

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



Kazuhiro Iida*, Yohji Ishii

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

ARTICLE INFO

Article history:

Received 12 March 2017
Received in revised form 20 June 2017
Accepted 1 August 2017

Keywords:

Head-related transfer function
Spectral peak
Localization
Median plane
Pinna

ABSTRACT

The parametric head-related transfer function (HRTF) recomposed of only a spectral peak (P1) and two spectral notches (N1 and N2), which are respectively generated by the first resonance and the first and second anti-resonances of the pinna, has been reported to provide approximately the same localization performance as the measured HRTF for the front and rear directions. However, for the upper direction, the localization performance for some of the subjects decreased. In the present study, we conducted two localization tests with four listeners and seven target angles in the upper median plane (0–180°) to investigate whether adding a spectral peak (P2), generated by the second resonance of the pinna, can resolve this performance decrease. The results suggested that (1) the mean vertical localization error of the parametric HRTF recomposed of N1, N2, and P1 was significantly larger than that of the measured HRTFs at the target vertical angles of 30° and 120°; (2) by adding P2 to N1N2P1, the mean vertical localization error decreased at the target vertical angles of 0°, 30°, 90°, and 120°, and no statistically significant difference was observed between N1N2P1 + P2 and the measured HRTFs at any target vertical angle; and (3) a sound image was hardly perceived in the upper direction by reproducing only P2, but the presence of P2 to improve the salience of N1 was discussed.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

It is widely accepted that the spectral notches and peaks in the human head-related transfer functions (HRTFs) in the frequency range above 5 kHz contribute to the perception of the vertical angle of a sound image. The frequencies of the prominent high-frequency notches systematically increase with elevation [25,4] and are related to the physical dimensions and shape of the pinna [23]. These notches are generated primarily by the transfer function of the pinna [6,22,14,8,26].

The importance of the outline of spectral notches and peaks, rather than the fine structures, in the HRTF has been reported [1,17,18,12,13,15]. Middlebrooks [17] hypothesized that the auditory system has knowledge of the directional filters of the pinnae and that the direction of a sound image is determined by the best-fitting directional filter.

Langendijk and Bronkhorst [13] reported that the most probable elevation cue, located in the middle 1-octave band

(5.7–11.3 kHz), is a spectral notch with a center frequency that increases as a function of elevation.

Iida *et al.* [9] proposed a parametric HRTF model for vertical sound localization. The parametric HRTF is recomposed of the spectral notches and peaks extracted from a listener's measured HRTF, regarding the peak around 4 kHz, which is independent of the vertical angle of the sound source [25], as the lower-frequency limit. The notches and peaks are labeled in order of frequency (e.g., P1, N1, P2, N2, and so on). The notches and peaks are expressed parametrically in terms of center frequency, level, and sharpness. They carried out sound localization tests in the upper median plane and demonstrated that (1) the parametric HRTF recomposed of all spectral notches and peaks provided approximately the same localization performance as the subject's own measured HRTF; (2) the parametric HRTF recomposed of only the first spectral peak around 4 kHz (P1) and the two lowest frequency notches (N1 and N2) provided approximately the same localization performance as the measured HRTFs for the front and rear directions; (3) for the upper directions, however, the localization performance of the parametric HRTF recomposed of N1, N2, and P1 for some of the subjects decreased as compared with the subject's own HRTFs; and (4) the frequencies of N1 and N2 were highly

* Corresponding author.

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

dependent on the vertical angle, whereas the frequency of P1 was approximately constant and was thus independent of the vertical angle. Based on these results, they concluded that N1 and N2 play an important role in the localization of, at least, the front and rear directions. They also reported that the human auditory system could use P1 as reference information in order to analyze N1 and N2 in ear-input signals.

Hebrank and Wright [7] carried out localization tests using narrow-band noise and claimed that an above cue is a 1/4-octave peak between 7 kHz and 9 kHz. This peak coincides with P2 and with the above directional band proposed by Blauert [3]. These findings infer that P2 might contribute the localization of upper directions. However, it is not clear whether the directional bands, which are obtained using a narrow-band signal, also act as a spectral cue for wide-band signals.

The present paper has two purposes. One is to examine whether the parametric HRTF recomposed of N1, N2, P1, and P2, by adding P2 to N1, N2, and P1, improves the localization performance at the upper directions. The other purpose is to clarify the role of P2 in the localization for the upper directions in the median plane.

2. General methods

2.1. HRTF acquisition

The HRTFs of four subjects (MKI, OIS, OTK, and YSD), who were 22–24 years of age with normal hearing sensitivity, were measured for seven vertical angles in the upper median plane (0–180° in 30° steps) in an anechoic chamber. The vertical angle, which ranges from 0° to 360°, is defined as the angle measured from front direction in the median plane, with 0° indicating front, 90° indicating above, and 180° indicating rear [21]. The test signal was a swept sine wave, the sampling frequency of which was 48 kHz. The test signal was presented in 30° steps by a loudspeaker having a diameter of 80 mm (FOSTEX FE83E) located in the upper median plane. The distance from the loudspeakers to the center of the subject's head was 1.2 m. Earplug-type microphones [11] were used to sense the test signals at the entrances of the ear canals of the subject.

The earplug-type microphones were placed into the ear canals of the subjects. The diaphragms of the microphones were located at the entrances of the ear canals. This condition is referred to as the blocked-entrances condition [25]. The HRTF was obtained as

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

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 anechoic chamber without a subject, and $G_{l,r}(\omega)$ is the Fourier transform of the impulse response, $g_{l,r}(t)$, measured at the entrance of the ear canal of the subject with the earplug-type microphones. Moreover, ω and t denote the angular frequency and time, respectively. Both $f(t)$ and $g(t)$ were 512 samples long.

2.2. Extraction of notches and peaks

For each subject, N1, N2, P1, and P2 for 0° and 180° were extracted from the early part of the head-related impulse response (HRIR) of the left and right ears because they are generated by the pinnae. The algorithm [11] for this extraction is as follows:

- (1) Detect the sample for which the absolute amplitude of the HRIR is maximum.
- (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, which replaces the derivative with the finite difference.
- (5) Define the lowest two frequencies of the local maxima above 3 kHz as P1 and P2, and the lowest two frequencies of the local minima above P1 as N1 and N2.

For 30–150°, N1 and N2 are shallow and unclear for some of the subjects. Macpherson and Sabin [15] reported findings that suggest that perceived location depends on a correlation-like spectral matching process that is sensitive to the relative, rather than absolute, across-frequency shape of the spectral profile. Then, N1 and N2 for 30–150° were extracted using the following algorithm.

- (1) Obtain the amplitude spectrum of the 512-point HRIR by FFT. Then, find the local minima of the amplitude using the difference method.
- (2) Estimate N1 and N2 frequencies using the following regression equations reported by Iida and Ishii [10]:

$$f_{N1}(\beta) = 1.001 \times 10^{-5} \times \beta^4 - 6.431 \times 10^{-3} \times \beta^3 + 8.686 \times 10^{-1} \times \beta^2 - 3.265 \times 10^{-1} \times \beta + f_{N1}(0) \text{ [Hz]} \quad (2)$$

$$f_{N2}(\beta) = 1.310 \times 10^{-5} \times \beta^4 - 5.154 \times 10^{-3} \times \beta^3 + 5.020 \times 10^{-1} \times \beta^2 + 2.563 \times 10 \times \beta + f_{N2}(0) \text{ [Hz]} \quad (3)$$

where f_{N1} and f_{N2} denote the N1 and N2 frequencies, respectively, and β is the vertical angle in degrees. Iida and Ishii reported that the behavior of the N1 and N2 frequencies as a function of vertical angle can be regarded as common among listeners, even though the frequencies of N1 and N2 for the front direction depend highly on the listener.

- (3) Search for the deepest local minima within 0.2 octaves of the estimated N1 and N2 frequencies, and define them as N1 and N2, respectively, because the just-noticeable difference in the N1 and N2 frequencies with regard to vertical localization are considered to range from 0.1 to 0.2 octaves [10].

Frequencies of P1 and P2 are considered to be direction independent and were only extracted for a vertical angle of 0°.

The four subjects' N1 and N2 frequencies ranged from 6750 to 11,438 Hz and from 9188 to 17,250 Hz, respectively, for seven vertical directions. The P1 and P2 frequencies ranged from 3656 to 4031 Hz and from 7688 to 8719 Hz, respectively.

2.3. Generation of parametric HRTFs

The parametric HRTFs for each subject, ear, and target vertical angle were generated by superposition of the notches and peaks, each of which was reproduced by a second-order IIR filter. The center frequency, level, and half-power bandwidth of the second-order IIR filters were adjusted to the extracted N1, N2, P1, and P2. The difference in the frequencies of the notches and peaks between the measured HRTFs and the parametric HRTFs were within 93.75 Hz, i.e., the unit of the frequency resolution (48,000 Hz/512 samples). This difference is sufficiently smaller than the just-noticeable differences, ranging from 0.1 to 0.2 octaves [10]. The dif-

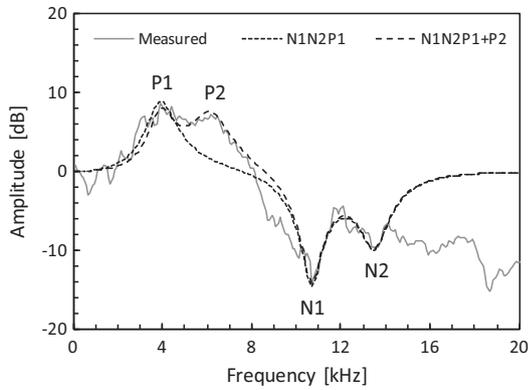


Fig. 1. Examples of a measured HRTF and parametric HRTFs (N1N2P1 and N1N2P1 + P2) for a vertical angle of 90°.

ferences in the levels of the notches and the peaks were within 1 dB.

Figure 1 shows examples of a measured HRTF and parametric HRTFs recomposed of N1, N2, and P1 (hereinafter N1N2P1) and of N1, N2, P1, and P2 (hereinafter N1N2P1 + P2). As shown in the figure, the parametric HRTF reproduced the recomposed spectral notches and peaks accurately. However, the amplitude spectrum of N1N2P1 differs from that of the measured HRTF for the frequency ranges from 4.5 to 10.5 kHz and higher than 14 kHz. The amplitude spectrum of N1N2P1 + P2 differs from that of the measured HRTF for the frequency ranges from 8 to 10.5 kHz and higher than 14 kHz.

2.4. Method of localization tests

The four subjects (MKI, OIS, OTK, and YSD) participated in the sound localization tests. The localization tests were conducted in a quiet sound-proof room. The working area of the room was 4.6 m (width) by 5.8 m (depth) by 2.8 m (height). 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), an A/D converter (Roland M-10MX), open-air headphones (AKG K1000), and earplug-type microphones were used for the localization tests.

The sound pressure at the eardrum for the open-ear-canal condition, P , can be obtained by processing the sound pressure at the entrance of the blocked ear canal with a compensation filter, G , through headphones [19], as follows:

$$P = S \times \text{HRTF} \times G, \quad (4)$$

$$G = \frac{1}{M \times \text{PTF}} \times \frac{Z_{\text{earcanal}} + Z_{\text{headphone}}}{Z_{\text{earcanal}} + Z_{\text{radiation}}} \quad (5)$$

$$\triangleq \frac{1}{M \times \text{PTF}} \times \text{PDR}, \quad (6)$$

where S denotes the sound source, M is the transfer function of the earplug-type microphones, $M \times \text{PTF}$ 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 impedances of the ear canal and the headphones, respectively, and $Z_{\text{radiation}}$ is the free-air radiation impedance as observed from the ear canal. The second term on the right-hand side of Eq. (5) is referred to as the pressure division ratio (PDR). Møller *et al.* defined free-air equivalent coupling to the ear (FEC) headphones as headphones for which the PDR reduces to unity. K1000 headphones (AKG), which were regarded as FEC headphones, were used in the localization tests.

The compensation of $M \times \text{PTF}$ was processed from 200 Hz to 17 kHz, and $M \times \text{PTF}$ was measured by the following procedure. The subjects sat at the center of the soundproof room. The earplug-type microphones were placed into the ear canals of the subjects. The diaphragms of the microphones were located at the entrances of the ear canals, in the same manner as in the HRTF measurements described Section 2.1. The subjects wore the headphones, and maximum-length sequence signals (48 kHz sampling, 12th-order, and no repetitions) were emitted through the headphones. The signals were received by the earplug-type microphones, and $M \times \text{PTF}$ was obtained. The earplug-type microphones were then removed without displacing the headphones because the pinnae of the subject were not enclosed by the headphones (Fig. 2). The typical peak-to-peak range of the transfer functions between the headphones and the earplug-type microphones from 200 Hz to 17 kHz was approximately 20 dB. This was reduced to 3 dB by the compensation filter, G , as shown in Fig. 3.

The source signal was a wideband Gaussian white noise from 200 Hz to 17 kHz. Stimuli were delivered at 63 dB SPL at the entrance of each ear. The duration of the stimuli was 1.2 s, including the rise and fall times, each of which was 0.1 s. The mapping method was adopted as a response method in order to respond on a continuous scale rather than selecting between distinct locations. A circle and a horizontal arrow through the center of the circle, which indicated the median plane and the front-back axis, were displayed on the screen of a laptop computer display. 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 measured and parametric HRTFs were tested separately. Møller *et al.* [19] demonstrated that no significant difference in localization performance was observed for the same set of HRTFs between separate tests and mixed tests under the following two conditions: (1) the HRTF set of one subject was randomized, and (2) the HRTF sets of several subjects were randomized.

3. Sound localization test 1

Sound localization tests in the upper median plane were carried out in order to examine whether N1N2P1 + P2, obtained by adding P2 to N1N2P1, improves the localization performance at the upper directions.

The following three types of HRTFs were used: (1) each subject's own measured HRTF, (2) each subject's N1N2P1, and (3) each subject's N1N2P1 + P2. The target vertical angles were seven directions, in steps of 30°, in the upper median plane.

In one test block, 35 stimuli (seven directions, five times) were randomized and presented to a subject. The duration of one block



Fig. 2. Headphones used in the experiments.

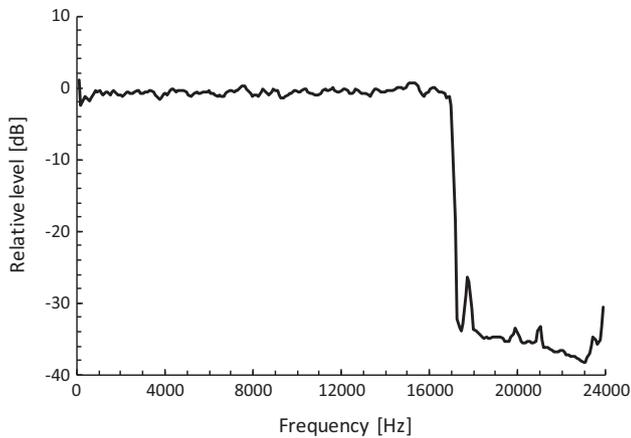


Fig. 3. Example of a transfer function between the headphones and the earplug-type microphone with a compensation filter, G.

was approximately 7 min. Each subject was presented two test blocks, which he/she examined using his/her own HRTFs and parametric HRTFs. Therefore, each subject responded to each stimulus 10 times. The localization tests were carried out using a double-blind method.

3.1. Individual responses

Figure 4 shows the responses to the subject's own measured HRTFs, N1N2P1, and N1N2P1 + P2 for the four subjects. The ordinate represents the responded vertical angle, and the abscissa represents the target vertical angle. The diameter of each circle is proportional to the number of responses with a resolution of 5° .

For the measured HRTF, most of the responses were distributed around the target vertical angles. However, the responses of subject OTK were distributed around 90° for the target vertical angles of 60° , 120° , and 150° . The responses of subject YSD shifted slightly upward at 0° and rearward at 90° .

For N1N2P1, the distribution of the responses was approximately the same as that for the measured HRTFs for subjects MKI and OIS. However, the localization performance was decreased at 120° for subject MKI and at 90° for subject OIS. The responses of subject OTK were distributed around both 0° and 90° at the target vertical angles of 60° , 90° , and 120° .

For N1N2P1 + P2, obtained by adding P2 to N1N2P1, the localization performance of all subjects was improved at certain target vertical angles. For subject MKI, the performance at 120° and 150° was improved. For subject OIS, the performance at 90° was improved. However, the responses shifted upward at 150° , the reason for which is unclear. For subject OTK, the separated distribution of the responses shown for N1N2P1 was not observed at the target vertical angles of 60° , 90° , and 120° . The distribution of responses to N1N2P1 + P2 was approximately the same as that for the measured HRTFs. For subject YSD, the localization performance at 0° was improved.

3.2. Mean vertical localization error

The mean vertical localization error for each HRTF and target vertical angle was calculated (Table 1). The mean vertical localization error is defined as the absolute difference between the responded and target vertical angles averaged over all of the responses.

For the measured HRTF, the mean vertical localization error tended to be small at the target vertical angles near the horizontal

plane (0° and 180°) and increased with the vertical angle, as reported by Carlile *et al.* [5] and Majdak *et al.* [16].

The mean vertical localization errors of N1N2P1 for most subjects and target vertical angles were larger than those of measured HRTFs. In particular, the errors at the target vertical angles of 30° , 60° , 90° , and 120° tended to be larger than those for the measured HRTFs.

For N1N2P1 + P2, the mean vertical localization errors were smaller than those for N1N2P1 for many of the subjects and target vertical angles. In particular, adding P2 to N1N2P1 decreased the error at the target vertical angle of 0° for subject YSD, at 90° for subjects OIS and OTK, at 120° for subjects MKI, OIS, and OTK, and at 150° for subject MKI. However, the error for N1N2P1 + P2 was larger than that for N1N2P1 at the target vertical angle of 150° for subject OIS. This is due to the upward shift of OIS's responses, the reason for which is unclear.

Adding P2 to N1N2P1, the mean vertical localization errors averaged across subjects decreased at target vertical angles of 0° , 30° , 90° , and 120° . In particular, the mean vertical localization error decreased by 10.8° at the target vertical angle of 90° . Then, the difference in mean vertical localization errors between N1N2P1 + P2 and the measured HRTFs became less than 10° for all seven vertical target angles.

Tukey's multiple comparison test was performed in order to determine whether a difference in the mean vertical localization error averaged across subjects among the measured HRTFs, N1N2P1, and N1N2P1 + P2 is statistically significant.

Table 2 shows the results of the statistical tests. The mean vertical localization errors of N1N2P1 were significantly larger than those of the measured HRTFs at target vertical angles of 30° ($p < 0.05$) and 120° ($p < 0.01$). On the other hand, no statistically significant difference was observed at any target vertical angles between the measured HRTFs and N1N2P1 + P2 or between N1N2P1 and N1N2P1 + P2.

These results imply that N1N2P1 + P2 provides approximately the same vertical localization performance to the measured HRTFs at any of the seven target vertical angles in the upper median plane, while the performance of N1N2P1 was significantly less than the measured HRTFs for the target vertical angles of 30° and 120° .

3.3. Ratio of front-back confusion

Table 3 shows the ratio of front-back confusion for each HRTF and the target vertical angle calculated from all of the responses of the four subjects. 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.

The ratio of front-back confusion for the measured HRTF, N1N2P1, and N1N2P1 + P2 at target vertical angles of 0° and 180° was 0%. At the other four target vertical angles, however, the ratios for N1N2P1 and N1N2P1 + P2 were higher than those for the measured HRTFs. The ratios for N1N2P1 were approximately double those for the measured HRTFs, except at 60° . The ratios for N1N2P1 + P2 tended to be slightly smaller than those for N1N2P1.

Statistical tests (chi-square tests) were performed in order to verify whether a difference in the ratio of front-back confusion among the measured HRTFs, N1N2P1, and N1N2P1 + P2 is statistically significant. The results show that statistically significant differences were observed at the target vertical angle of 120° between the measured HRTFs and N1N2P1, and between the measured HRTFs and N1N2P1 + P2. No significant difference was observed between N1N2P1 and N1N2P1 + P2 at any target vertical angle.

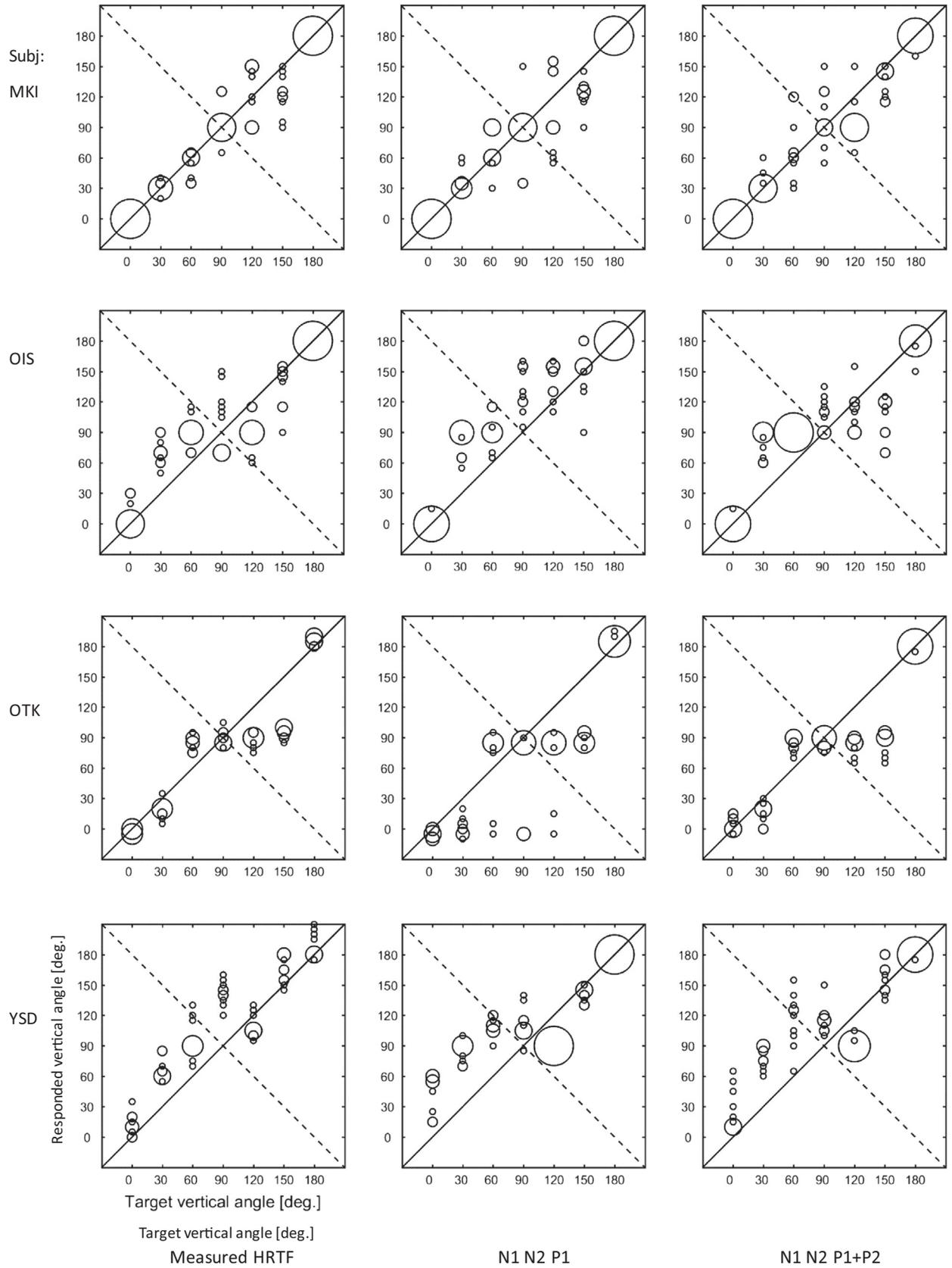


Fig. 4. Responses to the measured HRTF, N1N2P1, and N1N2P1 + P2. The ordinate represents the responded vertical angle, and the abscissa represents the target vertical angle. The diameter of each circle is proportional to the number of responses with a resolution of 5°.

Table 1
Mean vertical localization error (deg.)

Subject	HRTF	Target vertical angle (deg.)							
		0	30	60	90	120	150	180	Ave.
MKI	Measured	0.5	3.5	10.0	10.0	23.0	27.1	0.4	10.6
	N1N2P1	0.8	6.9	15.7	17.8	39.0	28.3	1.0	15.6
	N1N2P1 + P2	0.3	5.2	22.2	21.1	30.1	15.4	2.8	13.9
OIS	Measured	8.7	40.4	30.5	28.1	29.7	16.4	0.5	22.0
	N1N2P1	2.3	51.1	31.6	41.8	23.3	18.5	0.7	24.2
	N1N2P1 + P2	2.2	50.1	30.1	19.3	16.6	46.6	3.7	24.1
OTK	Measured	2.8	13.0	24.6	6.2	32.3	55.5	5.6	20.0
	N1N2P1	5.1	28.8	31.6	31.5	50.5	61.9	6.4	30.8
	N1N2P1 + P2	6.3	13.7	23.5	5.1	37.0	64.2	1.4	21.6
YSD	Measured	12.4	36.7	35.8	52.0	14.3	16.1	13.0	25.8
	N1N2P1	44.4	54.1	49.0	22.9	30.0	9.7	0.8	30.2
	N1N2P1 + P2	27.4	48.4	55.1	25.2	28.3	13.8	0.8	28.4
Average across subjects	Measured	6.1	23.4	25.2	24.1	24.8	28.8	4.9	19.6
	N1N2P1	13.2	35.2	32.0	28.5	35.7	29.6	2.2	25.2
	N1N2P1 + P2	9.0	29.3	32.7	17.7	28.0	35.0	2.2	22.0

Table 2
Results of Tukey's multiple comparison test among the measured HRTFs, N1N2P1, and N1N2P1 + P2 for mean vertical localization error.

Comparison between	Target vertical angle (deg.)							
	0	30	60	90	120	150	180	
Measured and N1N2P1								
Measured and N1N2P1 + P2					**			
N1N2P1 and N1N2P1 + P2								

** p < 0.01.

* p < 0.05.

Table 3
Ratio of front-back confusion for each HRTF and target vertical angle.

HRTF	Target vertical angle (deg.)							
	0	30	60	90	120	150	180	
Measured	0.00	0.05	0.35	–	0.23	0.10	0.00	
N1N2P1	0.00	0.13	0.45	–	0.50	0.20	0.00	
N1N2P1 + P2	0.00	0.05	0.43	–	0.50	0.15	0.00	

3.4. Ratio of inside-of-head localization

All four of the subjects reported to have never perceived a sound image inside their heads for either the subject's own measured HRTFs or the parametric HRTFs.

4. Sound localization test 2

Localization test 1 showed that N1N2P1 + P2, obtained by adding P2 to N1N2P1, improved the localization performance at the upper direction in the median plane. The maximum improvement in the mean vertical localization error was 10.8° at the target vertical angle of 90°. The purpose of localization test 2 is to clarify the role of P2 in the localization for the upper direction.

Three types of parametric HRTFs recomposed of (1) P1, (2) P2, and (3) P1 and P2 of the subject's HRTF at 90° were used. In other words, only P1, P2, and P1P2 of the measured HRTF at 90° were reproduced. The difference in the frequencies of the notches and peaks between the measured HRTFs and the parametric HRTFs were within 93.75 Hz, i.e., the unit of the frequency resolution (48,000 Hz/512 samples). This difference is sufficiently smaller than the just-noticeable differences, ranging from 0.1 to 0.2 octaves [10]. The differences in the levels of the notches and the peaks were within 1 dB.

Figure 5 shows examples of a measured HRTF and parametric HRTFs recomposed of P1, P2, and P1P2. As shown in the figure, the parametric HRTF accurately reproduced the recomposed spectral peaks and had flat spectrum characteristics in other frequency ranges. For example, the spectrum of P1P2 is similar to that of the measured HRTF for frequency ranges below 8 kHz. However, this differs from the measured HRTF for frequency ranges above 8 kHz.

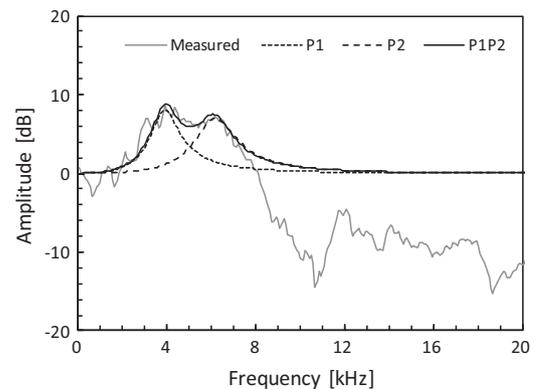


Fig. 5. Examples of a measured HRTF and parametric HRTFs (P1, P2, and P1P2) for a vertical angle of 90°.

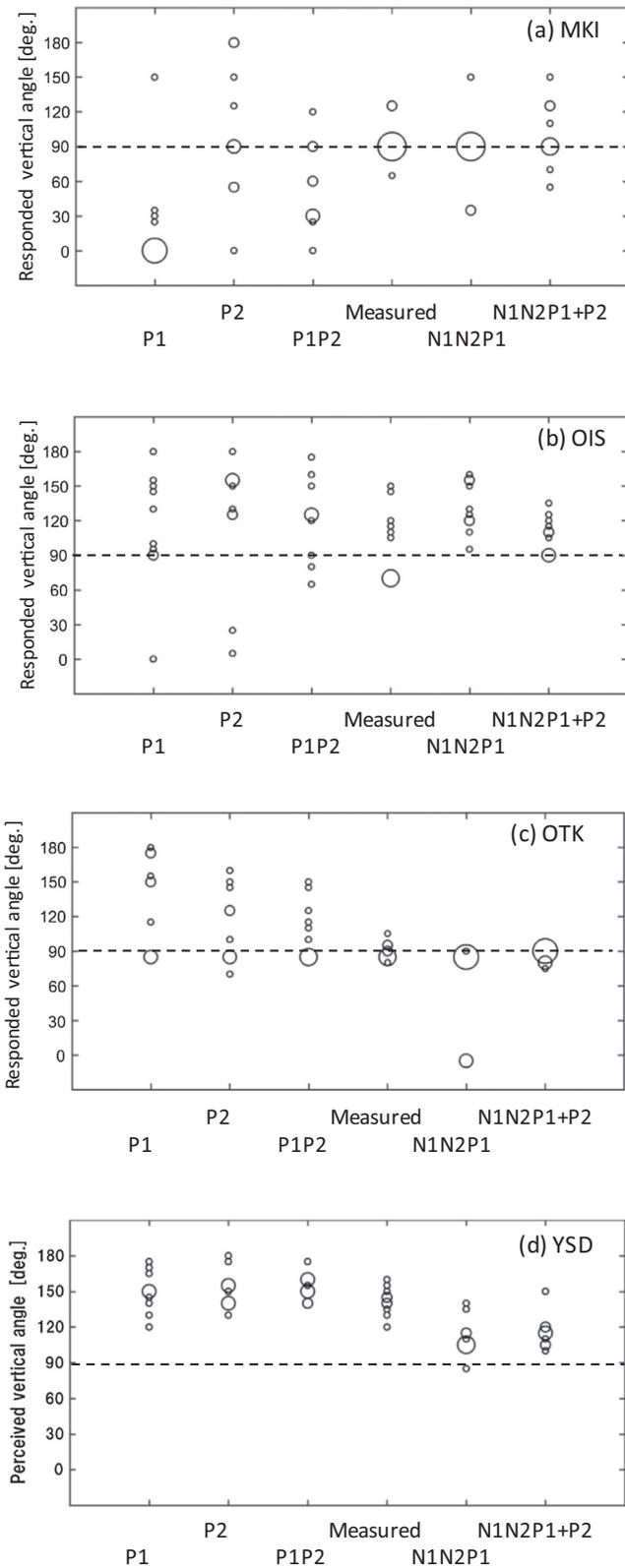


Fig. 6. Responses to P1, P2, and P1P2. For comparison, the responses to the measured HRTF, N1N2P1, and N1N2P1 + P2 for the target vertical angle of 90° obtained in localization test 1 are also shown.

4.1. Individual responses

Figure 6 shows the responses to P1, P2, and P1P2. Moreover, the responses to the measured HRTF, N1N2P1, and N1N2P1 + P2 for

the target vertical angle of 90° obtained in localization test 1 are shown in the figure. The distribution of the responses to N1N2P1 + P2 was approximately the same as that of the responses to the measured HRTFs, as mentioned in Section 3.1.

The distributions of the responses to P1, P2, and P1P2 were similar within each subject. Namely, the responses for MKI and OIS were distributed widely from forward to rearward, whereas those for OTK were distributed from upward to rearward, and those for YSD were distributed from diagonally rearward to rearward. The subjects hardly perceived a sound image in the upward direction by reproducing P1, P2, or both P1 and P2.

4.2. Mean vertical localization error

The mean vertical localization error averaged across subjects was calculated for the target vertical angle of 90° (Fig. 7).

The mean vertical localization error for N1N2P1 + P2 was similar to that for the measured HRTF. However, the mean vertical localization errors for P1, P2, and P1P2 were larger than that for the measured HRTF.

Tukey’s multiple comparison test was performed in order to verify whether the differences in the mean vertical localization error among the measured HRTF and parametric HRTFs are statistically significant.

Table 4 shows that the mean vertical localization errors of P1, P2, and P1P2 were significantly larger than that of the measured HRTF and that of N1N2P1 + P2, while the mean vertical localization error of N1N2P1 + P2 does not significantly differ from that of the measured HRTF.

These results imply that these spectral peaks were not sufficient in themselves for localization of the upper direction.

4.3. Ratio of inside-of-head localization

None of four subjects reported to have perceived a sound image inside his/her head for P1, P2, or P1P2, while they reported the distance of a sound image as being near.

5. Discussion

The results shown in Section 4 indicate that the broadband signal was hardly perceived in the upward direction by reproducing only the subject’s P2. In other words, our results do not support the claim of Hebrank and Wright [7] that an above cue is a 1/4-octave peak between 7 kHz and 9 kHz.

Then, let us discuss the possible role of P2 in the vertical localization. Moore *et al.* [20] examined whether the notches in the HRTF are detectable by listeners. They measured the thresholds for the detection of the spectral notches for center frequencies of

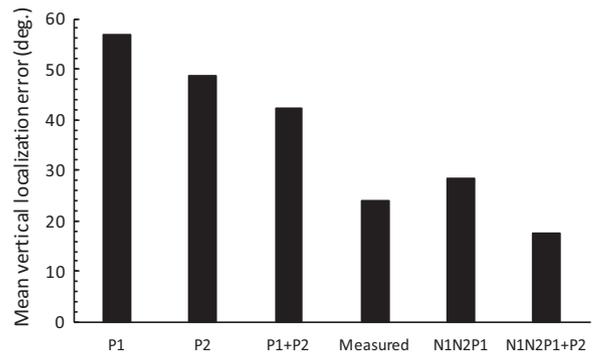


Fig. 7. Mean vertical localization error at the target vertical angle of 90°.

Table 4

Results of Tukey's multiple comparison test among the measured HRTF and parametric HRTFs for mean vertical localization error.

	P1	P2	P1P2	Measured	N1N2P1	N1N2P1 + P2
P1	–					
P2		–				
P1P2			–			
Measured	**	**		–		
N1N2P1	**	**			–	
N1N2P1 + P2	**	**	**			–

** $p < 0.01$.* $p < 0.05$.

1 and 8 kHz in a flat spectrum. Using the threshold of the notch detection at 8 kHz, we attempt to consider the reason why adding P2 to N1N2P1 improved the performance of the vertical localization.

Figure 8 shows the frequencies and the levels of N1, N2, P1, and P2 for subject OTK at the target vertical angles of 90° (closed circle) and 0° (open circle). As shown in Section 3, at the target vertical angle of 90°, the responses to N1N2P1 were distributed around both 0° and 90°, while the responses to N1N2P1 + P2 were distributed around 90°. On the other hand, at the target vertical angle of 0°, the responses to both N1N2P1 and N1N2P1 + P2 were distributed around the target vertical angle.

The two broken lines indicate the maximum and minimum thresholds of three subjects for the detection of the notch, the center frequency of which is 8 kHz and the bandwidth of which is 25% of the center frequency, as reported by Moore *et al.* [20]. This means that no subjects detected the notch when the level was more than –9 dB, and all of the subjects detected the notch when the level was less than –20 dB.

As a first attempt at explanation, we compare the levels of N1 and N2 with the threshold of the notch detection for 8 kHz, although the frequencies of N1 and N2 were not 8 kHz: they were 6656 Hz and 9564 Hz for 0°, and 10,500 Hz and 14,344 Hz for 90°. The levels of both N1 and N2 exceeded the threshold at the target vertical angle of 0°. However, at 90°, N1 was undetectable, and N2 was detectable by some of the subjects but undetectable by other subjects.

In the frequency domain, the notches and peaks were located in order of P1, P2, N1, and N2 from low frequency at 90°. Contrast effects are not expected because P1 (3844 Hz) is located far from

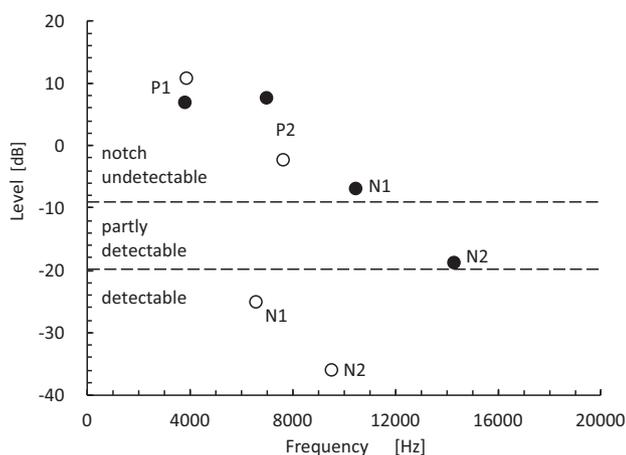


Fig. 8. Relationship between the frequencies and levels of N1, N2, P1, and P2 for subject OTK. Open and closed circles denote N1, N2, P1, and P2 for 0° and 90°, respectively. Broken lines indicate the maximum and minimum thresholds of three subjects for the detection of the notch, the center frequency of which is 8 kHz and the bandwidth of which is 25% of the center frequency, as reported by Moore *et al.* [20].

N1 (10,500 Hz) if P2 (7031 Hz) is not reproduced. However, reproducing P2, the relative level of N1 measured from P2 reaches –14.7 dB. At this level, some subjects can detect the notch.

Macpherson and Sabin [15] suggested that the perceived location depends on a correlation-like spectral matching process that is sensitive to the relative, rather than absolute, across-frequency shape of the spectral profile. These results appear to support our hypothesis. Reiss and Young [24] suggested that cats decode spectral cues by extracting the rising spectral edges (the upper edge of the spectral notch) of the HRTFs rather than the center frequencies of notches. Then, Baumgartner *et al.* [2] proposed a sagittal-plane localization model and showed that positive spectral gradient extraction is important for localization robustness to spectrally macroscopic variations of the source signal. However, as shown in Fig. 1, P2 could enhance the lower edge of N1, but could not enhance the upper edge for the upper direction.

These considerations suggest that P2 could play a role in improving the accuracy of localization for the upper direction in the median plane by enhancing N1. However, the condition in which a peak affects a notch is unclear and remains an important issue to be solved. Moreover, which part of P2, i.e., the summit or the foot, is more important is also a problem to be solved in the near future.

6. Conclusions

We conducted two localization tests with four listeners and seven target vertical angles in the upper median plane (0° to 180°, in 30° steps) in order to investigate whether adding P2 to N1N2P1 can improve the localization performance for upper target directions. The results suggested the following:

- (1) The mean vertical localization error of N1N2P1 was significantly larger than that of the measured HRTFs at the target vertical angles of 30° ($p < 0.05$) and 120° ($p < 0.01$).
- (2) By adding P2 to N1N2P1, the mean vertical localization error decreased at target vertical angles of 0°, 30°, 90°, and 120°. Then, the differences in mean vertical localization error between N1N2P1 + P2 and the measured HRTFs became less than 10° for all seven vertical target angles. No statistically significant difference was observed between N1N2P1 + P2 and the measured HRTFs at any of the seven target vertical angles.
- (3) No statistically significant difference in the ratio of front-back confusion was observed among measured HRTFs, N1N2P1, and N1N2P1 + P2 at any target vertical angles, except at the target vertical angle of 120° between the measured HRTFs and N1N2P1 and between the measured HRTFs and N1N2P1 + P2.
- (4) A sound image was hardly perceived in the upper direction by reproducing the subject's P2. In other words, P2 was not sufficient in itself for localization of the upper direction.

Acknowledgements

The present study was supported in part by MEXT KAKENHI through a Grant-in-Aid for Scientific Research (A) (Grant Number 15H01790). The authors would like to thank M. Morimoto, Professor Emeritus at Kobe University, for his fruitful discussions.

References

- [1] Asano F, Suzuki Y, Sone T. Role of spectral cues in median plane localization. *J Acoust Soc Am* 1990;88:159–68.
- [2] Baumgartner R, Majdak P, Laback B. Modeling sound-source localization in sagittal planes for human listeners. *J Acoust Soc Am* 2014;136:791–802.
- [3] Blauert J. Sound localization in the median plane. *Acustica* 1969/70;22:205–13.
- [4] Butler A, Belendiuk K. Spectral cues utilized in the localization of sound in the median sagittal plane. *J Acoust Soc Am* 1977;61:1264–9.
- [5] Carlile S, Leong P, Hyams S. The nature and distribution of errors in sound localization by human listeners. *Hear Res* 1997;114:179–96.
- [6] Gardner B, Gardner S. Problem of localization in the median plane: effect of pinna cavity occlusion. *J Acoust Soc Am* 1973;53:400–8.
- [7] Hebrank J, Wright D. Spectral cues used in the localization of sound sources on the median plane. *J Acoust Soc Am* 1974;56:1829–34.
- [8] Iida K, Yairi M, Morimoto M. Role of pinna cavities in median plane localization. In: *Proceedings of 16th international congress on acoustics*. p. 845–6.
- [9] 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.
- [10] 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.
- [11] Iida K, Nishioka S, Ishii Y. 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.
- [12] Kulkarni A, Colburn HS. Role of spectral detail in sound-source localization. *Nature* 1998;396:747–9.
- [13] Langendijk EHA, Bronkhorst AW. Contribution of spectral cues to human sound localization. *J Acoust Soc Am* 2002;112:1583–96.
- [14] Lopez-Poveda EA, Meddis R. A physical model of sound diffraction and reflections in the human concha. *J Acoust Soc Am* 1996;100:3248–59.
- [15] Macpherson EA, Sabin AT. Vertical-plane sound localization with distorted spectral cues. *Hear Res* 2013;306:76–92.
- [16] Majdak P, Goupell MJ, Laback B. 3-D localization of virtual sound sources: effects of visual environment, pointing method, and training. *Atten Percept Psychophys* 2010;72:454–69.
- [17] Middlebrooks JC. Narrow-band sound localization related to external ear acoustics. *J Acoust Soc Am* 1992;92:2607–24.
- [18] Middlebrooks JC. Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. *J Acoust Soc Am* 1999;106:1493–510.
- [19] Møller H, Hammershøi D, Jensen CJ, Sørensen MF. Transfer characteristics of headphones measured on human ears. *J Audio Eng Soc* 1995;43:203–17.
- [20] Moore BCJ, Oldfield R, Dooley GJ. Detection and discrimination of peaks and notches at 1 and 8 kHz. *J Acoust Soc Am* 1989;85:820–36.
- [21] Morimoto M, Aokata H. Localization cues of sound sources in the upper hemisphere. *J Acoust Soc Jpn (E)* 1984;5:165–73.
- [22] Musicant A, Butler R. The influence of pinnae-based spectral cues on sound localization. *J Acoust Soc Am* 1984;75:1195–200.
- [23] Raykar VC, Duraiswami R, Yegnanarayana B. Extracting the frequencies of the pinna spectral notches in measured head related impulse responses. *J Acoust Soc Am* 2005;118:364–74.
- [24] Reiss LAJ, Young ED. Spectral edge sensitivity in neural circuits of the dorsal cochlear nucleus. *J Neurosci* 2005;25:3680–91.
- [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;44:240–9.
- [26] 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.