한국소음진동공학회논문집 제17권 제7호, pp. 638~648, 2007.

Median HRIR Customization via Principal Components Analysis
주성분 분석을 이용한 HRIR 맞춤 기법#

Sungmok Hwang* and Youngjin Park†
황 성 목 · 박 영 진

(Received May 28, 2007 ; Accepted June 14, 2007)

Key Words : Head-related Transfer Function(머리전달함수), Customization(맞춤기법), Principal Components Analysis(주성분 분석)

ABSTRACT

A principal components analysis of the entire median HRIRs in the CIPIC HRTF database reveals that the individual HRIRs can be adequately reconstructed by a linear combination of several orthonormal basis functions. The basis functions represent the inter-individual and inter-elevation variations in median HRIRs. There exist elevation-dependent tendencies in the weights of basis functions, and the basis functions can be ordered according to the magnitude of standard deviation of the weights at each elevation. We propose a HRIR customization method via tuning of the weights of 3 dominant basis functions corresponding to the 3 largest standard deviations at each elevation. Subjective listening test results show that both front-back reversal and vertical perception can be improved with the customized HRIRs.

요 약

CIPIC HRTF database의 주성분 분석(PCA)을 통해 개인의 HRIR이 정규 직교화된 소수의 기저함수들의 선형결합으로 잘 묘사됨을 알 수 있다. 이 기저함수들은 음원의 고도각, 청취자 마다 달라지는 HRIR의 변화를 표현할 수 있다. 선형결합에 사용되는 기저함수들의 가중치들은 음원의 고도각에 따라 특이한 경향을 지닌다. 또한, 각각의 음원 위치에서 가중치의 표준편차 크기순으로 기저함수의 중요도를 결정할 수 있다. 이 논문에서는 각 음원 위치마다 중요한 3개 기저함수의 가중치를 청취자가 직접 조절하게 함으로써 맞춤형 HRIR을 생성하는 방법을 제안한다. 주관평가 결과, 청취자의 음원 고도각 인지 성능과 음원 앞-뒤 구분 성능이 향상됨을 확인하였다.

1. Introduction

The dominant determinants of the apparent direction of a sound are interaural time difference, interaural level difference, and spectral modification due to *pinnae*[1~3]. These are called *primary* sound cues and encrypted in the Head-related Transfer Functions(HRTFs), which is an acoustic transfer function from a sound source to a listener's eardrum. Thus, HRTFs play an important role in Virtual Auditory Display(VAD), and most VAD systems use the non-individualized HRTFs measured from a dummy

† Corresponding Author: Member, Department of Mechanical Engineering, KAIST
E-mail : yjpark@kaist.ac.kr
Tel : (042) 869-3036, Fax : (042) 869-8220
* Department of Mechanical Engineering, KAIST

이 논문은 2007 춘계 소음진동 학술대회에서 우수논문으로 추천되었음.

head microphone system. Non-individualized HRTFs, however, often cause problems such as inaccurate lateralization, poor vertical effects, and weak front-back distinction because HRTFs vary considerably from subject to subject. Although individual HRTFs can alleviate these problems, measurement of individual HRTFs for every listener is not practical due to the requirements of heavy and expensive equipments as well as a long measurement time. Thus, it is a priority to develop a customization method that provides the listener with proper sound cues without measurement of the individual HRTFs. Several methods for customization, such as HRTF clustering and selection of a few most representative ones[4], HRTF scaling in frequency[5], a structural model for composition and decomposition of HRTFs[6], and HRTF database matching[7], are already suggested. However, these previous methods have some practical limitations. For example, the method of HRTF scaling in frequency is based on the basic idea that HRTF will be shifted toward higher frequencies or lower frequencies if the size of pinna increases or decreases, respectively, while maintaining its shape. However, the pinnae of different listeners are different in many more aspects than just a size of pinna. Thus, a trivial change in the pinna shape can yield complex changes in HRTF. The method of HRTF database matching uses the individual HRTF database (CIPIC HRTF database) contains HRTFs of 45 subjects along with 7 anthropometric parameters about the subjects. The best matching set of individual HRTFs is selected by taking a picture of the listener's own ear and comparing the anthropometric parameters measured from the picture with the ones in the database. However, this method requires an additional imaging system to capture the listener's ear and compute the anthropometric parameters from the image. More recently, Shin and Park[8] suggested the Head-related Impulse Response (HRIR) customization method based on subjective tunings of the pinna responses in the time domain. HRIR is a time domain counter part of HRTF, and it is the Fourier transform pair of HRTF. The basic idea of their method is that the pinna response of any arbitrary listener can be reproduced by a linear combination of a set of basis function obtained by Principal Components Analysis (PCA) of the CIPIC HRTF database. However, they focused on the pinna response only. Although the pinna responses are important for listener to perceive sound direction, the shoulder or torso response also provides directional sound cue. In their PCA process, the pinna responses of 45 individuals at each elevation were included in a single analysis, and the set of basis functions is different from elevation to elevation, thus, the basis functions represent the inter-individual variation only.

Our customization method is similar with the Shin and Park's method, but we expand the HRIR dataset to be analyzed in PCA. Entire median HRIRs in the CIPIC HRTF database[9] are included in a single analysis. Thus, all median HRIRs share the same set of basis functions, and the basis functions represent not only the inter-individual variation but also the inter-elevation variation. The response of 1.5 msec since the arrival of direct pulse in HRIR, which contains the effects of pinna, shoulder, and torso, are included in PCA, whereas Shin and Park used the pinna response of 0.2 msec only.

2. Principal Components Analysis of median HRIRs

Principal Components Analysis is one of the statistical procedures that try to provide an efficient representation of a set of correlated data[10]. The basic idea of PCA is to simplify the dataset by reducing multidimensional dataset

to lower dimensions, while remaining the variation present in the dataset, as much as possible. Martens applied PCA to the problems of modeling of HRTFs[11]. Kistler and Wightman showed that the log magnitude of HRTF can be adequately approximated by a linear combination of five basis spectral shapes[12]. These previous works focus on the magnitude response of HRTF in the frequency domain. However, we apply PCA for modeling of HRIRs and customization in the time domain. The 2,205 median HRIRs in the CIPIC HRTF database are included in PCA.

Before PCA, we post-process HRIRs to remove the initial time delay and to extract the early response that lasts for 1.5 msec since the arrival of direct pulse as depicted in Fig. 1. The response of 1.5 msec includes the effects of pinna, head, shoulder, and torso. The first step in PCA is to make a matrix composed of the mean-subtracted HRIRs. The original data matrix ($\mathbf{X}: N \times M$) is composed of the post-processed median HRIRs. The each column of $\mathbf{X}$, $x_i(i=1,2,\cdots,M)$, indicates the post-processed HRIR, and the dimension of $\mathbf{X}$ is 67 $\times$ 2,205 in this case. The response of 1.5 msec corresponds 67 samples (sampling frequency: 44.1 kHz) and the number of median HRIRs is 2,205 (45 subjects $\times$ 49 elevations from -45° to 225° at 5.625° intervals).

The empirical mean of $\mathbf{X}$ is needed to obtain the mean-subtracted HRIRs, and the empirical mean vector ($\mathbf{u}$) of dimensions $N \times 1$ is given by

$$u[n] = \frac{1}{M}\sum_{m=1}^{M} \mathbf{X}[n,m]. \quad (1)$$

The mean-subtracted data matrix, $\mathbf{B}$, is computed by

$$\mathbf{B} = \mathbf{X} - \mathbf{u}\cdot\mathbf{h}, \quad (2)$$

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

$$\mathbf{C} = \mathrm{E}[\mathbf{B}\otimes\mathbf{B}] = \frac{1}{M-1}\mathbf{B}\cdot\mathbf{B}^{*}, \quad (3)$$

where $\otimes$ and $*$ indicate the outer product and the conjugate transpose operators, respectively. The basis functions (or basis vectors), v_q, are the q eigenvectors of the covariance matrix, $\mathbf{C}$, corresponding to the q largest eigen values. These basis functions are called "Principal Components (PCs)". If $q=N$, then the original HRIRs can be fully reconstructed by a linear combination of the q PCs. However, in many practical applications, $q \ll N$ because the object of PCA is to reduce the dimension of data set.

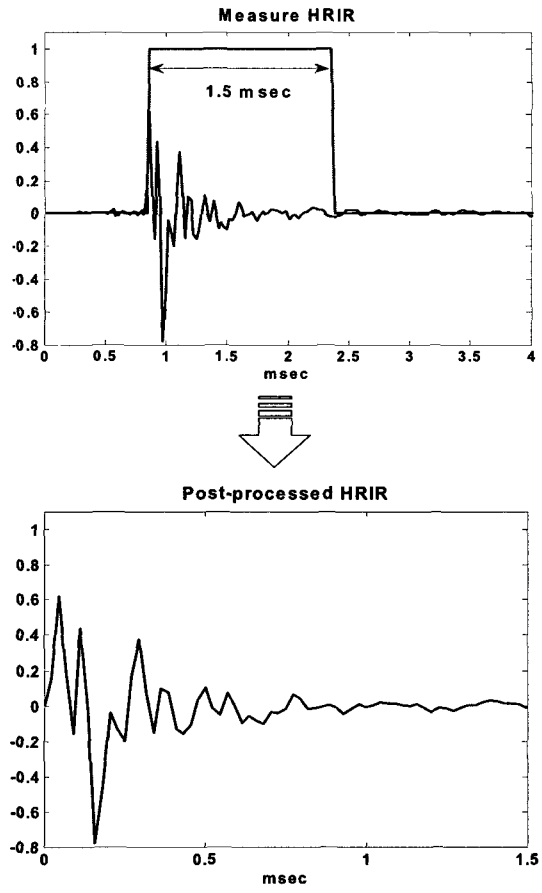


Fig. 1 Post-processing of HRIR

Thus, we can obtain only an estimate of the original dataset by using the $q(\ll N)$ PCs. The weights of PCs (PCWs) can be obtain as

$$\mathbf{W} = \mathbf{V}^{*} \cdot \mathbf{B}, \quad \mathbf{V} = [\mathbf{v}_1 \, \mathbf{v}_1 \cdots \mathbf{v}_q]. \tag{4}$$

PCWs represent the contribution of each basis function to the HRIRs. The estimate of HRIRs is given by

$$\tilde{\mathbf{X}} = \mathbf{V} \cdot \mathbf{W} + \mathbf{u} \cdot \mathbf{h} \, . \tag{5}$$

Then, we should determine how many PCs we use for the HRIR reconstruction. We define the reconstruction error in percentage as

$$e = \frac{\left\| \mathbf{X} - \tilde{\mathbf{X}} \right\|_F^2}{\left\| \mathbf{X} \right\|_F^2} \times 100 \ (\%), \tag{6}$$

where subscript F indicates the frobenius matrix norm. The more PCs are used, the more accurately HRIR can be reconstructed as depicted in Fig. 2. We arbitrary set the reconstruction error bound of 5%, and we retain 12 PCs.

Figure 3 shows the empirical mean and 12 PCs obtained from PCA of the left ear HRIRs (2,205 HRIRs). PC1, PC2, PC3, and PC4 have high energy in the pinna response up to 0.2 msec, whereas PC9 and PC10 have high energy in the shoulder response. Thus, it can be said that PC1, PC2, PC3, and PC4 mainly contribute to the effect of pinna and PC9 and PC10 mainly contribute to the shoulder/torso effects.

Figure 4 shows an example of the reconstruction process for the left ear HRIR of a representative subject (Subject 152 in the CIPIC HRTF database) for a source at -33.75° elevation in the median plane. When only PC1-8 used for reconstruction, the pinna response can be well reconstructed, however, the shoulder/torso response cannot be reproduced because these PCs have less energy in that response region. The shoulder/torso response can be reconstructed by PC9 and PC10. Of course, other PCs also contribute to the reconstruction, PC9 and PC10 are dominant components to recover the shoulder/torso effect.

PCWs also should be investigated because they represent the contribution of each PC in the reconstructed HRIR. Figure 5 shows the mean value and standard deviation of each PCW for all left ear HRIRs (45 subjects) with respect to the change of elevation in the median plane. There are some notable tendencies in PCWs, and each PCW provide a useful sound cue for front-back distinction and vertical perception. For example, PCW1 has positive mean value from about 30° to 150° and has negative value at other elevations. Furthermore, it increases monotonically from 0° to 90° and decreases monotonically from 90° to 180°. Thus, we can conclude that PC1 contributes to the vertical perception. PCW2 has positive mean value in the frontal region except for low sources and has negative mean value in the rear region. Mean of PCW3 is almost asymmetric about 90° of elevation. Therefore, PC2 and PC3 provide sound cue for front-back discrimination. The standard deviation of each PCW also provides

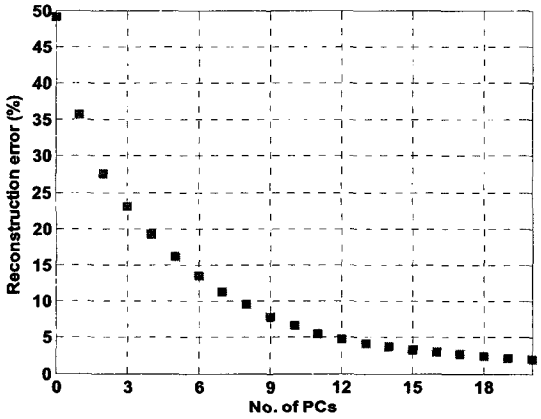


Fig. 2 Reconstruction errors with respect to the number of PCs

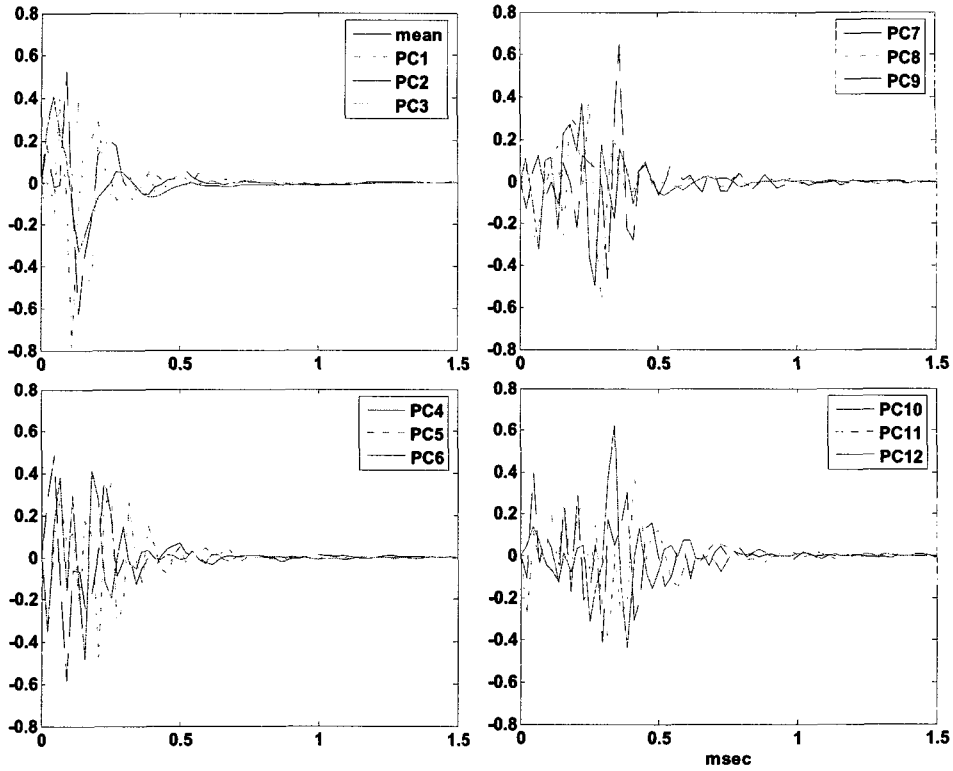


Fig. 3 The empirical mean and PCs

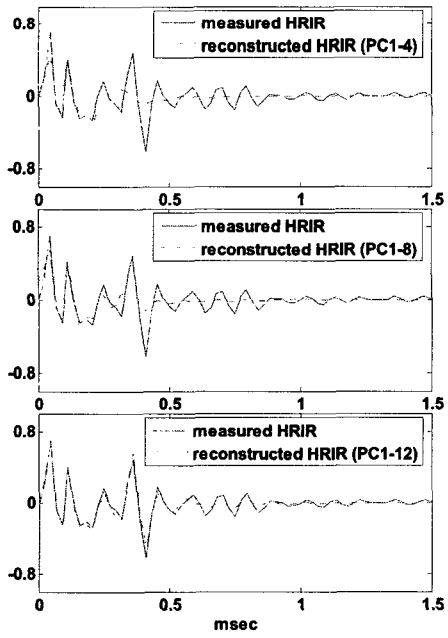


Fig. 4 Reconstruction result

how much inter-individual variation exists in HRIRs. The larger standard deviation means the larger inter-individual variation in HRIRs. PC1 has larger standard deviation for higher sources than lower sources. From the standard deviation of PCW5, it can be said that the inter-individual variation of PC5 decreases as elevation is high, and PC9 shows larger inter-individual variation for frontal low sources. In Fig. 4, PC9 contribute to the shoulder/torso response, thus the inter-individual variation is large at low elevation because the shoulder/torso effect is more prominent for lower sources.

3. HRIR Customization

As mentioned above, 12 PCs represent the inter-individual and inter-elevation variations in median HRIRs within the error bound of 5 %. Thus, by allowing a subject to tune the PCW on each PC, one can make customized HRIRs. However, tuning the 12 PCWs is very exhausting

task, thus the number of tuning PCWs should be reduced. At each elevation, we can arrange PCs with respect to the magnitude of standard deviation as depicted in Table 1. The order of PCs is different at each elevation. We need to pay attention to the order for customization. The PCs having small standard deviation don't contribute to the inter-individual variations, thus we let the subject tune the weights on dominant PCs (DPCs) having large standard deviation and take mean values for other PCWs. For customization, we chose 3 DPCs corresponding to the 3 largest standard deviations at each elevation. Above customization process is based on the MATLAB^{TM} GUI as depicted in Fig. 6. When a subject choose the elevation, the set of DPCs is automatically set. The maximum and minimum bounds of each weight of DPC were set to be mean ± 3 standard deviations. Customization is only carried out at 9 specific elevation angles from -30° to 210° at 30° intervals in the median plane. We substitute the mean values of DPCs at the elevation angles of -28.125°, 28.125°, 61.875°, 118.125°, 151.875°, and 208.125° for those at the elevation angles of -30°, 30°, 60°, 120°, 150°, and 210°, respectively, because the CIPIC HRTF database is available at the elevation angles from -45° to 230° at 5.625° intervals. We assumed that the left and right ears are symmetric and HRIRs for two ears are the same in the median plane. Thus, the left and right channels of a headphone (Sennheiser HD 250) were driven by the same signal.

For comparison of the localization performance, individual HRTFs of the two male participants (not including authors) were measured at the elevation angle where customization took place. A white noise with a bandwidth covering the entire audible frequency range (20 Hz ~ 20 kHz) was used as the general input to the speaker. The B&K binaural microphone type 4101 was mounted inside each subject's pinna. The distance between the speaker and the subject's head center was set to be 1 m. The input signal together with the resulting output signal from the microphone were each collected for a sampling duration of 1.5 seconds at a sampling rate of 44.1 kHz. Figure 7 shows the customized and individual HRIRs for a representative subject. The angle at the left or right of each panel indicates the source elevation, and PCs below the angle indicate the 3 DPCs used for customization. Of course there exist some discrepancies between the customized and individual HRIRs, the pinna response up to 0.2 msec and the shoulder/torso reflection (about 0.3 ~ 0.5 msec) can be well reproduced by customization.

4. Subjective Listening Test

Subjective listening tests using a pair of headphone (Sennheiser HD 250) were performed on the two subjects to evaluate the localization performances of non-individualized (Kemar), individual, and customized HRIRs. For convenient test procedure, another MATLAB^{TM} GUI was used for subjective listening test as depicted in Fig. 8. For correct headphone-presented stimulation of free-field listening when evaluating the individual HRIRs on their localization capabilities, the headphone dynamics was cancelled according to the method suggested by Wightman and Kistler[13,14]. When the subject registers his ID and date, a set of test signals containing 90 broadband stimuli for randomly selected HRIR set was generated. Each of the 9 elevation is stimulated 10 times in a random order yielding in total 90 stimuli. After subject listen each stimulus by pushing the "PLAY" button, he pushes one of the buttons corresponding to the perceived angle and "OK" button. Then, the number of sequence increases

by one. When the number of sequence hits 90, the test is completed and the test result is saved by pushing the "SAVE" button. Figures 9~11 show the subjective listening test results by the two subjects. In each panel, the horizontal and the vertical axes denote the actual source elevation and the perceived elevation, respectively. The radius of circle is directly proportional to the response frequency. The positive-sloped diagonal line indicates the perfect elevation perception and font-back non-reversal condition in which the subject is able to pinpoint the source elevation with perfect accuracy. The negative-sloped diagonal line indicates the perfect font-back reversal condition in which the subject reported the

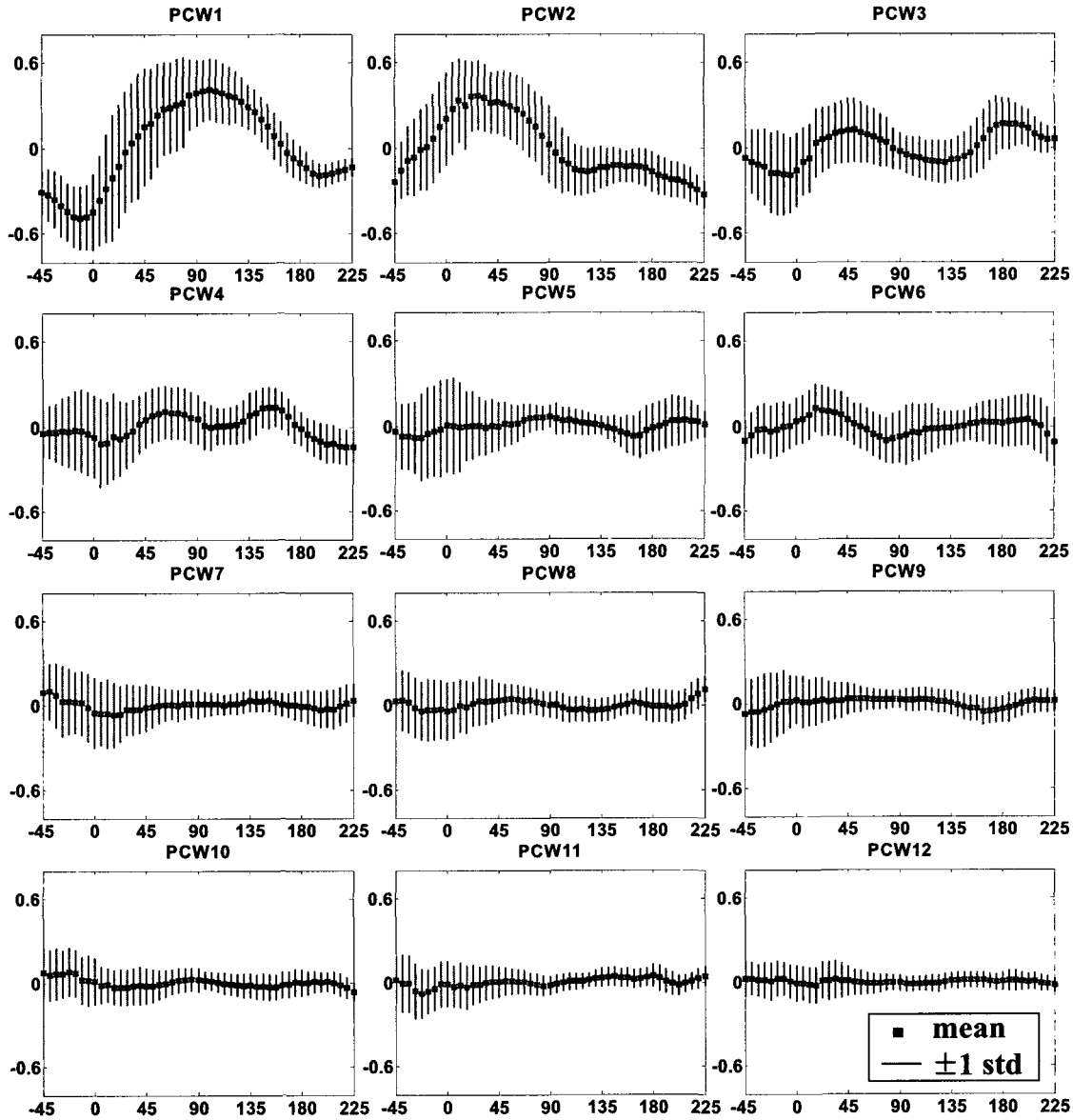


Fig. 5 Meanand ±1 standard deviation of each PCW for all left ear HRIRs (45 subjects) with respect to the change of median source elevation

accurate vertical perception but perceived all frontal sources as rear sources, and vice versa. All subjects showed prominent front-back reversal and poor vertical perception with the kemar HRIRs. With the individual HRIRs, subject **CH** showed front-back reversal at frontal low elevation but vertical perception was improved. Subject **KB** reported the improved localization performance for frontal sources with his own HRIRs, but the responses for rear sources were scattered. With the customized HRIRs, front-back reversal was frequent for subject **CH**. However, the responses were close to the two diagonal lines, and this means that vertical perception was improved with the customized HRIRs. Subject **KB** also reported the improved vertical perception with the customized HRIRs.

A quantitative error analysis for front-back reversal and vertical perception can be performed by a cluster analysis of the response data along the two diagonal lines. If the response is more close to the negative-sloped line than the positive-sloped line, we determined that the front-back reversal is occurred and the vertical perception error is the angular difference between the response and the negative-sloped line. Of course, if the response is more close to the positive-sloped line than the negative-sloped line, the front-back reversal is not occurred and the vertical perception error is the angular difference between the response and the positive-sloped line. Thus, we defined two kinds of errors, the front-back reversal error (e_{FBR}) and the vertical perception error(e_{VP}), as

$$e_{FBR} = \frac{\text{No. of responses satisfying } |P-(180°-A)| < |P-A|}{\text{No. of total responses}} \times 100\ (\%), \quad (7)$$

$$e_{VP} = \frac{1}{N}\sum_{i=1}^{N} \min(|P_i - A_i|, |P_i - (180° - A_i)|), \quad (8)$$

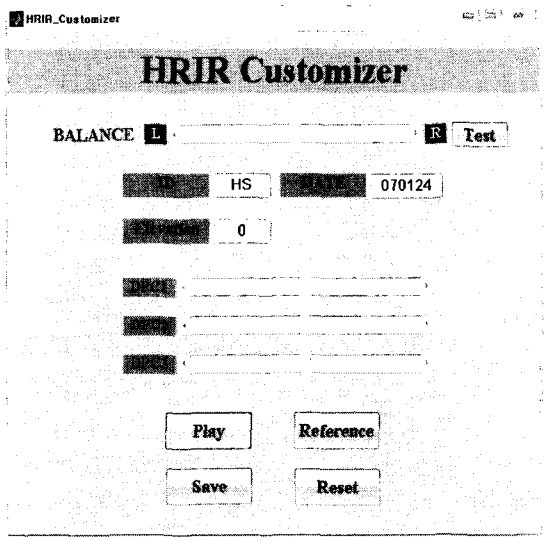


Fig. 6 MATLAB™ GUI for customization

Table 1 Order of PCs and standard deviations (below each PC) at each elevation

Elev.	Order of PCs and standard deviations(below each PC)											
-45°	PC9	PC8	PC5	PC3	PC7	PC2	PC10	PC11	PC4	PC1	PC6	PC12
	0.253	0.203	0.200	0.195	0.191	0.185	0.179	0.168	0.166	0.153	0.143	0.129
0°	PC5	PC2	PC4	PC1	PC3	PC7	PC8	PC9	PC10	PC11	PC6	PC12
	0.322	0.322	0.290	0.268	0.257	0.247	0.203	0.167	0.164	0.161	0.161	0.125
45°	PC1	PC3	PC2	PC4	PC5	PC7	PC6	PC10	PC8	PC9	PC11	PC12
	0.407	0.223	0.214	0.195	0.165	0.165	0.147	0.126	0.125	0.112	0.107	0.105
90°	PC2	PC1	PC6	PC4	PC3	PC7	PC8	PC10	PC5	PC11	PC9	PC12
	0.269	0.229	0.181	0.168	0.128	0.102	0.085	0.082	0.075	0.062	0.061	0.056
135°	PC4	PC1	PC2	PC3	PC6	PC10	PC9	PC8	PC7	PC11	PC5	PC12
	0.162	0.156	0.145	0.139	0.112	0.083	0.082	0.079	0.077	0.066	0.064	0.055
180°	PC3	PC5	PC6	PC2	PC1	PC4	PC8	PC7	PC10	PC9	PC11	PC12
	0.181	0.164	0.164	0.140	0.118	0.114	0.114	0.104	0.092	0.091	0.080	0.063
225°	PC6	PC&	PC4	PC2	PC8	PC5	PC3	PC9	PC11	PC10	OC1	PC12
	0.167	0.112	0.111	0.099	0.091	0.088	0.083	0.075	0.069	0.063	0.061	0.049

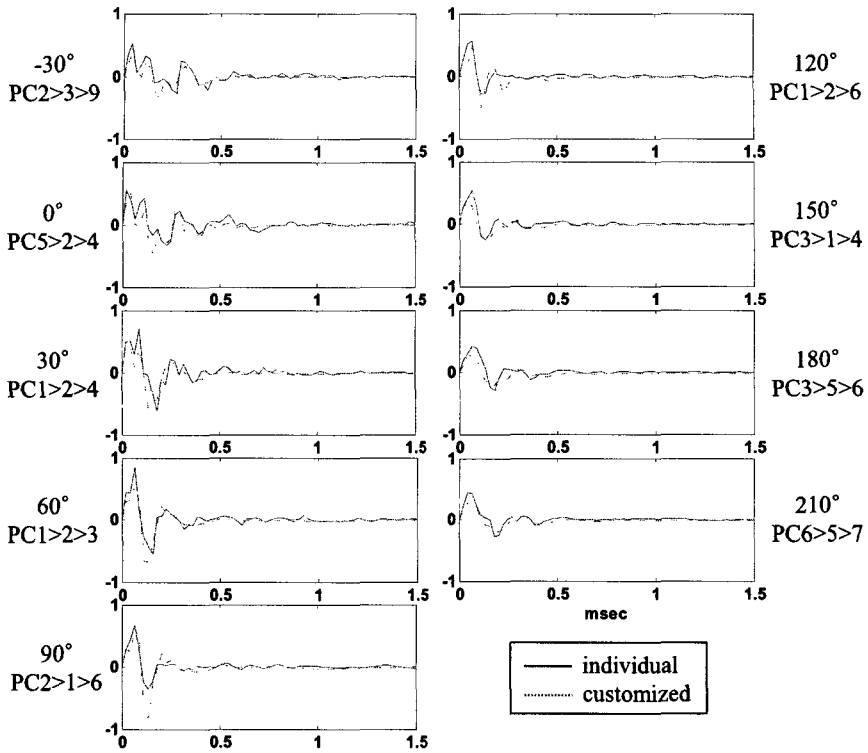


Fig. 7 Individual and customized HRIRs for subject CH

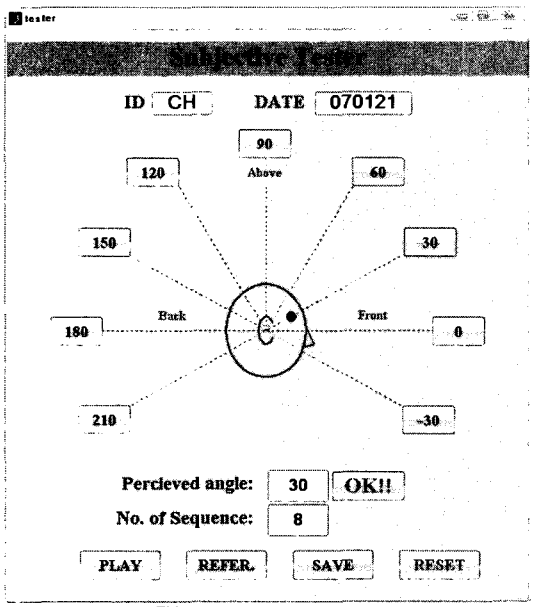
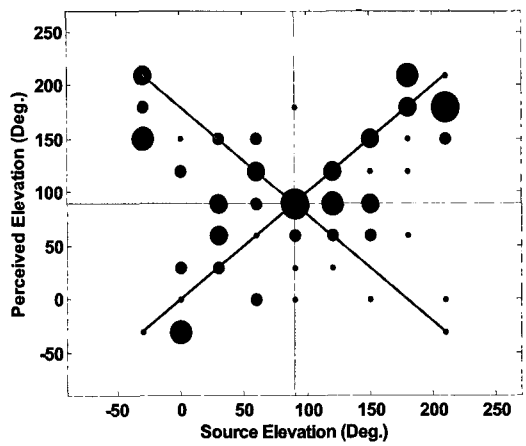


Fig. 8 MATLAB™ GUI for subjective listening test

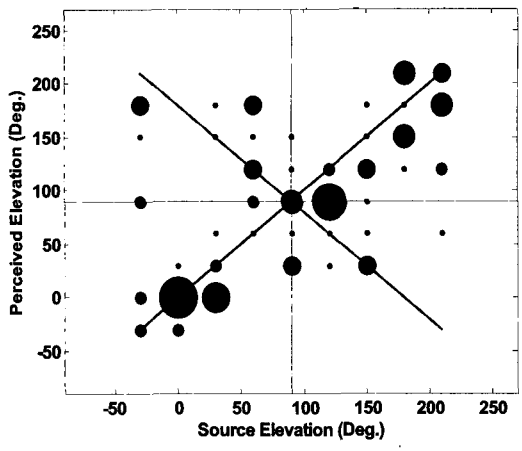
where, P and A indicate the perceived elevation and the actual source elevation, respectively, and A and$(180°-A)$ correspond to the position on the positive-sloped and negative-sloped lines, respectively. The two kinds of errors for each subject are summarized in Table 2. Comparison of the errors made with the individual HRIRs to those made with the kemar HRIRs reveals that both front-back reversal and vertical perception were improved for all subjects except for vertical perception of subject KB. All subjects showed the best performance for font-back discrimination and elevation perception with the customized HRIRs except for front-back reversal of subject KB.

Table 2 Quantitative error analysis for subjective listening test results

	Subject CH		Subject KB	
	e_{FBC}	e_{EP}	e_{FBC}	e_{EP}
Kemar	31.1 %	27.3°	22.2 %	27.0°
Individual	29.8 %	23.7°	12.2 %	32.3°
Customized	31.1 %	19.3°	14.4 %	21.3°

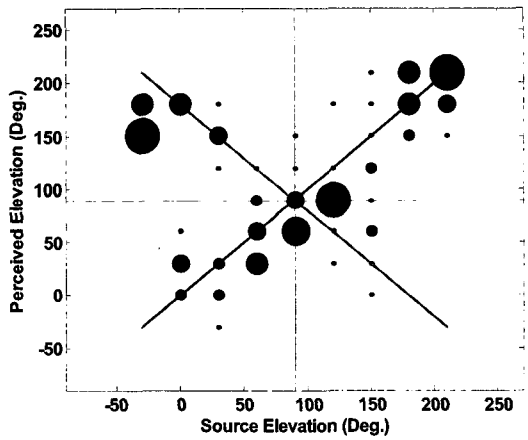


(a) Kemar HRIR(subject CH)

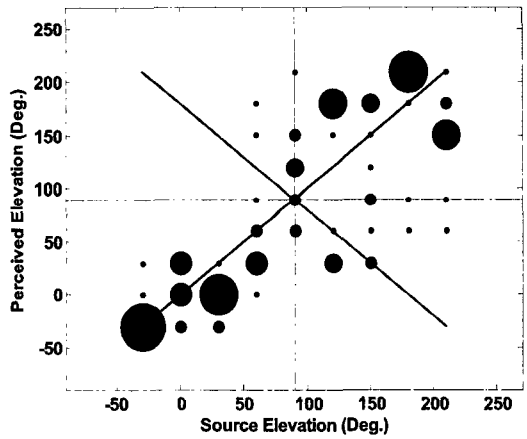


(b) Kemar HRIR(subject KB)

Fig. 9 Subjective test results for kemar HRIR

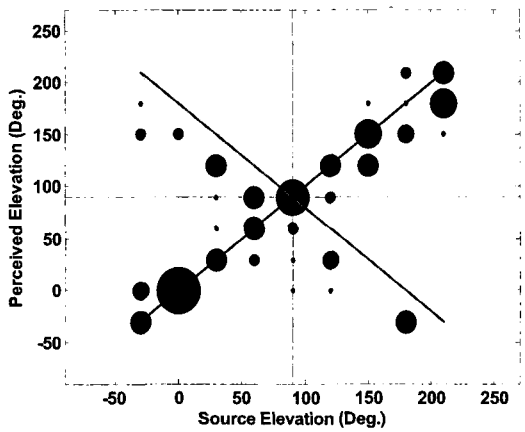


(a) Individual HRIR(subject CH)

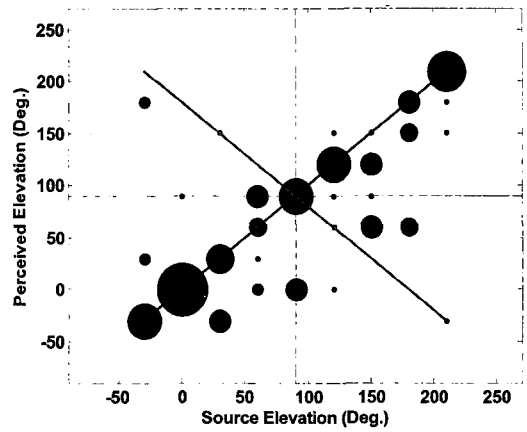


(b) Individual HRIR(subject KB)

Fig. 10 Subjective test results for individual HRIR



(a) Customized HRIR(subject CH)



(b) Customized HRIR(subject KB)

Fig. 11 Subjective test results for customized HRIR

5. Conclusion

Individual HRIRs can be adequately reconstructed by a linear combination of 12 basis functions in the time domain obtained from PCA of the entire median HRIRs in the CIPIC HRTF database. The basis functions represent the inter-individual and inter-elevation variations in median HRIRs. Each basis function contributes to effects of pinna, shoulder, and torso. There are elevation-dependent tendencies in PCWs, and the basis functions can be ordered according to the magnitude of standard deviation of the weights at each elevation. We proposed a HRIR customization method via tuning of the weights of 3 DPCs at each elevation. Subjective listening test results showed that all subjects perceive the elevation angles in the median plane more accurately with the customized HRIRs.

Acknowledgment

This work was supported by the National Research Laboratory Program(M10500000112-05J0000-11210), the Brain Korea 21 Project.

References

(1) Blauert, J., 1983, Spatial hearing, MIT, Cambridge, MA.

(2) Brungart, D. S. and Rabinowitz, W. M., 1999, "Auditory Localization of Nearby Sources. Head-related transfer functions", J. Acoust. Soc. Am., Vol. 106, pp. 1465~1479.

(3) Cheng, C. I. and Wakerfield, G. H., 2001, "Introduction to Head-related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space", J. Audio Eng. Soc., Vol. 49, pp. 231~248.

(4) Shimada, S., Hayashi, M. and Hayashi, S., 1994, "A Clustering Method for Sound Localization Transfer Functions", J. Audio Eng. Soc., Vol. 42, pp. 577~584.

(5) Middlebrooks, J. C., 1999, "Virtual Localization Improved by Scaling Non-individualized External-ear Transfer Functions in Frequency", J. Acoust. Soc. Am., Vol. 106, pp. 1493~1510.

(6) Algazi, V. R., Duda, R. O., Morrison, R. P., and Thompson, D. M., 2001, "Structural Composition and Decomposition of HRTFs", In Proc. WASPAA01, New Paltz, NY, pp. 103~106.

(7) Zotkin, D. N., Duraiswami, R. and Davis, L. S., 2002, "Customizable Auditory Display", In Proc. Int. Conf. on Auditory Display (ICAD), Kyoto, Japan.

(8) Shin, K. H. and Park, Y., 2006, "Customization of Head-related Transfer Functions Using Principal Components Analysis in the Time Domain (A)", J. Acoust. Soc. Am., Vol. 120, p. 3284.

(9) Algazi, V. R., Duda, R. O., Thompson, D. M. and Avendano, C., 2001, "The CIPIC HRTF database", In Proc. WASPAA01, New Paltz, NY, pp. 99~102.

(10) Dunteman, G. H., 1989, PRINCIPAL COMPONENTS ANALYSIS, Sage Publication, Inc.

(11) Martens, W. L., 1987, "Principal Components Analysis and Resynthesis of Spectral Cues to Perceive Direction", Proc. Int. Computer Music Conf., pp. 274~281.

(12) Kistler, D. J. and Wightman, F. L., 1992, "A Model of Head-related Transfer Functions Based on Principal Components Analysis and Minimum-phase Reconstruction", J. Acoust. Soc. Am., Vol. 91, pp. 1637~1647.

(13) Wightman, F. L. and Kistler, D. J., 1989, "Headphone Simulation of Free-field Listening. I: Stimulus Synthesis", J. Acoust. Soc. Am., Vol. 85, pp. 858~867.

(14) Wightman, F. L. and Kistler, D. J., 1989, "Headphone Simulation of Free-field Listening. II: Psychophysical Validation", J. Acoust. Soc. Am., Vol. 85, pp. 868~878.