



Estimation of the category of notch frequency bins of the individual head-related transfer functions using the anthropometry of the listener's pinnae



Kazuhiro Iida*, Ori Nishiyama, Tsubasa Aizaki

Faculty of Advanced Engineering, Chiba Institute of Technology, 2-17-1 Tsudanuma, Narashino, Chiba 275-0016, Japan

ARTICLE INFO

Article history:

Received 8 August 2020

Received in revised form 21 November 2020

Accepted 11 January 2021

Keywords:

Head-related transfer function

Individualization

Spectral notch

Pinna

Median plane

Virtual reality

ABSTRACT

An authentic approach to present three-dimensional acoustical sensation to everyone is to provide head-related transfer functions (HRTFs) that are exactly adapted to each listener. This approach is called the individualization of HRTFs. In order to generate individualized HRTFs, the present study proposes a modeling method of HRTFs that extracts the frequency bins in which spectral notches (N1 and N2) are included. The frequency bins for N1 and N2 were classified into two categories based on the just noticeable difference in notch frequency with regard to the vertical angle perception of a sound image. Then, discriminant analyses were carried out as objective variables of the category of each of the N1 and N2 frequency bins and as explanatory variables of ten anthropometric parameters of pinnae. The results show that the categories of N1 and N2 frequency bins for the front direction can be estimated with accuracies of 83.1% and 77.8%, respectively. For the rear direction, the accuracies of the category estimation of N1 and N2 were 74.3% and 77.1%, respectively.

© 2021 Elsevier Ltd. All rights reserved.

1. Introduction

Accurate three-dimensional sound image localization can be accomplished by reproduction of a listener's own head-related transfer functions (HRTFs) at the entrances of the ear canals [21]. However, there exist remarkable individual differences in HRTFs. The HRTFs of other listeners often cause front-back confusion of a sound image and inside-of-head localization. This is a serious problem, which prevents acoustic virtual reality (VR) from coming into widespread practical use. In order to address this problem, methods for obtaining individualized HRTFs of an unknown listener that do not require acoustical measurements have been investigated.

One such method uses principal component analysis (PCA), which resolves an HRTF into its principal components [14,18]. The weighting coefficient of each principal component depends on both the listener and the direction of a sound source. Attempts were made to estimate the weighting coefficients based on the anthropometry of the listener's pinnae using multiple regression analysis [24,1] or a deep neural network [2,16]. However, estima-

tion of the weighting coefficients for an unknown listener has not been successful.

Another method for HRTF individualization estimates the prominent spectral peaks and notches of the individual HRTFs. The minimum HRTF components, which provide approximately the same localization performance as the measured HRTFs, were demonstrated to be the two lowest-frequency notches (N1 and N2) and the two lowest-frequency peaks (P1 and P2) above 4 kHz [10,8]. The frequencies of the N1, N2, P1, and P2 for a sound source at the front direction were reported to be estimated based on the anthropometry of the listener's pinnae using multiple regression analysis [9,26,20]. Then, the HRTFs having frequencies of N1 and N2 that are closest to these frequencies were selected from the HRTF database as the best-matching HRTFs. Sound image localization tests for the seven directions in the upper median plane indicated that the best-matching HRTFs provided approximately the same performance as the subjects' own HRTFs for the target vertical angles of 0° (front) and 180° (rear) [9]. For the other five upper target directions, however, the performance of the localization for some of the subjects decreased compared to the subject's own HRTFs, and so there remains room for improvement.

A number of methods have been considered in which a listener chooses the appropriate HRTFs by performing a listening test. Inter-subject differences in directional transfer functions (DTFs),

* Corresponding author.

E-mail address: kazuhiro.iida@it-chiba.ac.jp (K. Iida).

which are the directional components of HRTFs, could be reduced by appropriately scaling the frequency of one set of DTFs [17]. Furthermore, it has been shown that the optimally frequency-scaled DTF halved the difference in quadrant error between other-ear and own-ear conditions [19]. However, one to three 20-min blocks of listening tests were required in order to find a listener's preferred scale factor. On the other hand, Iida [4] proposed typical HRTF data sets composed of six to ten HRTFs for each of seven directions in the upper median plane. Localization tests indicated that one of the six to ten typical HRTFs provides approximately the same vertical perception of a sound image as the listener's own HRTFs. However, the listener is required to perform a listening test to choose the appropriate HRTF from among typical HRTF data sets.

Numerical calculation of HRTFs has also been studied intensively. The boundary element method (BEM) has been used to calculate HRTFs in a number of studies [13,12,15]. 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 [28]. At present, however, neither the BEM nor the FDTD method is available for ordinary listeners because special equipment, e.g., a magnetic resonance imaging system, is required in order to digitize the complicated shape of the pinnae of an individual listener.

Thus, individual differences in HRTFs have not yet been overcome. Current acoustic VR, which can present three-dimensional acoustical sensation only to a specific listener, must be evolved into a universal system that can present three-dimensional acoustical sensation to everyone, even someone who does not have his/her own HRTF data.

The present study proposes a method by which to estimate the category of the notch frequency bins of the individual HRTF based on the anthropometry of the listener's pinnae using a band-divided notch-peak HRTF model and discriminant analysis.

2. Outline of the proposed method

An outline of the proposed method is as follows:

- (1) Modeling of HRTFs
 - (a) Obtain early HRTFs [6] from measured full-length HRTFs.
 - (b) Generate the modeled HRTFs by the band-divided notch-peak model.
- (2) Extraction and categorization of the N1 and N2 frequency bins
 - (a) Extract the frequency bins, in which N1 and N2 are included using the modeled HRTFs.
 - (b) Classify the N1 and N2 frequency bins into several categories based on the just noticeable difference (jnd) of notch frequency with regard to the vertical angle perception of a sound image [7].
- (3) Measurement of anthropometric parameters of pinnae
- (4) Discriminant analysis of the categories of N1 and N2 frequency bins

Objective variables: category of N1 and N2 frequency bins
Explanatory variables: anthropometric parameters of the pinnae
- (5) Category estimation of N1 and N2 frequency bins for naive ears

Estimate the category of N1 and N2 frequency bins of the ears, which were not involved in the discriminant analysis using the leave-two-out cross-validation method.

3. Modeling of HRTFs

3.1. Generation of early HRTFs

In order to extract the N1, N2, P1, and P2 from the amplitude spectrum of the measured HRTF, in which a number of small notches and peaks are included, the early HRTFs were generated from the measured HRTFs.

The notches and peaks are reported to be generated in the pinna. When the cavities of the pinna are occluded by clay, the notches and peaks vanish, and the front-back confusion of a sound image increases [3,23,11]. Peaks are considered to be generated by the resonances of the pinna [25]. At the notch frequency, the antinodes are generated at the cymba concha and the triangular fossa, and a node is generated at the cavum concha [28].

The effect of the pinnae is considered to be included in the early part of the head-related impulse response (HRIR), because the response from the pinna arrives at the receiving point (the entrance of the ear canal) earlier than that from the torso. Therefore, the information of the outline of N1, N2, P1, and P2 is supposed to be included in the early part of the HRIR.

Iida and Oota [6] reported that the outline of the amplitude spectrum of the early HRIR, the duration of which was 1 ms, was approximately the same as that of the full-length HRIR. They also showed that no statistically significant difference exists in the mean vertical localization error observed between the early HRIRs of 1 ms and the full-length HRIRs at seven target vertical angles in the upper median plane.

Then, early HRTFs were generated using the algorithm, as follows:

- (1) Detect the sample for which the absolute amplitude of the HRIR is maximum, S_{\max} .
- (2) Clip the HRIR using a four-term Blackman-Harris window, the duration of which is 2 ms, adjusting the temporal center of the window to S_{\max} .
- (3) Obtain the early HRTF from the clipped HRIR by discrete Fourier transform.

Fig. 1 shows the amplitude spectra of the full-length HRTFs and the early HRTFs of a subject for the front direction. The full-length HRTF (solid line) includes several prominent notches and peaks and a number of small notches and peaks. On the other hand, the early HRTF (broken line) includes only the prominent notches and peaks, while the small notches and peaks vanished.

3.2. Band-divided notch-peak HRTF model

In order to obtain the frequency bins in which the N1 and N2 are included, the band-divided notch-peak model was used. This model divides the amplitude spectra of the early HRTFs into 1/n-octave bands and decides the representative values of each band,

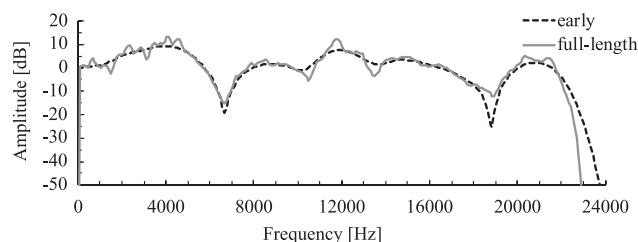


Fig. 1. Example of the amplitude spectrum of the early HRTF (broken line) and the full-length (usual) HRTF (solid line).

i.e., the center frequency and the band level. The band levels were decided using the algorithm, as follows:

- (1) Detect the local maxima and local minima in each band.
- (2) Adopt the amplitude level of the local maximum or the local minimum in the case that there exists either a local maximum or a local minimum in the band.
- (3) Adopt the amplitude level of the local maximum or the local minimum, the absolute amplitude level of which is maximum in the case there exist multiple local maxima or local minima in the band. This means that the model expresses each band by the level of a distinguishing notch or peak, avoiding cancelation of the multiple notches and peaks by taking their average level.
- (4) Adopt the amplitude level of the center frequency in the case there exists neither a local maximum nor a local minimum.

The amplitude spectrum of the modeled HRTF is obtained by linear interpolation of the band level. Fig. 2 shows examples of the amplitude spectra of the modeled HRTFs. Six bandwidths, 1/12, 1/6, 1/3, 1/2, and 1 octave and the critical bandwidth, were used for modeling. The amplitude spectra of the modeled HRTFs were approximately the same as those of the early HRTF, up to 4 kHz. Then, P1 was reproduced accurately.

The reproducibility of the notches and peaks for the larger bandwidth decreased for the frequency range above 4 kHz. N1 did not appear for the bandwidth of 1 octave. The frequencies of N1 for the bandwidths of 1/3 and 1/2 octave and critical bandwidth differed from that of the early HRTF. P2 did not appear for the bandwidth of 1 octave. The frequencies of P2 for the bandwidth of 1/2 octave differed from that of the early HRTF. N2 did not appear for the bandwidths of 1 and 1/2 octave. The frequencies of N2 for the critical bandwidth differed from that of the early HRTF.

3.3. Determination of the bandwidth

The optimal bandwidth for HRTF modeling, in which the prominent notches and peaks are reproduced, was verified. In order to reproduce all of the N1, N2, P1, and P2, these notches and peaks should be in different frequency bins. In other words, the bandwidth should be narrower than the notch-peak frequency intervals, which depend on the vertical angle of a sound source and also vary from person to person. Then, the notch-peak frequency intervals of 16 ears of Japanese adults were calculated for seven vertical angles in the upper median plane (0° to 180°, in 30° steps).

Table 1 shows the ratios for which the bandwidth is narrower than the notch-peak frequency interval. Since the P1-N1 frequency interval is relatively wide, P1 and N1 were included in different

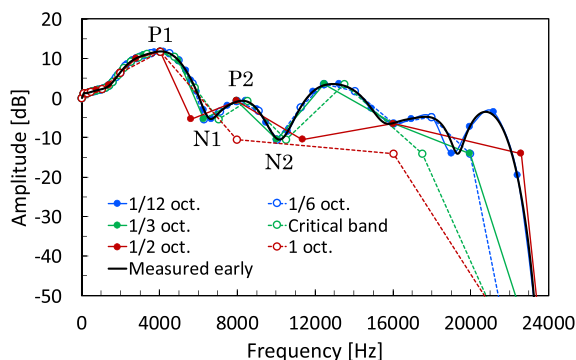


Fig. 2. Amplitude spectra of early HRTF and the modeled HRTFs.

bins for all bandwidths except 1 octave. However, the N1-P2 frequency interval is narrow, and the ratios in which N1 and P2 are included in different bins are low for 1/3, 1/2, and 1 octave.

In order to evaluate the similarity between the amplitude spectra of the modeled HRTFs and those of the early HRTFs, the spectral distortion (SD) up to 20 kHz was calculated by Eq. (1):

$$SD = \sqrt{\frac{1}{N} \sum_{i=1}^N \left[20 \log_{10} \left| \frac{HRTF_{model}(f_i)}{HRTF_{early}(f_i)} \right| \right]^2} \quad (1)$$

where $HRTF_{model}$ and $HRTF_{early}$ denote the modeled HRTF and the early HRTF, respectively. In addition, f denotes the discrete frequency (93.75-Hz steps) of the amplitude spectrum of the HRTFs, obtained by dividing 48,000 Hz by 512 samples.

Table 2 shows the SD for seven vertical angles in the upper median plane up to 20 kHz averaged over 62 ears of Japanese adults. The SD tended to be relatively large for the upper directions and small for the front and rear directions. For 1/12 octave, the SD was less than 1 dB for any direction. Then, the present study used 1/12 octave as the bandwidth for the HRTF modeling.

4. Category estimation of notch frequency bins

A previous study demonstrated that a sound image at any direction in the anterior half and posterior half of the horizontal plane can be localized using the listener's own HRTFs of the front (0°) and rear (180°) directions combined with the proper interaural difference, which corresponds to the lateral angle of a sound image [22]. Therefore, individualization of the HRTFs for the front and rear directions enables accurate sound image localization all around the horizontal plane for each listener.

In the present study, categories of notch frequency bins for the front and rear directions were estimated. From the CIT HRTF database [5], all the ears, for which both HRTFs and pinna photographs exist, were adopted for analysis. These were 76 ears of 57 donors. Of the 76 ears, 38 ears were either left or right ears of 19 Japanese adults. Of the remaining 38 ears, 19 were left ears of 19 Japanese adults, and 19 were right ears of another 19 Japanese adults.

4.1. Extraction of notch frequency bins

The early HRTFs were obtained from the measured HRTFs of the front and rear directions for 76 ears. Then, the modeled HRTFs were generated by the band-divided notch-peak model using the bandwidth of 1/12 octave. The two lowest-frequency bins above 4 kHz, in which notches are included, were extracted as N1 and N2 frequency bins.

4.1.1. Front direction

The distributions of the N1 and N2 frequency bins for the front direction are shown in Tables 3 and 4, respectively. The N1 frequency bins of 71 of 76 ears (93.4%) are distributed in six 1/12-octave bins, the frequency bin numbers of which range from 41 to 46. The center frequencies of the six bins range from 6.3 kHz to 8.4 kHz. The N2 frequency bins of 72 of 76 (94.7%) ears distributed in six 1/12-octave bins, the frequency bin numbers of which range from 47 to 52. The center frequencies of the six bins range from 9.0 kHz to 11.9 kHz.

4.1.2. Rear direction

The distributions of the N1 and N2 frequency bins for the rear direction are shown in Tables 5 and 6, respectively. Of 76 ears, four ears for N1 and one ear for N2 were excluded from the analysis because the notches of these ears were ambiguous and difficult to extract.

Table 1
Ratios for which the bandwidth is narrower than the notch-peak frequency interval.

Interval	Band width (oct.)				
	1/12	1/6	1/3	1/2	1
P1-N1	1.00	1.00	1.00	1.00	0.76
N1-P2	0.85	0.70	0.37	0.10	0.00
P2-N2	1.00	0.96	0.84	0.79	0.09

Table 2
Spectral distortion between the amplitude spectra of modeled HRTFs and those of early HRTFs up to 20 kHz averaged over 62 ears (dB).

Vertical angle (deg.)	Band width (oct.)					
	1/12	1/6	1/3	C.B.	1/2	1
0	0.9	2.2	4.5	4.7	6.5	12.3
30	0.4	1.3	3.0	3.5	4.2	10.7
60	0.7	1.5	3.2	4.1	4.5	9.9
90	0.3	0.7	1.7	2.5	2.6	8.0
120	0.6	1.5	3.1	4.1	4.1	10.1
150	0.8	2.0	4.1	5.2	5.9	12.5
180	0.9	2.1	4.7	5.6	6.5	13.2
Ave.	0.7	1.6	3.5	4.3	4.9	10.9

Table 3
Frequency distribution of N1 frequency bins for the front direction for 76 ears.

bin #	41	42	43	44	45	46	47	48	49
fc (kHz)	6.3	6.7	7.0	7.5	8.0	8.4	9.0	9.5	10.0
Number of Ears	8	8	10	23	12	10	2	3	0

Table 4
Frequency distribution of N2 frequency bins for the front direction for 76 ears.

bin #	46	47	48	49	50	51	52	53	54
fc (kHz)	8.4	9.0	9.5	10.0	10.6	11.3	11.9	12.6	13.4
Number of Ears	1	8	10	14	13	21	6	1	2

Table 5
Frequency distribution of N1 frequency bins for the rear direction.

bin #	43	44	45	46	47	48	49	50	51
fc (kHz)	6.7	7.0	7.5	8.0	8.4	9.0	9.5	10.0	10.6
Number of Ears	0	2	12	12	14	12	9	11	0

Table 6
Frequency distribution of N2 frequency bins for the rear direction.

bin #	51	52	53	54	55	56	57	58	59	60	61	62
fc (kHz)	10.6	11.3	11.9	12.6	13.4	14.2	15.0	15.9	16.9	17.9	18.9	20.0
Number of Ears	1	0	1	5	9	15	17	17	7	1	1	1

The N1 frequency bins of 70 of the 76 ears (92.1%) are distributed in six 1/12-octave bins, the frequency bin numbers of which range from 45 to 50. The center frequencies of the six bins range from 7.5 kHz to 10.0 kHz. The N2 frequency bins of 70 of the 76 ears (92.1%) are distributed in six 1/12-octave bins, the frequency bin numbers of which range from 54 to 59. The center frequencies of the six bins range from 12.6 kHz to 16.9 kHz.

4.2. Categorization of notch frequency bins

The jnd of the notch frequency on the vertical angle perception of a sound image is reported to be 0.1 to 0.2 octaves [7]. Considering a certain frequency bin, the value of the jnd is equivalent to the bandwidth obtained by adding frequency bins on both sides. Therefore, the three frequency bins are perceptually considered to be handled as one frequency bin.

4.2.1. Front direction

By combining the three frequency bins into one category, the N1 frequency bins for the front direction can be classified into two categories, as shown in Table 7. Of the 71 ears, 36.6% were included in category 1 (#41–43), and 63.4% were included in category 2 (#44–46). In the same way, N2 frequency bins can be classified into two categories, as shown in Table 8. Of the 72 ears, 44.4% were included in category 3 (#47–49), and 55.6% were included in category 4 (#50–52).

4.2.2. Rear direction

For the rear direction, the N1 frequency bins can be classified into two categories, as shown in Table 9. Of the 70 ears, 54.3% were included in category 1 (#45–47), and 45.7% were included in category 2 (#48–50). In the same way, N2 frequency bins can be classified into two categories, as shown in Table 10. Of the 70 ears,

Table 7
Categories of N1 frequency bins for the front direction.

Category	Category 1			Category 2		
bin #	41	42	43	44	45	46
fc (kHz)	6.3	6.7	7.0	7.5	8.0	8.4
Number of Ears	8	8	10	23	12	10

Table 8
Categories of N2 frequency bins for the front direction.

Category	Category 3			Category 4		
bin #	47	48	49	50	51	52
fc (kHz)	9.0	9.5	10.0	10.6	11.2	11.9
Number of Ears	8	10	14	13	21	6

Table 9
Categories of N1 frequency bins for the rear direction.

Category	Category 1			Category 2		
bin #	45	46	47	48	49	50
fc (kHz)	7.5	8.0	8.4	9.0	9.5	10.0
Number of Ears	12	12	14	12	9	11

Table 10
Categories of N2 frequency bins for the rear direction.

Category	Category 3			Category 4		
bin #	54	55	56	57	58	59
fc (kHz)	12.6	13.4	14.2	15.0	15.9	16.9
Number of Ears	5	9	15	17	17	7

41.4% were included in category 3 (#54–56), and 58.6% were included in category 4 (#57–59).

4.3. Measurement of anthropometry parameters of pinnae

Ten anthropometric parameters of the pinnae (x_1 through x_{10}), as shown in Fig. 3, were used. The origin of the coordinate system (p_0) was set at the front edge of tragus. Then, the two-dimensional coordinates of points p_1 through p_{10} were obtained. Points p_1

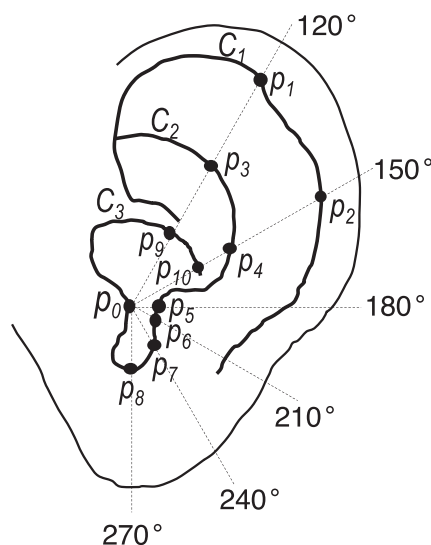


Fig. 3. Ten anthropometric parameters of the pinna.

through p_{10} are the intersections of the vertical lines (120° to 270° in 30° steps) and curves C_1 through C_3 .

Curves C_1 , C_2 , and C_3 denote the inner border of the helix, the outer border of the cymba concha, and the outer border of the cavum concha, respectively. Then, parameters x_1 through x_{10} were defined as the lengths from p_0 to p_1 through p_{10} .

These anthropometric pinna parameters were measured from a photograph showing an ear mold using homebrew software. The measured dimensions for 76 ears are listed in Table 11. Individual difference (Max/Min) ranged from 1.5 to 4.3.

4.4. Discriminant analysis of the category of N1 and N2 frequency bins

Discriminant analyses were carried out as objective variables of the category of the N1 and N2 frequency bins for the front and rear directions (two categories each) and as explanatory variables of ten anthropometric parameters of pinnae. The combination of explanatory variables that maximizes the percentage of correct classifications (PCCs) among all possible combinations was adopted for each number of explanatory variables.

4.4.1. Front direction

Fig. 4 shows the maximum PCCs of the N1 and N2 frequency bin categories for each number of explanatory variables for the front direction, using 71 and 72 ears for N1 and N2, respectively. The PCC for N1 converged to 94.4% at five variables (x_1, x_3, x_4, x_7 , and x_9). The maximum PCC of N2 was 90.3% for eight variables ($x_1, x_2, x_3, x_4, x_6, x_7, x_9$, and x_{10}).

4.4.2. Rear direction

Fig. 5 shows the maximum PCCs of N1 and N2 frequency bin categories for each number of explanatory variables for the rear direction, using 70 ears for N1 and N2. The maximum PCC for N1

Table 11
Ten measured pinnae dimensions for 76 ears (mm).

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
Ave	34.4	28.2	19.2	18.9	13.3	7.9	7.4	7.8	9.7	12.5
Max	43.6	37.9	23.5	24.5	21.6	19.6	11.9	11.0	13.5	17.5
Min	22.9	19.2	15.2	12.6	6.1	4.7	4.8	2.5	5.7	6.8
Max/Min	1.9	2.0	1.5	1.9	3.5	4.2	2.5	4.3	2.4	2.6

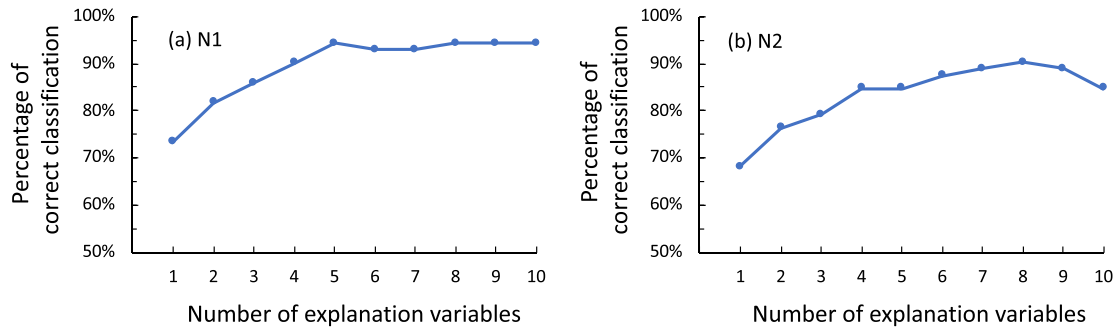


Fig. 4. Percentages of correct classification of two categories of N1 and N2 frequency bins for the front direction.

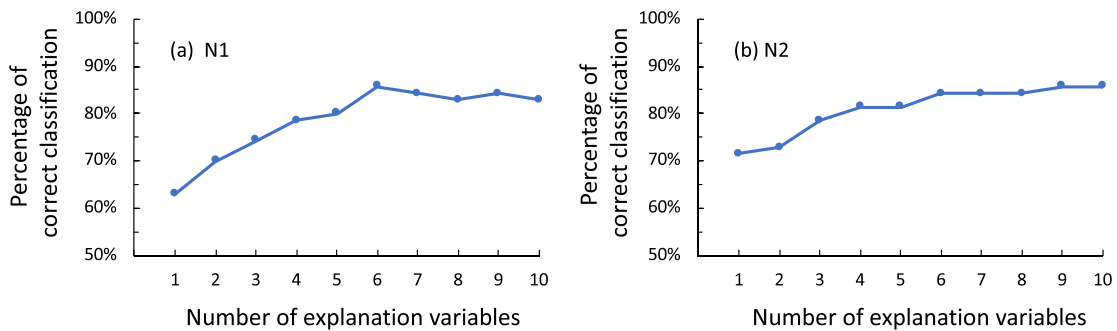


Fig. 5. Percentages of correct classification of two categories of N1 and N2 frequency bins for the rear direction.

was 85.7% for six variables ($x_1, x_2, x_3, x_5, x_7,$ and x_{10}). The PCC of N2 was 84.3% for six variables ($x_2, x_3, x_4, x_8, x_9,$ and x_{10}).

4.5. Category estimation of N1 and N2 frequency bins for naive ears

The category of the naive ear, which is not used in discriminant analysis, was estimated to be that with the smaller Mahalanobis' distance from the center of gravity of the two categories. The Mahalanobis' distance D is obtained by Eq. (2) using the pinna anthropometric parameter of the naive ear and the inverse matrix of the correlation matrix obtained by discriminant analysis of the training ears [27]:

$$D^2 = [x_1 x_2 \cdots x_9 x_{10}] \begin{bmatrix} r_{1,1} & \cdots & r_{1,10} \\ \vdots & \ddots & \vdots \\ r_{10,1} & \cdots & r_{10,10} \end{bmatrix}^{-1} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_9 \\ x_{10} \end{bmatrix} \quad (2)$$

where x_i denotes the pinna anthropometric parameters of the naive ear, and $r_{j,k}$ denotes the cross-correlation coefficients between the anthropometric parameters of the training pinna x_j and x_k obtained by the discrimination analyses of training ears.

The leave-two-out cross-validation method, which excludes a naive ear and another ear from the training data, was

applied in order to verify the accuracy of category estimation of N1 and N2 frequency bins for naive ears. The reason why the leave-two-out method, rather than the leave-one-out method, was adopted is that the PCC for a naive ear would become high when the opposite ear of the naive ear is included in the training data.

As described above, of the 76 ears, 38 ears were either left or right ears of 19 Japanese adults. For category estimation of one of the 38 ears, his/her opposite ear was excluded from the training data. For category estimation of the remaining 38 ears, which are the left or right ears of 38 Japanese adults, one randomly selected ear was excluded from the training data in order to match the number of training data for each naive ear.

4.5.1. Front direction

Fig. 6 shows the PCCs of N1 and N2 frequency bin categories for naive ears for the front direction obtained by the leave-two-out cross-validation method (solid line). For comparison, the PCC for the training ears is shown by the dotted line (same as Fig. 4). The maximum PCC for N1 was 83.1% for six variables ($x_1, x_3, x_5, x_7, x_9,$ and x_{10}). The maximum PCC of N2 was 77.8% for four variables ($x_1, x_4, x_6,$ and x_9). These six and four explanation variables had higher generalization performance than more variables. The percentage of ears, both the N1 and N2 categories of which were correctly classified, was 64.2%.

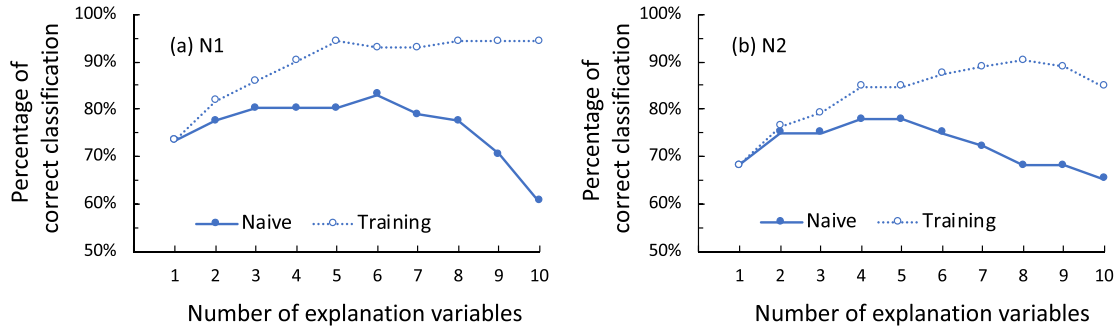


Fig. 6. Percentages of correct classification of two categories of N1 and N2 frequency bins for the front direction for naive ears (solid line). For comparison, the PCC for the training ears is shown by the dotted line (same as Fig. 4).

4.5.2. Rear direction

Fig. 7 shows the maximum PCCs of the N1 and N2 frequency bin categories for naive ears for the rear direction obtained by the leave-two-out cross-validation method. The maximum PCC for N1 was 74.3% for four variables (x_1 , x_2 , x_5 , and x_8), and the maximum PCC of N2 was 77.1% for three variables (x_3 , x_4 , and x_8). These four and three explanation variables had higher generalization performance than more variables. The percentage of ears, both the N1 and N2 categories of which were correctly classified, was 63.9%.

5. Discussions

5.1. Reason why some pinna anthropometric parameters are important in category estimation

Here, we discuss the relation between the pinna anthropometric parameters, which were included in the combination of explanatory variables of the maximum PCC, and the generation mechanism of the notches. For both the front and rear directions, x_1 and x_3 were included in the combinations of the explanatory variables of the maximum PCC.

For the generation mechanism of the notches, previous studies have proposed two hypotheses. In one study, the notches were generated at the entrance of the ear canal by the interference of the direct wave and the reflected wave from the concha wall (Raykar et al., 2005). The other hypothesis is that the notches would be generated by two resonances with opposite phase in the concha cavity and in the area between the cymba conchae and triangular fossa [28].

Here, x_1 corresponds to the distance between the tragus (p_0) and the triangular fossa, and x_3 corresponds to the distances between the tragus (p_0) and the concha wall and between the tra-

gus (p_0) and the cymba concha. Thus, the reason why x_1 and x_3 are important in category estimation of the notch frequency bin is that x_1 and x_3 are closely related to the generation mechanism of notches.

5.2. Comparison of percentage of correct classification

For comparison of PCCs, we consider the PCC when the notch frequency bin is fixed at a certain bin. The PCC for the fixed frequency bin is obtained from the proportion of ears that exist within the jnd of the notch frequency on the vertical angle perception of a sound image. As described in Section 4.2, the jnd is equivalent to three 1/12 octave bands. Therefore, PCC for a fixed bin is obtained from the proportion of the ears within three bins centered on the fixed bin. The best fixed notch frequency bin is the bin that contains the largest proportion of the ears within the three bins.

For the front direction, the distributions of N1 frequency bins shown in Tables 3 suggest that the maximum PCC for the N1 bin is 63.4% (45 of 71 ears) when the N1 frequency bin is fixed at #44 (7.5 kHz). In the same way, Table 4 suggests that the maximum PCC for the N2 bin is 66.7% (48 of 72 ears) when the N2 frequency bin is fixed at #50 (10.6 kHz).

For the rear direction, the maximum PCC for the N1 bin is 54.3% (38 of 70 ears) when the N1 frequency bin is fixed at #47 (8.4 kHz). The maximum PCC for the N2 bin is 70.0% (49 of 70 ears) when the N2 frequency bin is fixed at #57 (15.0 kHz).

Table 12 summarizes the PCC. The best fixed N1 and N2 frequency bins provide a 3.4–20.0% increase in correct classification compared to chance. The proposed method provides a 24.3–33.1% higher PCC than chance and a 7.1–20.9% higher PCC than the best fixed frequency bin.

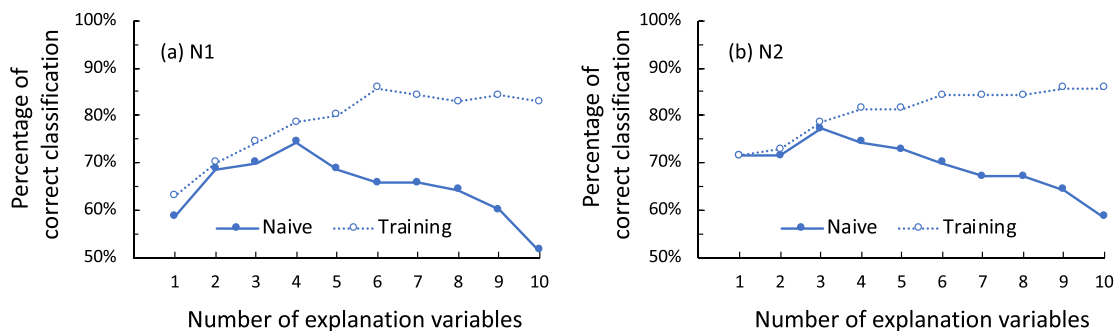


Fig. 7. Percentages of correct classification of two categories of N1 and N2 frequency bins for the rear direction for naive ears (solid line). For comparison, the PCC for training ears is shown by the dotted line (same as Fig. 5).

Table 12
Percentage of correct classification of N1 and N2 frequency bin categories for naive ears.

	Front		Rear	
	N1	N2	N1	N2
Chance rate for two categories	50.0%	50.0%	50.0%	50.0%
Best fixed frequency bin	63.4%	66.7%	53.4%	70.0%
Proposed estimation method	83.1%	77.8%	74.3%	77.1%

5.3. Pragmatic expansion to generation of individualized HRTFs

In the present study, we showed that the categories of N1 and N2 frequency bins of the HRTFs of an individual listener can be estimated by anthropometric parameters of pinnae with accuracies of 74.3 to 83.1% for the front and rear directions.

This suggests that the proposed method could be put to practical use depending on the application. The HRTF for an individual listener can be provided by the following two methods:

- A) Generate a parametric HRTF [10] using IIR notch filters, the frequencies of which are set at the center frequencies of the estimated categories of N1 and N2 frequency bins.
- B) Select the HRTFs having frequencies of N1 and N2 that are closest to the center frequencies of the estimated categories of N1 and N2 frequency bins from the HRTF database.

6. Conclusions

In order to generate individual HRTFs that are exactly adapted to each listener, the present study proposed a method by which to estimate the category of notch frequency bins of the individual HRTF based on the anthropometry of the listener's pinnae using the band-divided notch-peak HRTF model and discriminant analysis. The validity of the method was verified using the HRTFs of the front and rear directions. The results suggested the following:

- (1) For the front direction, the N1 frequency bins of 71 of 76 ears distributed in six 1/12-octave bins, the center frequencies of which range from 6.3 kHz to 8.4 kHz. The N2 frequency bins of 72 of 76 ears are distributed in six 1/12-octave bins, the center frequencies of which range from 9.0 kHz to 11.9 kHz.
- (2) For the rear direction, the N1 frequency bins of 70 of 76 ears distributed in six 1/12-octave bins, the center frequencies of which range from 7.5 kHz to 10.0 kHz. The N2 frequency bins of 70 of 76 ears are distributed in six 1/12-octave bins, the center frequencies of which range from 12.6 kHz to 16.9 kHz.
- (3) Based on the jnd of the N1 and N2 frequencies with regard to the vertical angle perception of a sound image, the six 1/12-octave bins for N1 and N2 were classified into two respective categories.
- (4) Percentages of correct classification of N1 and N2 for the front direction for naive ears were 83.1% and 77.8%, respectively. For the rear direction, PCCs of N1 and N2 were 74.3% and 77.1%, respectively.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The present study was supported in part by MEXT KAKENHI through a Grant-in-Aid for Scientific Research (C) (Grant Number 19K12068).

References

- [1] Bomhardt R, Braren H, Fels J. Individualization of head-related transfer functions using principal component analysis and anthropometric dimensions. Proc. of meetings on acoustics, 2016.
- [2] Chun CJ, Moon JM, Lee GW, Kim NK, Kim HK. Deep neural network based HRTF personalization using anthropometric measurements. Audio Eng Soc Convention 2017;143:9860.
- [3] Gardner B, Gardner S. Problem of localization in the median plane: effect of pinna cavity occlusion. J Acoust Soc Am 1973;53:400–8.
- [4] Iida K. Head-related transfer function and acoustic virtual reality. Singapore: Springer Singapore; 2019.
- [5] Iida K. Head-related transfer function and acoustic virtual reality. Springer; 2019. p. 171–7.
- [6] Iida K, Oota M. Median plane sound localization using early head-related impulse response. Appl Acoust 2018;139:14–23.
- [7] Iida K, Ishii Y. Individualization of the head-related transfer functions in the basis of the spectral cues for sound localization. In: Suzuki Y, Brungard D, Iwaya Y, Iida K, Cabrera D, Kato H, editors. Principles and Applications of Spatial Hearing. Singapore: World Scientific; 2011. p. 159–78.
- [8] Iida K, Ishii Y. Effects of adding a spectral peak generated by the second pinna resonance to a parametric model of head-related transfer functions on upper median plane sound localization. Appl Acoust 2018;129:239–47.
- [9] Iida K, Ishii Y, Nishioka S. Personalization of head-related transfer functions in the median plane based on the anthropometry of the listener's pinnae. J Acoust Soc Am 2014;136:317–33.
- [10] Iida K, Itoh M, Itagaki A, Morimoto M. Median plane localization using parametric model of the head-related transfer function based on spectral cues. Appl Acoust 2007;68:835–50.
- [11] Iida K, Yairi M, Morimoto M. Role of pinna cavities in median plane localization. In: Proc. 16th Int Congress Acoustics. p. 845–6.
- [12] Kahana Y, Nelson PA. Numerical modeling of the spatial acoustic response of the human pinna. J Sound Vib 2006;292:148–78.
- [13] Katz BFG. Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation. J Acoust Soc Am 2001;110:2440–8.
- [14] Kistler DJ, Wightman FL. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. J Acoust Soc Am 1992;91:1637–47.
- [15] Kreuzer W, Majdak P, Chen Z. Fast multipole boundary element method to calculate head-related transfer functions for a wide frequency range. J Acoust Soc Am 2009;126:1280–90.
- [16] Miccini R, Spagno S. HRTF Individualization using Deep Learning. In IEEE 5th VR Workshop on Sonic Interactions in Virtual Environments, Mar. 2020.
- [17] Middlebrooks JC. Virtual localization improved by scaling non-individualized external-ear transfer functions in frequency. J Acoust Soc Am 1999;106:1493–510.
- [18] Middlebrooks JC, Green DM. Observations on a principal components analysis of head-related transfer functions. J Acoust Soc Am 1992;92:597–9.
- [19] Middlebrooks JC, Macpherson EA, Onsan ZA. Psychophysical customization of directional transfer functions for virtual sound localization. J Acoust Soc Am 2000;108:3088–91.
- [20] Mokhtari P, Takemoto H, Nishimura R, Kato H. Frequency and amplitude estimation of the first peak of head-related transfer functions from individual pinna anthropometry. J Acoust Soc Am 2015;137:690–701.
- [21] Morimoto M, Ando Y. On the simulation of sound localization. J Acoust Soc Jpn (E) 1980;1:167–74.
- [22] Morimoto M, Iida K, Itoh M. Upper hemisphere sound localization using head-related transfer functions in the median plane and interaural differences. Acoust Sci Tech 2003;24:267–75.
- [23] Musicant AD, Butler RA. The influence of pinnae-based spectral cues on sound localization. J Acoust Soc Am 1984;75:1195–200.

- [24] Reddy CS, Hegde RM. A joint sparsity and linear regression based method for customization of median plane HRIR. 49th Asilomar conference on signals, systems and computers. Nov. 2015:785–9.
- [25] Shaw EAG, Teranishi R. Sound pressure generated in an external-ear replica and real human ears by a nearby point source. *J Acoust Soc Am* 1968;4:240–9.
- [26] Spagnol S, Avanzini F. Frequency estimation of the first pinna notch in head-related transfer functions with a linear anthropometric model. In Proc. the 18th Int. conference on digital audio effects (DAFx-15), 2015. Trondheim, Norway.
- [27] Taguchi G, Jugulum R. In: *The Mahalanobis–Taguchi strategy: a pattern technology system*. John Wiley & Sons; 2002. p. 24.
- [28] Takemoto H, Mokhtari P, Kato H, Nishimura R, Iida K. Mechanism for generating peaks and notches of head-related transfer functions in the median plane. *J Acoust Soc Am* 2012;132:3832–41.