Personalization of head-related transfer functions in the median plane based on the anthropometry of the listener's pinnae[a)]

Kazuhiro Iida,[b)] Yohji Ishii, and Shinsuke Nishioka
Faculty of Engineering, Chiba Institute of Technology, 2-17-1 Tsudanuma, Narashino, Chiba 275-0016, Japan

(Received 13 September 2013; revised 1 May 2014; accepted 16 May 2014)

A listener's own head-related transfer functions (HRTFs) are required for accurate three-dimensional sound image control. The HRTFs of other listeners often cause front-back confusion and errors in the perception of vertical angles. However, measuring the HRTFs of all listeners for all directions of a sound source is impractical because the measurement requires a special apparatus and a lot of time. The present study proposes a method for estimating the appropriate HRTFs for an individual listener. The proposed method estimates the frequencies of the two lowest spectral notches (N1 and N2), which play an important role in vertical localization, in the HRTF of an individual listener by anthropometry of the listener's pinnae. The best-matching HRTFs, of which N1 and N2 are the closest to the estimates, are then selected from an HRTF database. In order to examine the validity of the proposed method, localization tests in the upper median plane were performed using four subjects. The results revealed that the best-matching HRTFs provided approximately the same performance as the listener's own HRTFs for the target directions of the front and rear for all four subjects. For the upper target directions, however, the performance of the localization for some of the subjects decreased. © *2014 Acoustical Society of America*.
[http://dx.doi.org/10.1121/1.4880856]

PACS number(s): 43.66.Pn, 43.66.Qp [ELP] Pages: 317–333

I. INTRODUCTION

Accurate sound image control can be accomplished by reproducing a listener's own head-related transfer functions (HRTFs) at the entrances of the ear canals (Morimoto and Ando, 1980). Measurements of the HRTFs for an arbitrary listener for an arbitrary direction are, however, impractical because the measurements require a special apparatus and a great deal of time.

In 1999, at an ASA meeting, Sottek and Genuit (1999) talked about Blauert's vision for the personalization of an HRTF, whereby a person who enters a multimedia shop is scanned by a camera, and a few moments later his/her individual HRTF set is ready for use in advanced 3D applications. However, this scenario has not yet been realized.

A number of studies have been conducted in an attempt to establish methods for obtaining personalized HRTFs that do not require acoustical measurements. One such method uses principal component analysis (PCA), which resolves an HRTF into its principal components (Kistler and Wightman, 1992; Middlebrooks and Green, 1992). The coefficients of each principal component are then estimated based on the anthropometry of the listener's pinnae (Hu *et al.*, 2008; Xu *et al.*, 2008; Hugeng and Gunawan, 2010; Zhang *et al.*, 2011). However, the results of sound localization tests were not reported in these studies.

Numerical calculation of HRTFs has been studied intensively. The boundary element method (BEM) has been used to calculate HRTFs in a number of studies (Katz, 2001; Kahana and Nelson, 2006; Kreuzer *et al.*, 2009). The results of numerical calculations by the finite-difference time-domain (FDTD) method, which is much faster than the BEM, revealed that the fundamental spectral feature of the HRTF of an individual listener can be calculated from the baffled pinna (Takemoto *et al.*, 2012). At the present, however, neither the BEM nor the FDTD method is available for ordinary listeners because special equipment, e.g., a functional magnetic resonance imaging system, is required to digitize the complicated shape of the pinnae of an individual listener.

A number of methods have been considered in which a listener chooses the appropriate HRTFs by performing a listening test. Middlebrooks (1999a,b) reported that intersubject differences in directional transfer functions (DTFs), which are the directional components of HRTFs, could be reduced by appropriately scaling the frequency of one set of DTFs. Middlebrooks *et al.* (2000) then showed that optimally frequency-scaled DTF halved the difference in quadrant error between other-ear and own-ear conditions. The quadrant error was defined as errors larger than 90° in the vertical and/or front-back dimension. However, one to three 20-min blocks of listening tests were required to find a listener's preferred scale factor. Iwaya (2006) claimed that tournament-style listening tests required approximately 15 min to select the most appropriate HRTF set from among 32 sets of HRTFs. The time required to choose the appropriate HRTFs becomes a more serious problem as the size of the database increases.

Zotkin *et al.* (2003) proposed a method that measures the anthropometric data of a listener's pinna and selects the

[a)]Portions of this work were presented in "Estimation of spectral notch frequencies of the individual head-related transfer function from anthropometry of listener's pinna," Proceedings of Meetings on Acoustics **19**, 050174 (2013).
[b)]Author to whom correspondence should be addressed. Electronic mail: kazuhiro.iida@it-chiba.ac.jp

HRTF of the most similar pinna from a database. They carried out localization tests in the frontal hemisphere using the selected HRTFs and those of a KEMAR dummy head. The results revealed that the improvement in the localization accuracy by the proposed method averaged over eight subjects was only approximately 1.9°. Their method appears not to be effective because the contributions of all anthropometric data to the HRTF are considered to be equal. Considering the contribution of each anthropometric datum to the localization cues in HRTFs would enable selection of the HRTFs, which provide high localization performance, from the database.

The present study proposes a method that does not require any acoustical measurements or listening tests in order to determine an individual's appropriate HRTFs. The method estimates the listener's spectral cues for vertical localization from the anthropometry of the listener's pinnae. These estimates are then used to select from a database the HRTFs for which the spectral cues are the closest match. The validity of this method was evaluated based on both physical and perceptual aspects.

II. INDIVIDUAL DIFFERENCE IN SPECTRAL CUES

A. Spectral cues to be estimated

The spectral peaks and notches in the frequency range above 5 kHz prominently contribute to the perception of the vertical angle of a sound source (Hebrank and Wright, 1974; Butler and Belendiuk, 1977; Mehrgardt and Mellert, 1977; Musicant and Butler, 1984). Kulkarni and Colburn (1998) carried out localization tests in the horizontal plane (azimuth of 0°, ±45°, 180°) using four subjects. The magnitude spectra of the subjects' own HRTFs were systematically smoothed in seven levels. The results indicate that the fine spectral structure is relatively unimportant for sound localization, as compared to the outline of the peaks and notches. They also reported that the elevation of a sound image shifted upward under the extreme smoothing condition.

Iida *et al.* (2007) extracted the spectral peaks and notches from a listener's measured HRTFs, regarding the peak around 4 kHz, which is independent of the vertical angle of the sound source (Shaw and Teranishi, 1968), as the lower-frequency limit. The peaks and notches were then labeled in order of frequency. They carried out sound localization tests in the upper median plane using three subjects and demonstrated that the simplified HRTFs, which are composed of either only the first spectral peak around 4 kHz (P1) and the two lowest spectral notches (N1 and N2) above the P1 frequency or only N1 and N2, provided approximately the same localization performance as the measured HRTFs for the front and rear directions. For the upper directions, the simplified HRTFs provided approximately the same localization performance as the measured HRTFs for some subjects, but not for others.

Furthermore, they showed that the frequencies of N1 and N2 are highly dependent on the vertical angle, whereas the frequency of P1 is approximately constant and is thus independent of the vertical angle. Figure 1 shows the amplitude of the HRTFs of a subject in the median plane measured in 10° steps. The lines of N1 and N2 are fitted by fourth-order polynomial approximation, and P1 is fitted by linear approximation. A method by which to obtain the N1, N2, and P1 frequencies is described in Sec. II B 2.

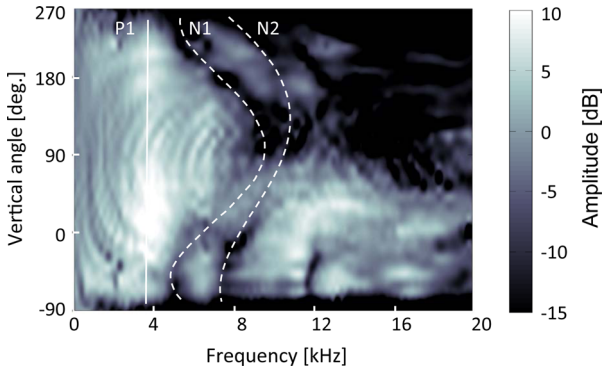

FIG. 1. (Color online) Amplitude spectrum of HRTF in the median plane.

Based on these results, it can be concluded that N1 and N2 play an important role in the localization of, at least, the front and rear directions and that the hearing system of a human being could use P1 as reference information in order to analyze N1 and N2 in ear-input signals.

In the extreme smoothing condition of Kulkarni and Colburn (1998) mentioned above, in which the elevation of a sound image shifted upward, N1 and P1 of the subjects' own HRTFs were preserved but N2 vanished. This result also implies the importance of N2.

B. Distribution range of the frequencies of N1, N2, and P1

As described above, the information to be reproduced could be focused on N1, N2, and P1. However, there are large individual differences in the N1, N2, and P1 frequencies. Therefore, the HRTFs of other listeners often cause front-back confusion, errors in vertical perception, and inside-of-head localization. Individual differences in the N1, N2, and P1 frequencies for the front direction were measured.

1. Measurements of HRTFs

The HRTFs of the seven directions in the upper median plane in 30° steps for 28 Japanese male adult subjects were measured in an anechoic chamber. The test signal was a swept sine wave (2^{18} samples), the sampling, start, and stop frequencies of which are 48 kHz, 2 Hz, and 23 998 Hz, respectively. The test signal was presented by one of the loudspeakers of 80 mm in diameter (FOSTEX FE83E) located in the upper median plane in 30° steps (seven directions). The distance from the loudspeakers to the center of the subject's head was 1.2 m. No frequency equalization was performed. Ear microphones (Iida *et al.*, 2007) were used to pick up the test signals at the entrances of the ear canals of the subject.

The ear microphones were fabricated using the subject's ear molds. The ear mold was constructed by the following procedure: (1) an inverse mold was formed by occluding the

pinna with silicon, (2) the inverse mold was encased in plaster, and (3) the silicon mold was removed. A miniature electret condenser microphone element of 5 mm in diameter (Panasonic WM64AT102) was embedded in the silicon resin at the entrance of the ear canal of the ear mold. The microphone and silicon resin were then removed from the ear mold in order to be used as an ear microphone [Fig. 2(a)].

In the HRTF measurements, the ear microphones were placed into the ear canals of the subjects [Fig. 2(b)]. The diaphragms of the microphones were located at the entrances of the ear canals. This condition is referred to as the blocked-entrances condition (Shaw and Teranishi, 1968). The HRTF was obtained as

$$\mathrm{HRTF}_{l,r}(\omega) = G_{l,r}(\omega)/F(\omega), \quad (1)$$

where $F(\omega)$ is the Fourier transform of the impulse response, $f(t)$, measured at the point corresponding to the center of the subject's head in the free field without a subject, and $G_{l,r}(\omega)$ is that measured at the entrance of the ear canal of the subject with the ear microphones.

2. Extraction of N1, N2, and P1

Then, N1, N2, and P1 for the front direction (vertical angle: 0°) of 56 ears of 28 subjects were extracted. Since N1, N2, and P1 are generated by the pinnae (Shaw and Teranishi, 1968; Lopez-Poveda and Meddis, 1996; Takemoto *et al.*, 2012), they were extracted from the early part of the head-related impulse response (HRIR) using software, of which the algorithm is as follows:

(1) Detect the sample for which the absolute amplitude of the HRIR is maximum.

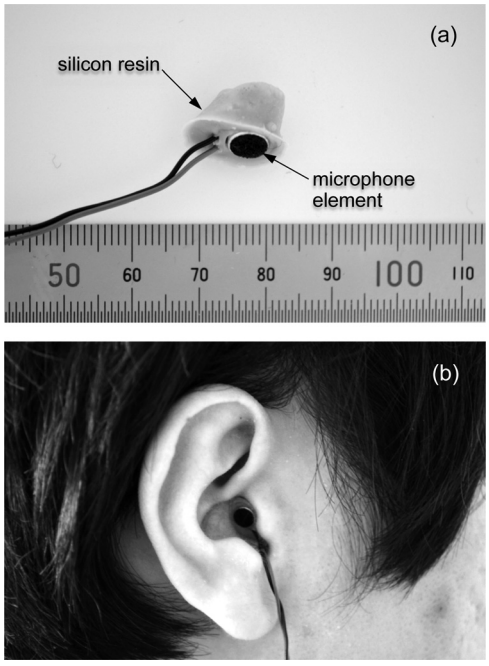

FIG. 2. Photographs of (a) ear microphone and (b) its placement into the ear canal of a subject.

(2) Clip the HRIR using a four-term, 96-point Blackman-Harris window, adjusting the temporal center of the window to the maximum sample detected in (1).
(3) Prepare a 512-point array, all of the values of which are set to zero, and overwrite the clipped HRIR in the array, where the maximum sample of the clipped HRIR should be placed at the 257th point in the array.
(4) Obtain the amplitude spectrum of the 512-point array by FFT. Then, find the local maxima and local minima of the amplitude using the difference method.
(5) Define the lowest frequency of the local maxima above 3 kHz as P1 and the lowest two frequencies of the local minima above P1 as N1 and N2, respectively.

As a result, N1, N2, and P1 of 54 ears (two ears in 26 subjects and one ear in two subjects) were obtained. Of the 56 ears, 2 have only one notch in the spectrum of HRTFs.

The N1, N2, and P1 frequencies were distributed from 5719 to 9563 Hz (0.74 octaves), 8250 to 13 500 Hz (0.71 octaves), and 3469 to 4313 Hz (0.31 octaves), respectively. These results indicate that the individual differences in the P1 frequency are much smaller than those in the N1 and N2 frequencies, and the distributions of N1 and N2 overlap.

The just-noticeable differences (JNDs) in the P1 frequency for the front direction with regard to vertical localization are 0.35 and 0.47 octaves for higher and lower frequencies, respectively (see the Appendix). Therefore, the individual difference in the P1 frequency can be considered to have little effect on vertical localization, and as such, it is not considered hereinafter. On the other hand, the JNDs of N1 and N2 can be considered to range from 0.1 to 0.2 octaves (Iida and Ishii, 2011b). Therefore, individual differences in N1 (0.74 octaves) and N2 (0.71 octaves) are considered to have remarkable effects on vertical localization.

III. INDIVIDUAL DIFFRENCE IN PINNA SHAPE

Takemoto *et al.* (2012) calculated the HRTFs from four subjects' head shapes using the FDTD method. They reported that the N1 and N2 frequencies of four subjects differ from each other, and that the basic peak-notch pattern of the HRTFs originated from the pinnae. Moreover, they showed that one or two anti-nodes and a node appear in the pinna cavities at the N1 frequency. These findings imply that individual differences in the N1 and N2 frequencies could be attributed to individual differences in the shape and size of the listener's pinnae.

Thus, ten anthropometric parameters of the pinna to be analyzed (Fig. 3) were adopted after Algazi *et al.* (2001). However, we replaced their $\theta 2$ (pinna flare angle) by x_4 (width of the helix), because $\theta 2$ is difficult to measure.

Nine anthropometric parameters (x_1 through x_8 and x_d) of the pinna for 54 ear molds were measured using a vernier caliper. The tilt of the pinna (x_a) was measured from a photograph of the profile of the subject.

The measured dimensions for 54 ears are listed in Table I. The range of values for each dimension spanned 10 to 25 mm, and the angle of tilt, x_a, ranged widely from 4° to 40°.

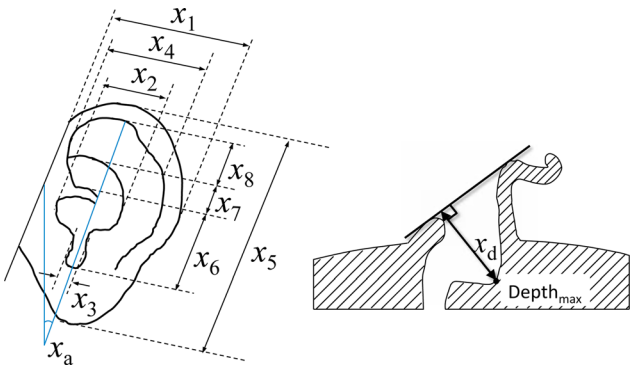

FIG. 3. (Color online) Ten anthropometric parameters of the pinna.

IV. ESTIMATION OF THE N1 AND N2 FREQUENCIES FROM THE ANTHROPOMETRY OF THE PINNA

A. Multiple regression model

Multiple regression analyses were carried out using 54 ears as objective variables of the N1 and N2 frequencies of the front direction and as explanatory variables of ten anthropometric parameters of the pinnae, using the linear least squares solution, as follows:

$$f(S)_{\mathrm{N1,N2}} = a_1x_1 + a_2x_2 + \cdots + a_nx_n + \mathrm{b}\ \ [\mathrm{Hz}], \tag{2}$$

where S, a_l, b, and x_i denote the subject, the regression coefficients, a constant, and the anthropometric parameters, respectively.

B. Accuracy of multiple regression

The results show that the multiple correlation coefficients between the frequencies of N1 and N2 extracted from the measured HRIRs and those estimated from ten anthropometric parameters of the pinnae were 0.84 and 0.87, respectively. However, the p-values of x_1, x_4, and x_7 for N1 exceeded 0.05. For N2, the p-values of x_1, x_2, x_3, x_4, and x_a exceeded 0.05. These parameters are considered not to be relevant to N1 and N2. Therefore, we performed multiple regression analysis for all combinations of parameters, varying the number of parameters to be used. Then, we adopted the combination of parameters for which the correlation coefficient was the highest under the conditions that all of the p-values were less than 0.05. As a result, six parameters (x_2, x_3, x_6, x_8, x_d, and x_a) were adopted for N1, and three parameters (x_6, x_8, and x_d) were adopted for N2. This means that the width, length, and depth of the pinna cavities and the tilt of the pinna correlated to N1, and the length and depth of the cavities correlated to N2. The multiple regression coefficients, p-values, and 95% confidence intervals are listed in Table II.

The relationship between the N1 and N2 frequencies extracted from the measured HRIRs and those estimated from the listener's anthropometric parameters are shown in Fig. 4. The statistics of the multiple regression models are shown in Table III. The multiple correlation coefficients of N1 and N2 were 0.81 and 0.82, respectively. The average absolute residual errors were 0.07 and 0.08 octaves, respectively. The probability for both N1 and N2 that the absolute residual error was within the JND was 91%. The JND is regarded to be 0.15 octaves because the JNDs can be considered to range from 0.1 to 0.2 octaves, as mentioned in Sec. II B.

Then, in order to confirm the multicollinearities among x_2, x_3, x_6, x_8, x_d, and x_a for N1, and among x_6, x_8, and x_d for N2, variance inflation factors (VIF) were calculated, where VIF is defined as follows:

$$\mathrm{VIF}(j) = \frac{1}{(1 - R(j)^2)}, \tag{3}$$

where $R(j)^2$ denotes the determination coefficient of the multiple regression analysis using the jth explanation variable as the objective variable and other explanation variables as the explanation variables. All of the VIFs, i.e., six VIFs for N1 and three VIFs for N2, were less than 10, which means that there was no multicollinearity between the explanatory variables (Chatterjee and Hadi, 2012).

These results indicate that the proposed multiple regression model can estimate the N1 and N2 frequencies with an accuracy that is almost within the JND.

C. Accuracy of estimation of the N1 and N2 frequencies for naive subjects

In order to confirm the validity of the multiple regression model, the N1 and N2 frequencies of naive subjects, who were not involved in the multiple regression analysis, were estimated and then compared with the extracted N1 and N2 frequencies.

The subjects were three males (OIS, TCY, and MTZ) and a female (CKT), 21 to 25 yr of age, with normal hearing sensitivity. Six anthropometric parameters were measured from their actual pinnae. The measured parameters are shown in Table IV. The N1 and N2 frequencies for the front direction were then estimated using Eq. (2) and were also extracted from the HRIRs using the algorithm described in Sec. II B.

The estimated and extracted frequencies and the residual error are listed in Table V. The residual errors of N1 and N2 were less than the JND for all eight ears. Relatively large errors were observed in N1 of subject CKT's left ear (0.10 octaves), MTZ's left ear (0.09 octaves), and TCY's right ear (0.09 octaves).

V. SELECTION OF THE BEST-MATCHING HRTFS

In the present section, the authors propose a method for selecting the best-matching HRTFs from an HRTF database. The best-matching HRTF is defined as the HRTF for which the N1 and N2 frequencies are the closest to the estimated N1 and N2 frequencies. Furthermore, the validity of the proposed method is clarified by sound localization tests.

A. Method for selecting the best-matching HRTFs in the median plane

The notch frequency distance (NFD) (Iida and Ishii, 2011a) was used as a physical measure to select the

TABLE I. Ten measured pinnae dimensions for 54 ears (mm).

	Width of				Length of					
Pinna	pinna x_1	concha x_2	incisura intertragica x_3	helix x_4	pinna x_5	concha x_6	cymba conchae x_7	scapha x_8	Depth of concha x_d	Tilt of pinna (°) x_a
1	37.0	20.1	8.3	29.7	69.1	21.0	9.3	16.8	15.2	18
2	35.7	19.6	6.5	29.4	72.1	21.5	9.8	15.5	14.9	24
3	33.0	19.7	5.9	27.0	66.3	21.8	6.4	19.8	15.8	29
4	37.5	15.8	5.5	24.7	75.8	19.3	9.9	18.8	13.8	30
5	35.9	18.6	9.0	24.7	73.8	22.9	6.2	19.6	11.9	16
6	34.2	19.5	9.8	26.3	72.0	22.3	9.3	17.9	12.1	17
7	34.7	14.9	5.3	20.1	71.3	21.4	5.9	20.8	16.7	7
8	31.9	16.8	6.8	25.2	70.1	20.2	4.5	21.4	17.2	12
9	34.4	17.8	7.0	25.7	71.4	22.3	5.3	19.4	13.3	17
10	35.0	18.8	9.0	26.1	63.4	22.2	4.7	20.8	14.1	18
11	34.4	18.2	8.3	24.8	64.4	20.4	7.1	20.8	15.0	19
12	38.0	17.6	8.3	28.3	67.0	17.7	8.8	19.2	12.5	23
13	35.1	19.1	8.4	26.3	68.5	17.1	8.1	19.8	13.8	7
14	34.8	18.3	8.7	22.2	66.4	18.8	6.6	22.5	14.1	23
15	35.2	18.7	8.3	27.1	67.0	18.6	8.3	22.0	13.8	24
16	36.1	16.5	5.5	27.3	64.3	19.3	5.5	17.3	14.0	30
17	37.5	16.6	5.5	29.7	66.4	18.7	8.0	17.2	13.8	40
18	36.6	21.2	8.4	27.0	67.2	19.3	7.5	18.4	11.3	27
19	36.1	19.8	7.5	28.1	66.4	19.2	8.5	18.3	11.1	21
20	35.7	15.6	5.8	25.7	69.2	20.5	8.6	20.3	13.7	28
21	35.8	15.4	6.4	26.8	70.8	20.9	9.4	21.2	13.1	36
22	36.1	20.5	6.7	27.0	64.0	21.0	6.0	20.2	14.8	32
23	33.4	16.5	5.9	27.5	63.8	22.5	4.3	21.1	13.8	27
24	43.8	20.2	8.3	31.6	78.3	21.9	8.0	24.1	14.1	27
25	42.1	19.2	7.3	31.8	77.8	19.0	9.3	23.1	14.4	20
26	36.9	14.8	6.7	27.5	72.8	24.1	7.5	19.4	13.0	23
27	36.1	17.4	7.5	24.3	70.4	22.9	9.3	17.4	13.8	31
28	31.7	16.6	8.9	19.0	69.2	23.3	7.7	15.9	16.0	25
29	32.8	16.7	8.7	18.1	69.3	20.4	9.1	17.1	15.7	24
30	36.0	20.1	10.8	28.3	64.0	22.0	5.8	16.7	13.8	27
31	34.0	21.8	11.9	26.9	64.4	22.0	10.3	14.2	13.1	36
32	40.6	19.1	10.2	28.0	75.8	25.1	2.6	21.1	14.1	23
33	36.4	18.4	7.6	26.2	83.2	24.9	4.7	22.1	13.8	29
34	31.2	19.3	10.9	21.6	63.5	20.4	5.9	15.4	13.3	13
35	32.5	19.9	10.1	23.0	65.1	21.3	8.2	15.9	11.7	14
36	33.1	21.0	9.0	23.1	58.2	18.4	3.5	14.0	9.9	32
37	32.7	20.4	7.2	26.6	59.1	17.8	5.1	16.9	9.7	16
38	35.7	19.7	7.4	29.0	66.9	20.2	5.3	17.6	11.8	19
39	35.0	21.8	7.3	28.8	68.9	19.5	7.7	20.0	12.5	28
40	34.5	17.5	8.4	24.3	66.1	23.0	5.9	16.5	14.9	29
41	34.5	19.9	8.1	25.3	68.2	23.1	6.3	17.9	15.0	17
42	35.2	19.0	10.0	26.6	69.9	21.6	4.0	20.3	12.6	18
43	34.5	19.1	6.1	28.3	72.5	21.0	6.0	20.7	15.2	20
44	37.4	20.9	7.5	29.4	73.8	19.4	9.7	23.2	13.6	33
45	35.3	20.0	11.2	25.2	63.4	24.1	3.5	18.5	15.6	14
46	33.2	18.5	9.9	22.8	68.7	24.5	3.9	16.5	17.6	10
47	32.9	18.5	10.4	24.0	62.4	21.2	5.8	18.9	14.8	4
48	35.9	19.2	9.0	26.1	63.2	20.1	7.1	17.1	14.0	17
49	32.9	20.0	9.3	23.2	64.5	23.5	2.9	13.2	17.6	14
50	37.3	18.1	8.3	23.9	65.7	21.8	4.7	14.9	17.4	15
51	35.7	18.1	8.8	26.2	66.7	22.3	6.1	18.8	13.4	20
52	38.4	17.9	10.1	26.4	65.4	20.4	5.6	20.8	11.3	9
53	36.9	19.9	11.5	28.8	66.8	21.0	6.5	19.4	13.6	39
54	38.0	21.7	8.1	28.2	68.8	22.8	7.5	21.4	14.8	28
Min	31.2	14.8	5.3	18.1	58.2	17.1	2.6	13.2	9.7	4
Max	43.8	21.8	11.9	31.8	83.2	25.1	10.3	24.1	17.6	40

TABLE II. Multiple regression coefficients, p-values, and 95% confidence intervals of N1 and N2 for the front direction.

	Regression coefficient		p-value		95% confidence intervals			
					N1		N2	
	N1	N2	N1	N2	lower	upper	lower	upper
a_1								
a_2	116.9		1.6E-02		22.9	210.9		
a_3	−157.5		4.7E-03		−264.2	−50.8		
a_4								
a_5								
a_6	−183.4	−327.0	8.3E-05	2.9E-07	−269.1	−97.8	−438.0	−216.0
a_7								
a_8	−93.2	−245.0	2.3E-03	4.4E-08	−151.5	−34.9	−321.3	−168.6
a_d	−131.4	−172.8	4.0E-03	3.7E-03	−218.7	−44.2	−286.9	−58.7
a_a	−48.7		7.2E-07		−65.8	−31.6		
b	14906.4	23903.1	9.2E-14	2.0E-22	12019.9	17792.9	21079.9	26726.3

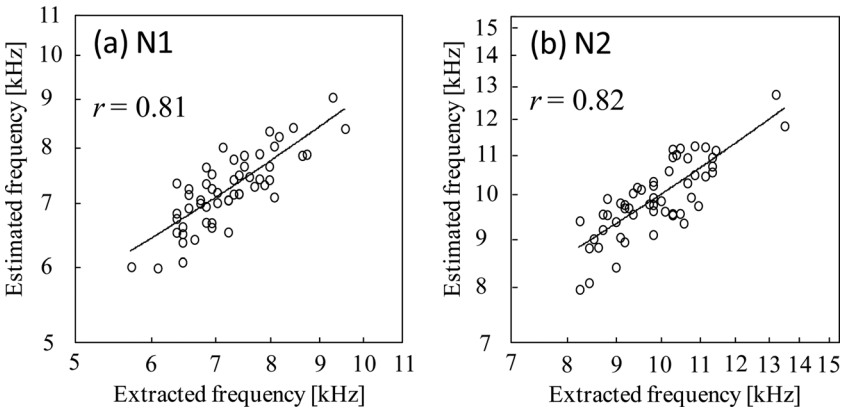

FIG. 4. Relationship between the frequencies extracted from the measured HRIR and the frequencies estimated from the listener's anthropometric parameters for 54 ears. (a) N1; (b) N2. r denotes the correlation coefficient.

TABLE III. Statistics of the multiple regression models of N1 and N2 for the front direction.

	Correlation coefficient	Significance level	Absolute mean residual error		Probability that residual error less than 0.15 octaves [%]
			[Hz]	[oct.]	
N1	0.81	1.1E-09	357	0.07	91
N2	0.82	5.4E-12	550	0.08	91

TABLE IV. Six measured pinnae dimensions of four subjects (mm).

		Width of		Length of			
Subject	Ear	concha x_2	incisura intertragica x_3	concha x_6	scapha x_8	Depth of concha x_d	Tilt of pinna (°) x_a
OIS	L	19.2	7.9	23.3	19.5	12.9	28
	R	17.0	8.7	21.0	20.1	13.5	18
TCY	L	14.4	8.0	23.0	20.3	12.5	29
	R	14.3	7.6	22.8	20.4	12.9	23
CKT	L	15.7	7.4	19.7	18.3	13.9	22
	R	15.0	8.2	19.6	18.3	12.1	21
MTZ	L	17.6	7.6	22.6	17.2	11.2	28
	R	18.1	7.2	21.5	18.0	13.9	24

TABLE V. Estimated and extracted frequencies of N1 and N2 and the residual errors for four subjects.

		Estimated frequency [Hz]		Extracted frequency [Hz]		Residual error [oct.]	
Subject	Ear	N1	N2	N1	N2	N1	N2
OIS	L	6749	9273	6938	9375	−0.04	−0.02
	R	7147	9779	6938	9281	0.04	0.08
TCY	L	6163	9249	6094	9656	0.02	−0.06
	R	6481	9221	6094	9375	0.09	−0.02
CKT	L	7358	10 576	6844	10 406	0.10	0.02
	R	7454	10 920	7219	11 250	0.05	−0.04
MTZ	L	7182	10 364	6750	10 875	0.09	−0.07
	R	7271	10 061	7313	10 125	−0.01	−0.01

best-matching HRTF from the HRTF database. The NFD expresses the distance between HRTF_j of subject j and HRTF_k of subject k in the octave scale, as defined by the following equations (L1-norm):

$$\mathrm{NFD}_1 = \log_2\{f_{N1}(\mathrm{HRTF}_j)/f_{N1}(\mathrm{HRTF}_k)\}\ [\mathrm{oct.}], \quad (4)$$

$$\mathrm{NFD}_2 = \log_2\{f_{N2}(\mathrm{HRTF}_j)/f_{N2}(\mathrm{HRTF}_k)\}\ [\mathrm{oct.}], \quad (5)$$

$$\mathrm{NFD} = |\mathrm{NFD}_1| + |\mathrm{NFD}_2|\ \ [\mathrm{oct.}], \quad (6)$$

where f_{N1} and f_{N2} denote the frequencies of N1 and N2, respectively.

The HRTFs with the smallest NFD were selected from the database as the best-matching HRTFs for the front direction when the estimated N1 and N2 frequencies for the front direction were substituted into Eqs. (4)–(6). The best-matching HRTFs for the left and right ears were selected independently. Then, the HRTFs for the other directions in the upper median plane were provided by the donor, for which the HRTF for the front direction was selected as the best match.

The database consists of the HRTFs for 120 ears of Japanese adults, measured in seven directions in the upper median plane in steps of 30° and the N1, N2, and P1 frequencies for the front direction (see http://www.iida-lab.it-chiba.ac.jp/e/). Among the 120 ears, the 54 listed in Table I were used in the multiple regression analysis. The other 66 ears are from another 33 (10 females and 23 males) Japanese adults. The HRTFs of these 66 ears were not used in the multiple regression analysis because the anthropometric parameters were not known.

B. Accuracy of the best-matching HRTF with respect to physical aspects

The best-matching HRTFs of the four subjects (OIS, TCY, CKT, and MTZ) for the front direction were selected based on the estimated N1 and N2 frequencies.

Figure 5 shows the amplitude spectrum of the best-matching HRTFs and the subjects' own HRTFs. Here, the frequencies of N1 and N2 of the best-matching HRTF (closed symbols) were similar to those of the subjects' own HRTFs (open symbols). Similar structural features were observed both in the best-matching HRTFs (broken lines) and in the subjects' own HRTFs (solid lines) for almost all of the ears. However, the spectrum of the best-matching HRTF of CKT's left ear was not similar to the subject's own HRTF. N1 and N2 of the best-matching HRTF are shallow and deep, respectively, compared with those of the subject's own HRTF. This is attributed to the notch level not being considered in the estimation method.

Table VI shows the N1 and N2 frequencies of the best-matching HRTFs and the subjects' own HRTFs for the front direction. The residual errors were superpositions of the errors due to the estimations of the N1 and N2 frequencies based on the anthropometry of each subject's pinnae and the errors due to the selection of the best-matching HRTFs from the database. The residual errors of N1 and N2 were less than the JND for all eight ears. Relatively large errors were observed in N1 of subject CKT's left ear (0.11 octaves), MTZ's left ear (0.10 octaves), and TCY's right ear (0.09 octaves). This tendency is the same as that shown in Table V. Figure 6 shows a scatterplot of the best-matching HRTFs and the subjects' own HRTFs on the N1-N2 plane. For each subject, the best-matching HRTF and the subject's own HRTF are located in close proximity to each other among the 120 widely scattered HRTFs, whereas relatively large distances were observed for the left ears of CKT and MTZ.

Next, the residual errors in the N1 and N2 frequencies between the best-matching HRTFs and the subjects' own HRTFs for each of the seven directions in the upper median plane were calculated (Table VII). As shown in Table VI, the residual errors were within the JND for 0°. For the other six directions, the residual errors for most cases were also within the JND. However, for N1, residual errors greater than the JND were observed for 180° for the left ear of subject OIS (0.20 octaves), 150° for the right ear of OIS (0.16 octaves), and 120° for the right ear of CKT (−0.16 octaves).

Therefore, the best-matching HRTFs can be considered to have spectral features similar to those of each subject's own HRTFs for not only the front direction, but also most of the other directions in the upper median plane.

C. Accuracy of the best-matching HRTF with respect to perceptual aspect

In order to examine the validity of the best-matching HRTFs, sound localization tests in the upper median plane were carried out.

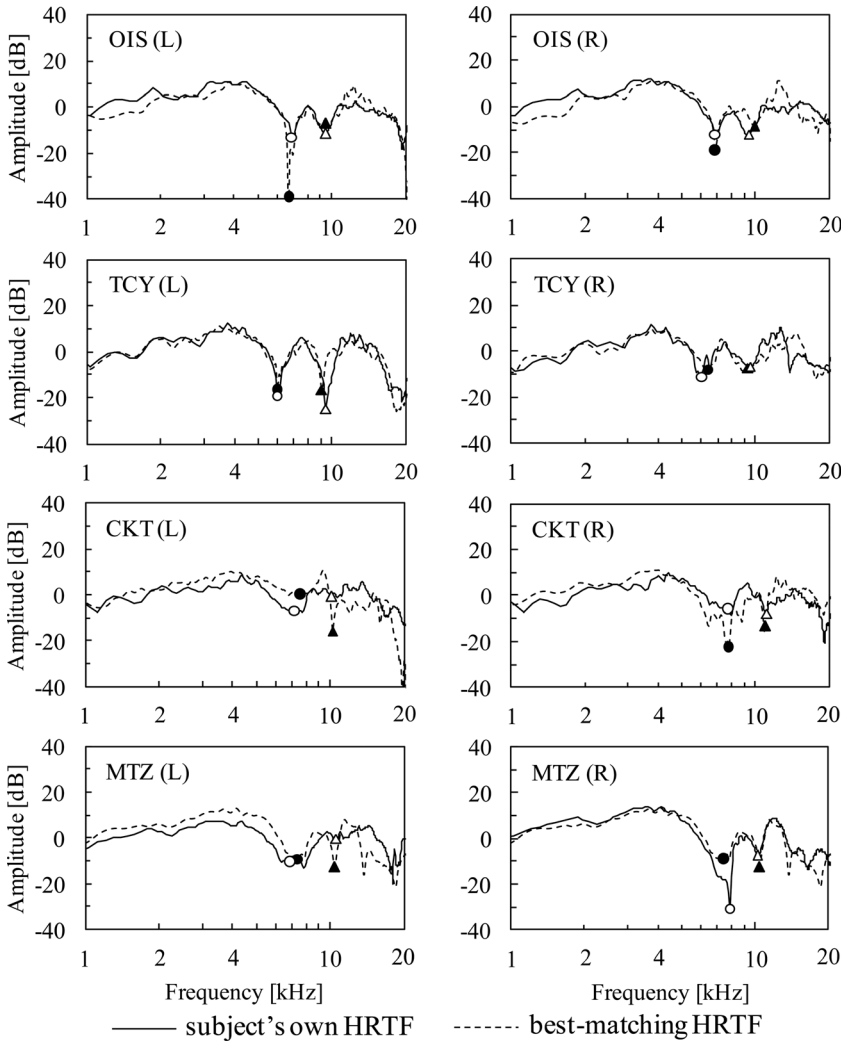

FIG. 5. Amplitude spectrum of best-matching HRTFs (broken line) and subjects' own HRTFs (solid line) for the front direction. ●, N1 (best-matching); ○, N1 (own); ▲, N2 (best-matching); △, N2 (own).

1. Method of sound localization tests

a. Sound localization tests using HRTFs. Four subjects (OIS, TCY, CKT, and MTZ) participated in the sound localization tests. The following two types of HRTFs were used: (1) each subject's own measured HRTFs and (2) each subject's best-matching HRTFs.

The localization tests were conducted in a quiet soundproof room. The working area of the room was 4.6 m wide, 5.8 m deep, and 2.8 m high. The background A-weighted sound pressure level (SPL) was 19.5 dB. A notebook computer (DELL XPS M1330), an audio interface (RME Fireface 400), an amplifier (Marantz PM4001), open-air headphones (AKG K1000), the ear microphones described in Sec. II B, and an A/D converter (Roland M-10MX) were used for the localization tests.

The subjects sat at the center of the soundproof room. The ear microphones were placed into the ear canals of the subject. The diaphragms of the microphones were located at the entrances of the ear canals in the same manner as in the HRTFs measurements described in Sec. II B. The subjects wore the open-air headphones, and the maximum length

TABLE VI. N1 and N2 frequencies of best-matching HRTFs and subjects' own HRTFs for the front direction.

		Best-matched frequency [Hz]		Extracted frequency [Hz]		Residual error [oct.]	
Subject	Ear	N1	N2	N1	N2	N1	N2
OIS	L	6844	9375	6938	9375	−0.02	0.00
	R	6938	9844	6938	9281	0.00	0.08
TCY	L	6094	9188	6094	9656	0.00	−0.07
	R	6469	9188	6094	9375	0.09	−0.03
CKT	L	7406	10 594	6844	10 406	0.11	0.03
	R	7500	10 875	7219	11 250	0.06	−0.05
MTZ	L	7219	10 313	6750	10 875	0.10	−0.08
	R	7219	10 313	7313	10 125	−0.02	0.03

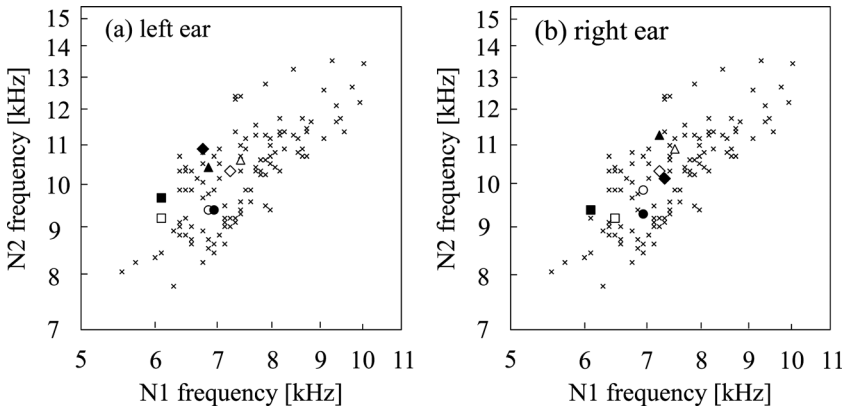

FIG. 6. Scatterplot of best-matching HRTFs and subjects' own HRTFs on the N1-N2 plane for the front direction. (a) left ear; (b) right ear. ●, OIS (best-matching); ○, OIS (own); ■, TCY (best-matching); □, TCY (own); ▲, CKT (best-matching); △, CKT (own); ◆, MTZ (best-matching); ◇, MTZ (own); ×, other HRTFs in the database.

sequence signals (48 kHz sampling, 12th order, and no repetitions) were emitted through the headphones. The signals were received by the ear microphones, and the transfer functions between the open-air headphones and the ear microphones were obtained. The sound pressure at the eardrum for the open-ear-canal condition can be obtained by processing the sound pressure at the entrance of blocked ear canal with the compensation, G (Møller *et al.*, 1995),

$$G = \frac{1}{M \times PTF} \frac{Z_{\text{ear canal}} + Z_{\text{headphone}}}{Z_{\text{ear canal}} + Z_{\text{radiation}}}, \tag{7}$$

where M is the transfer function of the microphone, PTF is the electroacoustic transfer function of the headphones measured at the entrance of the blocked ear canal, $Z_{\text{ear canal}}$ and $Z_{\text{headphone}}$ denote the impedance of the ear canal and headphones, respectively, and $Z_{\text{radiation}}$ is the free-air radiation impedance seen from the ear canal. The second term of G is referred to as the pressure division ratio (PDR). They also showed that the PDR of the headphones (AKG K1000), which were used in the sound localization tests, can be regarded as unity.

The ear microphones were then removed without displacing the headphones because the pinnae of the subject were not enclosed by the headphones. The stimuli $P_{l,r}(\omega)$ were delivered through the headphones as follows:

$$P_{l,r}(\omega) = S(\omega) \times HRTF_{l,r}(\omega)/(M_{l,r}(\omega) \times PTF_{l,r}(\omega)), \tag{8}$$

where $\boldsymbol{S}(\omega)$, l, and r denote the source signal, the left ear, and the right ear, respectively. The source signal was a wideband Gaussian white noise from 200 Hz to 17 kHz. The compensation was processed from 200 Hz to 17 kHz with a frequency resolution of 11.7 ($48\,000/2^{12}$) Hz. The typical peak-to-peak range of the transfer functions between the open-air headphones and the ear microphones from 200 Hz to 17 kHz was approximately 20 dB. This was reduced to 3 dB by the compensation. No regularization was needed in the division process.

The target vertical angles were seven directions, in steps of 30°, in the upper median plane. Stimuli were delivered at 63 dB SPL at the entrance of each ear (interaural level difference = 0). The interaural time difference of the stimuli was also set to 0. The duration of the stimuli was 1.2 s, including the rise and fall times, each of which were 0.1 s.

TABLE VII. Residual errors in the N1 and N2 frequencies between best-matching HRTFs and subjects' own HRTFs for each ear for seven vertical angles in the upper median plane (oct.).

			Vertical angle (°)						
Subject	Ear	Notch	0	30	60	90	120	150	180
OIS	L	N1	−0.02	0.10	−0.09	−0.03	−0.04	0.13	0.20
		N2	0.00	0.14	0.07	−0.04	0.06	0.11	0.08
	R	N1	0.00	0.05	0.00	0.08	0.06	0.16	0.00
		N2	0.08	0.06	0.01	0.03	0.04	0.00	0.00
TCY	L	N1	0.00	−0.04	−0.02	0.00	−0.07	0.01	−0.05
		N2	−0.07	0.01	−0.03	0.06	0.02	−0.02	0.08
	R	N1	0.09	0.00	0.00	−0.07	−0.05	−0.01	−0.03
		N2	−0.03	0.08	0.00	0.08	0.00	−0.10	0.05
CKT	L	N1	0.11	0.13	−0.03	−0.08	−0.09	−0.07	−0.09
		N2	0.03	−0.01	0.06	0.11	0.04	−0.07	−0.01
	R	N1	0.06	0.02	−0.12	−0.04	−0.16	−0.03	−0.06
		N2	−0.05	−0.08	−0.14	−0.03	−0.04	0.12	0.04
MTZ	L	N1	0.10	0.07	0.08	−0.06	−0.02	0.07	0.08
		N2	−0.08	0.08	−0.03	−0.07	−0.06	0.10	0.06
	R	N1	−0.02	0.09	0.00	−0.06	0.05	−0.09	−0.08
		N2	0.03	0.03	0.01	−0.01	−0.03	0.05	−0.05

The mapping method was adopted as a response method in order to prevent the listener from estimating the target direction. A circle and an arrow, which indicated the median plane, were shown on the display of a laptop computer. The subject's task was to click on the perceived vertical angle on the circle on the computer display using a stylus pen. Each subject was also instructed to check the box on the display when he/she perceived a sound image inside his/her head.

The subject's own and best-matching HRTFs were tested separately. Møller *et al.* (1995) demonstrated that no significant difference was observed in localization performance for the same set of HRTFs between separate tests and mixed tests comparing the following two conditions: (1) the HRTF set of one subject was randomized, and (2) the HRTF sets of several subjects were randomized.

In a test trail, 35 stimuli (7 directions × 5 times) were randomized and presented to a subject. The duration of one trial was approximately 7 min. Each subject carried out two trials using his/her own and best-matching HRTFs. Therefore, each subject responded to each stimulus 10 times. The localization tests were carried out using a double-blind method.

b. Sound localization tests using real sound sources. Sound localization tests in the upper median plane using real sound sources were carried out in advance of the tests using the HRTFs. The purpose of these tests was to confirm the subject's basic ability with regard to the upper median plane localization. The tests were carried out in an anechoic chamber. The source signal was a wide-band white noise from 200 Hz to 17 kHz. The stimuli were presented by one of the loudspeakers of 80 mm in diameter (FOSTEX FE83E) located in the upper median plane in 30° steps (seven directions) in random order. The distance from the loudspeakers to the center of the subject's head was 1.2 m. The one-third octave band levels of the loudspeakers were equalized in the range of 1 dB from the center frequency of 250 Hz to 16 kHz. The mapping method was adopted as a response method. The loudspeakers were not visible in the localization tests because the anechoic chamber was darkened except for a small light that was necessary in order to allow the subjects to draw the perceived vertical angle on the response sheet. Each subject responded the vertical angle for each stimulus 10 times.

2. Results of the localization tests

a. Responses to real sound sources and HRTFs. Figures 7–10 show the responses to the real sound sources, the subject's own HRTFs, and the subject's best-matching HRTFs for the four subjects. The ordinate represents the perceived vertical angle, and the abscissa represents the target vertical angle. The diameter of each circle is proportional to the histograms of responses with a resolution of 5°.

For subject OIS (Fig. 7), the responses to the real sound sources (a) were distributed as an s-shaped curve centered over a diagonal line. As in the case of the real sound sources, the responses with the subject's own measured HRTFs (b) also produced an s-shaped curve. The latter, however, tended to shift slightly upward for the target vertical angles of 60° and 120°. For the best-matching HRTFs (c), the perceived vertical angles were approximately the same as those for the subject's own HRTFs at the target vertical angle of 0°, for which the N1 and N2 frequencies were estimated. The distribution of the responses was approximately the same as that of the subject's own HRTFs at the target vertical angles of 30°, 60°, and 180°. However, the responses tended to localize to between 120° and 150° for the target vertical angles of 90° and 120°, and upward for a target vertical angle of 150°.

For subject TCY (Fig. 8), the responses to the real sound sources (a) were distributed along a diagonal line, however, some of the responses tended to localize to the rear for the target vertical angles of 90° and 120°, and upward for a target vertical angle of 150°. For the subject's own measured HRTFs (b), most of the responses were distributed along a diagonal line, whereas the variances of the responses were larger than those of the real sound sources at the target vertical angles of 120° and 150°. For the best-matching HRTFs (c), the perceived vertical angles were approximately the same as those for the subject's own HRTFs at the target vertical angle of 0°, for which the N1 and N2 frequencies were estimated. The responses were distributed along a diagonal line at the target vertical angles of 30°, 60°, and 180°. For 90°, the variance of the responses for the best-matching HRTFs was larger than that for the subject's own HRTFs, but was approximately the same as that for the real sound source. For 120°, the variance was approximately the same as that for the subject's own HRTFs, however, the responses tended

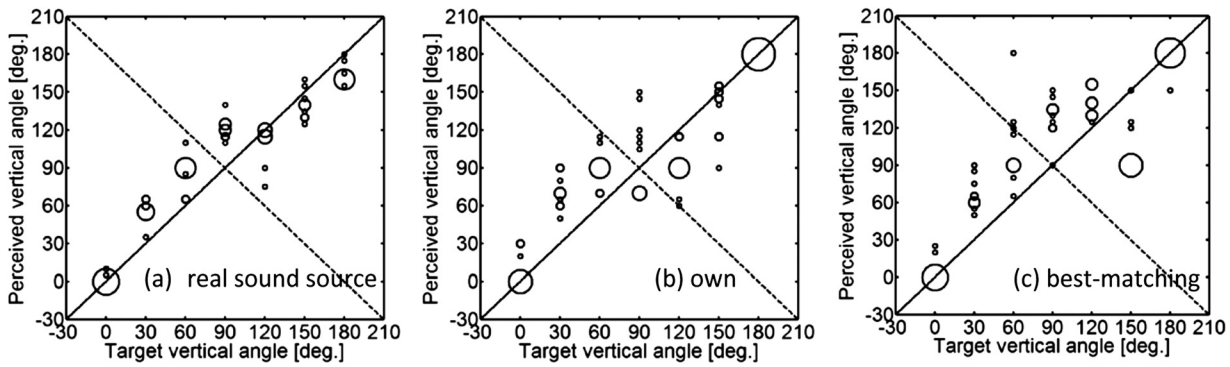

FIG. 7. Responses of subject OIS to (a) the real sound sources, (b) the subject's own HRTFs, and (c) the best-matching HRTFs. The ordinate represents the perceived vertical angle, and the abscissa represents the target vertical angle. The diameter of each circle is proportional to the histograms of responses with a resolution of 5°.

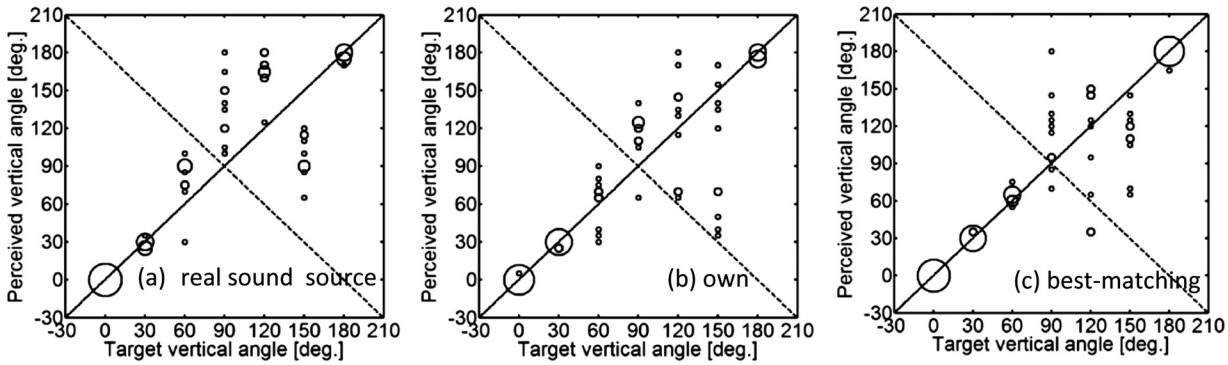

FIG. 8. Responses of subject TCY to (a) the real sound sources, (b) the subject's own HRTFs, and (c) the best-matching HRTFs.

to shift forward. For 150°, the variance of the responses was smaller than that for the subject's own HRTFs.

The responses of subject CKT (Fig. 9) to the real sound sources (a) were distributed along a diagonal line. However, in one instance she localized upward for the target vertical angle of 0°. For the subject's own measured HRTFs (b), most of the responses were distributed along a diagonal line, whereas the variance of the responses was larger than that for the real sound sources at the target vertical angle of 150°. For the best-matching HRTFs (c), the perceived vertical angles were approximately the same as those for the subject's own HRTFs at the target vertical angle of 0°, for which the N1 and N2 frequencies were estimated. The distribution of the responses was approximately the same as that of the subject's own HRTFs at target vertical angles of 150° and 180°. For a target angle of 30°, the responses tended to shift upward. At the other three target vertical angles (60°, 90°, and 120°), the variances of the responses were larger than those of the subject's own HRTFs.

For subject MTZ (Fig. 10), the responses to the real sound sources (a) were distributed as an s-shaped curve centered over a diagonal line. The responses with the subject's own measured HRTFs (b) also produced an s-shaped curve, as did the real sound sources. The latter, however, tended to shift slightly rearward for the target vertical angles of 90° and 120°. For the best-matching HRTFs (c), the perceived vertical angles were approximately the same as those for the subject's own HRTFs at the target vertical angle of 0°, for which the N1 and N2 frequencies were estimated. The distribution of the responses was approximately the same as that for the subject's own HRTFs at the target vertical angles of 60°, 120°, 150°, and 180°. However, the variances of the responses were larger than those of the subject's own HRTFs at target vertical angles of 30° and 90°.

b. Mean localization error. The mean localization errors for each subject, HRTF, and target vertical angle were calculated (Table VIII). The localization error is defined as the absolute difference between the perceived and target vertical angles averaged over the number of repetitions. Regardless of the HRTFs, the mean localization error tended to be small for the directions near the horizontal plane (0° and 180°) and large for the elevated directions, as reported by Carlile *et al.* (1997) and Majdak *et al.* (2010).

For subject OIS, the mean localization error of the best-matching HRTF for target vertical angles of 0° and 180° were 5.2° and 3.4°, respectively. These values are comparable to those of the subject's own HRTFs. However, the error for the target vertical angle of 150° was 47.3°, which is approximately three times that of the subject's own HRTFs. This might be caused by the large difference in the N1 frequency of the right ear (0.16 octaves) and the left ear (0.13 octaves), as shown in Table VII. Another large difference in the N1 frequency of the left ear (0.20 octaves) for 180° is shown in Table VII. However, the N1 and N2 frequencies of the best-matching HRTFs of the right ear were exactly same as the subject's own HRTFs. This may explain why the mean localization error was small (3.4°) for 180°.

For subject TCY, the mean localization error of the best-matching HRTF for the target vertical angles of 0° and 180° were 0.3° and 2.4°, respectively. These values are comparable to those of the subject's own HRTFs. The mean localization errors of the best-matching HRTFs for the other

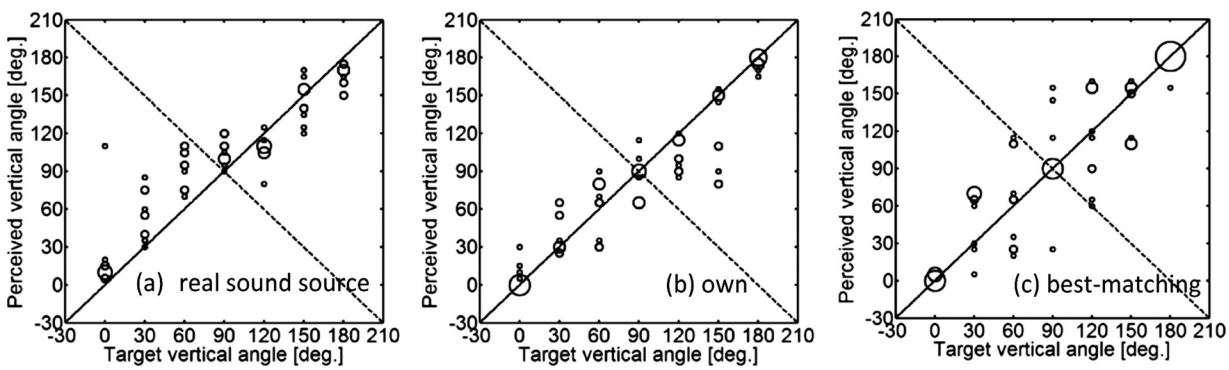

FIG. 9. Responses of subject CKT to (a) the real sound sources, (b) the subject's own HRTFs, and (c) the best-matching HRTFs.

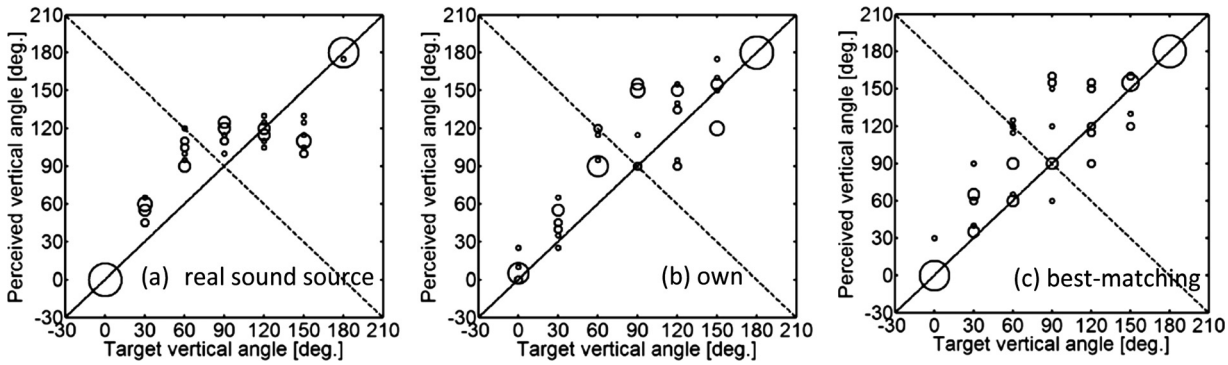

FIG. 10. Responses of subject MTZ to (a) the real sound sources, (b) the subject's own HRTFs, and (c) the best-matching HRTFs.

five directions were also comparable to or smaller than those of the subject's own HRTFs.

For subject CKT, the mean localization error of the best-matching HRTF for the target vertical angles of 0° and 180° were 2.1° and 3.5°, respectively. These values are comparable to those for the subject's own HRTFs. However, the error for the target vertical angle of 30° was 29.2°, which is approximately twice the error of the subject's own HRTFs. This might be caused by the large difference in the N1 frequency of the left ear (0.13 octaves), as shown in Table VII. The error for the target vertical angle of 120° was 31.9°. This is also approximately twice the error of the subject's own HRTFs. This might be caused by the large difference in the N1 frequency of the right ear (−0.16 octaves).

For subject MTZ, the mean localization error of the best-matching HRTF for the target vertical angles of 0° and 180° were 3.3° and 0.5°, respectively. These values are comparable to those for the subject's own HRTFs. The mean localization errors of the best-matching HRTFs for the other five directions were comparable to or smaller than those of the subject's own HRTFs, except for a target vertical angle of 30°.

c. Ratio of front-back confusion. Table IX shows the ratio of front-back confusion for each subject, HRTF, and target vertical angle. The ratio of front-back confusion is defined as the ratio of the responses for which the subjects localized a sound image in the quadrant opposite that of the target direction in the upper median plane.

For all four subjects, the ratio of front-back confusion of the best-matching HRTFs for target vertical angles of 0° and 180° were 0%. These values are same to those for the subject's own HRTFs. The ratio of front-back confusion of the best-matching HRTFs for the other five directions were comparable to those for the subject's own HRTFs. However, the ratios of best-matching HRTFs were higher for 150° for OIS, 60° for CKT, and 120° for MTZ.

Chi-square tests were then performed to clarify whether the differences in the ratio of front-back confusion averaged over seven target directions between the real sound source, the subject's own HRTF, and the subject's best-matching HRTF are statistically significant. Table X shows that all of the p-values were larger than 0.05. Namely, no statistically significant differences in the ratio of front-back confusion appeared between the real sound source, the subject's own HRTF, and the subject's best-matching HRTF.

d. Ratio of inside-of-head localization. All four of the subjects reported never to have perceived a sound image inside their heads for either the subject's own HRTFs or the subject's best-matching HRTFs.

e. Conclusions of the sound localization tests. The results mentioned above demonstrate that the best-matching

TABLE VIII. Mean localization errors for each subject, HRTF, and target vertical angle (°).

		Target vertical angle (°)							
Subject	HRTF	0	30	60	90	120	150	180	Ave.
OIS	real sound source	2.5	25.9	27.4	32.0	9.5	12.4	17.5	18.2
	own HRTF	8.7	40.4	30.5	28.1	29.7	16.4	0.5	22.0
	best-matched HRTF	5.2	36.7	44.3	39.8	20.4	47.3	3.4	28.2
TCY	real sound source	0.4	2.8	25.7	45.7	44.0	52.2	2.8	24.8
	own HRTF	0.7	1.8	17.3	29.5	34.2	56.8	3.0	20.5
	best-matched HRTF	0.3	1.9	4.9	30.5	36.7	40.0	2.4	16.7
CKT	real sound source	20.8	25.2	32.9	14.8	13.3	13.2	15.4	19.4
	own HRTF	5.9	13.6	19.1	12.1	17.4	29.0	4.1	14.4
	best-matched HRTF	2.1	29.2	30.8	21.3	31.9	17.3	3.5	19.5
MTZ	real sound source	0.9	25.1	41.6	26.9	5.4	39.0	0.8	20.0
	own HRTF	5.9	16.6	39.2	46.2	25.4	17.1	1.1	21.6
	best-matched HRTF	3.3	24.8	27.9	38.5	20.6	13.1	0.5	18.4

TABLE IX. Ratio of front-back confusion for each subject, HRTF, and target vertical angle (%).

Subject	HRTF	Target vertical angle (°) 0	30	60	90	120	150	180	Ave.
OIS	real sound source	0	0	40	-	10	0	0	8.3
	own HRTF	0	20	70	-	20	10	0	20.0
	best-matched HRTF	0	10	60	-	0	30	0	16.7
TCY	real sound source	0	0	10	-	0	20	0	5.0
	own HRTF	0	0	10	-	30	50	0	15.0
	best-matched HRTF	0	0	0	-	30	20	0	8.3
CKT	real sound source	10	0	60	-	10	0	0	13.3
	own HRTF	0	0	10	-	20	30	0	10.0
	best-matched HRTF	0	0	30	-	20	0	0	8.3
MTZ	real sound source	0	0	100	-	0	0	0	16.7
	own HRTF	0	0	100	-	0	0	0	16.7
	best-matched HRTF	0	10	50	-	20	0	0	13.3

HRTFs provided approximately the same performance for the perception of the vertical angle as the subject's own HRTFs for the target vertical angle of 0°, for which the N1 and N2 frequencies were estimated. For the target vertical angle of 180°, the best-matching HRTFs provided approximately the same performance of vertical perception as for the target vertical angle of 0°. For the upper target directions, however, the performance of the localization for some of the subjects decreased as compared with the subject's own HRTFs.

D. Comparison of localization accuracy between best-matching HRTFs and non-individualized HRTFs

In order to verify the improvement in localization accuracy by the proposed method for the personalization of HRTFs, sound localization tests in the upper median plane using non-individualized HRTFs were carried out.

The HRTFs of subjects TCY and OIS were chosen as the non-individualized HRTFs. The other three persons participated in the tests as subjects for each non-individualized HRTF. The test method was the same as that described in Sec. V C 1 a. Figures 11 and 12 show the responses of three subjects to the HRTFs of OIS and TCY, respectively.

TABLE X. Results of chi-square tests for the ratio of front-back confusion.

Comparison between	Subject	p-value
own HRTF and best-matching HRTF	OIS	0.64
	TCY	0.26
	CKT	0.75
	MTZ	0.61
real sound source and best-matching HRTF	OIS	0.17
	TCY	0.71
	CKT	0.38
	MTZ	0.61
real sound source and own HRTF	OIS	0.067
	TCY	0.068
	CKT	0.57
	MTZ	1.0

For the HRTFs of subject OIS, the responses of subject TCY [Fig. 11(a)] localized around target vertical angles of 0° and 180°, as for the best-matching HRTFs of subject TCY [Fig. 8(c)]. For target vertical angles of 30°, 60°, 90°, and 150°, however, the variances of the responses were larger than those for the best-matching HRTFs of subject TCY. The responses of subject CKT [Fig. 11(b)] localized to 0°, 120°, and 150° for a target vertical angle of 0°, whereas the responses for the best-matching HRTFs of subject CKT [Fig. 9(c)] localized around 0°. The responses were distributed between 60° and 150° for target vertical angles of 60°, 90°, 120°, and 150°. For a target angle of 180°, the distribution of the responses was approximately the same as that for the best-matching HRTFs of subject CKT. Subject MTZ [Fig. 11(c)] sometimes localized to the front and at other times localized to the rear for target vertical angles of 0° and 180°, whereas the responses for the best-matching HRTFs of subject MTZ were distributed around the target vertical angles [Fig. 10(c)]. The responses were distributed between 45° and 150° for the target vertical angle of 30° and between 90° and 180° for target vertical angles of 60°, 90°, and 120°.

For the HRTFs of TCY, most of the responses of subject OIS [Fig. 12(a)] were distributed along a diagonal line as for the best-matching HRTFs of subject OIS [Fig. 7(c)]. However, in one instance, subject OIS localized to the rear for a target vertical angle of 0°. Subject CKT [Fig. 12(b)] localized to rear for a target vertical angle of 0° and localized between 60° and 90° for target vertical angles of 60°, 90°, 120°, and 150°. For a target vertical angle of 180°, the distribution of the responses was approximately the same as that for the best-matching HRTFs of subject CKT [Fig. 9(c)]. Subject MTZ [Fig. 12(c)] localized to the rear for a target vertical angle of 0°. The responses were distributed between 90° and 180° for target vertical angles of 90°, 120°, and 150°. For a target vertical angle of 180°, the responses were widely distributed between 0° and 180°.

Table XI shows the mean localization errors for the HRTFs of subjects OIS and TCY and those for the subjects' best-matching HRTFs transcribed from Table VIII, and Table XII shows the ratios of front-back confusion for the

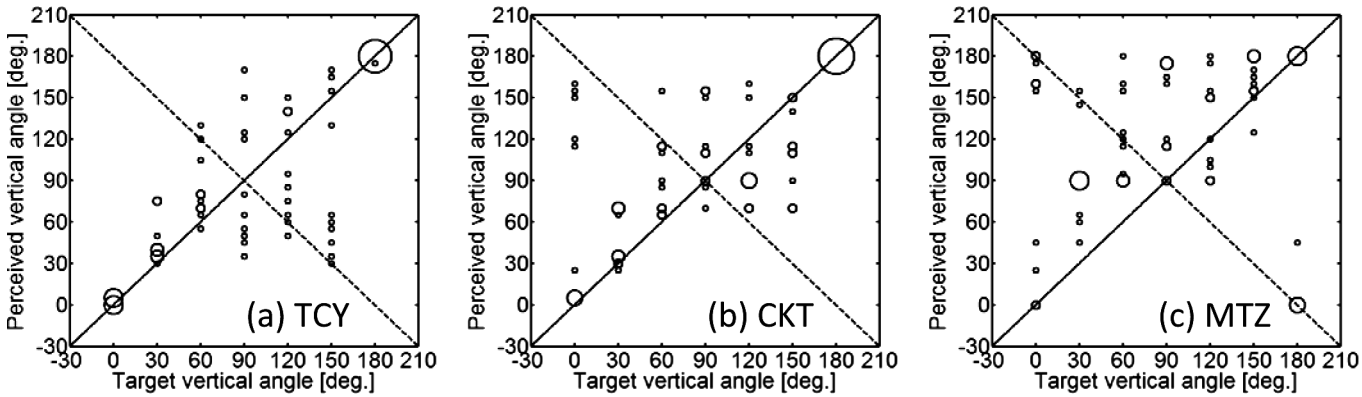

FIG. 11. Responses of subjects (a) TCY, (b) CKT, and (c) MTZ to HRTFs of subject OIS. The ordinate represents the perceived vertical angle, and the abscissa represents the target vertical angle. The diameter of each circle is proportional to the histograms of responses with a resolution of 5°.

HRTFs of subjects OIS and TCY and those for the subjects' best-matching HRTFs transcribed from Table IX.

For subject OIS, the mean localization errors for the HRTFs of subject TCY were larger than those for the best-matching HRTFs for target vertical angles of 0°. This was due to a front-back confusion. For 120° and 180°, the errors for the HRTFs of subject TCY were comparable to those for the best-matching HRTFs. For 30°, 60°, 90°, and 150°, the errors for the HRTFs of subject TCY were smaller than those for the best-matching HRTFs. As shown in Fig. 12(a), the HRTFs of subject TCY provided approximately the same localization performance for subject OIS as the best-matching HRTFs of subject OIS.

For subject TCY, the mean localization errors for the HRTFs of subject OIS were larger than those for the best-matching HRTFs for target vertical angles of 30°, 60°, 90°, and 150°. For the other target vertical angles, however, the errors for the HRTFs of subject OIS were comparable to those for the best-matching HRTFs.

For subject CKT, the mean localization errors for the HRTFs of subject OIS were larger than those for the best-matching HRTFs for target vertical angles of 0°, 90°, and 150°. The large error for 0° was due to the high ratio of front-back confusion (50%), as shown in Table XII. For 30°, however, the error for the HRTFs of subject OIS was smaller than that of the best-matching HRTFs. The mean localization errors for the HRTFs of subject TCY were larger than those for the best-matching HRTFs for target angles of 0°, 120°, and 150°. These large errors were also due to high ratios of front-back confusion (100%, 100%, and 90%, respectively). For 30°, 60°, and 90°, however, the errors for the HRTFs of subject OIS were smaller than those for the best-matching HRTFs.

For subject MTZ, all of the mean localization errors for the HRTFs of subjects OIS and TCY were larger than those for the best-matching HRTFs, except for a target angle of 90° for the HRTFs of subject TCY. The large errors for the HRTFs of both subjects OIS and TCY for 0° and 180° were due to high ratios of front-back confusion [OIS: 60% (0°) and 50% (180°), TCY: 100% (0°) and 50% (180°)].

The values of the mean localization errors averaged over seven target directions for the HRTFs of other subjects were larger than the subject's best-matching HRTFs, except for subject OIS. Moreover, for all subjects, the values of the ratios of front-back confusion averaged over seven target directions for the HRTFs of other subjects were larger than those for the subject's best-matching HRTFs.

Chi-square tests were then performed in order to clarify whether the differences in the ratio of front-back confusion averaged over seven target directions between the subject's best-matching HRTFs and the HRTFs of other subjects are statistically significant. Table XIII shows the results of the tests. The ratios of front-back confusion of the best-matching HRTFs were significantly ($p < 0.05$) smaller than those of the other's HRTFs for all combinations of the subject's best-matching HRTFs and the HRTFs of other subjects, except for the combination of the best-matching HRTFs of subject OIS and the HRTFs of subject TCY.

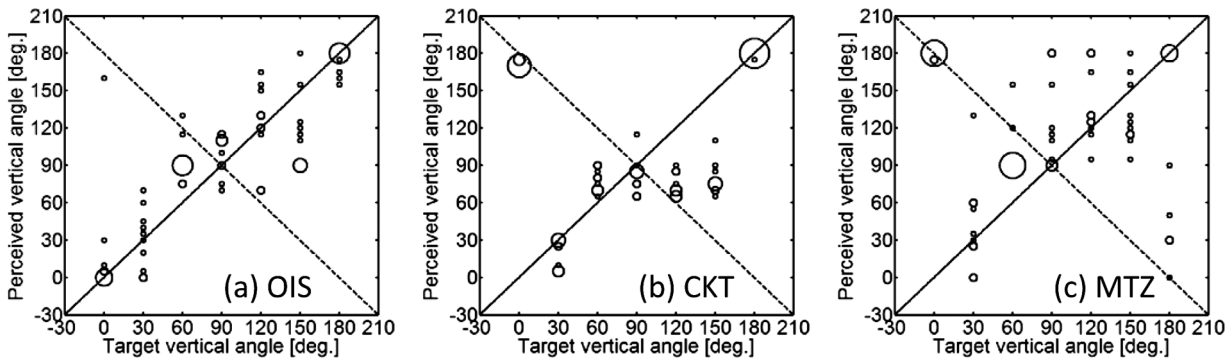

FIG. 12. Responses of subjects (a) OIS, (b) CKT, and (c) MTZ to HRTFs of subject TCY.

TABLE XI. Mean localization errors for HRTFs of OIS and TCY and those for subjects' best-matching HRTFs transcribed from Table VIII (°).

		Target vertical angle (°)							
Subject	HRTF	0	30	60	90	120	150	180	Ave.
OIS	TCY	20.6	19.9	33.0	16.3	23.3	39.5	7.1	22.8
	best-matching	5.2	36.7	44.3	39.8	20.4	47.3	3.4	28.2
TCY	OIS	2.5	15.2	26.3	42.2	36.0	66.0	1.2	27.0
	best-matching	0.3	1.9	4.9	30.5	36.7	40.0	2.4	16.7
CKT	OIS	75.0	17.9	33.8	27.4	30.7	38.4	1.0	32.0
	TCY	172.1	10.2	18.5	12.3	46.2	71.2	1.3	47.4
	best-matching	2.1	29.2	30.8	21.3	31.9	17.3	3.5	19.5
MTZ	OIS	108.2	62.4	61.7	48.3	30.6	17.9	85.9	59.3
	TCY	178.6	25.7	40.2	32.0	22.8	29.4	70.6	57.0
	best-matching	3.3	24.8	27.9	38.5	20.6	13.1	0.5	18.4

All the subjects reported never to have perceived a sound image inside their heads for the HRTFs of other subjects. Subject CKT, however, reported the distance of the sound image to be very near, only just outside of her head, for other subjects' HRTFs.

The above-mentioned results indicate that the best-matching HRTFs provide a substantial improvement compared to non-individualized HRTFs in performance with regard to vertical perception.

VI. DISCUSSION

A. Comparison with the previous HRTF personalization method

In the proposed method, six anthropometric parameters of the listener's actual pinnae are measured. The N1 and N2 frequencies are then estimated using the multiple regression equation, and the best-matching HRTFs are selected from the database.

Here, we compare the proposed method with previous methods with respect to the duration of the procedure. For the proposed method, all of the personalization processes require only 3 min to perform. This is remarkably faster than previous methods. Middlebrooks *et al.* (2000) reported that finding a listener's preferred scale factor required one to three 20-min blocks of listening tests. Seeber and Fastl (2003) proposed a two-step selection procedure to find the appropriate HRTFs from the non-individualized HRTFs. They reported their method took approximately 10 min. Iwaya (2006) claimed that selecting the most appropriate HRTF set among 32 sets of HRTFs in tournament-style listening tests required approximately 15 min.

The proposed method requires neither trained personnel nor special equipment. Using Fig. 3 as a reference, any untrained individual can measure the anthropometric parameters of a listener.

B. Expansion to an arbitrary direction in three-dimensional space

In the present study, we proposed a method by which to provide the personalized HRTFs for the front direction. The HRTFs for the other directions in the upper median plane were then provided by a donor, for which the HRTF for the front direction was selected as the best match.

Furthermore, we attempted to control the sound image to an arbitrary direction in three-dimensional space. Morimoto and Aokata (1984) showed that a listener perceives the lateral and vertical angles of a sound image independently and that a listener uses the interaural difference cues to perceive the lateral angle and the spectral cues to perceive the vertical angle. Furthermore, Morimoto *et al.* (2003) demonstrated that sound image control for an

TABLE XII. Ratios of front-back confusion for HRTFs of OIS and TCY and those for subjects' best-matching HRTFs transcribed from Table IX (%).

		Target vertical angle (°)							
Subject	HRTF	0	30	60	90	120	150	180	Ave.
OIS	TCY	10	0	60	-	20	20	0	18.3
	best-matching	0	10	60	-	0	30	0	16.7
TCY	OIS	0	0	30	-	50	60	0	23.3
	best-matching	0	0	0	-	30	20	0	8.3
CKT	OIS	50	0	40	-	30	20	0	23.3
	TCY	100	0	20	-	100	90	0	51.7
	best-matching	0	0	30	-	20	0	0	8.3
MTZ	OIS	60	60	90	-	0	0	50	43.3
	TCY	100	10	80	-	0	0	50	40.0
	best-matching	0	10	50	-	20	0	0	13.3

TABLE XIII. Results of chi-square tests for the ratio of front-back confusion between the subject's best-matching HRTFs and the other subject's HRTFs.

Subject's best-matching HRTF	Other's HRTF	p-value
OIS	TCY	0.81
TCY	OIS	0.024
CKT	OIS	0.024
	TCY	2.2E-07
MTZ	OIS	2.7E-04
	TCY	9.6E-04

arbitrary direction in three-dimensional space can be accomplished by reproducing the HRTFs in the median plane combined with the interaural time difference. These results infer the feasibility of sound image control for an arbitrary direction in three-dimensional space by combining the best-matching HRTFs in the median plane with the interaural time difference. This remains as an area for future research.

VII. CONCLUSIONS

In order to accomplish accurate sound image control, we herein proposed a method for estimating the frequencies of N1 and N2 in the HRTF, which play an important role in vertical localization, for an individual listener based on the anthropometry of the listener's pinna. A method was also proposed for selecting the best-matching HRTFs, for which N1 and N2 are closest to the estimates, from a database. The validity of these methods was examined based on both physical and perceptual aspects. The conclusions are as follows:

(1) The N1, N2, and P1 frequencies of 54 ears of Japanese adults were extracted for the front direction. The individual differences were 0.74, 0.71, and 0.31 octaves, respectively.
(2) Multiple regression analyses were carried out using 54 ears as objective variables of the listener's N1 and N2 frequencies for the front direction and as explanatory variables of the six anthropometric parameters of the listener's pinna. The multiple correlation coefficients for N1 and N2 were 0.81 and 0.82, respectively.
(3) For four naive subjects, the N1 and N2 frequencies for the front direction were estimated based on the six anthropometric parameters. The residual errors of N1 and N2 were less than the JND for all eight ears.
(4) Localization tests in the upper median plane were carried out for the four subjects. The results indicate that the best-matching HRTFs provided approximately the same performance with respect to the perception of vertical angle as the subjects' own HRTFs for the target vertical angle of 0° (front), for which the N1 and N2 frequencies were estimated.
(5) For 180° (rear), the best-matching HRTFs provided approximately the same performance as for the target vertical angle of 0°.
(6) For the other five upper target directions, the performance of the localization for some of the subjects decreased as compared with the subject's own HRTFs.
(7) All of the personalization processes require only three minutes to perform. This is remarkably faster than previous methods. Neither trained personnel nor special equipment is required for the measurement of anthropometric parameters.

ACKNOWLEDGMENTS

The present study was supported in part by MEXT through the Foundation for Strategic Research at Private Universities (S1311003). The authors would like to thank M. Morimoto, Professor Emeritus at Kobe University, for his fruitful discussions and T. Okamatsu, S. Sakaguchi, and H. Tsuchiya for their cooperation in conducting the measurements of the anthropometry of the pinnae.

APPENDIX

The JNDs of the P1 frequency with regard to vertical angle perception were measured through listening tests (Nishioka *et al.*, 2013). The tests were carried out by the constant method in the same room using the same apparatus, as described in Sec. V C 1 a. The reference HRTFs were simplified HRTFs, which are composed of N1, N2, and P1 extracted from the subject's own HRTF for the front direction (vertical angle: 0°). The comparison HRTFs were the simplified HRTFs, the P1 frequency of which was shifted by ±0.05, ±0.1, ±0.2, and ±0.5 octaves. The source signal was a wide-band Gaussian white noise from 200 Hz to 17 kHz. The reference stimulus and one of the comparison stimuli were presented to a subject as a pair. Each stimulus was presented for 1.2 s, and the interval between the reference and the comparison stimulus was 0.5 s. The subject's task was to indicate whether the perceived vertical angles of the two stimuli were the same. Eight kinds of comparison stimuli were presented in random order. The subjects responded ten times for each pair. The subjects were four males, 23 to 26 yr of age.

The JNDs of the lower and the upper side were obtained as follows. The probability that the perceived vertical angles of the paired stimuli were not the same was calculated for each comparison stimulus using the responses of all of the subjects. The linear regression line was then obtained by z-transformation. The correlation coefficients for the lower and upper sides were 0.94 and 0.85, respectively. The JNDs were obtained as the frequency shift, at which the probability is 50% ($z = 0$). The JNDs for the lower and upper sides were 0.47 and 0.35 octaves, respectively.

The results indicate that the JNDs of the P1 frequencies are larger than the range of individual difference in P1 frequency (0.31 octaves).

Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (**2001**). "The CIPIC HRTF database," in *Proc. IEEE Workshop on Application of Signal Processing to Audio and Acoustics*, New Paltz, NY, pp. 99–102.
Butler, A., and Belendiuk, K. (**1977**). "Spectral cues utilizes in the localization of sound in the median sagittal plane," J. Acoust. Soc. Am. **61**, 1264–1269.
Carlile, S., Leong, P., and Hyams, S. (**1997**). "The nature and distribution of errors in sound localization by human listeners," Hear. Res. **114**, 179–196.

Chatterjee, S. and Hadi, A. S. (**2012**), *Regression Analysis by Example*, 5th ed. (John Wiley and Sons, Hoboken, NJ).
Hebrank, J., and Wright, D. (**1974**). "Spectral cues used in the localization of sound sources on the median plane," J. Acoust. Soc. Am. **56**, 1829–1834.
Hu, H., Chen, L., and Wu, Z. (**2008**). "The estimation of personalized HRTFs in individual VAS," in *Proc. Fourth International Conference on Natural Computation*, Jinan, China, pp. 203–207.
Hugeng, W. W., and Gunawan, D. (**2010**). "Improved method for individualization of head-related transfer functions on horizontal plane using reduced number of anthropometric measurements," J. Telecommun. **2**, 31–41.
Iida, K., and Ishii, Y. (**2011a**). "3D sound image control by individualized parametric head-related transfer functions," in *Proc. Inter-Noise 2011*, Osaka, Japan, 428959.
Iida, K., and Ishii, Y. (**2011b**). "Individualization of the head-related transfer functions in the basis of the spectral cues for sound localization," in *Principles and Applications of Spatial Hearing*, edited by Y. Suzuki, D. Brungard, Y. Iwaya, K. Iida, D. Cabrera, and H. Kato (World Scientific, Singapore), pp. 159–178.
Iida, K., Itoh, M., Itagaki, A., and Morimoto, M. (**2007**). "Median plane localization using parametric model of the head-related transfer function based on spectral cues," Appl. Acoust. **68**, 835–850.
Iwaya, Y. (**2006**). "Individualization of head-related transfer functions with tournament-style listening test: Listening with other's ears," Acoust. Sci. Technol. **27**, 340–343.
Kahana, Y., and Nelson, P. A. (**2006**). "Numerical modeling of the spatial acoustic response of the human pinna," J. Sound Vib. **292**, 148–178.
Katz, B. F. (**2001**). "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation," J. Acoust. Soc. Am. **110**, 2440–2448.
Kistler, D. J., and Wightman, F. L. (**1992**) "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," J. Acoust. Soc. Am. **91**, 1637–1647.
Kreuzer, W., Majdak, P., and Chen, Z. (**2009**). "Fast multipole boundary element method to calculate head-related transfer functions for a wide frequency range," J. Acoust. Soc. Am. **126**, 1280–1290.
Kulkarni, A., and Colburn, H. S. (**1998**). "Role of spectral detail in sound-source localization," Nature **396**, 747–749.
Lopez-Poveda, E. A., and Meddis, R. (**1996**). "A physical model of sound diffraction and reflections in the human concha," J. Acoust. Soc. Am. **100**, 3248–3259.
Majdak, P., Goupell, M. J., and Laback, B. (**2010**). "3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training," Atten. Percept. Psychophys. **72**, 454–469.
Mehrgardt, S., and Mellert, V. (**1977**). "Transformation characteristics of the external human ear," J. Acoust. Soc. Am. **61**, 1567–1576.
Middlebrooks, J. C. (**1999a**). "Individual differences in external-ear transfer functions reduced by scaling in frequency," J. Acoust. Soc. Am. **106**, 1480–1492.
Middlebrooks, J. C. (**1999b**). "Virtual localization improved by scaling non-individualized external-ear transfer functions in frequency," J. Acoust. Soc. Am. **106**, 1493–1510.
Middlebrooks, J. C., and Green, D. M. (**1992**). "Observations on a principal components analysis of head-related transfer functions," J. Acoust. Soc. Am. **92**, 597–599.
Middlebrooks, J. C., Macpherson, E. A., and Onsan, Z. A. (**2000**). "Psychophysical customization of directional transfer functions for virtual sound localization," J. Acoust. Soc. Am. **108**, 3088–3091.
Møller, H., Hammershøi, D., Jensen, C. J., and Sørensen, M. F. (**1995**). "Transfer characteristics of headphones measured on human ears," J. Audio Eng. Soc. **43**, 203–217.
Morimoto, M., and Ando, Y. (**1980**). "On the simulation of sound localization," J. Acoust. Soc. Jpn. (E) **1**, 167–174.
Morimoto, M., and Aokata, H. (**1984**). "Localization cues of sound sources in the upper hemisphere," J. Acoust. Soc. Jpn. (E) **5**, 165–173.
Morimoto, M., Iida, K., and Itoh, M. (**2003**). "Upper hemisphere sound localization using head-related transfer functions in the median plane and interaural differences," Acoust. Sci. Technol. **24**, 267–275.
Musicant, A., and Butler, R. (**1984**). "The influence of pinnae-based spectral cues on sound localization," J. Acoust. Soc. Am. **75**, 1195–1200.
Nishioka, S., Ishii, Y., and Iida, K. (**2013**). "Just noticeable difference of frequency of the first peak of head-related transfer functions with regard to vertical angle perception," in *Proc. Autumn Annual Meeting of the Acoustical Society of Japan*, Toyohashi, Japan, pp. 859–860 (in Japanese).
Seeber, B. U., and Fastl, H. (**2003**). "Subjective selection of non-individual head-related transfer functions," in *Proc. 2003 International Conference on Auditory Display*, Boston, MA, pp. 1–4.
Shaw, E. A. G., and Teranishi, R. (**1968**). "Sound pressure generated in an external-ear replica and real human ears by a nearby point source," J. Acoust. Soc. Am. **44**, 240–249.
Sottek, R., and Genuit, K. (**1999**). "Physical modeling of individual head-related transfer functions," J. Acoust. Soc. Am. **105**(2), 1162.
Takemoto, H., Mokhtari, P., Kato, H., Nishimura, R., and Iida, K. (**2012**). "Mechanism for generating peaks and notches of head-related transfer functions in the median plane," J. Acoust. Soc. Am. **132**, 3832–3841.
Xu, S., Li, Z., and Salvendy, G. (**2008**). "Improved method to individualize head-related transfer function using anthropometric measurements," Acoust. Sci. Technol. **29**, 388–390.
Zhang, M., Kennedy, R. A., Abhayapala, T. D., and Zhang, W. (**2011**). "Statistical method to identify key anthropometric parameters in HRTF individualization," in *Proc. IEEE workshop on hands-free speech communication and microphone arrays*, Edinburgh, UK, pp. 213–218.
Zotkin, D. N., Hwang, J., Duraiswami, R., and Davis, L. S. (**2003**). "HRTF personalization using anthropometric measurements," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, pp. 157–160.