Proceedings of Meetings on Acoustics

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Psychological and Physiological Acoustics
Session 4aPPb: Binaural Hearing (Poster Session)

4aPPb36.   Estimation of spectral notch frequencies of the individual head-related transfer function from anthropometry of listener's pinna

Yohji Ishii and Kazuhiro Iida*

 ***Corresponding author's address: Faculty of Engineering, Chiba Institute of Technology, 2-17-1 Tsudanuma, Narashino, 275-0016, Chiba, Japan, kazuhiro.iida@it-chiba.ac.jp**

  Listener's own head-related transfer functions (HRTFs) are necessary for accurate sound image reproduction. The HRTFs of other listeners often cause the front-back confusion and the errors in elevation perception. It is, however, impractical to measure the HRTFs of any listener for any sound source direction because the measurement requires special apparatus and much time. On the other hand, the estimation of the entire spectrum information of listener's own HRTF still remains as an unsolved difficult issue. One of the authors has shown that the simplified HRTFs, which is recomposed only of the first spectral peak around 4 kHz (P1) and the lowest two spectral notches (N1 and N2) above P1, extracted from the listener's measured HRTFs in the median plane, provide almost the same localization accuracy as the measured HRTFs. While the frequency of P1 is almost constant independent of the sound source elevation and the listener, those of N1 and N2 are highly dependent on both the elevation and the listener. The present study proposes a method, which estimates the frequencies of N1 and N2 in the median plane for the individual listener from the anthropometry of the listener's pinna, and examines the validity of the method.

INTRODUCTION

The listener's own head-related transfer functions (HRTFs) are necessary for accurate three-dimensional sound image control [1]. The HRTFs of other listeners often cause front-back confusion and errors in elevation perception. However, measuring the HRTFs of a listener for any sound source direction is impractical because the measurement requires a special apparatus and a great deal of time.

A scenario for estimating the HRTFs of the individual listener from the anthropometry of his/her pinna, which is obtained, e.g., by photographic means, has been proposed [2]. However, the estimation of the entire spectrum information of the listener's own HRTF is difficult and has not yet been accomplished.

Middlebrooks reported that inter-subject differences in external-ear transfer functions could be reduced by appropriately scaling one set of external-ear transfer functions in frequency relative to the other [3]. However, the reduction of errors in other-ear localization using the frequency scaling was not sufficient for accurate three-dimensional sound image control [4].

Iida *et al.* [5] have shown that simplified HRTFs, which are composed of only the first spectral peak around 4 kHz (P1) and the lowest two spectral notches (N1 and N2) above P1 (Fig. 1) extracted from the listener's measured HRTFs in the median plane, provide almost the same localization accuracy as the measured HRTFs (Fig. 2). Although the frequency of P1 is approximately constant independent of the sound source elevation and the listener, the frequencies of N1 and N2 are highly dependent on both the elevation and the listener (Fig. 3). Based on these results, they concluded that N1 and N2 can be regarded as spectral cues and that the hearing system of a human being could use P1 as reference information to analyze N1 and N2 in ear-input signals.

The present study proposes a method of estimating the frequencies of N1, N2, and P1 in the median plane for the individual listener from the anthropometry of his/her pinnae. The most suitable HRTF, the N1, N2, and P1 frequencies of which are the closest to the estimated values, can then be extracted from the HRTF database for the individual listener without acoustical measurements.
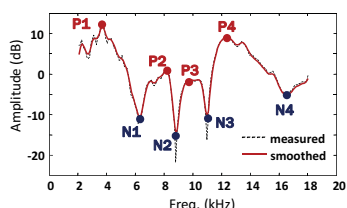

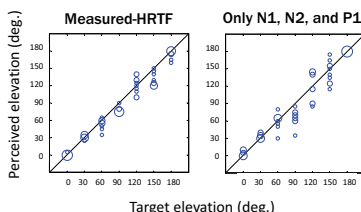**FIGURE 1.** Example of extracted spectral peaks and notches from a measured HRTF [3].

FIGURE 2. Localization in the median plane by measured HRTF and simplified HRTF recomposed of only N1, N2, and P1 [3].
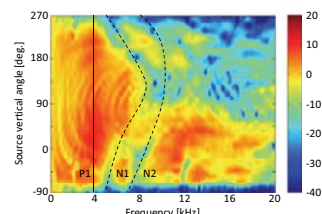
FIGURE 3. Distribution of frequencies of N1, N2, and P1 in the median plane.

DATA USED IN THE ANALYSIS

Anthropometric parameters of pinna

Based on the findings that N1, N2, and P1 are generated by the resonance of three major cavities of the pinna, i.e., concha, fossa, and scapha [6], 10 anthropometric parameters (x_1 through x_{10}) of the subject's ear mold (Fig. 4 and Table 1) were measured using a digital vernier caliper. The tilt of the pinna (x_a) was also measured from a photograph of the subject's pinna.

Figure 5 shows the measured 11 anthropometric parameters for 46 ears. The distribution ranges of parameters x_1 through x_{10} were from 10 to 20 mm and that of x_a was approximately 40 degrees.

N1, N2, and P1 frequencies

The median plane HRTFs of the 46 ears were measured in an anechoic chamber. The distance from the sound sources to the entrance of the ear canal was 1.2 m. The ear-microphones [5] were put into the ear canals of the

subjects. The diaphragms of the microphones were located at the entrances of the ear canals. Therefore, this is so called the "blocked entrances condition". The HRTF was obtained by

$$HRTF_{l,r}(\omega) = \frac{G_{l,r}(\omega)}{F(\omega)} \qquad (1)$$

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 anechoic chamber without a subject using a swept-sine signal (2^{18} sample), and $G_{l,r}(\omega)$ is that measured at the entrance of the ear canal of the subject with the ear-microphones.

N1, N2, and P1 were extracted from the HRTF, which is obtained by FFT of the early part of the head-related impulse response (within 1 ms) [7] because the peak and notches are generated by the pinna cavities.

Figure 6 shows the distributions of N1, N2, and P1 frequencies of 46 ears for the front direction. N1, N2, and P1 frequencies were distributed from 5.5 to 9.5 kHz (0.79 oct.), 7.5 to 12.0 kHz (0.68 oct.), and 3.5 to 4.5 kHz (0.36 oct.), respectively. This indicates that the individual difference of P1 frequency is much smaller than those of N1 and N2 frequencies.
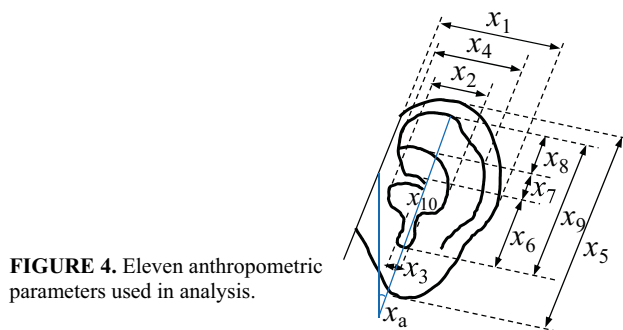

FIGURE 4. Eleven anthropometric parameters used in analysis.

TABLE 1. Definition of anthropometric parameters.

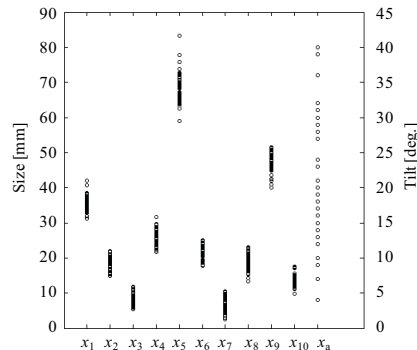|	
x_1	width of pinna
x_2	width of concha
x_3	width of incisura intertragica
x_4	width of helix
x_5	length of pinna
x_6	length of concha
x_7	length of cymba conchae
x_8	length of scapha
x_9	length of cavity
x_{10}	depth of concha
x_a	tilt of pinna

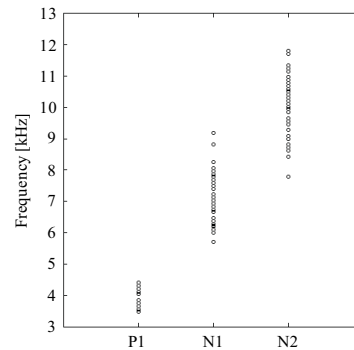FIGURE 5. Distribution of 11 measured anthropometric parameters (46 ears).

FIGURE 6. Distribution of N1, N2, and P1 frequencies for the front direction (46 ears).

ESTIMATION OF INDIVIDUAL N1, N2, AND P1 FREQUENCIES OF THE FRONT DIRECTION

Multiple regression analyses were carried out using 46 ears, as objective variables of N1, N2, and P1 frequencies and as explanatory variables of 11 anthropometric parameters of the pinnae:

$$f(S, \beta)_{N1,N2,P1} = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n + b \text{ [Hz]} \qquad (2)$$

where S, β, a_i, b, and x_i denote the subject, the vertical angle in the median plane, regression coefficients, a constant, and anthropometric parameters, respectively. There was no multicollinearity between the explanatory variables.

TABLE 2. Statistics of regression models

	model	correlation coefficient	significance level	absolute residual error [Hz]	[oct.]	probability of residual error is less than 0.15 octaves [%]
N1	A: all params	0.78	0.0002	389	0.080	87
	B: except x_{10}	0.74	0.0006	409	0.084	87
N2	A: all params	0.82	0.0000	449	0.066	96
	B: except x_{10}	0.81	0.0000	458	0.067	96
P1	A: all params	0.62	0.0719	148	0.054	96
	B: except x_{10}	0.50	0.3349	156	0.057	96
	C: x_2, x_6, x_{10}	0.56	0.0011	160	0.059	96

TABLE 3. Regression coefficients

	N1 A	N1 B	N2 A	N2 B	P1 A	P1 B	P1 C
a_1	-10.15	-21.57	-59.76	-65.77	-5.83	-11.21	-
a_2	173.42	179.32	25.69	28.80	-55.75	-52.97	-39.51
a_3	-188.44	-166.42	-107.49	-95.90	14.71	25.09	-
a_4	-23.62	-4.42	72.22	82.32	13.63	22.68	-
a_5	53.56	53.58	69.86	69.87	3.94	3.95	-
a_6	-240.26	-321.36	-227.46	-270.14	21.55	-16.70	66.31
a_7	-148.75	-176.97	93.11	78.26	-40.51	-53.82	-
a_8	-213.35	-251.41	-118.06	-138.09	-58.51	-76.46	-
a_9	32.11	80.73	-154.18	-128.60	27.52	50.45	-
a_{10}	-131.26	-	-69.07	-	-61.91	-	-59.27
a_a	-31.55	-28.43	-19.87	-18.23	0.84	2.31	-
b	13890	11953	20871	19852	4908	3994	4079

FIGURE 5. Relation between the measured and estimated N1, N2, and P1 frequencies.
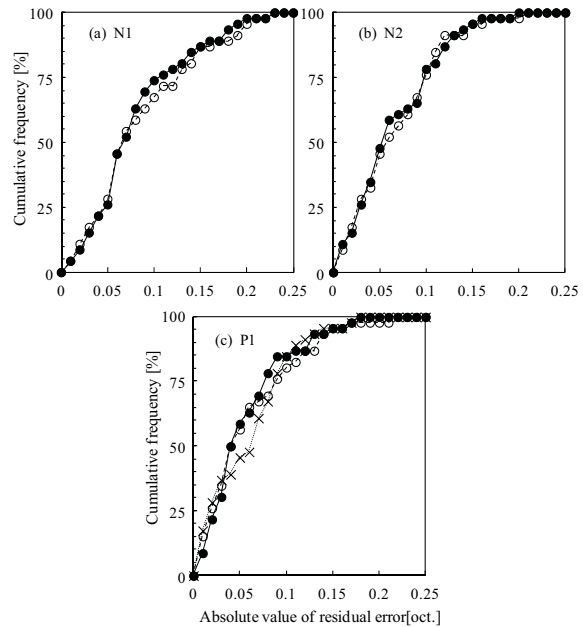

FIGURE 6. Cumulative frequency of absolute residual error. ●: regression model A, ○: B, ×: C.

In the present study, N1, N2, and P1 frequencies were estimated by two regression models:

Model A: All anthropometric parameters were used as explanatory variables.

Model B: The anthropometric parameters, which can be obtained using a photograph of the listener's pinna, i.e., x_1 through x_9 and x_a, were used as explanatory variables. The depth of concha (x_{10}) is difficult to obtain from a photograph.

Tables 2 and 3 show the statistics of the regression models and the regression coefficients, respectively. Since the just noticeable difference in the frequencies of N1 and N2 on vertical perception is considered to be between approximately 0.1 and 0.2 octaves [8], the percentage of pinnae for which the absolute estimation error is less than 0.15 octaves was calculated. Fig. 5 shows the relation between the measured and estimated N1, N2, and P1 frequencies.

For the N1 frequency, the multiple correlation coefficients for model A and B were 0.78 and 0.74, respectively. The average absolute residual errors were 389 and 409 Hz. The percentages of pinnae for which the absolute residual error was less than 0.15 octaves were 87% for both models.

For the N2 frequency, the multiple correlation coefficients for model A and B were 0.82 and 0.81, respectively. The average absolute residual errors were 464 and 468 Hz. The percentages of pinnae for which the absolute residual error was less than 0.15 octaves were 96% for both models.

For the P1 frequency, an extra regression model (model C) was analyzed in addition to models A and B, because the significance level (p) of model B was very high (0.3349). The explanatory variables of model C were chosen based on the results of stepwise backward regression and the fact that P1 is generated by resonance in the concha [9]. As a result, model C was composed only of x_2, x_6, and x_{10}. The multiple correlation coefficients for models A, B, and C were 0.62, 0.50, and 0.56, respectively. The average absolute residual errors were 148, 156, and 160 Hz. The percentages of pinnae for which the absolute residual error is less than 0.15 octaves were 96% for all models.

Then, the cumulative frequency curves of the absolute residual error were obtained. For the N1 frequency (Fig. 6(a)), the cumulative frequency curves of models A and B were almost the same. The absolute residual error, at which the cumulative frequency reaches 75%, was approximately 0.13 octaves. The cumulative frequency reaches almost 100% at an absolute residual error of approximately 0.20 octaves.

For the N2 frequency (Fig. 6(b)), the cumulative frequency curves of models A and B were also almost the same. The absolute residual error, at which the cumulative frequency reaches 75%, was approximately 0.10 octaves. The cumulative frequency reaches almost 100% at an absolute residual error of approximately 0.15 octaves. Therefore, the estimation accuracy of the N2 frequency is higher than that of the N1 frequency.

For the P1 frequency (Fig. 6(c)), the cumulative frequency curves of models A, B, and C were almost the same. The absolute residual error, at which the cumulative frequency reaches 75%, was approximately 0.08 octaves. The cumulative frequency reaches almost 100% at an absolute residual error of approximately 0.15 octaves.

EXTRACTION OF THE MOST SUITABLE FRONT DIRECTION HRTFS FOR AN INDIVIDUAL LISTENER

The most suitable front direction HRTF for an individual listener can be extracted from the HRTF database as follows:

Step 1: N1, N2, and P1 frequencies for the front direction are estimated from the anthropometry of the listener's pinna using the regression equations mentioned above.

Step 2: The HRTFs, the N1, N2, and P1 frequencies of which are closest to the estimated frequencies, are extracted from the HRTF database. The distance between the estimated notch frequencies and the notch frequencies of actual HRTFs in the database can be evaluated based on notch frequency distance (NFD) [8]. NFD is a physical measure expressing the distance between $HRTF_j$ and $HRTF_k$ in the octave scale, as defined by Eqs. 3 through 5. The front direction HRTF for which the NFD is the minimum and is within the jnd is extracted from the HRTF database as the most suitable HRTF for the listener.

$$NFD = |NFD_1| + |NFD_2| \quad [\text{oct.}] \tag{3}$$
$$NFD_1 = \log_2\{f_{N1}(HRTF_j)/f_{N1}(HRTF_k)\} \quad [\text{oct.}] \tag{4}$$
$$NFD_2 = \log_2\{f_{N2}(HRTF_j)/f_{N2}(HRTF_k)\} \quad [\text{oct.}] \tag{5}$$

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

SOUND IMAGE CONTROL FOR THE ARBITRARY THREE-DIMENSIONAL DIRECTION USING THE EXTRACTED FRONT DIRECTION HRTF

A sound image control method for the arbitrary three-dimensional direction using the extracted front direction HRTF is divided into following two stages.

Expansion from the front direction to the median plane

There are two alternative methods for estimating N1, N2, and P1 frequencies for arbitrary vertical angle in the median plane:

> Method 1: Use the median plane HRTFs of the subject whose front direction HRTF is extracted from the HRTF database by the estimation method mentioned above.

> Method 2: Calculate N1 and N2 frequencies for the arbitrary vertical angle, β, in the median plane from the N1 and N2 frequencies of the front direction, b_{N1} and b_{N2}, using the regression curves (Fig. 7). The N1 and N2 frequencies of 80% of subjects have been reported to be strongly correlated ($0.8 < r$) with the regression curves [10]. The frequency of P1 can be regarded as constant because P1 is independent of vertical angle, as shown in Fig. 3.

Expansion from the median plane to the arbitrary three-dimensional direction

Then, a sound image control in the median plane can be expanded into the arbitrary three-dimensional direction by combining the HRTFs in the median plane obtained by the above-mentioned method with interaural time difference (ITD), which corresponds to the lateral angle of a sound image [11].
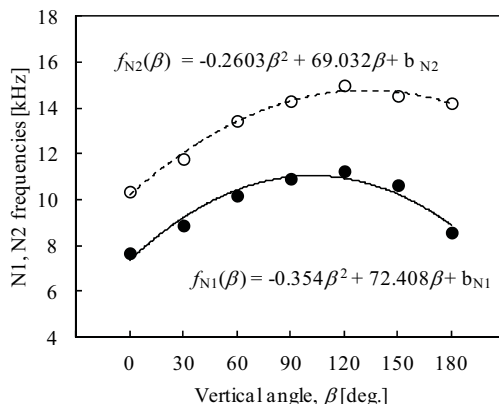

$$f_{N2}(\beta) = -0.2603\beta^2 + 69.032\beta + b_{N2}$$

$$f_{N1}(\beta) = -0.354\beta^2 + 72.408\beta + b_{N1}$$

FIGURE 7. Relation between vertical angle, β, in the median plane and average frequencies of N1 and N2 among 156 ears.

CONCLUSIONS

In the present study, multiple regression analyses were carried out using 46 ears, as objective variables of the listener's N1, N2, and P1 frequencies and as explanatory variables of the 11 anthropometric parameters of the listener's pinna. The results indicate the following:

1. The N1 and N2 frequencies of the front direction have strong correlations with the anthropometric parameters. The multiple correlation coefficients for N1, N2, and P1 were 0.78, 0.82, and 0.62, respectively.
2. The percentages of pinnae for which the absolute residual error was less than 0.15 octaves for N1, N2, and P1 were 87%, 96%, and 96%, respectively.

These results indicate that the N1, N2, and P1 frequencies of an individual listener can be estimated accurately from his/her anthropometric parameters.

ACKNOWLEDGMENT

REFERENCES

1. M. Morimoto and Y. Ando, "On the simulation of sound localization," J. Acoust. Soc. Jpn. (E), **1**, 167-174, (1980).
2. R. Sottek and K. Genuit, "Physical modeling of individual head-related transfer functions," Proc. DAGA, (1999).
3. J. C. Middlebrooks, "Individual differences in external-ear transfer functions reduced by scaling in frequency," J. Acoust. Soc. of Am., **106**, 1480-1492, (1999).
4. J. C. Middlebrooks, "Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency," J. Acoust. Soc. of Am., **106**, 1493-1510, (1999).
5. K. Iida, M. Itoh, A. Itagaki, and M. Morimoto "Median plane localization using parametric model of the head-related transfer function based on spectral cues," Applied Acoustics, **68**, 835-850, (2007).
6. H. Takemoto, P. Mokhtari, H. Kato, R. Nishimura, and K. Iida, "Mechanism for generating peaks and notches of head-related transfer functions in the median plane," J. Acoust. Soc. of Am., **132**, 3832-3841, (2012).
7. K. Iida, N. Gamoh, Y. Ishii, and M. Morimoto, "Contribution of the early part of the head-related impulse responses to the formation of two spectral notches of vertical localization cues," Proc. FORUM ACUSTICUM 2011 (Aalborg, Denmark, 2011).
8. K. Iida and Y. Ishii, "3D sound image control by individualized parametric head-related transfer functions," Proc. inter-noise 2011 (Osaka, Japan, 2011).
9. E. A. G. Shaw and R. Teranishi, "Sound Pressure Generated in an External-Ear Replica and Real Human Ears by a Nearby Point Source," J. Acoust. Soc. Am. **44**, 240-249, (1968).
10. Y Ishii, S. Nishioka, and K. Iida, "Consideration on the individual difference of the pinna shape and the spectral notches on the median plane - Construction of the HRTF database with the quantitative individual difference information -," (in Japanese), Proc. Mtg. Acoust. Soc. Jpn., 463-466 (Oct. 2012).
11. M. Morimoto, K. Iida and M. Itoh, "Upper hemisphere sound localization using head-related transfer functions in the median plane and interaural differences," Acoustical Science and Technology, **24**, 267-275, (2003).