

HRIR CUSTOMIZATION IN THE MEDIAN PLANE VIA PRINCIPAL COMPONENTS ANALYSIS

SUNGMOK HWANG¹ AND YOUNGJIN PARK²

¹ Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology, Republic of Korea
tjdahr78@kaist.ac.kr

² Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology, Republic of Korea
yjpark@kaist.ac.kr

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 cover the inter-individual and inter-elevation variations in median HRIRs. There are elevation-dependent tendencies in the weights of basis functions. 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 test results show that both front-back reversal and vertical perception can be improved with the customized HRIRs.

INTRODUCTION

The dominant determinants of the apparent direction of a sound are interaural time difference, interaural level difference, and spectral modification due to the pinnae [1,2,3]. These are called the primary sound cues and encrypted in the Head-Related Transfer Functions (HRTFs), which is the direction-dependent acoustic transfer function from a sound source to a listener's eardrum. Thus, HRTFs play an important role in Virtual Auditory Display (VAD) systems, and most VAD systems are based on the non-individualized HRTFs measured from a dummy head microphone system. Non-individualized HRTFs, however, often cause problems such as inaccurate lateralization, poor vertical effects, and weak front-back distinction. These problems are mainly caused by the considerable variability of HRTFs from subject to subject and should be settled to the production of high-fidelity VAD. It is well-known that these problems can be alleviated when a listener's own HRTFs are used [4,5]. However, 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 a listener with proper sound cues without measurement of the individual HRTFs.

Several HRTF customization methods, such as HRTF clustering and selection of a few most representative ones [6], HRTF scaling in frequency [7], a structural model for composition and decomposition of HRTFs [8], and HRTF database matching [9], are already developed. 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 the higher frequencies or lower

frequencies when the size of pinna decreases or increases, respectively, while keeping its shape. However, the pinnae of different listeners are different in many aspects than just a size of pinna. Therefore, a trivial change in the pinna shape can yield complex changes in HRTFs. The method of HRTF database matching uses the individual HRTF database (CIPIC HRTF database [10]) containing 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 [11] suggested the Head-Related Impulse Response (HRIR) customization method based on subjective tuning of the pinna responses in the time domain. HRIR is a time domain counterpart of HRTF, and it is the Fourier transform pair of HRTF. The basic idea of their method is that the pinna response of an arbitrary listener can be reproduced by a linear combination of a set of basis functions, which is computed by Principal Components Analysis (PCA) of the CIPIC HRTF database. Thus, they obtained the customized pinna response by letting a subject tune the weight on each basis function, and they showed an improved performance for the elevation perception by a series of subjective test. They, however, focused on the pinna response only. Although the pinna responses are important for the listener to perceive sound direction, the shoulder and torso responses also provide directional sound cues especially for the up-down distinction. In their PCA process, the pinna responses of

45 individuals at each elevation were included in a single analysis. Therefore, the set of basis function is different from elevation to elevation, and the basis functions cover the inter-individual variation only.

Our customization method is similar with Shin and Park's method, but we expand the HRIR dataset to be analyzed in PCA. Entire median HRIRs (except 45 HRIRs at 230.625°) in the CIPIC HRTF database are included in a single analysis. Thus, all median HRIRs share the same set of basis functions, and the basis functions cover 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.

1 PCA OF HRIRS IN THE TIME DOMAIN

PCA is one of the statistical procedures that try to provide an efficient representation of a set of correlated data [12]. The basic idea of PCA is to simplify the dataset by reducing multidimensional dataset to lower dimensions, while keeping as much as possible the variation present in the original dataset. Martens applied PCA to the problems of modelling of HRTFs [13]. Kistler and Wightman showed that the magnitude of HRTF in dB scale can be adequately approximated by a linear combination of five basis spectral shapes [14]. These previous works focus on the magnitude response of HRTF in the frequency domain. We, however, apply PCA for modelling of HRIRs and customization in the time domain.

1.1 Post-processing of HRIRs

Before PCA, we did post-process the median 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 depict in Fig. 1. The initial time delay indicates the propagation time of sound from a source to a listener's eardrum, and this can be re-inserted afterwards if it is needed. The response of 1.5 msec includes the effects of pinna, head, shoulder, and torso. Thus, we can reduce the size of dataset to be analyzed in PCA without loss of meaningful information by the post-processing.

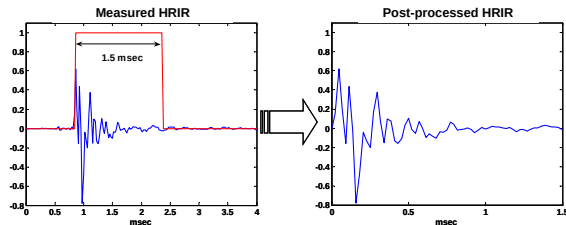


Figure 1: Post-processing of HRIR

1.2 PCA of median HRIRs

The first step in PCA is to make a matrix composed of directional impulse responses (DIRs). The original data matrix, $\mathbf{X}$ ($N \times M$), is composed of the post-processed median HRIRs. The each column of $\mathbf{X}$, $\mathbf{x}_i$ ($i=1,2,\dots,M$), indicates the post-processed HRIR. The response of 1.5 msec corresponds to 67 samples (sampling frequency: 44.1 kHz) and the number of median HRIRs at each ear is 2205 (45 subject $\times$ 49 elevations from -45° to 225° at 5.625° intervals). Thus, the dimension of $\mathbf{X}$ is 67×2205 in our case. The empirical mean of $\mathbf{X}$ is needed to obtain DIRs, and the empirical mean vector ($\mathbf{u}$: $N \times 1$) is given by

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

The mean-subtracted data matrix is the DIR matrix ($\mathbf{B}$).

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

where, $\mathbf{h}$ is a $1 \times M$ row vector of all 1's. Actually, DIR is defined as sound pressure caused by the wave from a specific direction when a sound source radiates an impulse signal, however, in this study the mean-subtracted HRIR is called DIR for convenience because the mean-subtracted HRTF is called directional transfer function (DTF) [14].

The next step is to compute a covariance matrix, $\mathbf{C}$.

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

where, $\otimes$ and $*$ indicate the outer product and the conjugate transpose operators, respectively. The basis functions (or basis vectors), $\mathbf{v}_i$ ($i=1,2,\dots,q$), are the q eigenvectors of the covariance matrix, $\mathbf{C}$, corresponding to the q largest eigenvalues. These basis functions are called "Principal Components (PCs)". If $q=N$, then the DIRs can be fully reconstructed by a linear combination of the q PCs. However, in many practical applications, q is smaller than N because the goal of PCA is to reduce the dimension of dataset. Thus, we can obtain an estimate of the original dataset only by using the q ($\ll N$) PCs. The weights of PCs (PCWs), $\mathbf{W}$, can be obtain as

$$\mathbf{W} = \mathbf{V}^* \cdot \mathbf{B}, \quad \mathbf{V} = [\mathbf{v}_1 \mathbf{v}_2 \dots \mathbf{v}_q] \quad (4)$$

PCWs represent the contribution of each basis function to the DIRs. The matrix composed of the estimates of HRIRs, $\tilde{\mathbf{X}}$ is given by

$$\tilde{\mathbf{X}} = \mathbf{V} \cdot \mathbf{W} + \mathbf{u} \cdot \mathbf{h} \quad (5)$$

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

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

where, subscript F indicates the matrix Frobenius norm. Previous studies decided the number of PCs based on the variance in the mean-subtracted data in the frequency domain or in the time domain (see details in [14,11]), however, we determine the number of PCs with considering of the empirical mean because we want to reconstruct not DIRs but HRIRs. The more PCs are used, the more accurately HRIRs can be reconstructed. Fig. 2 shows the reconstruction error with respect to the number of PCs. The case of zero for the number of PCs indicates that $\tilde{X}$ is composed of the empirical mean only. As the number of PCs increases, the reconstruction error decays exponentially.

We arbitrary set the reconstruction error bound of 5%, and we retain 12 PCs. Fig. 3 shows the empirical mean and 12 PCs obtained from PCA of the left ear HRIRs in the median plane (2205 HRIRs). PC1, PC2, PC3, and PC4 have higher energy in the pinna response up to 0.2 msec, whereas PC9 and PC10 have higher energy in the shoulder/torso 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 effect. Fig. 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 up to 0.2 msec can be well reconstructed, however, the shoulder/torso response after 0.2 msec cannot be reproduced because these PCs have less energy in the shoulder/torso response. The shoulder/torso response can be reconstructed by PC9 and PC10. Of course, other PCs also contribute to the reconstruction, it can be said that PC9 and PC10 are dominant components to recover the shoulder/torso effect.

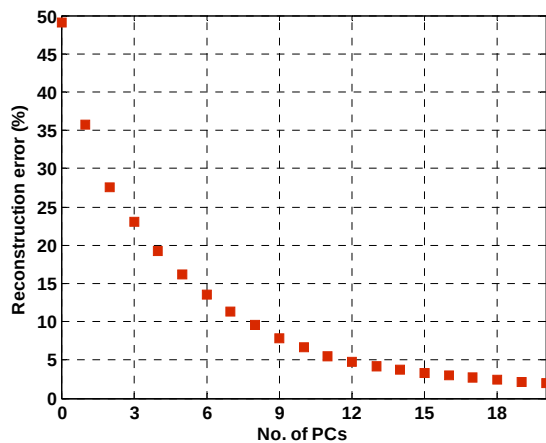


Figure 2: Reconstruction error with respect to the number of PCs.

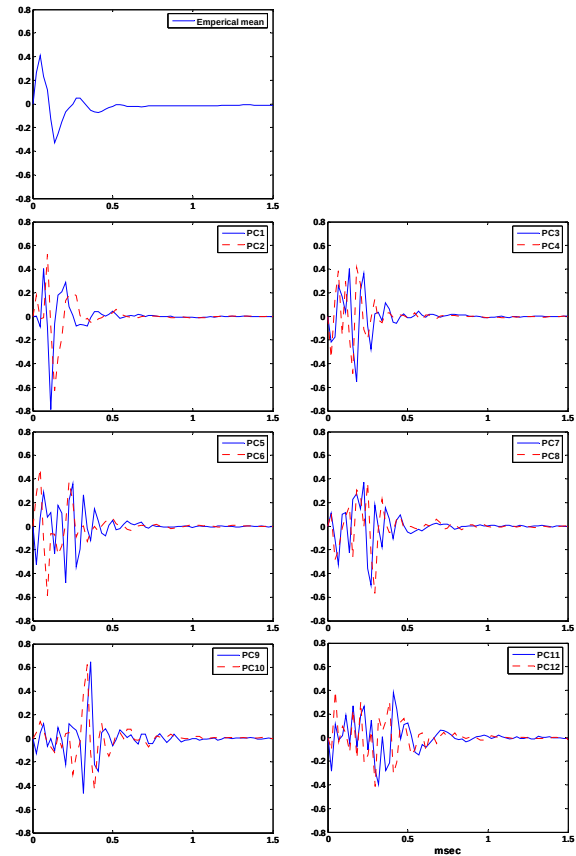


Figure 3: The empirical mean and PCs obtained from PCA of the left ear HRIRs in the median plane.

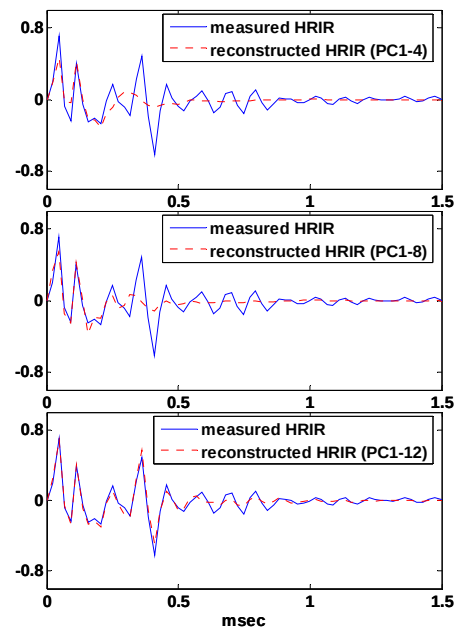


Figure 4: HRIR reconstruction process for subject 152 in the CIPIC HRTF database (-33.75° elevation in the median plane).

PCWs also should be investigated because they represent the contribution of each PC in the reconstructed HRIR. Fig. 5 shows the mean value and standard deviation of each PCW for all left ear HRIRs (45 subjects) from sources in the median plane.

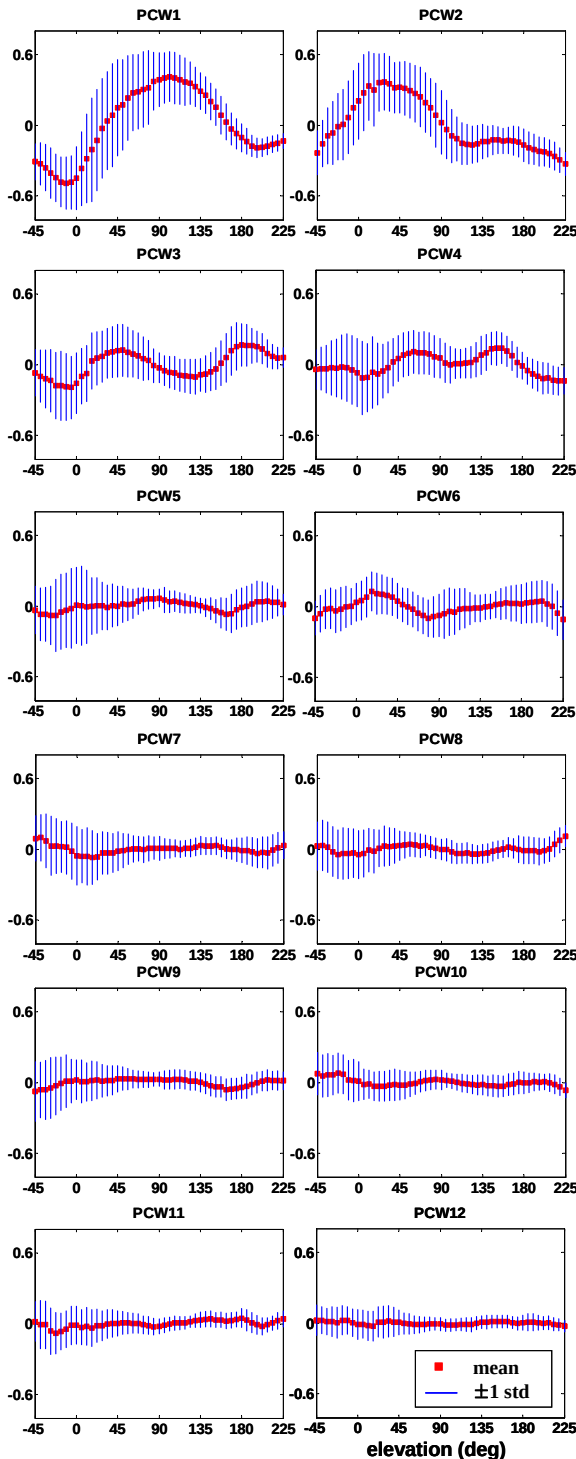


Figure 5: Mean and ± 1 standard deviation of each PCW for all left ear HRIRs (45 subjects) from sources in the median plane.

There are some notable tendencies of PCWs, and each PCW provide a useful sound cue for front-back distinction and vertical perception. For example, PCW1 have positive mean value from about 30° to 150° of elevation and have negative value for other elevations. Furthermore, it increases monotonically from 0° to 90° and decreases monotonically from 90° to 180° . Thus, we can conclude that PC1 contribute to the vertical perception. PCW2 have positive mean value in the frontal region except for low sources and have 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 distinction. PCW7-12 have almost zero mean from sources in the median plane. Thus, we can say that these PCs cover the inter-individual variation only, whereas the others having some tendency of mean value with respect to the source elevation cover not only the inter-elevation variation but also the inter-individual variation. The standard deviation of each PCW also provides 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 in the frontal region than in the rear region, thus the inter-individual variation of PC1 is large for the frontal sources. From the standard deviation of PCW5, it can be said that the inter-individual variation of PC5 decreases as source elevation is higher, and PC9 shows larger inter-individual variation for the frontal low sources. In Fig. 3, 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 low sources.

2 HRIR CUSTOMIZATION

2.1 Dominant PCs

As mentioned above, 12 PCs cover 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 PCWs to be tuned should be reduced. At each elevation, we can arrange PCs with respect to the magnitude of standard deviation as depict 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. The number of DPCs to be tuned depends on each subject. Of course, more DPCs are tuned, more accurately HRIRs can be customized with more complex tuning process.

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

Table 1: Order of PCs and standard deviations (below each PC) at each elevation in the median plane.

2.2 Subjective tuning of weights on 3 DPCs

In this study, we choose that the number of DPCs to be tuned is 3 at each elevation. Thus, letting a subject tune the weights of 3 DPCs only bring a customized HRIR at each elevation. 2 male subjects, 23 and 27 years of age, with normal hearing sensitivity participated in making customized HRIRs. They were experienced in the headphone-based localization test. Customization process is based on the MATLAB™ GUI as depict in Fig. 6.



Figure 6: MATLAB™ GUI for customization.

Customization was conducted in a office room having 45 dB of background noise (A-weighted). The room size was 7 m × 7 m × 4 m. 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 carried out at 9 specific elevation angles from -30° to 210° at 30° intervals in the median

plane. We substitute the mean values of weights 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 at each source position in the median plane. Thus, the left and right channels of a headphone (Sennheiser HD 250 linear II) were driven by the same signal. Actually, the left and right ear HRIRs are slightly different even in the median plane, however, Morimoto showed that the elevation angles in the median plane are correctly perceived although either of two ears is occluded [15]. Therefore, our assumption doesn't affect the elevation perception. Fig. 7 shows the customized and individual HRIRs for a representative subject (details for the measurement of individual HRIRs are drawn in Section 3).

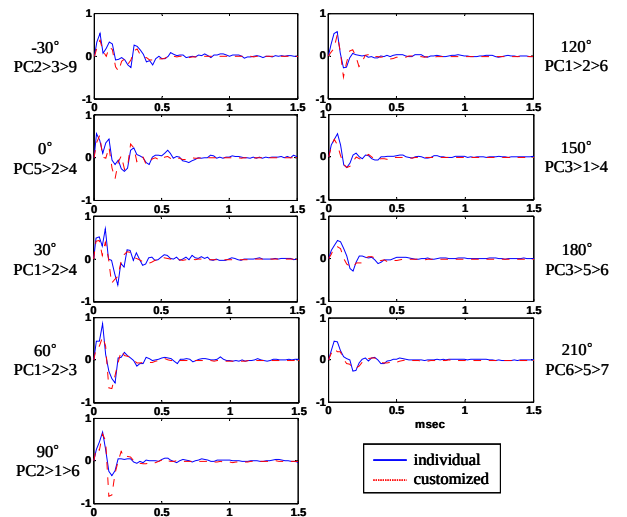


Figure 7: Customized and individual HRIRs for subject CH.

The angle at the left or right of each panel indicates the source elevation, and PCs below the each angle denote the 3 DPCs used for customization. Although there exist slight discrepancies between the customized and individual HRIRs, the pinna response up to 0.2 msec and the shoulder/torso response can be well reproduced by customization.

3 INDIVIDUAL HRIR MEASUREMENT

For comparison of the localization performance, individual HRIRs of the two male participants were measured at the elevation angle where customization took place (9 elevation angles). Experimental setup is depicted in Fig. 8.



Figure 8: Experimental setup for the measurement of individual HRIRs in an anechoic chamber.

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 duration of 1.5 seconds at a sampling rate of 44.1 kHz. The individual HRTF was computed from a 1024-point FFT with a Hanning window and a 50 % overlap, and the individual HRIR was obtained from IFFT of HRTF.

4 SUBJECTIVE EVALUATION

Subjective listening tests using a pair of headphone (Sennheiser HD 250 linear II) were performed on the two subjects to evaluate the localization performance of 3 HRIR set including the non-individualized HRIRs (Kemar HRIRs in the CIPIC HRTF database), the individual HRIRs, and the customized HRIRs. For convenient test procedure, another MATLABTM GUI was used for subjective listening test as depicted in Fig. 9. For correct headphone-presented stimulation of free-field listening when evaluating the individual HRIRs on their localization capabilities, the headphone-to-ear-canal

transfer function (HpTF) was cancelled according to the method suggested by Wightman and Kistler [4,16].

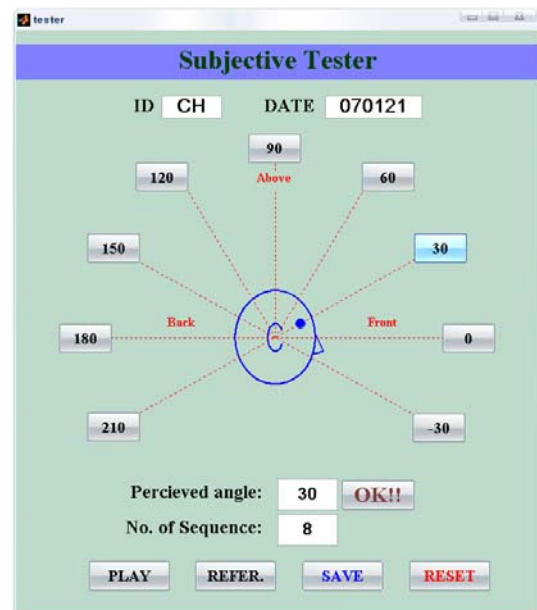


Figure 9: MATLABTM GUI for subjective listening test.

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 angles is stimulated 10 times in a random order yielding in total 90 stimuli. After listen each stimulus by pushing the "PLAY" button, subject 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.

4.1 Subjective listening test results

Figs. 10-12 show the subjective listening test results by the two subjects. The horizontal axis and the vertical axis denote the actual source elevation angle and the perceived elevation angle, respectively, in each panel. The size of circle is directly proportional to the response frequency. The positive-sloped diagonal line indicates the perfect elevation perception and front-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 front-back reversal condition in which the subject reported the 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 non-individualized HRIRs (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

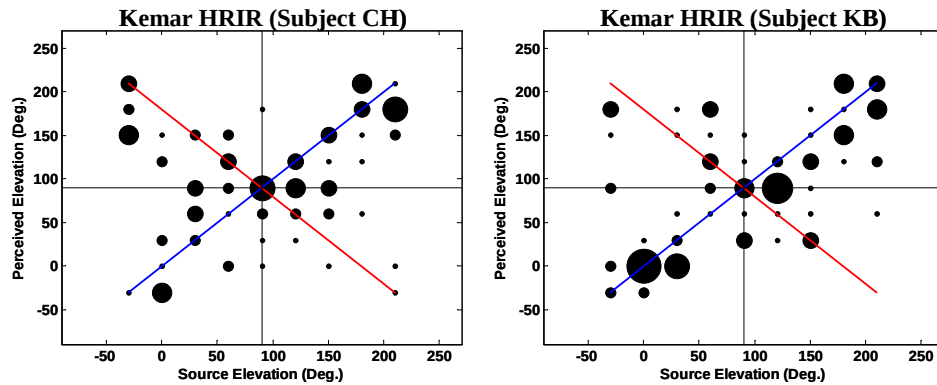


Figure 10: Subjective listening test results on the non-individualized HRIRs (Kemar HRIRs).

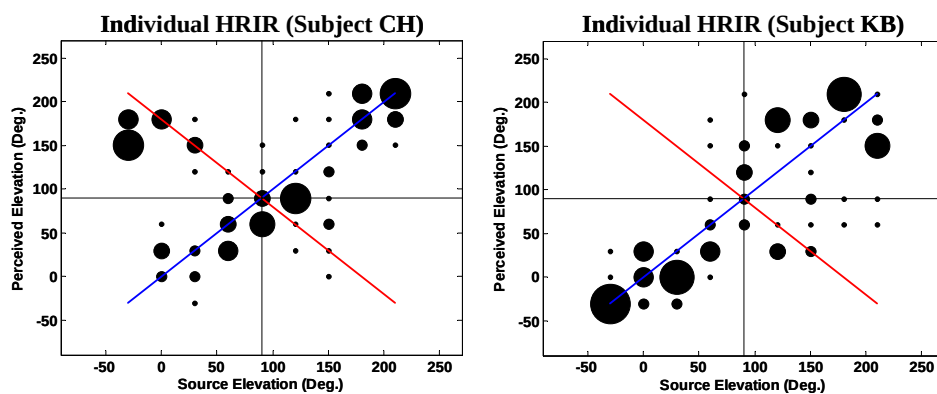


Figure 11: Subjective listening test results on the individual HRIRs.

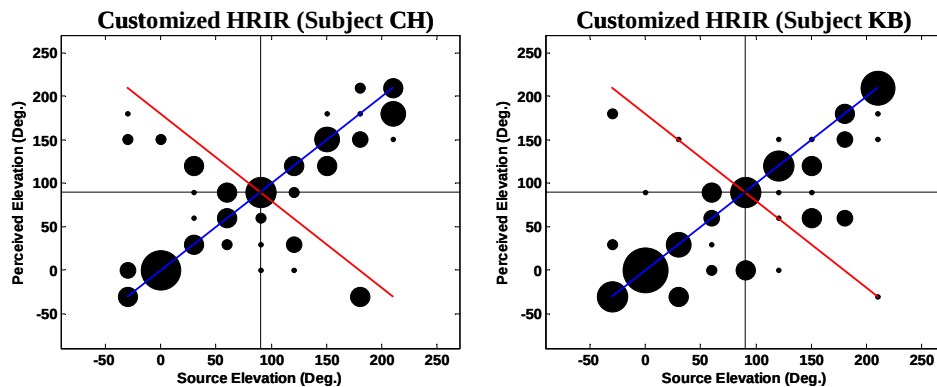


Figure 12: Subjective listening test results on the customized HRIRs.

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 very 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.

The localization performance with the individual HRIRs in this study is less accurate than the previous works conducted by Wightman and Kistler and Iida et al. [4,5]. We think that this is mainly due to the HpTF. The correction of HpTF is very closely related with the localization performance, and it is known that open-air headphones are more effective for binaural synthesis than closed-air headphones [17]. Wightman and Kistler

and Iida et al. also used open-air headphones. However, in this study, we used a closed-air headphone because it cannot be guaranteed that every commercial users use open-air headphones. Furthermore, Kulkarni and Colburn have claimed that making appropriate corrections will be difficult, as the variability of HpTF is also high across headphone placements for a given listener [18]. Thus, to achieve an accurate correction, the HpTF should be measured and cancelled whenever a listener wears a headphone. However, we measured the HpTF for each subject just one time in an anechoic chamber. The inaccurate correction of HpTF can be compensated in the customization process, and it may be the reason why the localization performance with the customized HRIRs is better than that with the individual HRIRs.

4.2 Quantitative error analysis

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 diagonal line than the positive-sloped diagonal 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 diagonal line. Of course, if the response is more close to the positive-sloped diagonal line than the negative-sloped diagonal line, the front-back reversal is not occurred and the vertical perception error is the angular difference between the response and the positive-sloped diagonal line. Thus, we defined two kinds of errors, the front-back reversal error (e_{FBR}) and the vertical perception error (e_{VP}), as

$$e_{FBR} = \frac{\text{No. of responses more close to the negative-sloped diagonal line}}{\text{No. of total responses}} \quad (7)$$

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

where, P and A indicate the perceived elevation angle and the actual source elevation angle, respectively, and A and $(180^\circ - A)$ correspond to the position on the positive-sloped and negative-sloped diagonal 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 front-back distinction and elevation perception with the customized HRIRs except for front-back reversal of subject KB.

HRIR set	Subject CH		Subject KB	
	e_{FBC}	e_{VP}	e_{FBC}	e_{VP}
Kemar	0.311	27.3°	0.222	27.0°
Individual	0.298	23.7°	0.122	32.3°
Customized	0.189	19.3°	0.144	21.3°

Table 2: Quantitative error analysis of subjective listening test results.

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 cover 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.

6 ACKNOWLEDGMENT

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

REFERENCES

- [1] J. Blauert, *Spatial hearing*, MIT, Cambridge, MA (1983).
- [2] D. S. Brungart and W. M. Rabinowitz, "Auditory localization of nearby sources. Head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 106, no. 3, pp. 1465-1479 (1999).
- [3] C. I. Cheng and G. H. Wakerfield, "Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space," *J. Audio Eng. Soc.*, vol. 49, no. 4, pp. 231-248 (2001).
- [4] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. I: Stimulus synthesis," *J. Acoust. Soc. Am.*, vol. 85, no. 2, pp. 858-867 (1989).
- [5] K. Iida, M. Itoh, A. Itagaki, and M. Morimoto, "Median plane localization using parametric model of the head-related transfer function based on spectral cues," *Applied Acoustics*, In press (2006).

- [6] S. Shimada, M. Hayashi, and S. Hayashi, "A clustering method for sound localization transfer functions," *J. Audio Eng. Soc.*, vol. 42, no. 7/8, pp. 577-584 (1994).
- [7] J. C. Middlebrooks, "Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency," *J. Acoust. Soc. Am.*, vol. 106, no. 3, pp. 1493-1510 (1999).
- [8] V. R. Algazi, R. O. Duda, R. P. Morrison, and D. M. Thompson, "Structural composition and decomposition of HRTFs," *Proc. WASPAA01, New Paltz, NY*, pp. 103-106 (2001).
- [9] D. N. Zotkin, R. Duraiswami, and L. S. Davis, "Customizable auditory display," *Proc. Int. Conf. on Auditory Display (ICAD)*, Kyoto, Japan (2002).
- [10] CIPIC HRTF Database Files, Released 1.1, August 21, CIPIC Interface Laboratory, U. C. Davis, <http://interface.cipic.ucdavis.edu/> (2001).
- [11] Ki-hoon Shin and Youngjin Park, "Customization of head-related transfer functions using principal components analysis in the time domain," *JASA 4th Joint meeting of ASA/ASJ* (2006).
- [12] G. H. Duntleman, *PRINCIPAL COMPONENTS ANALYSIS*, Sage Publication, Inc (1989).
- [13] W. L. Martens, "Principal components analysis and resynthesis of spectral cues to perceive direction," *Proc. Int. Computer Music Conf.*, pp. 274-281 (1987).
- [14] D. J. Kistler and F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *J. Acoust. Soc. Am.*, vol. 91, no. 3, pp. 1637-1647 (1992).
- [15] M. Morimoto, "The contribution of two ears to the perception of vertical angle in sagittal planes," *J. Acoust. Soc. Am.*, vol. 109, no. 4, pp. 1596-1603 (2001).
- [16] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. II: Psychophysical validation," *J. Acoust. Soc. Am.*, vol. 85, no. 2, pp. 868-878 (1989).
- [17] H. Møller, "Fundamentals of Binaural Technology," *Applied Acoustics*, vol. 36, issue 3-4, pp. 171-218 (1992).
- [18] A. Kulkarni and H. S. Colburn, "Variability in the characterization of the headphone transfer-function," *J. Acoust. Soc. Am.*, vol. 107, no. 2, pp. 1071-1074 (2000).