transfer functions"*. J. Acoust. Soc. Amer., 92(1):597-599, 1992

10. H. Hu, J. Zhou, H. Ma and Z. Wu. *"Head related transfer function personalization based on multiple regression analysis"*. In Proceedings of IEEE Int. Conf. on Computational Intelligence and Security, Guangzhou, China, vol. 2, pp.1829 – 1832, 2006

11. H. Hu, L. Chen and Z. Wu. *"The estimation of personalized HRTFs in individual VAS"*. In Proceedings of IEEE 4th Int. Conf. on Natural Computation, Jinan, China, pp. 203-207, 2008

12. J. Chen, B. D. Van Veen and K. E. Hecox. *"A spatial feature extraction and regularization model for the head-related transfer function"*. J. Acoust. Soc. Amer., 97(1):439-452, 1995

13. N. Inoue, T. Kimura, T. Nishino, K. Itou and K. Takeda. *"Evaluation of HRTFs estimated using physical features"*. Acoust. Sci. & Tech., 26(5):453-455, 2005

14. Hugeng, W. Wahab and D. Gunawan. *"Improved method for individualization of HRTFs on horizontal plane using reduced number of anthropometric measurements"*. J. of Telecommunications, 2(2):31-41, May 2010

15. S. Xu, Z. Li and G. Salvendy. *"Improved method to individualize head-related transfer function using anthropometric measurements"*. Acoust. Sci. & Tech., 29(6):388-390, 2008

16. J. Sodnik and S. Tomazic. *"Directional information in head related transfer functions"*. In Proceedings of TENCON 2004 – 2004 IEEE Region 10 Conf., 2:100-103. Chiang Mai, Thailand, 2004

17. Z. Wu, F. H. Y. Chan, F. K. Lam and J. C. K. Chan. *"A time domain binaural model based on spatial feature extraction for the head-related transfer function"*. J. Acoust. Soc. Amer., 102(4):2211-2218, 1997

18. K. H. Shin and Y. Park. *"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

19. S. Hwang and Y. Park. *"HRIR customization in the median plane via principal components analysis"*. In Proceedings of AES 31st Int. Conf. New Directions in High Resolution Audio. London, UK, June 2007

20. Hugeng, W. Wahab and D. Gunawan. "*Enhanced individualization of head-related impulse response model in horizontal plane based on multiple regression analysis*". In Proceedings of IEEE 2010 2nd Int. Conf. on Computer Engineering and Applications (ICCEA 2010), 2:226-230. Bali Island, Indonesia, March 2010

21. V. R. Algazi, R. O. Duda, D. P. Thompson and C. Avendano, *"The CIPIC HRTF Database"*. In Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'01). New Paltz, NY, USA, Oct. 2001