한국소음진동공학회논문집 제 18 권 제 6 호, pp. 642~653, 2008.

Comparison of Head-related Transfer Function Models Based on Principal Components Analysis

주성분 분석법을 이용한 머리전달함수 모형화 기법의 성능 비교

Sungmok Hwang†, Youngjin Park* and Youn-sik Park*

황 성 목 · 박 영 진 · 박 윤 식

(Received March 10, 2008 ; Accepted May 20, 2007)

Key Words : Head-related Transfer Function(머리전달함수), Principal Components Analysis(주성분분석법), Virtual Auditory Display(가상음향)

ABSTRACT

This study deals with modeling of head-related transfer functions(HRTFs) using principal components analysis(PCA) in the time and frequency domains. Four PCA models based on head-related impulse responses(HRIRs), complex-valued HRTFs, augmented HRTFs, and log-magnitudes of HRTFs are investigated. The objective of this study is to compare modeling performances of the PCA models in the least-squares sense and to show the theoretical relationship between the PCA models. In terms of the number of principal components needed for modeling, the PCA model based on HRIR or augmented HRTFs showed more efficient modeling performance than the PCA model based on complex-valued HRTFs. The PCA model based on HRIRs in the time domain and that based on augmented HRTFs in the frequency domain are shown to be theoretically equivalent. Modeling performance of the PCA model based on log-magnitudes of HRTFs cannot be compared with that of other PCA models because the PCA model deals with log-scaled magnitude components only, whereas the other PCA models consider both magnitude and phase components in linear scale.

요 약

이 연구는 중앙면 상에서 주성분 분석법을 이용하여 시간 및 주파수 영역에서 머리전달함수의 모형화 기법들을 다룬다. 시간영역의 머리전달함수, 복소수 값의 머리전달함수, 확장된 머리전달함수, 로그 크기의 머리전달함수에 기반하여 각각 주성분 분석법을 수행하여 얻은 네 가지 머리전달함수 모형들에 대해서 최소자승오차 관점에서 모형화 성능을 비교하고, 모형들간의 이론적인 관계를 살펴보는 것이 이 연구의 목적이다. 모형화에 사용되는 기저함수의 수가 동일하다면, 시간영역에서의 머리전달함수 혹은 확장된 머리전달함수에 기반한 모형이 복소수 값의 머리전달함수에 기반한 모형보다 최소자승오차 관점에서 더 효율적인 모형화 성능을 지닌다. 시간영역에서의 머리전달함수에 기반한 모형과 확장된 머리전달함수에 기반한 모형은 이론적으로 동일한 모형이며 서로 푸리에 변환 관계가 있다. 로그 크기의 머리전달함수에 기반한 모형은 다른 모형들과 모

† Corresponding Author: Member, KAIST
E-mail : tjdahr78@kaist.ac.kr
Tel : (042)869-3076, Fax : (042)869-8220
* Member, KAIST

형화 성능 및 이론적인 관계를 비교할 수가 없는데, 이는 로그 크기의 머리전달함수에 기반한 모형은 머리전달함수의 크기 정보만을 로그 크기로 다루는 반면에 다른 모형들은 선형 크기로 머리전달함수의 크기와 위상 정보를 모두 다루기 때문이다.

1. Introduction

Human can perceive a sound direction as one of the crucial auditory abilities with head-related transfer functions(HRTFs)[1] which is defined as the sound pressure at the listener's eardrum divided by the sound pressure measured at the position of the head center with the head absent. The physical structures of a listener, such as head, external ear(pinna), shoulder, and torso, transform the spectrum of sound waves when they reach to the listener's eardrum. This physical transformation of sound waves is characterized by the general concept of HRTF, and the HRTFs contain all information on the primary sound cues including interaural time difference, interaural level difference, and spectral modification.

In the headphone-based simulation, if sounds are filtered with HRTFs and delivered to a listener through a pair of headphones, then a virtual acoustic environment can be produced and the listener feels the spatialized sounds appear to originate from well-designed directions in the 3-dimensional space surrounding him/her. Systems or techniques generating spatialized sounds and conveying them to a listener is referred virtual auditory display(VAD)[2,3]. VAD has many promising applications such as entertainment including PC games, military training, virtual audio including home theatre, teleconference, auditory navigation, etc. Therefore, it is expected that more attention will be paid on VAD and the study on VAD will has a promising future.

Since VAD generates convincing 3-dimensional sounds by real-time convolution of audio source with the HRTFs corresponding to the desired source positions, VAD requires a large library containing the HRTFs corresponding to source positions densely distributed in the 3-dimensional space. In other words, many HRTFs must be empirically measured and stored in order to generate well-spatialized sounds by VAD, and this constraint makes real-time implementation difficult and requires large storage capacity. Thus, it is necessary to model the HRTFs using only a few parameters while keeping the perceptual relevant features of the HRTFs intact.

The modeling of spatial dependence of HRTFs or head-related impulse responses (HRIRs), which are the time-domain counterpart of the HRTFs, by a set of basis functions has been attempted by several authors. Especially, many authors utilized principal components analysis(PCA) or Karhunen-Loève expansion (KLE), which is mathematically equivalent to PCA, to obtain the basis functions[4~11]. PCA is one of the statistical procedures that try to provide an efficient representation of a set of correlated data. In the frequency domain, Martens[4] first applied the PCA to model HRTFs. He computed a set of basis functions from the critical-band-filtered HRTFs in the horizontal plane. Kistler and Wightman[5] performed a PCA of the HRTFs data, which were measured from 256 source positions distributed around the subject's head(10 subjects), for two ears on each subject, yielding 5300 HRTFs. Then, they showed that the log-magnitudes of HRTFs can be adequately approximated by a linear combination of 5 basis spectral shapes,

and they showed the systematic tendencies in the weights of the basis functions according to the source position. They also successfully demonstrated a technique for low-order HRTF representation with a series of psychophysical validations. Chen et al.[7] obtained the orthogonal basis functions from the KLE of the complex-valued(both magnitude and phase) HRTFs. In the time domain, Wu et al.[8] performed the KLE of HRIRs. Shin and Park[9] performed the PCA of HRIRs of 45 subjects. Hwang and Park[10,11] also performed the PCA of HRIRs and showed the systematic elevation-dependency in the weights of the basis functions.

Although many PCA models have already been developed in time and frequency domains, the systematic comparison of their modeling performance and analysis of their theoretical relationship are not reported yet. Thus, the goal of this study is to compare the modeling performance of several PCA models in the least-squares sense and to show the theoretical relationship between the PCA models. In the time and frequency domains, we deal with four PCA models based on HRIRs, complex-valued HRTFs, augmented HRTFs, and log-magnitudes of HRTFs.

2. Principal Components Analysis for HRTFs Modeling

The basic idea of PCA is to simplify the dataset by reducing multidimensional dataset to lower dimensions, while keeping the variation present in the original dataset as much as possible. In this study, the PCA is carried out using the median-plane HRIRs in the CIPIC HRTF database[12] containing 45 subjects' individual HRIRs. Since the CIPIC HRTF database is available at the elevation angles from -45° to 230° at 5.625° intervals, 2205 median-plane HRIRs(45 subjects × 49 elevations) are included in the PCA.

2.1 General Procedure of PCA

For convenience of further explanation, the general procedure of PCA is briefly introduced. The first step in PCA is to make a mean-subtracted data matrix from the original data matrix ($\mathbf{Y}: N\times M$). The empirical mean vector ($\mathbf{u}: N\times 1$) is given by

$$\mathbf{u}[n]=\frac{1}{M}\sum_{m=1}^{M}\mathbf{Y}[n,m] \quad (n=1,2,\ldots,N). \tag{1}$$

The mean-subtracted data matrix ($\mathbf{G}: N\times M$) is given by

$$\mathbf{G}=\mathbf{Y}-\mathbf{u}\cdot\mathbf{h}, \tag{2}$$

where $\mathbf{h}(1\times M)$ is a row vector of all 1's. The next step is to compute a covariance matrix ($\mathbf{C}$) of the mean-subtracted data matrix as

$$\mathbf{C}=\mathrm{E}[\mathbf{G}\otimes\mathbf{G}]=\frac{1}{M-1}\mathbf{G}\cdot\mathbf{G}^{*}, \tag{3}$$

where $\otimes$ and $*$ indicate the outer product and the hermitian operators, respectively. The basis vectors (basis functions), v_i (i=1,2,$\cdots$,q), are the q eigenvectors of the covariance matrix corresponding to the first q largest eigenvalues, and they are orthonormal to each other. These basis vectors are called principal components (PCs). If q=N, then the original dataset can be fully reproducible by a linear combination of the N PCs with the empirical mean. However, in many practical applications, q is smaller than N because the objective of PCA is to reduce the dimension of dataset. Thus, one can obtain an approximate of the original dataset only by using the $q(\ll N)$ PCs. The weights of PCs (PCWs), $\mathbf{W}$, can be computed by

$$\mathbf{W}=\mathbf{V}^{*}\cdot\mathbf{G}, \tag{4}$$

where $\mathbf{V}=[\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_q]$. The matrix composed of the approximates ($\hat{\mathbf{Y}}$), i.e. the modeled data matrix, is computed by

$$\hat{\mathbf{Y}} = \hat{\mathbf{G}} + \mathbf{u} \cdot \mathbf{h}, \tag{5}$$

where $\hat{\mathbf{G}} = \mathbf{V} \cdot \mathbf{W}$. Then, the number of PCs, which are needed for modeling, should be determined. Reasonable measures to determine the number of PCs are the percentage variance and the percentage modeling error in the least-squares sense. The percentage variance (*% var*) is expressed as

$$\%\ var(k) = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{N} \lambda_i} \times 100 \quad (\lambda_1 \geq \cdots \geq \lambda_k \geq \cdots \geq \lambda_N) \tag{6a}$$

$$= \left(1 - \frac{\left\|\mathbf{G} - \hat{\mathbf{G}}_k\right\|_F^2}{\left\|\mathbf{G}\right\|_F^2}\right) \times 100, \tag{6b}$$

where λ_i is the ith largest eigenvalue of the covariance matrix and $\hat{\mathbf{G}}_k$ indicates the modeled mean-subtracted data matrix from the first k PCs. The percentage modeling error (*% error*) is defined as

$$\%\ error(k) = \frac{\left\|\mathbf{Y} - \hat{\mathbf{Y}}_k\right\|_F^2}{\left\|\mathbf{Y}\right\|_F^2} \times 100, \tag{7}$$

where $\mathbf{Y}$ and $\hat{\mathbf{Y}}_k$ indicate the original data matrix and the modeled data matrix from the first k PCs, respectively. Subscript F indicates the matrix Frobenius norm. The variance (*% var*) excludes the empirical mean, whereas the modeling error (*% error*) includes it.

2.2 PCA of HRIRs

In 1997, Wu et al.[8] tried to model HRIRs based on several basis vectors obtained from the KLE. However, they obtained the basis vectors from only single subject's data. More recently, Shin and Park[9] obtained a set of basis vectors at each source position from the PCA of the pinna responses of multiple subjects. However, they included the pinna response only corresponding to the early response that lasts for 0.2 ms since the arrival of direct pulse in the HRIR, i.e., the basis vectors obtained by them are for the pinna response only. In this section, we expand on these previous works. We extract a set of basis functions not only for the pinna response but for the shoulder and torso responses from the median-plane HRIRs of 45 subjects.

Prior to the PCA, a pre-processing on the median-plane HRIRs is carried out to remove the initial time delay, defined as the time at which the pulse in HRIR first exceeds 12 % of its maximum amplitude, and to extract the early response that lasts for 1.5 ms since the arrival of direct pulse[10,11]. The initial time delay indicates the propagation time of sound wave from a source to a listener's eardrum, and this can be reinserted later if needed. This early response over the 1.5 ms following the arrival of the direct pulse is the data on which PCA is performed, and this response includes the effects of pinna, head, shoulder, and torso[13]. Thus, the size of dataset to be analyzed in PCA can be reduced without loss of meaningful information by the pre-processing. In this case, the dimension of the original data matrix, $\mathbf{Y}$, is 67×2205. 2205 corresponds to the total number of median-plane HRIRs and 67 corresponds to the number of samples of the pre-processed HRIR (sampling frequency: 44.1 kHz). The empirical mean ($\mathbf{u}$), the mean-subtracted HRIRs ($\mathbf{G}$), and the covariance matrix ($\mathbf{C}$) can be computed from Eqs. (1), (2), and (3). The PCs, $\mathbf{v}_{i,T}$ (i=1,2,$\cdots$,q), are orthonormal to each other

$$< \mathbf{v}_{k,T}, \mathbf{v}_{l,T} > = \mathbf{v}_{l,T}^{*} \cdot \mathbf{v}_{k,T} = \delta_{kl} = \begin{cases} 1\ (k = l) \\ 0\ (k \neq l) \end{cases}, \tag{8}$$

where $<,>$ indicates the inner product operator of two vectors. Subscript T means that the PCs are obtained in the time domain. The PCWs can be computed from Eq. (4) and the HRIRs can be modeled by Eq. (5). The number of PCs for modeling of HRIRs can be determined from *% var* and *% error* in Eqs. (6) and (7), respectively. Table 1 summarizes *% var* and *% error* for the left-ear median-plane HRIRs as a function of the number of PCs. As the number of PCs increases, *% var* increases and *% error* decreases exponentially. We arbitrarily set the modeling error bound of 5 % (comparable to 90 % of the variance in HRIRs), yielding 12 PCs.

Table 1 Percentage variance (*% var*) and modeling error (*% error*) for the left-ear HRIRs as a function of the number of PCs (k)

k	*% var*	*% error*	k	*% var*	*% error*
1	25.1	36.7	11	88.7	5.5
2	42.4	28.2	12	90.2	4.8
3	52.3	23.4	13	91.4	4.2
4	59.8	19.7	14	92.4	3.7
5	66.5	16.4	15	93.2	3.3
6	72.0	13.7	16	93.9	3.0
7	76.9	11.3	17	94.5	2.7
8	80.5	9.6	18	95.1	2.4
9	83.9	7.9	19	95.6	2.2
10	86.4	6.7	20	96.0	2.0

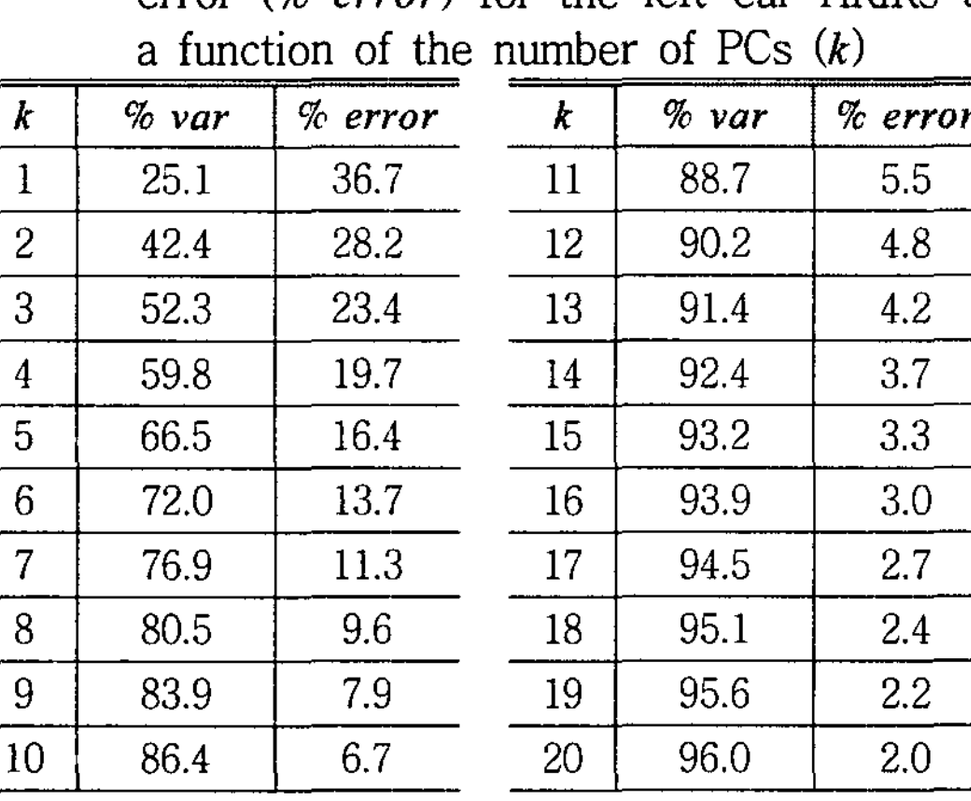

Fig. 1 Histogram of modeling error of each HRTF when 12 PCs obtained in the time domain are used

We can obtain several physical and meaningful interpretations on the PCs and PCWs, and these interpretations are described in detail in the authors' previous work[11].

Modeled HRTF for each subject and for each elevation can be obtained by the Fourier transform of the modeled HRIR, and the modeling error can be computed by Eq. (7). In this case, $\mathbf{Y}$ and $\hat{\mathbf{Y}}$ are replaced with $\mathbf{y}$ and $\hat{\mathbf{y}}$, which are computed by the 1024-point discrete Fourier transform (DFT) of measured and modeled HRIR (67×1) with zero-padding, respectively. Fig. 1 shows the histogram of modeling error of each HRTF when the 12 PCs obtained in the time domain are used. μ and σ at the upper-right corner of the figure indicate the mean value and the standard deviation of the 2205 modeling errors.

2.3 PCA of Complex-valued HRTFs

The HRTFs also can be modeled from PCA of HRTFs in the frequency domain. Note that HRTFs are complex valued, i.e., they are composed of real and imaginary parts (or magnitude and phase). Chen et al.[7] first tried to model the complex-valued HRTFs from a set of orthogonal basis vectors, which are obtained from the KLE of single subject's complex-valued HRTFs. In this section, we expand on their approach. 2205 complex-valued HRTFs of 45 subjects, which are obtained from the 1024-point DFT of the median-plane pre-processed HRIRs with zero-padding, are included in the PCA. The overall procedure is the same with the one in the time domain described in section 2.2. The empirical mean ($\mathbf{u}$), the mean-subtracted complex-valued HRTFs ($\mathbf{G}$), and the covariance matrix ($\mathbf{C}$) computed from Eqs. (1), (2), and (3) are also complex-valued vector or matrix. The complex-valued PCs, $\mathbf{v}_{i,C}(i=1,2,\cdots,q)$, are orthonormal to each other as

$$< \mathbf{v}_{k,C}, \mathbf{v}_{l,C} > = (\mathbf{v}_{l,R} + j\mathbf{v}_{l,I})^* \cdot (\mathbf{v}_{k,R} + j\mathbf{v}_{k,I}) \quad (9a)$$

$$= (\mathbf{v}_{l,R}^T - j\mathbf{v}_{l,I}^T) \cdot (\mathbf{v}_{k,R} + j \cdot \mathbf{v}_{k,I}) \quad (9b)$$

$$= (\mathbf{v}_{l,R}^T \cdot \mathbf{v}_{k,R} + \mathbf{v}_{l,I}^T \cdot \mathbf{v}_{k,I}) + j(\mathbf{v}_{l,R}^T \cdot \mathbf{v}_{k,I} - \mathbf{v}_{l,I}^T \cdot \mathbf{v}_{k,R}) = \delta_{kl}, \quad (9c)$$

where subscript C mean that the PCs are complex-valued, and subscripts R and I indicate the real and imaginary parts of the complex-valued PC, respectively. Superscript T indicates the transpose operator. The PCWs can be computed by Eq. (4) and the complex-valued HRTFs can be modeled by Eq. (5). Table 2 summarizes % var and % error for the left-ear complex-valued HRTFs in the median plane as a function of the number of PCs. When the first 6 complex-valued PCs obtained from PCA of the complex-valued HRTFs in the frequency domain are used for modeling, the complex-valued HRTFs can be modeled with *% var* of 89.4 % and *% error* of 5.2 %. Note that these modelling performances are comparable with those based on the real-valued 12 PCs obtained from PCA of HRIRs in section 2.2. Histogram of modeling error of each complex-valued HRTF based on the 6 complex-valued PCs is shown in Fig. 2.

2.4 PCA of Augmented HRTFs

The complex-valued HRTFs matrix ($\mathbf{Y}_C$: $N \times M$) can be augmented to the real-valued matrix ($\mathbf{Y}_A$: $2N \times M$), which are composed of two real-valued matrices corresponding to the real and imaginary parts of the complex-valued HRTFs. This can be expressed as

$$\mathbf{Y}_A = \begin{bmatrix} \mathbf{Y}_R \\ \mathbf{Y}_I \end{bmatrix}. \quad (10)$$

As like PCA of HRIRs or PCA of complex-valued HRTFs, the HRTFs also can be modeled from PCA of augmented HRTFs in the frequency domain. Note that the augmented HRTFs are real valued. The original data matrix, $\mathbf{Y}$, is composed of the augmented HRTFs in the median plane. The overall procedure of the PCA is the same with the one described in section 2.2. The empirical mean ($\mathbf{u}$), the mean-subtracted HRTFs ($\mathbf{G}$), and the covariance matrix ($\mathbf{C}$) computed from Eqs. (1), (2) and (3) are real-valued vector or matrix. Note that the first half components ($N \times 1$) of the real-valued PCs ($2N \times 1$) correspond to the real part of complex-valued basis vector, and the other components correspond to the imaginary part of complex-valued basis vector. The orthonormality condition is expressed by

Table 2 Percentage variance (*% var*) and modeling error (*% error*) for the left-ear complex-valued HRTFs as a function of the number of PCs (k)

k	*% var*	*% error*	k	*% var*	*% error*
1	41.9	28.5	6	89.4	5.2
2	59.0	20.1	7	91.8	4.0
3	70.2	14.6	8	93.3	3.3
4	79.4	10.1	9	94.6	2.7
5	85.5	7.1	10	95.5	2.2

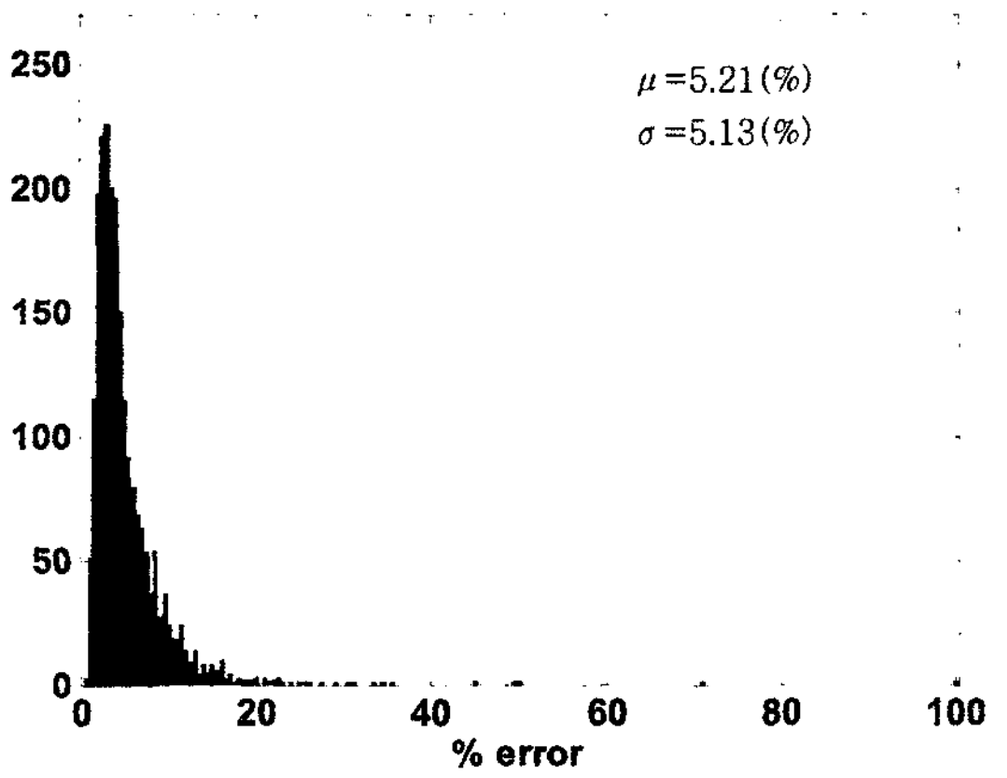

Fig. 2 Histogram of modeling error of each HRTF when 6 complex-valued PCs obtained in the frequency domain are used

$$< \mathbf{v}_{k,A}, \mathbf{v}_{l,A} > = \begin{bmatrix} \mathbf{v}_{l,R} \\ \mathbf{v}_{l,I} \end{bmatrix}^{*} \cdot \begin{bmatrix} \mathbf{v}_{k,R} \\ \mathbf{v}_{k,I} \end{bmatrix} \quad (11a)$$

$$= \begin{bmatrix} \mathbf{v}_{l,R}^{T} & \mathbf{v}_{l,I}^{T} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{v}_{k,R} \\ \mathbf{v}_{k,I} \end{bmatrix} \quad (11b)$$

$$= \mathbf{v}_{l,R}^{T} \cdot \mathbf{v}_{k,R} + \mathbf{v}_{l,I}^{T} \cdot \mathbf{v}_{k,I} = \delta_{kl} . \quad (11c)$$

However, the PCA model based on augmented HRTFs in the frequency domain is theoretically equivalent to the PCA model based on HRIRs in the time domain. In other words, the HRTF, which is the Fourier transform pair of HRIR, can be modeled by linear combination of PCs, which are the Fourier transform pair of PCs obtained from PCA of HRIRs in the time domain. Furthermore, the PCs in the frequency domain, which are obtained by the Fourier transform of the PCs in the time domain, are orthonormal to each other and their weights are the same with the PCWs in the time domain except for just scaling due to the number of DFT. A derivation of the theoretical equivalence of the two PCA models is provided in Appendix. *% var* and *% error* for the left-ear augmented HRIRs as a function of the number of PCs is the same with the ones in the time domain shown in Table 1. Histogram of modeling error of each augmented HRTF based on 12 PCs is identical to the one in the time domain shown in Fig. 1, as expected.

2.5 PCA of on Log-magnitudes of HRTFs

Sections 2.3 and 2.4 deal with HRTF modeling from PCA of complex-valued or augmented HRTFs in the frequency domain. However, many previous studies on human HRTFs have focused on the magnitude components[14~17]. Especially, Kistler and Wightman[5] applied PCA of directional transfer functions (DTFs : mean-subtracted HRTFs) of 10 subjects to model only the magnitude components of the HRTFs. However, synthesis of stimuli requires both magnitude and phase components. Thus, they constructed a simple model of the phase components of HRTFs, based on an assumption that HRTFs are minimum-phase functions[18]. Of course, it is unfair to compare the modeling performance of the PCA model based on log-magnitudes of HRTFs with the one of other PCA models because the PCA model based on log-magnitudes of HRTFs deals with log-scaled magnitude components only whereas the other PCA models deal with both magnitude and phase components in linear scale. However, it is meaningful to expand the previous study by considering more subjects' data.

The log-magnitude of HRTF ($\mathbf{H}_{log}$) are computed by

$$\mathbf{H}_{log} = 20\log10|\mathbf{H}_C|, \quad (12)$$

where $\mathbf{H}_C$ is the complex-valued HRTF. The original data matrix, $\mathbf{Y}$, is composed of the log-magnitudes of HRTFs in the median plane. The overall procedure of the PCA is the same with the one described in section 2.2. However, the dimension of $\mathbf{Y}$ is 458×2205 in this case. 458 corresponds to the number of frequency components of the log-magnitude of HRTF (from 300 Hz to 20 kHz with 43.07 Hz intervals) included in the PCA. The empirical mean ($\mathbf{u}$), the mean-subtracted HRTFs ($\mathbf{G}$), and the covariance matrix ($\mathbf{C}$) computed from Eqs. (1), (2), and (3) are real-valued vector or matrix. The PCs, $\mathbf{v}_{i,log}(i=1,2,\cdots,q)$, are orthonormal to each other. The PCWs can be computed by Eq. (4) and the log-magnitudes of HRTFs can be modeled by Eq. (5). Table 3 summarizes *% var* and *% error* for the left-ear log-magnitudes of HRTFs in the median plane as a function of the number of PCs. When the first 12 PCs are used for modeling, the log-magnitudes of HRTFs can be modeled with *% var* of 91.3% and *% error* of 4.0 %. Histogram of modeling error of each

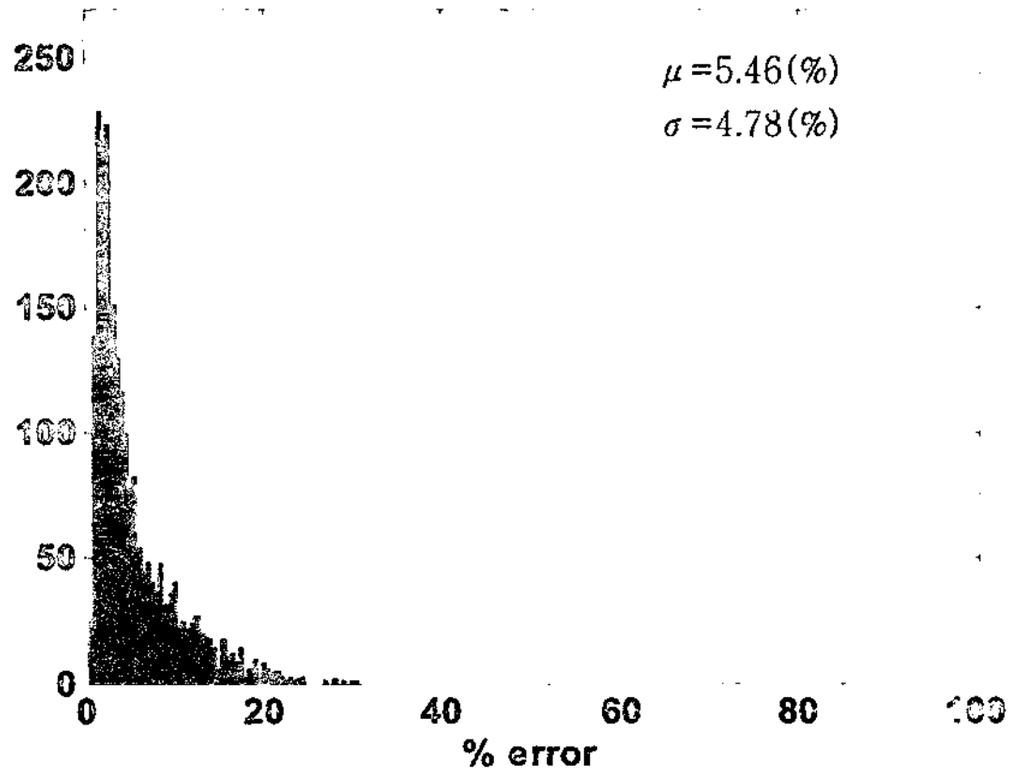

Fig. 3 Histogram of modeling error of each log-magnitude of HRTF when 12 PCs obtained in the frequency domain are used

Table 3 Percentage variance (*% var*) and modeling error (*% error*) for the left-ear log-magnitude of HRTF as a function of the number of PCs (*k*)

k	*% var*	*% error*	*k*	*% var*	*% error*
1	42.5	26.7	11	90.3	4.5
2	57.1	19.9	12	91.3	4.0
3	66.3	15.7	13	92.2	3.6
4	71.6	13.2	14	92.9	3.3
5	76.5	10.9	15	93.5	3.0
6	80.0	9.3	16	94.0	2.8
7	83.1	7.8	17	94.5	2.6
8	85.6	6.7	18	94.9	2.4
9	87.6	5.8	19	95.2	2.2
10	89.1	5.1	20	95.6	2.0

log-magnitude of HRTF based on the 12 PCs is shown in Fig. 3.

3. Discussions on PCA models

In section 2, four kinds of PCA models in the time and frequency domains are investigated. Each PCA model has advantages and disadvantages. The PCA models based on log-magnitudes of HRTFs in the frequency domain under the assumption of a minimum phase property of HRTFs can reasonably represent human's sense of hearing because it is roughly logarithmic, but the minimum phase assumption leads to HRIR approximation that is much shorter than the original HRIR and in which the pinna and torso contributions are merged. The PCA model based on log-magnitudes of HRTFs deals with log-scaled magnitude components only, whereas the other PCA models consider both magnitude and phase components in linear scale.

A noticeable difference between PCA models in the time and frequency domains(except for the PCA model based on augmented HRTFs) is that it is not clear which PCs in the frequency domain are due to the pinna or to the shoulder/torso only because the shoulder/torso response and pinna response are closely coupled, whereas PCs in the time domain can be classified into the effects of pinna or shoulder/torso because HRIR can be decomposed into a series of temporal sound events[11].

The PCA models based on HRIRs in the time domain and based on augmented HRTFs in the frequency domain are theoretically equivalent, i.e., the Fourier transform pair of the PCA model based on HRIRs is the PCA model based not on complex-valued HRTFs but on augmented HRTFs.

From the modeling results in section 2, in terms of the number of PCs one might conclude that the PCA model based on complex-valued HRTFs is more efficient to model HRTFs than the PCA model based on HRIRs or augmented HRTFs because the half of the number of PCs are required in the PCA model based on complex-valued HRTFs in comparison with the other PCA models. Although this sounds reasonable, a complex-valued vector can be dealt with a 2-dimensional vector composed of two 1-dimensional real-valued vectors corresponding to the real and imaginary parts of the complex-valued vector. Therefore, each complex-valued

PC can be dealt with two real-valued vectors, and the 6 complex-valued PCs obtained from PCA of complex-valued HRTFs are comparable to the 12 real-valued PCs obtained from PCA of HRIRs or augmented HRTFs. By comparing the results in Table 1 and Table 2, it can be found that the modeling performance of PCA model based on complex-valued HRTFs is lower than that of PCA model based on HRIRs or augmented HRTFs. Fig. 1 and Fig. 2 also show the same result. We can find the reason in the orthonormality condition of PCs. The orthonormality condition of PCA model based on augmented HRTFs in Eq. (11) is similar to the one of PCA model based on complex-valued HRTFs in Eq. (9). In this case, however, the only real part in Eq. (9) is needed, i.e., the orthonormality condition in Eq. (9) is stricter than that in Eq. (11). In other words, the PCA model based on complex-valued HRTFs requires one more constraint for PCs than the PCA model based on augmented HRTFs. This is the main difference between the two PCA models and the reason why the modeling performance of the PCA model based on complex-valued HRTFs is lower than that of the PCA model based on augmented HRTFs or HRIRs.

The storage capacity needed for modeling is directly proportional to the number of PCs. Thus, one can conclude that the four PCA models require almost equal storage capacity because the four PCA models are based on 6 complex-valued PCs (section 2.3) or 12 real-valued PCs (sections 2.2, 2.4, and 2.5). However, the PCA model based on augmented HRTFs requires larger storage capacity than other PCA models because the PCs are augmented by the real and imaginary parts of the complex-valued ones.

4. Summary and Conclusions

The present study dealt with modeling of HRTFs using PCA based on HRIRs, complex-valued HRTFs, augmented HRTFs, and log-magnitudes of HRTFs. Modeling procedure and performance in the least-squares sense for each PCA model were investigated. In terms of the number of PCs needed for modeling, the PCA model based on HRIR or augmented HRTFs showed more efficient modeling performance than the PCA model based on complex-valued HRTFs. This is due to the difference of orthonormality condition between the models. And the PCA models based on HRIRs in the time domain and based on augmented HRTFs in the frequency domain are theoretically equivalent. However, the PCA model based on augmented HRTFs may need larger storage capacity because the dimension of data is doubled. Therefore, we recommend the PCA model based on HRIRs in the time domain if the target is to model HRTFs in the least-squares sense. Modeling performance of the PCA model based on log-magnitudes of HRTFs cannot be compared with that of other PCA models directly because the model deals with log-scaled magnitude components only, whereas the other PCA models consider both magnitude and phase components in linear scale.

Acknowledgment

This study was supported by the BK21 project and the Korea Science and Engineering Foundation(KOSEF) through the National Research Laboratory Program by the Ministry of Science and Technology(R0A-2005-000-10112-0).

References

(1) Blauert, J., 1983, Spatial Hearing, MIT,

Cambridge, MA.

(2) Honda, A., Shibata, H., Gyoba, J., Saitou, K., Iwaya, Y. and Suzuki, Y., 2007, "Transfer Effects on Sound Localization Performances from Playing a Virtual Three-dimensional Auditory Game," Applied Acoustics, Vol. 68, pp. 885~896.

(3) Iwaya, Y., 2006, "Individualization of Head-related Transfer Functions with Tournament-style Listening Test: Listening with Other's Ears", Acoustical Science and Technology, Vol. 27, pp. 340~343.

(4) Martens, W. L., 1987, "Principal Components Analysis and Resynthesis of Spectral Cues to Perceived Direction", in Proceedings of International Computer Music Conference, San Francisco, CA, pp. 274~281.

(5) Kistler, D. J. and Wightman, F. L., 1992, "A Model of Head-related Transfer Functions Based on Principal Components Analysis and Minimum-phase Reconstruction", Journal of the Acoustical Society of America, Vol. 91, pp. 1637~1647.

(6) Middlebrooks, J. C. and Green, D. M., 1992, "Observations on a Principal Components Analysis of Head-related Transfer Functions", Journal of the Acoustical Society of America, Vol. 92, pp. 597~599.

(7) Chen, J., van Veen, B. D. and Hecox, K. E., 1995, "A Spatial Feature Extraction and Regularization Model for the Head-related Transfer Function", Journal of the Acoustical Society of America, Vol. 97, pp. 439~542.

(8) Wu, Z., Chan, F. H. Y., Lam, F. K. and Chan, J. C. K., 1997, "A Time Domain Binaural Model Based on Spatial Feature Extraction for the Head-related Transfer Function", Journal of the Acoustical Society of America, Vol. 102, pp. 2211~2218.

(9) Shin, K. and Park, Y., 2008, "Enhanced Vertical Perception Through Head-related Impulse Response Customization Based on Pinna Response Tuning in the Median Plane", IEICE Transactions on Fundamentals, Vol. E91-A, pp. 345~356.

(10) Hwang, S. and Park, Y., 2007, "Median HRIR Customization via Principal Components Analysis", Transactions of the Korea Society for Noise and Vibration Engineering, Vol. 17, pp. 638~648.

(11) Hwang, S. and Park, Y., 2008, "Interpretations on Principal Components Analysis of Head-related Impulse Responses in the Median Plane", Journal of the Acoustical Society of America, Vol. 123, pp. EL65-71.

(12) Algazi, V. R., Duda, R. O., Thompson, D. M. and Avendano, C., 2001, "The CIPIC HRTF Database", In Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, pp. 99~102.

(13) Brown, C. P. and Duda, R. O., 1998, "A Structural Model for Binaural Sound Synthesis", IEEE Transactions on Speech and Audio Processing, Vol. 6, pp. 476~488.

(14) Shaw, E. A. G., 1974, "Transformation of Sound Pressure Level from the Free Field to the Eardrum in the Horizontal Plane", Journal of the Acoustical Society of America, Vol. 56, pp. 1848~1861.

(15) Mehrgardt, S. and Mellert, V., 1977, "Transformation Characteristics of the External Human Ear", Journal of the Acoustical Society of America, Vol. 61, pp. 1567~1576.

(16) Middlebrooks, J. C., Makous, J. C. and Green, D. M., 1989, "Directional Sensitivity of Sound-pressure Levels in the Human Ear Canal", Journal of the Acoustical Society of America, Vol. 86, pp. 89~108.

(17) Wightman, F. L. and Kistler, D. J., 1989, "Headphone Simulation of Free-field Listening. I: Stimulus Synthesis", Journal of the Acoustical Society of America, Vol. 85, pp. 858~867.

(18) Oppenheim, A. V., Schafer, R. W. and Buck, J. R., 1999, "Discrete-time Signal Processing", Prentice-Hall, Inc., New Jersey, 2nd edition.

Appendix

The theoretical equivalence of the PCA model based on HRIRs in the time domain and that based on augmented HRTFs in the frequency domain can be proven by showing the equivalence of linear combination of PCs and the equivalence of orthogonality of PCs.

A.1 Equivalence of Linear Combination of PCs

Let $\mathbf{h}$ denotes the HRIR in the time domain. $\mathbf{h}$ can be modeled by a linear combination of L PCs and empirical mean, which are obtained from PCA of HRIRs in the time domain as

$$\mathbf{h}[n]=\sum_{i=1}^{L}\left(w_i\cdot\mathbf{v}_i[n]\right)+\mathbf{u}[n]\,, \tag{A.1}$$

where $\mathbf{v}_i$ and w_i indicate the ith PC and its PCW(real valued), respectively. $\mathbf{u}$ is the empirical mean of the HRIRs. The Fourier transform pair of $\mathbf{v}_i$ and $\mathbf{u}$ can be obtained by the N-point DFT as

$$\mathbf{V}_i[k]=\sum_{n=0}^{N-1}\mathbf{v}_i[n]\cdot W_N^{-kn}\,, \tag{A.2a}$$

$$\mathbf{U}[k]=\sum_{n=0}^{N-1}\mathbf{u}[n]\cdot W_N^{-kn}\,, \tag{A.2b}$$

where $W_N=e^{j2\pi/N}(k=1,2,...,N)$. $\mathbf{V}_i$ and $\mathbf{U}$ in Eq. (A.2) can be augmented as

$$\mathbf{V}_{i,A}[k]=\begin{bmatrix}\sum_{n=0}^{N-1}\mathbf{v}_i[n]\cdot\cos\left(\frac{2\pi n}{N}k\right)\\ -\sum_{n=0}^{N-1}\mathbf{v}_i[n]\cdot\sin\left(\frac{2\pi n}{N}k\right)\end{bmatrix}, \tag{A.3a}$$

$$\mathbf{U}_A[k]=\begin{bmatrix}\sum_{n=0}^{N-1}\mathbf{u}[n]\cdot\cos\left(\frac{2\pi n}{N}k\right)\\ -\sum_{n=0}^{N-1}\mathbf{u}[n]\cdot\sin\left(\frac{2\pi n}{N}k\right)\end{bmatrix}, \tag{A.3b}$$

The HRTF ($\mathbf{H}$), which is the Fourier transform pair of the HRIR, can be obtained by the N-point DFT as

$$\mathbf{H}[k]=\sum_{n=0}^{N-1}\mathbf{h}[n]\cdot W_N^{-kn}\ \left(k=0,1,\ldots,N\text{-}1\right). \tag{A.4}$$

By substituting of Eq. (A.1) into Eq. (A.4) and augmenting it, the augmented HRTF ($\mathbf{H}_A$) can be written as

$$\mathbf{H}_A[k]=\begin{bmatrix}\sum_{n=0}^{N-1}\left\{\mathbf{h}[n]\cdot\cos\left(\frac{2\pi n}{N}k\right)\right\}\\ -\sum_{n=0}^{N-1}\left\{\mathbf{h}[n]\cdot\sin\left(\frac{2\pi n}{N}k\right)\right\}\end{bmatrix} \tag{A.5a}$$

$$=\begin{bmatrix}\sum_{n=0}^{N-1}\left\{\left(\sum_{i=1}^{L}\left(w_i\cdot\mathbf{v}_i[n]\right)+\mathbf{u}[n]\right)\cdot\cos\left(\frac{2\pi n}{N}k\right)\right\}\\ -\sum_{n=0}^{N-1}\left\{\left(\sum_{i=1}^{L}\left(w_i\cdot\mathbf{v}_i[n]\right)+\mathbf{u}[n]\right)\cdot\sin\left(\frac{2\pi n}{N}k\right)\right\}\end{bmatrix} \tag{A.5b}$$

$$=\sum_{i=1}^{L}w_i\cdot\begin{bmatrix}\sum_{n=0}^{N-1}\left\{\mathbf{v}_i[n]\cdot\cos\left(\frac{2\pi n}{N}k\right)\right\}\\ -\sum_{n=0}^{N-1}\left\{\mathbf{v}_i[n]\cdot\sin\left(\frac{2\pi n}{N}k\right)\right\}\end{bmatrix} +\begin{bmatrix}\sum_{n=0}^{N-1}\left\{\mathbf{u}[n]\cdot\cos\left(\frac{2\pi n}{N}k\right)\right\}\\ -\sum_{n=0}^{N-1}\left\{\mathbf{u}[n]\cdot\sin\left(\frac{2\pi n}{N}k\right)\right\}\end{bmatrix} \tag{A.5c}$$

$$=\sum_{i=1}^{L}w_i\cdot\mathbf{V}_{i,A}[k]+\mathbf{U}_A[k]. \tag{A.5d}$$

Thus, the augmented HRTF in the frequency domain can be modeled by linear combination of PCs ($\mathbf{V}_{i,A}$) and empirical mean ($\mathbf{U}_A$), which are directly computable by the Fourier transform of the PCs ($\mathbf{v}_i$) and empirical mean ($\mathbf{u}$) in the time domain, respectively. Note that the weights of vectors (w_i) in the frequency domain are identical to those in the time domain.

A.2 Equivalence of Orthogonality of PCs

The vectors, obtained from the Fourier transform of PCs in the time domain, also

should be orthogonal to each other to be the basis vectors or PCs in the frequency domain. Let $\mathbf{v}_p$ and $\mathbf{v}_q$ are the pth and qth PCs in the time domain and they are orthogonal to each other as

$$\left\langle \mathbf{v}_p, \mathbf{v}_q \right\rangle = \sum_{n=0}^{N-1} \mathbf{v}_q^*[n] \cdot \mathbf{v}_p[n] = \delta_{pq} \,. \tag{A.6}$$

Then, by using Eq. (A.3) the inner product of $\mathbf{V}_{p,A}$ and $\mathbf{V}_{q,A}$ can be written as

$$\left\langle \mathbf{V}_{p,A}, \mathbf{V}_{q,A} \right\rangle = \sum_{k=0}^{N-1} \left(\mathbf{V}_{q,A}^*[k] \cdot \mathbf{V}_{p,A}[k] \right) \tag{A.7a}$$

$$= \sum_{k=0}^{N-1} \left[\sum_{l=0}^{N-1} \left\{ \mathbf{v}_q[l] \cdot \cos\left(\frac{2\pi l}{N} k \right) \right\} \cdot \sum_{n=0}^{N-1} \left\{ \mathbf{v}_p[n] \cdot \cos\left(\frac{2\pi n}{N} k \right) \right\} + \sum_{l=0}^{N-1} \left\{ \mathbf{v}_q[l] \cdot \sin\left(\frac{2\pi l}{N} k \right) \right\} \cdot \sum_{n=0}^{N-1} \left\{ \mathbf{v}_p[n] \cdot \sin\left(\frac{2\pi n}{N} k \right) \right\} \right] \tag{A.7b}$$

$$= \sum_{n=0}^{N-1} \sum_{l=0}^{N-1} \left[\mathbf{v}_p[n] \cdot \mathbf{v}_q[l] \cdot \sum_{k=0}^{N-1} \left\{ \cos\left(\frac{2\pi n}{N} k \right) \cdot \cos\left(\frac{2\pi l}{N} k \right) + \sin\left(\frac{2\pi n}{N} k \right) \cdot \sin\left(\frac{2\pi l}{N} k \right) \right\} \right] \tag{A.7c}$$

$$= \sum_{n=0}^{N-1} \sum_{l=0}^{N-1} \left[\mathbf{v}_p[n] \cdot \mathbf{v}_q[l] \cdot \sum_{k=0}^{N-1} \cos\left(\frac{2\pi (n-l)}{N} k \right) \right] \tag{A.7d}$$

$$= N \sum_{n=0}^{N-1} \left(\mathbf{v}_p[n] \cdot \mathbf{v}_q[n] \right) = N \delta_{pq} \,. \tag{A.7e}$$

Therefore, it is proven that the vectors obtained from the Fourier transform of PCs in the time domain are also orthogonal to each other. This orthogonality is resulted from

$$\sum_{k=0}^{N-1} \cos\left(\frac{2\pi (n-l)}{N} k \right) = \begin{cases} N & (l - n = mN) \\ 0 & (\text{otherwise}) \end{cases}, \tag{A.8}$$

where m is an integer.