

COVER PAGE

Document downloaded by @DAEL

Thu Nov 12 23:40:29 2026
For personal use

When automatic English translation is provided, only the original document is authentic.
The EAA cannot be held responsible of any translation error

Bibliographical reference

Customization of Spatially Continuous Head-Related Impulse Responses in the Median Plane, Sungmok Hwang, Youngjin Park and Youn-Sik Park, Acta Acustica **vol. 96** (Number 2), 2010, pp. 351-363

DOI
https://doi.org/10.3813/AAA.918284

ACTA ACUSTICA UNITED WITH ACUSTICA
Vol. 96 (2010) 351 – 363
DOI 10.3813/AAA.918284

Customization of Spatially Continuous Head-Related Impulse Responses in the Median Plane[0]

Sungmok Hwang[1)], Youngjin Park[2)], Youn-sik Park[2)]
[1)] Marine Research Institute, Samsung Heavy Industries co., ltd. Geoje Shipyard, Sinhyeon-Eup, Geoje-Si, 656-710, Republic of Korea. sungmok.Hwang@samsung.com,
[2)] Center for Noise and Vibration Control, KAIST, Science Town, Daejeon 305-701, Republic of Korea. yjpark@kaist.ac.kr, yspark@kaist.ac.kr

Summary
The authors already have proposed a novel method of HRIR customization for synthesizing stationary sounds in the median plane based on subjective tuning of three parameters at each static source position. For synthesizing both stationary and dynamic sounds, this study extends the previous study to customize spatially continuous HRIRs in the median plane. The median plane in the upper hemisphere was systematically divided into two sectors, and three parameters were tuned at three positions (0°, 70°, and 180°), which are the endpoints of the two sectors, with listening to moving sounds synthesized by filtering the obtained spatially continuous HRIRs. Thus, the entire median-plane HRIRs in the upper hemisphere can be customized by subjective tuning of nine parameters only. Subjective localization listening tests were carried out to assess the customization performance using both stationary and moving sounds. When the stationary sounds were presented to nine subjects, the localization performance with the customized HRIRs was comparable to that with the individual HRIRs, but it was significantly better than that with the Kemar HRIRs. When the moving sounds were presented to four subjects, all subjects perceived the center position of the target trajectory more accurately with the customized HRIRs than with the Kemar HRIRs. The error with the customized HRIRs was comparable to that with the individual HRIRs. Therefore, it can be concluded that to make desired perception of both stationary and moving sounds is possible by the proposed customization methods.

PACS no. 43.66.Ba, 43.66.Pn, 43.66.Qp

1. Introduction

The Virtual Auditory Display (VAD), which is defined as systems or technologies generating spatialized virtual sounds and conveying them to a listener [1, 2, 3], is an important and state-of-the-art technology in many application fields [4, 5, 6, 7, 8, 9]. The key technology of VAD is the reproduction of sound image in 3-dimensional space by filtering the sound source with the head-related transfer functions (HRTFs), which describe the physical transformation of sound waves due to physical structures of a listener [10]. The HRTF depends on the direction of a sound source, and it also varies with subjects. In other words, the HRTF is a function of both the sound direction and the subject, and the HRTF contains information about the direction of a sound source and the listener. Thus, the HRTFs or Head-Related Impulse Responses (HRIRs), which are the time domain counterparts of the HRTFs, play a key role for the rendering of high-fidelity VAD.

Most VAD systems generally based on non-individualized HRTFs measured from a dummy head microphone system because it is practically impossible to measure individual HRTFs for every listener due to the requirements of heavy and expensive equipments as well as long measurement time. However, many previous reports showed that non-individualized HRTFs may cause high error rate in sound source localization because of the mismatching between the non-individualized HRTFs and the listener's ones [11, 12, 13, 14, 15]. Therefore, it is necessary to find a HRTF customization method that provides a listener with proper sound cues without direct measurement of individual HRTFs.

A novel method to customize HRIR at each static source position based on subjective tuning of the general basis functions, which are principal components (PCs) extracted from Principal Components Analysis of median-plane HRIRs in the CIPIC HRTF database [16], have already been proposed by the authors [17]. At each static source position, customized HRIR was obtained by letting

Received 31 August 2009,
accepted 21 November 2009.

[0] Portions of this work were presented at the 10th Western Pacific Acoustic Conference in September 2009.

© S. Hirzel Verlag · EAA

Document downloaded by @DAEL, 2026-11-12 23:40:31 - personal use only

a subject tune only three tuning parameters, i.e. weights of PCs (PCWs). In the results of preliminary subjective listening tests by three subjects with normal hearing sensitivity, all subjects reported dramatically improved performances for the vertical perception and the front-back discrimination with the customized HRIRs compared to those with the non-individualized HRIRs.

In many practical applications, however, a moving source in the trajectory of interest is considered. In this case, the customized HRIRs corresponding to all source positions on the trajectory of interest must be obtained. In other words, customization of spatially continuous HRIRs in the trajectory of interest is needed. Thus, this study expands the previous study to customize spatially continuous HRIRs in the trajectory simultaneously by tuning of a few parameters. Section 2 describes the general methodology. The localization performance of the proposed method is evaluated by a series of subjective listening tests using a pair of headphones, and the evaluation results are included in section 3.

The only median-plane HRIRs are dealt with in this study because the sagittal plane can be mainly manipulated by introducing proper interaural time difference to the median-plane HRIRs [18, 19]. In the median plane, the 0° is ahead of the listener, the 90° is above, and the 180° is behind.

2. Methodology

2.1. Data preparation

Principal Components Analysis (PCA) is one of the statistical techniques that can be used to provide an efficient representation of correlated data set. PCA have been used to model the HRTFs or HRIRs [17, 19, 20, 21, 22]. This study also performed PCA of the median-plane HRIRs in the CIPIC HRTF database.

Before PCA was performed, a pre-processing on the median-plane HRIRs was carried out to align the HRIRs in time and to extract the early response that lasts for 1.5 msec since the arrival of direct pulse. Basic assumption used in this study for time-alignment of HRIRs is that the time delay between the HRIRs at two closely adjacent source positions must be zero in the median plane. Although this assumption is not valid in the horizontal plane, it is reasonable in the median plane because the distance from a source to head center is constant in the median plane. The time delay between the median-plane HRIRs corresponding to two sparsely located source positions can be non-zero because of the offset of eardrum position from the head center, but this might be negligible if a source position is sufficiently close to another source position. Therefore, the HRIRs of each subject in the CIPIC HRTF database were aligned in time by compensating the time delay between two adjacent HRIRs with 5.625° intervals. Before time-alignment of HRIRs, the HRIRs were upsampled by a factor of 10 to enhance the time resolution. The time delay was estimated by the time lag argument that maximizes the cross-correlation function between the two upsampled HRIRs. After time-alignment of HRIRs, the HRIRs were downsampled again by a factor of 10. The pre-processing on HRIRs deals with the time-alignment of HRIRs of a single subject, i.e. inter-elevation alignment. However, time-alignment of HRIRs of different subjects, i.e. inter-subject alignment, is also needed[1]. In this study, inter-subject alignment was carried out by compensating the time delay between the HRIRs of different subjects at 90° of elevation. The inter-subject time delay was estimated from the maximum peak positions in the two HRIRs. The reason why the HRIRs at 90° of elevation was used for inter-subject alignment is that the HRIR at 90° of elevation is very simple and single dominant peak exists in the HRIRs. After time-alignment and inter-subject alignment of HRIRs, the arrival of direct pulse or the propagation time delay in the HRIR was found by searching the time corresponding to 20% of the maximum peak value of the HRIR. The propagation time delay indicates the propagation time of sound from a source to a listener's eardrum, and this can be reinserted later if needed.

PCA was carried out using the median-plane HRIRs in the CIPIC HRTF database after the pre-processing. Overall procedure of PCA was the same with the one described in the authors' previous study [17]. The twelve Principal Components (PCs) were extracted from PCA. The pre-processed HRIRs can be reconstructed by a weighted linear summation of PCs, and this can be expressed as

$$\boldsymbol{h} \simeq \sum_{k=1}^{12} w_k \boldsymbol{v}_k + \boldsymbol{u}, \tag{1}$$

where $\boldsymbol{h}$ is the pre-processed HRIR, w is a weight of PC (PCW), and $\boldsymbol{v}$ and $\boldsymbol{u}$ are PC and the empirical mean, respectively. Physical meaning and contributions to inter-subject and inter-elevation variations of PCs and PCWs can be interpreted by the similar way described in [22], but this is not included in this manuscript for brevity.

2.2. Selection of tuning parameters

In equation (1), the HRIRs can be modeled by a weighted linear summation of the twelve PCs, thus customization of HRIRs using the PCs is turned into the proper selection of each PCW for a specific subject to provide sound cues for the vertical perception and the front-back discrimination. In the author's previous study [17], the selection of PCWs at each static source position completely depends on each subject, i.e. customization is carried out by letting a subject tune the weight on each PC brings an ac tual change in HRIRs and listen to a broadband stationary sound filtered with the resulting customized HRIRs. In this study, the same approach is applied to customize the

[1] The inter-subject alignment of HRIRs is for compensation of the measurement error in the CIPIC HRTF database. Based on the analysis of maximum peak position in the HRIRs at 90° of elevation, maximum inter-subject difference in head centre is approximately 10 cm. This is because that a device to fix head position was not used when the HRIRs were measured by the U.C. Davis.

Document downloaded by @DAEL, 2026-11-12 23:40:31 - personal use only

spatially continuous HRIRs in the median plane. In this case, all PCW values in the sector of interest in the median plane are tuned simultaneously. In other words, the HRIRs in the sector of interest can be customized simultaneously not by tuning of PCW at static point but by tuning of PCW curve in the sector of interest. However, tuning the 12 PCW curves is practically impossible because it is very exhausting and time consuming task, thus the number of tuning parameters should be systematically reduced as much as possible.

To select the tuning parameters systematically, the relationship between PCWs are considered in this study. If $(12 - L)$ PCWs can be effectively estimated by a linear summation of L PCWs and the relationship between L PCWs and other $(12 - L)$ PCWs is known or systematically estimated, then the customization can be achieved by tuning of the L PCWs only. The relationship can be estimated from the multiple regression analysis using the CIPIC HRTF database. Assume that PCWs are given by PCA of the 2205 median-plane HRIRs (45 subjects × 49 elevations) in the CIPIC HRTF database. $\boldsymbol{W}_{in}$ $(\in \mathbf{R}^{45N\times L})$ is a matrix composed of the L PCWs of 45 subjects as

$$\boldsymbol{W}_{in} = \begin{bmatrix} \boldsymbol{w}_{1,1} & \boldsymbol{w}_{1,2} & \cdots & \boldsymbol{w}_{1,L} \\ \boldsymbol{w}_{2,1} & \boldsymbol{w}_{2,2} & \cdots & \boldsymbol{w}_{2,L} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{w}_{45,1} & \boldsymbol{w}_{45,2} & \cdots & \boldsymbol{w}_{45,L} \end{bmatrix}, \tag{2}$$

where $\boldsymbol{w}_{i,k}$ $(\in \mathbf{R}^{N\times 1})$ is the kth PCW vector for the ith subject. N is the number of elevations in the sector of interest and $\boldsymbol{W}_j$ $(\in \mathbf{R}^{45N\times 1})$ is a vector of the jth PCWs of 45 subjects as

$$\boldsymbol{W}_j = \begin{bmatrix} \boldsymbol{w}_{1,j} \\ \boldsymbol{w}_{2,j} \\ \vdots \\ \boldsymbol{w}_{45,j} \end{bmatrix} \qquad (j = L+1, \ldots, 12). \tag{3}$$

Then, $\boldsymbol{W}_j$ can be represented by a weighted linear summation of the L PCWs with estimation error, $\boldsymbol{E}_j$, by

$$\boldsymbol{W}_j = \boldsymbol{W}_{in}\boldsymbol{a}_j + \boldsymbol{E}_j, \tag{4}$$

where $\boldsymbol{a}_j = [a_{1,j}\ a_{2,j}\ \ldots\ a_{L,j}]$. $a_{k,j}$ is a weight for the kth PCW vector to reproduce the jth PCW. a_j is estimated to minimize $\boldsymbol{E}_j$ in the least-squares sense by

$$\begin{aligned} \frac{\partial \boldsymbol{E}_j^T \boldsymbol{E}_j}{\partial \boldsymbol{a}_j} &= \frac{\partial}{\partial \boldsymbol{a}_j}\left(\boldsymbol{W}_j^T\boldsymbol{W}_j - \boldsymbol{a}_j^T\boldsymbol{W}_{in}^T\boldsymbol{W}_j + \boldsymbol{a}_j^T\boldsymbol{W}_{in}^T\boldsymbol{W}_{in}\boldsymbol{a}_j\right) \\ &= 2\boldsymbol{W}_{in}^T\boldsymbol{W}_{in}\boldsymbol{a}_j - 2\boldsymbol{W}_{in}^T\boldsymbol{W}_j = 0. \end{aligned} \tag{5}$$

Equation (5) gives

$$\boldsymbol{a}_j = \left(\boldsymbol{W}_{in}^T\boldsymbol{W}_{in}\right)^{-1}\boldsymbol{W}_{in}^T\boldsymbol{W}_j. \tag{6}$$

From equation (6), the multiple regression model is obtained to estimate the jth PCWs using the L PCWs only.

If the number of tuning parameters, L, is given, there exist many combinations of L PCWs, i.e. ${}_{12}C_L$. Thus, to find the optimal combination of L PCWs is important. It is reasonable to select the combination of L PCWs for minimizing the HRIR reconstruction error in the least-squares sense. If L is given, the optimal combination of L PCWs is determined by minimizing the percentage modeling error, $\%error$, as

$$\%error = \frac{||\boldsymbol{X} - \hat{\boldsymbol{X}}_L||_F^2}{||\boldsymbol{X}||_F^2} \cdot 100\%, \tag{7}$$

where $\boldsymbol{X}$ is a matrix composed of the 2205 median-plane HRIRs and $\hat{\boldsymbol{X}}_L$ is its reconstructed version from L PCWs and estimated $(12 - L)$ PCWs. Subscript F indicates the matrix Frobenius norm. The optimal combination of PCWs may be different from sector to sector. For example, if $L = 5$ and the median plane from −45° to 225° of elevation is uniformly divided into three sectors (frontal sector: −45°–45°, above sector: 45°–135°, and rear sector: 135°–225°), the optimal combination is {PCW1, PCW2, PCW3, PCW4, PCW5}, {PCW1, PCW2, PCW3, PCW5, PCW6}, and {PCW1, PCW2, PCW3, PCW6, PCW7} in the frontal, above, and rear sectors, respectively. The more PCWs are used for customization, i.e. the larger L, the better customization performance can be achieved but the more tedious process with long tuning time is required to complete customization. In this study, the number of tuning PCWs is set to be 3 ($L = 3$), and in this case the optimal combination of PCWs is {PCW1, PCW2, PCW3} for all sectors.

2.3. Development of tuning method

The tuning parameters are selected in section 2.2, then the next step is to clarify how the parameters can be systematically tuned. Tuning of PCWs at all source positions in the sector of interest is very complex problem. In this study, we focus on the inter-subject variations in PCWs. Figure 1 shows distributions of PCWs 1–3 of subjects at left ear in the CIPIC HRTF database in the median plane from −45° to 225° of elevation. The bold line indicates the mean of PCWs (mPCW) of all subjects. It can be seen that PCWs of all subjects show similar tendency with mPCW which represents the common elevation dependency of PCW across all subjects. Difference between PCW of each subject and mPCW, i.e. mean-subtracted PCW, represents the inter-subject variation in PCW (ΔPCW) of each subject. Thus, PCW of each subject can be represented by

$$w = \overline{w} + \Delta w, \tag{8}$$

where w, $\overline{w}$ and Δw represent PCW, mPCW, and ΔPCW, respectively. Figure 2 shows distributions of ΔPCWs 1–3 of subjects at left ear in the CIPIC HRTF database in the median plane from −45° to 225° of elevation. Except for low elevations, ΔPCWs of all subjects vary with a similar trend as a simple function of elevation. Especially, ΔPCW varies smoothly according to the source elevation in the median plane. Therefore, ΔPCW, i.e. inter-subject variation in PCW, can be modeled as a simple function of elevation. In this study, Δw is simply modeled as a piecewise

Document downloaded by @DAEL, 2026-11-12 23:40:31 - personal use only

linear function of elevation. In other words, Δw between two source positions is approximated by

$$\Delta w \simeq f(x) = ax + b, \tag{9}$$

where f is a linear function of elevation, x. a and b are coefficients to be estimated. For example, assume that PCWs at two source positions at x_i and x_f of elevation are w_i and w_f, respectively. Inter-subject variations, Δw_i and Δw_f corresponding to x_i and x_f of elevation, can be easily obtained by subtracting mPCW from each PCW at x_i and x_f of elevation. Then, the coefficients of the linear function in equation (9) can be estimated by

$$a = \frac{(w_f - \overline{w_f}) - (w_i - \overline{w_i})}{x_f - x_i}$$

$$\text{and} \qquad b = \frac{x_f(w_i - \overline{w_i}) - x_i(w_f - \overline{w_f})}{x_f - x_i}. \tag{10}$$

From equations (8)–(10), ΔPCW at all elevations between x_i and x_f can be easily estimated by simple linear interpolation of ΔPCW. This means that the HRIRs at any source position in the sector of interest can be obtained if the coefficients are properly selected to provide sound cues for the vertical perception and the front-back discrimination. This approach is different from the simple linear interpolation of PCWs because the mean PCWs are nonlinear functions and the only ΔPCW is considered in this study. In this case, interpolation performance is better than conventional linear interpolation of PCWs [23]. Based on this approach, tuning of PCW curves of the sector of interest is possible by tuning of ΔPCWs at two source positions only. The other PCWs are estimated from the PCWs 1–3 by the multiple regression model obtained in section 2.2.

The customization performance might be enhanced if the median plane is divided into many sectors and the PCW curves are tuned at each sector. However, the number of tuning sectors and the effort and time needed to complete customization are in the trade-off relationship. Thus, systematic determination of the tuning sectors is necessary. The simple method is to divide the median plane uniformly as like the frontal, above, and rear sectors described in section 2.2. In this study, however, we consider the perceptual sensitivity to PCWs (or HRIRs) to determine the tuning sectors in the median plane. In the author's previous study [24], Just Noticeable Difference (JND) in PCWs (or HRIRs) was investigated by a series of subjective listening tests using a pair of headphones, and it was shown that the common elevation-dependent tendency for JND in the mean-subtracted HRIRs (Directional Impulse Responses: DIRs) across all subjects exists. It can be said that there is no perceptual difference if the difference between the actual and estimated (or tuned) PCWs is within a bound, i.e. JND. Therefore, it is reasonable to select the sector in which PCWs, which are obtained from the tuned ΔPCWs at the endpoints of the sector of interest, are generally within the bound for all subjects. The bound is different not only from subject to subject but also from elevation to elevation. Based on the 85% of probability that

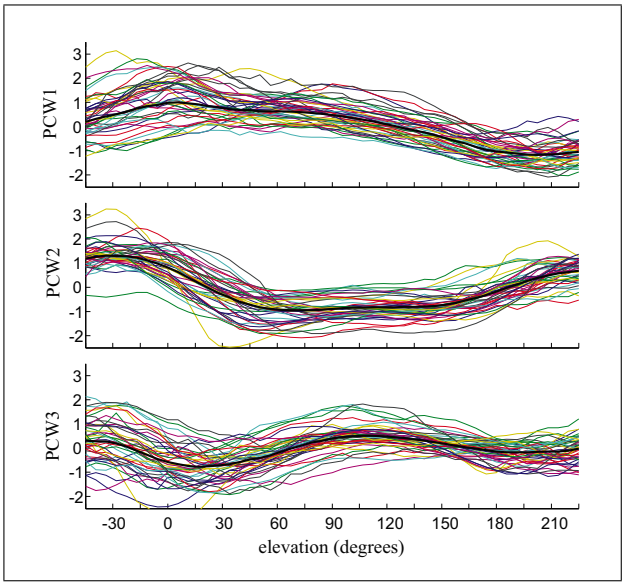


Figure 1. PCWs 1–3 of subjects at left ear in the CIPIC HRTF database and their mean (mPCW, bold line) in the median plane.

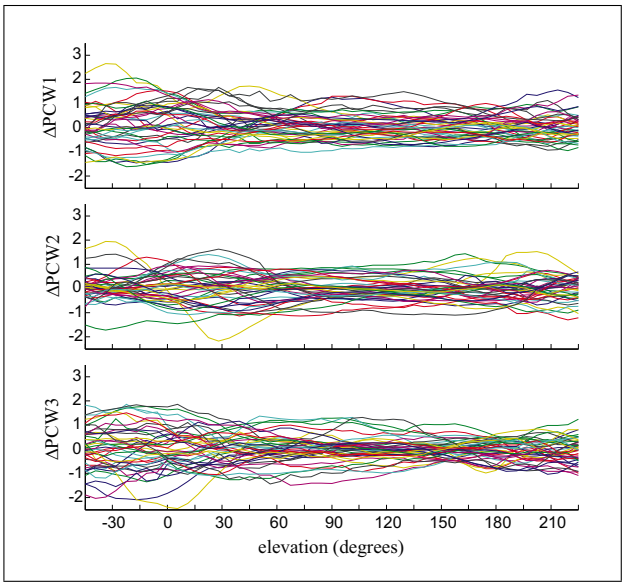


Figure 2. ΔPCWs 1–3 of subjects at left ear in the CIPIC HRTF database in the median plane.

the difference between the actual and modeled PCWs is within the bound on average across all subjects and all elevations, the median plane in the upper hemisphere is divided into two sectors, 0°–70° and 70°–180°, as depicted in Figure 3. As an example, Figure 4 shows PCW1 (bold solid line) of Subject 065 in the CIPIC HRTF database, its bound (light dashed line), mean of PCW1 of all subjects (light solid line), and the modeled PCWs (bold dash-dotted line) based on linear interpolation of inter-subject variation. This subject shows significant discrepancy between actual PCW1 and mPCW1, but the inter-subject variation can be sufficiently represented by the linear interpolation of the inter-subject variation. In the figure, the modeled PCW1 shows a good correspondence with the

Document downloaded by @DAEL, 2026-11-12 23:40:31 - personal use only

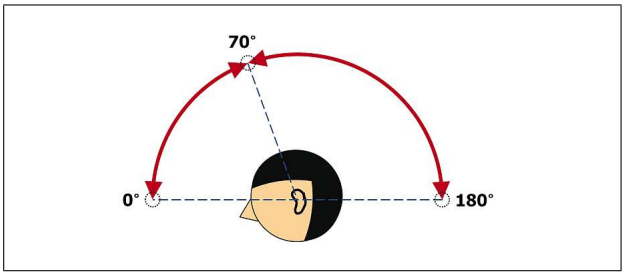


Figure 3. Selected tuning sectors for customization of spatially continuous median-plane HRIRs in the upper hemisphere.

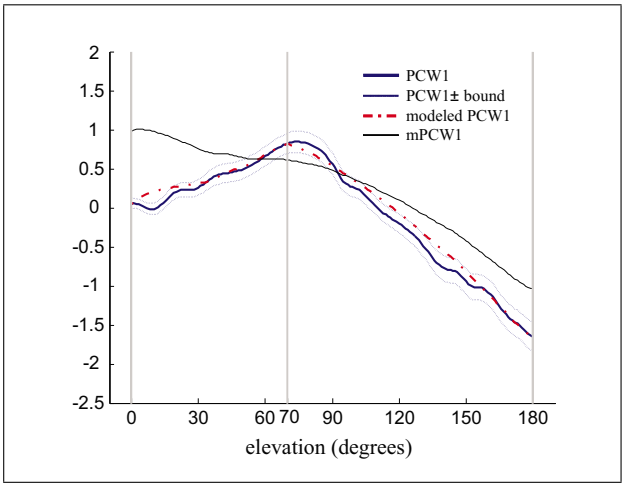


Figure 4. PCW1 (bold solid line) of Subject 065 in the CIPIC HRTF database, its bound (light dashed line), mean of PCW1 of all subjects (light solid line), and the modeled PCWs (bold dash-dotted line) based on linear interpolation of inter-subject variation.

actual one, and the difference between actual PCW1 and modeld PCW1 is generally within the bound. Thus, the median-plane PCWs in the upper hemisphere can be accurately modeled from linear interpolation ff ΔPCWs at three static positions (0°, 70°, and 180°) with 85% of accuracy on average across all subjects and all elevations.

2.4. Customization of spatially continuous HRIRs

Customization is carried out by letting a subject tune ΔPCWs 1–3 at each endpoint of the sector of interest (0°, 70°, and 180°) brings an actual change in HRIRs and listen to a broadband stationary and moving sounds filtered with the resulting customized HRIRs. In this study, ear symmetry is assumed although the HRIRs at left and right ears are slightly different even in the median plane. As a result, customization is performed by tuning ΔPCWs on only one ear and the left and right channels of the headphone are driven by the same signal. The customization process is based on the MATLAB™ GUI as depicted in Figure 5. Sectors in the GUI are divided by boxes and labeled per function in the figure. In the registration sector, subjects input his ID and date, and balance control adjusts gains to be applied to the left and right channels. In the tuning sector, the slider on each slide-bar represents each ΔPCW. The max-

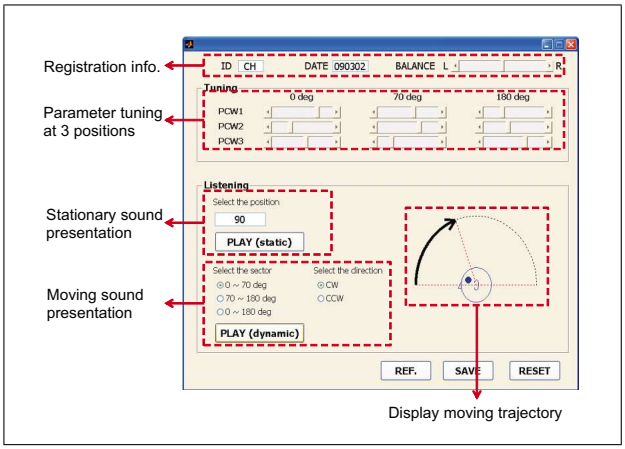


Figure 5. MATLAB™ GUI for HRIR customization.

imum and minimum bounds of each slide-bar are set to be ±2 standard deviations of each ΔPCW for all subjects in the CIPIC HRTF database. Subjects tune three parameters using the slider bars at three static source positions, 0°, 70°, and 180° of elevation. PCWs 1–3 is computed from mPCWs 1–3 and the tuned inter-subject variations, whereas the other PCWs are estimated by multiple regression model obtained in section 2.2. Then, the new median-plane HRIRs are created by weighted linear combination of the twelve PCs. The HRIRs are automatically updated whenever subject tunes any slider in the tuning sector. In the listening sector, subjects listen to broadband stimuli (1 kHz– 18 kHz) filtered by the newly created HRIRs, i.e. the customized HRIRs, by pushing the "Play" button. The filtered stimuli are generated by the MATLAB™ program, and a pair of open-air headphones (AKG K1000) is driven by them through the computer sound card (Creative SB X-Fi Elite Pro) and the audio amplifier (Audio Analog Verdi Settanta). The headphone to ear-carnal transfer function (HpTF) is very closely related to the localization performance and the inter-individual variability of the HpTF is significant and individualized equalization of the HpTF is recommended [25, 26, 27, 28]. Thus, it was measured and equalized whenever each listener wore the headphone. Detailed procedure is explained in the authors' previous paper [17]. Both stationary and moving sounds can be presented. A subject may select any elevation from 0° to 180° in the median plane, and listens to stationary sound at the selected static position. For listening to moving sound, a subject selects the sector and direction of moving source. The selected trajectory of moving sound is displayed in the GUI. The "Reference" button is for listening to the white noise which is not filtered by any HRIRs. This stimulus plays a role for subject to remove the memory of previous stimulus and to rouse a subject's attention to the next stimulus. After a subject finds that the vertical perceptions produced by the customized HRIR is good enough, the customized HRIRs are saved by pushing the "Save" button.

Nine male subjects (Subject ID: CH, CY, HS, KB, KD, KY, LS, LY, PJ) with normal hearing sensitivity participated in making their customized HRIRs. Their median-

Document downloaded by @DAEL, 2026-11-12 23:40:31 - personal use only

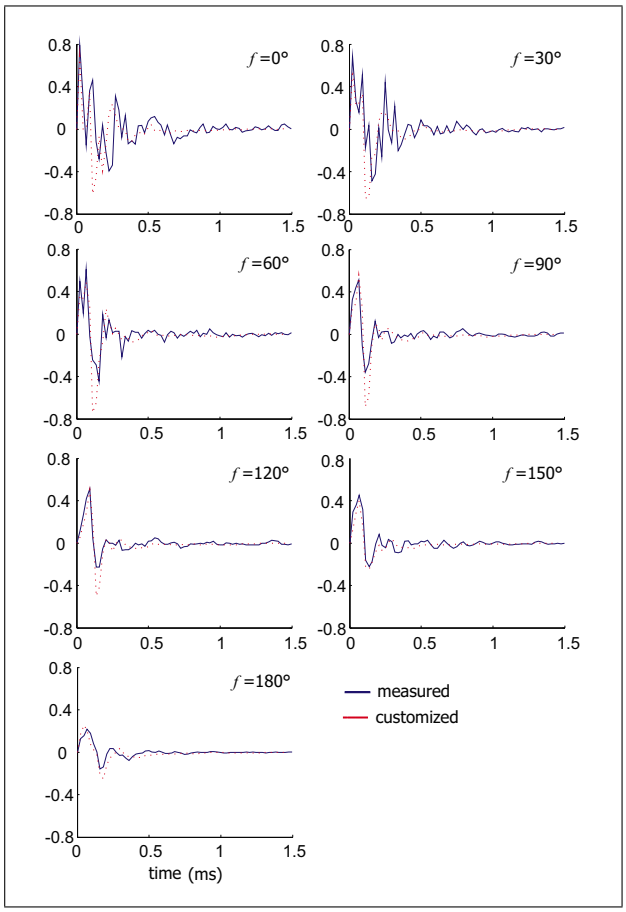


Figure 6. Individual (solid line) and customized (dashed line) HRIRs of Subject CH at left ear in the median plane.

plane HRIRs were measured in an anechoic chamber to compare the individual and customized HRIRs. Detailed description on the apparatus and procedure of measurement can be found in the authors' previous paper [17]. The customization took 17.7 minutes to complete on average across the nine subjects; however the complete time was different from subject to subject and ranged from 12 to 26 minutes. Figure 6 shows the individual and customized median-plane HRIRs of a representative subject (Subject CH) from 0° to 180° with 30° intervals. The only main part of HRIR without the propagation time delay is shown. The customized HRIRs shows correspondence with the individual HRIRs for rear sources above 90° of elevation, whereas there are some discrepancies between the customized and individual HRIRs for frontal sources. In the frontal region, however, the first and second peaks in the individual HRIRs, which mainly represent pinna response, are well reproduced in the customized HRIRs. The discrepancies between the customized and individual HRIRs might be resulted from the incompleteness of customization due to the limit of tuning parameters and the difference between the empirical means, which is not direction-dependent, of the nine subjects and the 45 subject in the CIPIC HRTF database. However, the objective of this paper is not to recover the exact individual HRIRs but to customize HRIRs that can improve vertical perceptions as

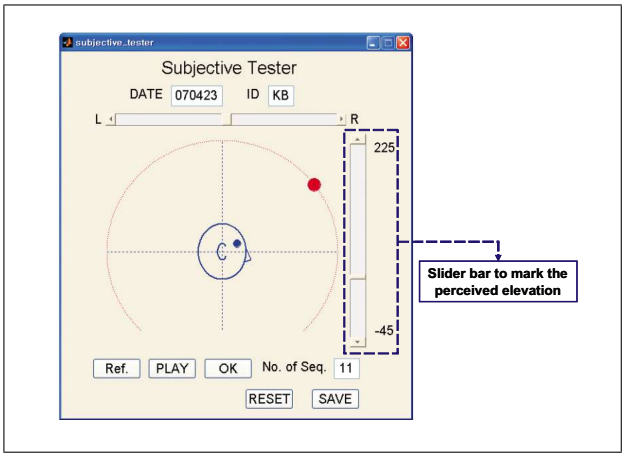


Figure 7. MATALAB™ GUI for the subjective listening tests using stationary sound. This figure was originally published in [17].

compared to non-individualized HRIRs, and reduce front-back confusion. Thus, it is necessary to assess the customization performance by a series of subjective listening tests, and this is dealt with in the next section.

3. Evaluation of customization performance

Subjective listening tests were performed to evaluate the performance for synthesizing both stationary and moving sounds on 3 sets of HRIRs including the individual, customized, and non-individualized (Kemar) HRIRs. The experiment was carried out in the audiometric room, where sound insulation is above approximately 50 dB in the frequency range above 1 kHz.

3.1. Stationary sound presentation

The nine subjects, who participated in making customized HRIRs in section 2.4, participated again in the listening tests. The localization tests using stationary sounds were carried out at 7 elevation positions in the median plane from 0° to 180° with 30° intervals. The apparatus for subjective listening test was the same with the one used in section 2.4. For convenient test procedure, the MATLAB™ GUI in Figure 7 was used for subjective listening test. When the subject registered his ID and date, a set of test signals containing 70 stimuli was generated. Each of the 7 elevation angles was stimulated 10 times in a random order. Three blocks including the stimuli synthesized by the individual, customized, and Kemar HRIRs, respectively, were separately presented to each subject. The order of blocks was randomized and which block was presented was not informed to subjects. There are a lot of different answering methods for indicating the perceived location of a sound. For example, Kan *et al.* used an electromagnetically-tracked sensor, mounted on the tip of a wand, for sound localization test [29], and it was found that the method is the least biased and more accurate than

Document downloaded by @DAEL, 2026-11-12 23:40:31 - personal use only

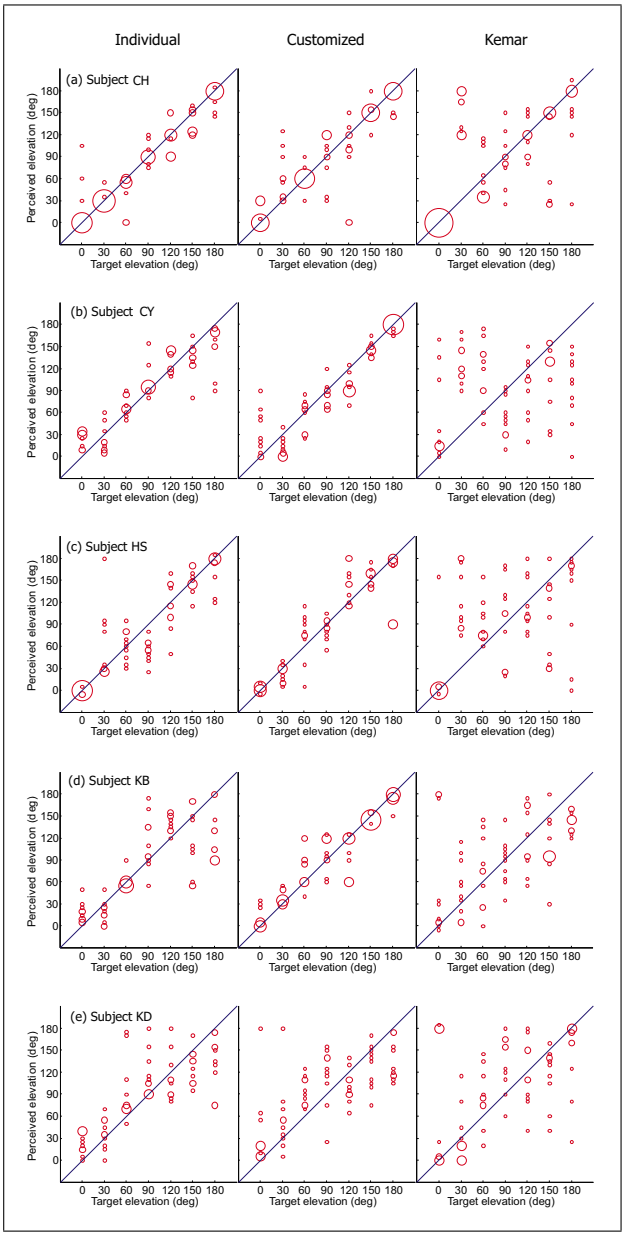


Figure 8. Subjective listening test results for nine subjects using stationary sounds synthesized by the individual, customized, and Kemar HRIRs at 7 static source positions. (a) Responses of Subject CH. (b) Responses of Subject CY. (c) Responses of Subject HS. (d) Responses of Subject KB. (e) Responses of Subject KD.

other types of response methods for indicating the perceived location of a sound [30]. This method, however, requires a sophisticated system including an electromagnetic tracking system to measure the perceived sound source location, a device including handheld pushbuttons to trigger the computer, an LED display to provide head-orientation feedback to the listener prior to each stimulus presentation, etc. On the other hand, in this study the subject's task was to locate the source (solid circle in the GUI) on the perceived elevation by moving the slider bar after he listens to each stimulus by pushing the "PLAY" button in the GUI. This method is simple and no device, except for a computer, is necessary. The resolution of the slider bar was 1°.

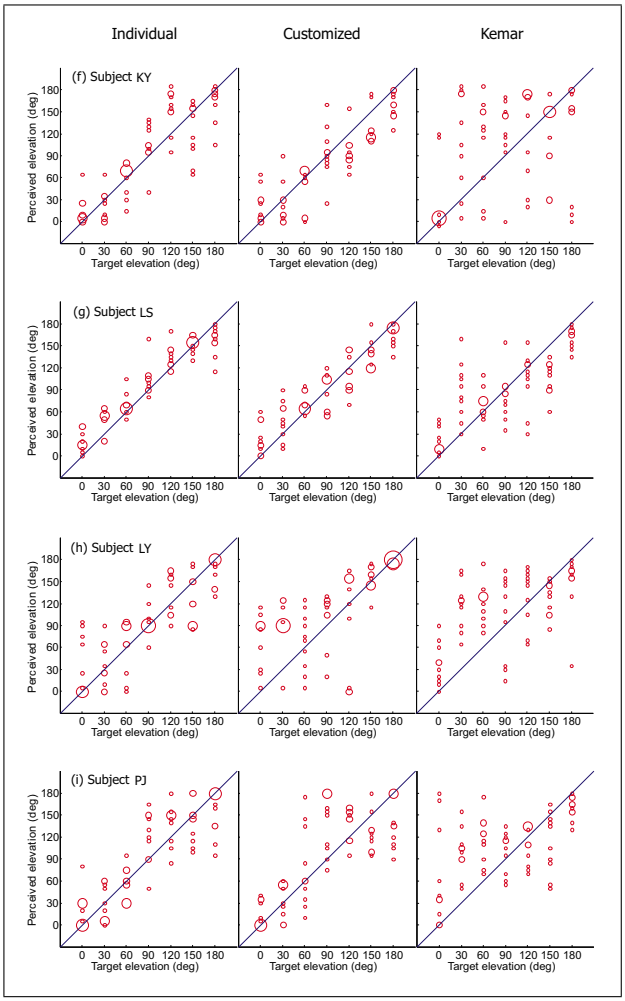


Figure 8. Continuation. (f) Responses of Subject KY. (g) Responses of Subject LS. (h) Responses of Subject LY. (i) Responses of Subject PJ.

Then, subjects pushed the "OK" button. The input signal was broadband white noise (1 kHz–18 kHz), and duration was 1 sec, including the cosine-squared rise and fall times of 20 ms, respectively. Test of each block was conducted in a different day for the subjects. One block took 15.1 minutes on average. Prior to test of each block, the HpTF of each subject was measured and equalized as described in section 2.4, and 5–10 minutes were given to subjects to complete pre-test for familiarization and practice.

Figure 8 shows the subjective listening test results using stationary sound of the nine subjects with the individual, customized, and Kemar HRIRs. Each row corresponds to a single subject's responses and subject ID is denoted at the left-top corner at each row. Each column corresponds to the responses of the subjects with the individual, customized, and Kemar HRIRs as denoted on the top of each column. In each panel, the horizontal and the vertical axes denote the target and perceived elevations, respectively. The radius of circle is directly proportional to the response frequency within 5°. In general, as shown in the first column, most subjects perceived the elevation of a sound source accurately and discriminated whether the

Document downloaded by @DAEL, 2026-11-12 23:40:31 - personal use only

Table I. Mean value and standard deviation (in parentheses) of the localization errors in degrees for each subject with the individual, customized, and Kemar HRIRs.

	Individual	Customized	Kemar
CH	12.4 (19.3)	18.4 (26.6)	38.8 (47.5)
CY	18.7 (17.8)	16.8 (17.1)	57.9 (45.1)
HS	20.4 (25.0)	17.4 (21.5)	47.4 (47.7)
KB	27.7 (28.1)	13.9 (17.2)	40.9 (37.8)
KD	29.2 (26.8)	35.1 (31.2)	42.6 (46.2)
KY	23.9 (21.4)	24.8 (19.0)	56.5 (49.5)
LS	16.9 (14.8)	20.5 (15.0)	30.6 (27.6)
LY	26.6 (23.9)	39.1 (34.5)	46.6 (34.0)
PJ	25.1 (21.3)	32.9 (27.5)	43.4 (38.2)

source is in frontal region or in rear region for almost stimuli with the individual HRIRs. In case of Subject LY, the responses were more scattered for the frontal sources than for the rear sources whereas other subjects including Subjects KB, KD, KY, and PJ, showed the opposite results. In the second column, most subjects showed comparable localization performance with the customized HRIRs to that with the individual HRIRs. In case of Subject KB, localization performance for the rear sources was even slightly enhanced with the customized HRIRs. In case of Subject LY, the responses for the frontal sources were still scattered and a few front-back confusions did occur with the customized HRIRs. With the Kemar HRIRs, in the third column, all subjects showed the worst localization performance, and the responses were distributed over all directions.

For quantitative analysis of the localization performances, the localization error, e, is defined as the absolute difference between the target and perceived elevations as

$$e_i = |T_i - P_i| \quad (i = 1, 2, \ldots, 70), \tag{11}$$

where T_i and P_i are the target and perceived elevations for the ith stimulus, respectively. Table I summarizes the mean value and the standard deviation (in parentheses) of the localization errors in degrees for each subject with the individual, customized, and Kemar HRIRs. All subjects showed the smaller localization errors on average with the customized HRIRs than with the Kemar HRIRs. They also showed the best localization performance with the individual HRIRs except for Subjects CY, HS, and KB. In case of Subjects CY and HS, the mean of localization errors with the customized HRIRs were slightly smaller than the ones with the individual HRIRs. In case of Subject KB, however, the mean of localization errors with the customized HRIRs was nearly half of that with the individual HRIRs, and this is mainly due to that the localization performance for the rear sources was enhanced with the customized HRIRs.

Statistical tests based on the *t-test* were performed to determine whether a difference in localization errors between the individual HRIRs and the customized or Kemar HRIRs are statistically significant or not. Table II summarizes the results of the *t-test* using the localization errors across all elevations for each subject. There was no statistically significant difference in the localization errors between the individual HRIRs and the customized HRIRs except for Subjects KB and LY. In case of Subject KB, the reason why statistically significant difference at the 0.01 level of significance was observed between the individual and customized HRIRs is that the localization performance for the rear sources was enhanced with the customized HRIRs. In case of Subject LY, statistically significant difference at the 0.05 level of significance was also observed between the individual and customized HRIRs, and this might be due to a few front-back confusions for the frontal sources with the customized HRIRs. There was statistically significant difference between the individual HRIRs and the Kemar HRIRs for all subjects.

Many previous studies dealt with the localization error and the front-back confusion error separately [17, 19, 31]. In this study, a quantitative error analysis for the front-back confusion and the vertical perception is performed. The front-back confusion error (e_{FBC}) is defined as

$$e_{FBC} = \frac{\text{No. of front-back confusions}}{\text{No. of total responses}} \times 100\%. \tag{12}$$

The vertical perception error (e_{VP}) is the same with the localization error in equation (11) unless the front-back confusion does not occur. If the front-back confusion occurs, the perceived elevation P is entered in equation (11) after reflecting the response about the vertical interaural plane. Thus, e_{VP} is just for the vertical perception regardless of whether a subject shows the front-back confusion or not. More detailed description can be found in [17]. Front-back confusion was determined based on the probabilistic method proposed by the authors [32]. Table III summarizes mean value and standard deviation (in parentheses) of vertical perception errors (e_{VP}) in degrees and front-back confusion error (e_{FBC}) in percentage for the nine subjects on 3 sets of HRIRs including individual, customized, and Kemar HRIRs. All subjects reported frequent front-back confusions and unsatisfactory vertical perceptions with the Kemar HRIRs. When the customized HRIRs were presented, however, all subjects showed dramatically improved localization performance compared to that with the Kemar HRIRs. Although Subjects KD, LY, and PJ showed slightly larger e_{VP} with the customized HRIRs than with the Kemar HRIRs, most subjects reported improved vertical perception with the customized HRIRs, and especially e_{FBC} was significantly reduced for all subjects with the customized HRIRs compared with the results of the Kemar HRIRs. Compared with the results of the customized HRIRs, all subjects, except for Subjects CY, HS, and KB, reported the better vertical perception with the individual HRIRs, and all subjects showed slightly improved e_{FBC} with the individual HRIRs except for Subject HS.

Based on the results obtained in this section, it can be concluded that stationary sounds can be effectively synthesized from the customized HRIRs, which are obtained

Document downloaded by @DAEL, 2026-11-12 23:40:31 - personal use only

Table II. Statistically significant difference in localization error for stationary sound between individual HRIRs and two HRIR sets including customized and Kemar HRIRs across all elevations. *: $p < 0.05$, **: $p < 0.01$.

	CH	CY	HS	KB	KD	KY	LS	LY	PJ
Kemar HRIRs	**	**	**	*	*	**	**	**	**
Customized HRIRs	-	-	-	**	-	-	-	*	-

Table III. Mean value and standard deviation (in parentheses) of vertical perception errors (e_{VP}) in degrees and front-back confusion error (e_{FBC}) in percentage for the nine subjects on 3 sets of HRIRs including individual, customized, and Kemar HRIRs.

	Individual HRIRs		Customized HRIRs		Kemar HRIRs	
	e_{VP}	e_{FBC}	e_{VP}	e_{FBC}	e_{VP}	e_{FBC}
CH	12.0 (17.5)	1.4	15.2 (18.1)	5.7	15.5 (16.2)	27.1
CY	18.4 (17.2)	1.4	16.2 (16.6)	1.4	28.3 (22.9)	40.0
HS	17.4 (17.9)	7.1	16.3 (20.7)	4.3	24.1 (22.5)	25.7
KB	24.9 (24.4)	4.3	11.4 (14.3)	4.3	26.3 (19.1)	17.1
KD	26.1 (20.3)	7.1	26.9 (20.8)	12.9	23.7 (21.7)	20.0
KY	22.6 (19.0)	2.9	23.4 (18.5)	5.7	28.5 (23.9)	31.4
LS	16.9 (14.8)	0.0	19.8 (14.5)	2.9	23.4 (17.8)	14.3
LY	26.0 (23.2)	5.7	31.7 (25.4)	14.3	30.8 (21.7)	24.3
PJ	24.9 (21.3)	1.4	30.4 (24.5)	4.3	23.9 (17.3)	30.0

from subjective tuning of only nine parameters at the three static positions, and the customized HRIRs can provide effective sound cues for both vertical perception and front-back discrimination for most of the subjects. The localization performance with the customized HRIRs is almost comparable to that with the individual HRIRs, but it is significantly better than that with the Kemar HRIRs for all subjects.

3.2. Moving sound presentation

In this section, performances for synthesizing moving sounds in the median plane with the individual, customized, and Kemar HRIRs are evaluated by a series of subjective listening tests. Four subjects (Subjects CH, HS, KB and LS), who participated in making customized HRIRs in section 2.4, participated again in the listening tests. Performance for synthesizing moving sound was evaluated based on localization performance, which represents how accurately the perceived moving source moves in the target trajectory.

The successful rendering of spatial trajectories may depend heavily on the spectral content of the sound source, and it is intuitively reasonable that spatialization of broadband sounds, such as white noise, is generally more convincing than the one of narrowband sounds. Thus, broadband white noise (1 kHz–8 kHz) was used as the general sound source. The three HRIR sets were built up at static source positions from 0° to 180° with 1° interval. The individual and Kemar HRIRs were interpolated by spline method using the HRIRs experimentally measured at elevations from −30° to 210° with 5° intervals and from −45° to 225°with 5.625° intervals, respectively. For synthesizing of moving sound, the HRIR after the pre-processing described in section 2.1 was updated at every 1° in the median plane, i.e. 1° of update rate in space, and this update rate is sufficient to prevent the artifact, which is audible to a listener due to switching of two adjacent HRIRs, because it was found that sensitivity to switching of the spectral differences require spatial resolution of 2.4°–11° depending on direction [33]. The moving sound image schemes using headphones are based on switching HRIRs, and several switching strategies have been studies in many literatures. In this study, the overlap-save method [34], which is known as simple but effective, was applied for synthesizing moving sounds. ITD is zero for all sources in the median plane, thus ITD does not need to be updated.

Table IV. Center position (α_T) and width (β_T) of each of the 18 trajectories of moving sound in degrees.

β_T	α_T					
30	15	45	75	105	135	165
60	30	60	90	120	150	
90	45	75	105	135		
120	60	90	120			

In the median plane, there were 18 different target trajectories. The target trajectory is defined using center position (α_T) and width (β_T) in degrees as

$$\alpha_T - \beta_r/2 \le \varphi \le \alpha_T + \beta_r/2, \tag{13}$$

where φ indicates elevation of moving source. α_T and β_T of each of the 18 trajectories of moving sound are summarized in Table IV. For example, if α_T and β_T are 15° and 30°, respectively, the elevation of moving source varies from 0° to 30° in the median plane. The elevation of moving source did change with a constant speed (v) of 60° or 90° per sec. Moving sound at each of the 18 trajectories was presented with a speed of both 60° and 90° per sec. in a random order yielding in total 36 stimuli for a block.

Document downloaded by @DAEL, 2026-11-12 23:40:31 - personal use only

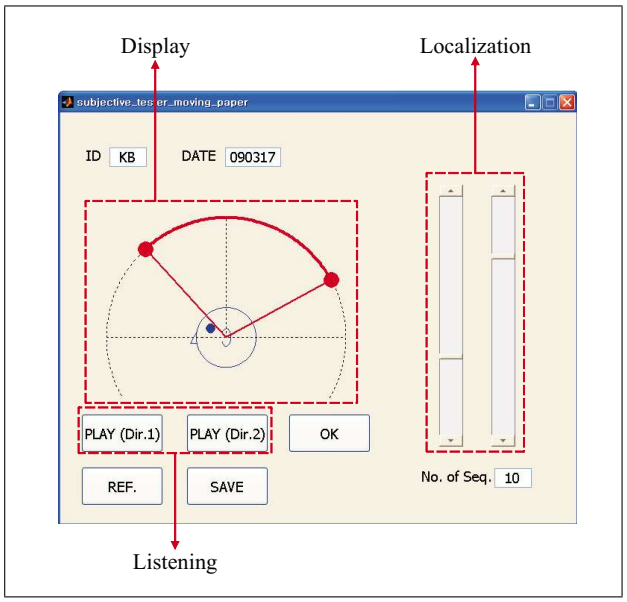


Figure 9. MATLAB™ GUI for subjective listening tests using moving sound.

Table V. Mean values of the localization errors (e_α and e_β) in degrees of each subject for moving sound synthesized with the individual, customized, and Kemar HRIRs.

HRIR set	Subject	e_α	e_β
Individual HRIRs	CH	17.6	19.8
	HS	18.3	19.3
	KB	23.0	31.1
	LS	23.9	31.5
Customized HRIRs	CH	20.2	19.4
	HS	18.7	17.4
	KB	23.7	33.1
	LS	23.7	27.1
Kemar HRIRs	CH	34.1	18.6
	HS	34.7	24.9
	KB	36.1	35.7
	LS	25.3	31.4

Duration of stimulus is totally depends on both the trajectory and speed of moving source. It is ranged from 0.33 sec. ($v = 90°$/sec, $\beta_T = 30°$) to 2 sec. ($v = 60°$/sec, $\beta_T = 120°$) including the cosine-squared rise and fall times of 20 ms, respectively. 5 blocks were synthesized by each of the 3 HRIR sets including the individual, customized, and Kemar HRIRs, thus total 15 blocks were presented to each subject. Each session was composed of randomly selected two or three blocks, and test of each session was conducted in a different day. One block took 16.7 minutes on average, and a pause of several hours was used between blocks. Prior to test of each block, the HpTF of each subject was measured and equalized as described in section 2.4. Each block was presented at comfortable listening levels as determined by the subject. The apparatus of the listening test was the same with the one used in section 2.4. For convenient test procedure, the MATLAB™ GUI in Figure 9 was used for subjective listening test of moving sounds. To evaluate the localization performance, the subject's task was to locate the solid circles in the GUI on the endpoints of the perceived trajectory by moving the slider bars after he listens to each stimulus by pushing the "PLAY" buttons. The moving source moves clockwise or counterclockwise in the target trajectory, and subjects can listen to both directions by pushing the "PLAY (Dir.1)" or "PLAY (Dir.2)", but which button corresponds to clockwise or counterclockwise was randomized for each stimulus and was not informed to subjects. After subjects move the slider bars to record the perceived trajectory, they pushed the "OK" button. Then, the center position (α_P) and width (β_P) in degrees of the perceived trajectory is automatically saved, and the number of sequence is also increased by one. When the number of sequence hit 36, the test was completed and the result was saved by pushing the "SAVE" button.

Figure 10 shows the distribution of responses of the representative subject (Subject CH) with the individual, customized, and Kemar HRIRs when the speed of moving source (v) is 60°/sec. The responses are separated by bold solid vertical lines according to β_T, which is depicted at the top of each row. Solid and empty squares represent α_T and α_P, respectively. Solid and dashed bars represent β_T and β_P, respectively. Five responses for each target trajectory are displayed at the right of the target trajectory, and they are separated by faint dashed vertical lines. Vertical axis indicates the elevation in degrees. In the first raw, the perceived trajectories are generally in correspondence to the target trajectory with the individual HRIRs. In the second row, the subject showed somewhat scattered responses with the customized HRIRs than those with the individual HRIRs, and he showed the worst localization performance and the responses were distributed over all directions with the Kemar HRIRs. Especially, the front-back confusion is dominant in the responses with the Kemar HRIRs. The other subjects also showed the similar results with that of Subject CH, thus their response distributions are not introduced here.

To assess the localization performance in detail, a quantitative error analysis is carried out. In this study, two kinds of errors are defined as

$$e_\alpha = |\alpha_T - \alpha_P| \quad \text{and} \quad e_\beta = |\beta_T - \beta_P|. \tag{14}$$

e_α is the error for the center position of trajectory, and e_β is the error for width of trajectory. Table V summarizes mean values of e_α and e_β in degrees of each subject for moving sounds synthesized with the individual, customized, and Kemar HRIRs. Based on the *t-test* results, it was found that there was no statistically significant difference between the responses with $v = 60°$/sec and $v = 90°$/sec, thus all responses are included in the Table V. All subjects showed that e_α with the Kemar HRIRs is larger than that with the individual or customized HRIRs, and they, except for Subject LS, also showed that e_α is dramatically reduced with the customized HRIRs compared to that with the Kemar HRIRs. In case of Subject LS, the improvement of e_α with

Document downloaded by @DAEL, 2026-11-12 23:40:31 - personal use only

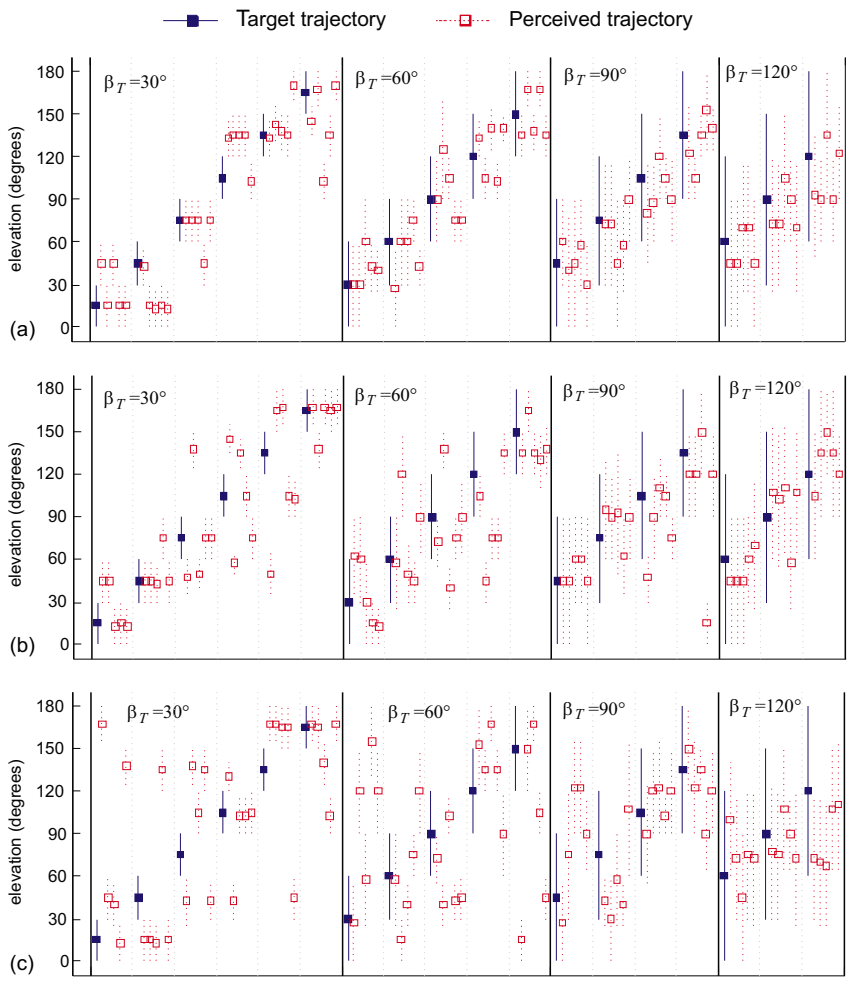


Figure 10. Responses of the representative subject (Subject CH) when the speed of moving source v is 60° per sec. Solid and empty squares represent the center positions of target and perceived trajectories, respectively. Solid and dashed bars represent the widths of target and perceived trajectories, respectively. (a) Responses with the individual HRIRs. (b) Responses with the customized HRIRs. (c) Responses with the Kemar HRIRs.

Table VI. Statistically significant difference in localization errors, e_α and e_β, for moving sound between individual HRIRs and two HRIR sets including customized and Kemar HRIRs across all trajectories. *: $p < 0.05$, **: $p < 0.01$.

	CH		HS		KB		LS	
	e_α	e_β	e	e_β	e_α	e_β	e_α	e_β
Kemar HRIRs	**	-	**	**	**	-	-	-
Cust. HRIRs	-	-	-	-	-	-	-	*

the customized HRIRs was slight, and he also showed the slightly larger e_α with the individual HRIRs than with the customized HRIRs. However, the difference in e_α between the three HRIR sets was negligible. In case of other subjects, e_α with the customized HRIRs was slightly larger than that with the individual HRIRs. On the other hands, the significant difference in e_β between the three HRIR sets was not observed for all subjects.

Statistical tests based on the *t-test* were performed to determine whether a difference in errors, e_α and e_β, between the individual HRIRs and the customized or Kemar HRIRs are statistically significant or not. Table VI summarizes the results of the *t-test* using e_α and e_β across all trajectories for each subject. There was no statistically significant difference in e_α between the individual and customized HRIRs for all subjects except for Subject LS, whereas statistically significant difference at the 0.01 level of significance in e_α between the individual and Kemar HRIRs was observed for all subjects except for Subject LS. This result tells us that the customized HRIRs can provide effective sound cues for perceiving the center position of trajectory of moving sounds as much as the individual HRIRs for most people. On the other hand, statistically significant difference between e_β with the individual and customized HRIRs was not observed for all subjects except for Subject LS, and there was no statistically significant difference in e_β between the individual and Kemar HRIRs except for Subject HS.

Based on the results obtained in this section, it can be concluded that moving sounds can be effectively synthesized from the customized HRIRs, which are obtained from subjective tuning of only nine parameters at the three static positions, and the customized HRIRs can provide effective sound cues for perceiving the center position and width of the trajectory. Therefore, it can be concluded that to make desired moving sounds for arbitrary subjects without measurement of their HRIRs is possible by the pro-

Document downloaded by @DAEL, 2026-11-12 23:40:31 - personal use only

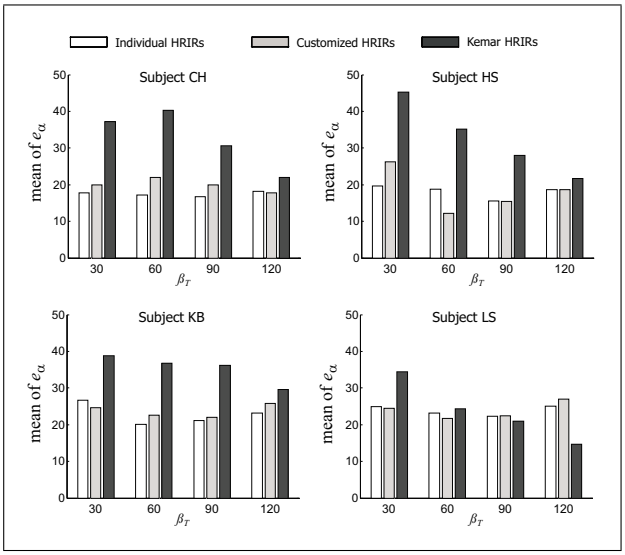


Figure 11. Mean of e_α (error for the center position of trajectory) of each subject as a function of β_T in degrees.

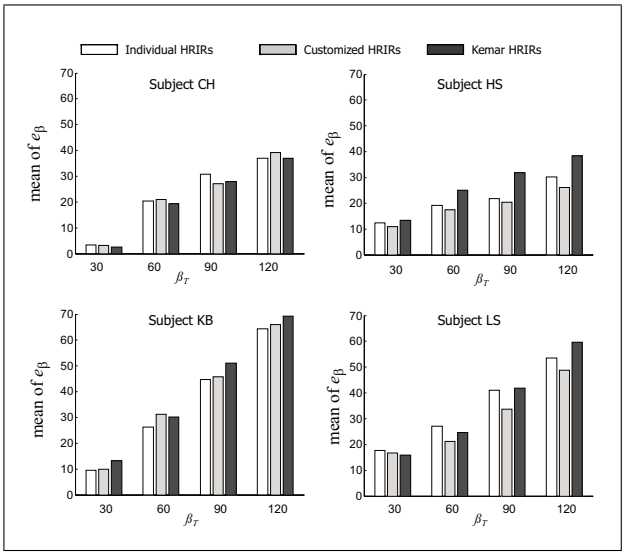


Figure 12. Mean of e_β (error for the width of trajectory) of each subject as a function of β_T in degrees.

posed customization methods based on simple tuning of a few parameters for the median plane.

4. Discussion

It is necessary to analyze the results in section 3.2 as a function of each variable, center position and width of the target trajectory and speed of moving source. For all subjects, there was no statistically significant differences in localization errors between $v = 60°$/sec and $v = 90°$/sec. This is an unexpected result because it is intuitively reasonable that the perception of moving sounds depends on the speed of moving source. The reason why the responses are independent of the speed of moving source may be that the difference between the speeds used in this study is not significant in viewpoint of perception. It is not possible to guarantee that the same results will be obtained if very slow and fast speeds are used in the experiment. Figure 11 and Figure 12 show the mean of e_α (error for the center position) and e_β (error for the width) of each subject as a function of width of the target trajectory, β_T, respectively. In Figure 11, it is interesting that e_α decreases as β_TT increases with the Kemar HRIRs whereas it is almost the same for all β_T with the individual or customized HRIRs. When β_T is 120°, e_α with the Kemar HRIRs is comparable to that with the individual or customized HRIRs. From this result, it can be inferred that subjects can catch sound cues for perception of the movement of sound source as the width of trajectory increases, even if the Kemar HRIR is used for synthesizing the moving sound. In Figure 12, all subjects show the same tendency that e_β is proportional to β_T for the three HRIR sets.

5. Summary and conclusions

This study extends the author's previous study to customize spatially continuous HRIRs for synthesizing both stationary and moving sounds. Customization was carried out by letting a subject tune ΔPCWs 1–3 at each endpoint of the sector of interest. The median plane in the upper hemisphere was divided into two sectors by considering the perceptual sensitivity to HRIRs, and the three parameters, i.e. ΔPCWs, were tuned at each of three positions (0°, 70°, and 180°), which are the endpoints of sectors. Thus, the entire median-plane HRIRs in the upper hemisphere can be customized by tuning of nine parameters only.

Subjective listening tests were carried out to assess the customization performance using both stationary and moving sounds. When the stationary sounds were presented to nine subjects, the localization performance with the customized HRIRs was comparable to that with the individual HRIRs, but it was significantly better than that with the Kemar HRIRs. There was no statistically significant difference in localization errors between the individual and customized HRIRs for seven of the nine subjects whereas statistically significant difference between the individual and Kemar HRIRs was observed for all subjects. In addition, when the moving sounds were presented to four subjects, all subjects accurately perceived the center position of the target trajectory with the customized HRIRs than with the Kemar HRIRs. The error for the center position with the customized HRIRs was comparable to that with the individual HRIRs. There was no statistically significant difference in the errors for center position between the individual and customized HRIRs for all subjects, whereas statistically significant difference between the individual and Kemar HRIRs was observed for three subjects.

Based on the results in the listening tests, it can be concluded that to make desired perception of both stationary and moving sounds for a subject without measurement of one's HRIRs is possible for most people by the proposed customization methods based on simple tuning of a few parameters.

Document downloaded by @DAEL, 2026-11-12 23:40:31 - personal use only

Acknowledgement

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

References

[1] R. D. Shilling, B. Shinn-Cunningham: Virtual auditory displays. – In: Handbook of virtual environments. K. M. Stanney (ed.). Erlbaum, Mahwah, NJ, 2002.

[2] Y. Iwaya: Individualization of head-related transfer functions with tournament-style listening test: Listening with other's ears. Acoustical Science and Technology **27** (2006) 340–343.

[3] A. Honda, H. Shibata, J. Gyoba, K. Saitou, Y. Iwaya, Y. Suzuki: Transfer effects on sound localization performances from playing a virtual three-dimensional auditory game. Applied Acoustics **68** (2007) 885–896.

[4] D. R. Begault, E. M. Wenzel: Techniques and applications for binaural sound manipulation in human-machine interfaces. International Journal of Aviation Psychology **2** (1992) 1–22.

[5] J. M. Loomis, R. G. Golledge, R. L. Klatzky: Navigation system for the blind: Auditory display modes and guidance. Presence: Teleoperators and Virtual Environments **7** (1998) 193–203.

[6] S. Holland, D. Morse, H. Gedenryd: AudioGPS: Spatial audio navigation with a minimal attention interface. Personal and Ubiquitous Computing **6** (2002) 253–259.

[7] A. B. Oving, J. A. Veltman, A. W. Bronkhorst: Effectiveness of 3-D audio for warnings in the cockpit. International Journal of Aviation Psychology **14** (2004) 257–276.

[8] R. S. Bolia: Special issues: Spatial audio displays for military aviation. International Journal of Aviation Psychology **14** (2004) 233–238.

[9] T. Lokki, M. Gröhn: Navigation with auditory cues in a virtual environment. IEEE Multimedia **12** (2005) 80–86.

[10] J. Blauert: Spatial hearing. The MIT Press, Cambridge, MA, 1983.

[11] E. Wenzel, M. Arruda, D. Kistler, F. Wightman: Localization using nonindividualized head-related transfer functions. Journal of the Acoustical Society of America **94** (1993) 111–123.

[12] H. Møller, M. F. Sørensen, D. Hammershøi, C. B. Jensen: Head-related transfer function of human subjects. Journal of the Audio Engineering Society **43** (1995) 300–321.

[13] H. Møller, M. F. Sørensen, C. B. Jensen: Binaural techniques: Do we need individual recordings. Journal of the Audio Engineering Society **44** (1996) 451–469.

[14] J. C. Middlebrooks: Virtual localization improved by scaling non-individualized external-ear transfer functions in frequency. Journal of the Acoustical Society of America **106** (1999) 1493–1510.

[15] P. J. Minnaar, S. K. Olesen, F. Christensen, H. Møller: Localization with binaural recordings from artificial and human heads. Journal of the Audio Engineering Society **49** (2001) 323–336.

[16] V. R. Algazi, R. O. Duda, D. M. Thompson, C. Avendano: The CIPIC HRTF database. Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, New York, 2001.

[17] S. Hwang, Y. Park, Y. Park: Modeling and customization of head-related impulse responses based on general basis functions in time domain. Acta Acustica united with Acustica **94** (2008) 965–980.

[18] M. Morimoto, K. Iida, M. Itoh: Upper hemisphere sound localization using head-related transfer functions in the median plane and interaural differences. Acoustical Science and Technology **24** (2003) 267–275.

[19] K. Shin, Y. Park: Enhanced vertical perception through head-related impulse response customization based on pinna response tuning in the median plane. IEICE Transactions on Fundamentals E91-A (2008) 345–356.

[20] W. L. Martens: Principal components analysis and resynthesis of spectral cues to perceived direction. Proceedings of International Computer Music Conference **San Francisco, CA** (1987).

[21] D. J. Kistler, F. L. Wightman: A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. Journal of the Acoustical Society of America **91** (1992) 1637–1647.

[22] S. Hwang, Y. Park: Interpretations on principal components analysis of head-related impulse responses in the median plane. Journal of the Acoustical Society of America **123** (2008) EL65–71.

[23] S. Hwang: Modeling, customization, and interpolation of head-related impulse responses based on principal components analysis. KAIST, Ph.D. Dissertation, 2009.

[24] S. Hwang, Y. Park, Y. Park: Analysis on perceptual sensitivity to head-related impulse responses in the median plane. Journal of Mechanical Science and Technology **23** (2010) 3340–3348.

[25] H. Møller: Fundamentals of binaural technology. Applied Acoustics **36** (1992) 171–218.

[26] H. Møller, D. Hammershøi, C. B. Jensen, M. F. Sørensen: Transfer characteristics of headphones measured on human ears. Journal of the Audio Engineering Society **43** (1995) 203–217.

[27] H. Møller, C. B. Jensen, D. Hammershøi, M. F. Sørensen: Design criteria for headphones. Journal of the Audio Engineering Society **43** (1995) 218–232.

[28] A. Kulkarni, H. S. Colburn: Variability in the characterization of the headphone transfer-function. Journal of the Acoustical Society of America **107** (2000) 1071–1074.

[29] A. Kan, C. Jin, A. van Schaik: A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function. Journal of the Acoustical Society of America **125** (2009) 2233–2242.

[30] D. Brungart, W. Rabinowitz, N. Durlach: Evaluation of response methods for the localization of nearby objects. Perception and Psychophysics **62** (2000) 48–65.

[31] F. Chen: Localization of 3-D sound presented through headphone. Duration of sound presentation and localization accuracy. Journal of the Audio Engineering Society **51** (2003) 1163–1171.

[32] S. Hwang, Y. Park: A probabilistic method for analysis of sound localization performance. Applied Acoustics **70** (2009) 771–776.

[33] P. F. Hoffmann, H. Møller: Audibility of differences in adjacent head-related transfer functions. Acta Acustica united with Acustica **94** (2009) 945–954.

[34] A. V. Oppenheim, R. W. Schafer, J. R. Buck: Discrete-time signal processing. Prentice-Hall, New Jersey, 1999.

Document downloaded by @DAEL, 2026-11-12 23:40:31 - personal use only