Applied Acoustics 229 (2025) 110370

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Cue for rear sound image localization in head-related transfer function below 4 kHz

Fuka Nakamura [a], Kazuhiro Iida [b,*]

[a] Graduate School, Chiba Institute of Technology, 2-17-1 Tsudanuma, Narashino, Chiba 275-0016, Japan
[b] Faculty of Advanced Engineering, Chiba Institute of Technology, 2-17-1 Tsudanuma, Narashino, Chiba 275-0016, Japan

ARTICLE INFO

PACS numbers:
43.66.Pn
43.66.Qp

Keywords:
Head-Related Transfer Function
Localization
Notch
Peak
Directional Band
Boosted Band

ABSTRACT

It is known that notches and peaks at frequencies above 5 kHz in the head-related transfer function (HRTF) act as cues for median plane sound localization. However, it has also been shown that front-back discrimination of the direction of a sound image can be achieved even with only the components below 4 kHz. In the present study, we investigated the cues for rear sound image localization below 4 kHz. First, we analyzed the HRTFs for 118 ears (59 subjects) in the median plane and showed that the sound pressure around 1 kHz in the rear HRTF was larger than that in the front HRTF. This boosted band (hereinafter referred to as P0) coincided with Blauert's directional band. Next, we proposed a hypothesis that can comprehensively explain the two types of cues that have been proposed in the past: spectral notches and peaks, and directional bands. In order to verify the effects of P0 for rear-direction localization, three preliminary psychoacoustic experiments were conducted. The results showed that eliminating P0 tends to increase localization errors at frequencies below 4 kHz. For wide-band signals, adding P0 to a previous parametric HRTF model (N1N2P1P2) tends to reduce the mean vertical localization error, and made the auditory source width approximately the same as that for the measured HRTF. These preliminary results support our hypothesis and imply that P0 acts as a cue for rear sound image localization.

1 Introduction

Since the 1960s, a number of studies have examined cues for the perception of the vertical angle of a sound image. It is widely known that spectral notches and peaks above 5 kHz contribute to the perception of the vertical angle of a sound image [1–4]. The frequency of notches increases as the sound source moves from the front of the subject to above the subject in the median plane [2,5]. The notches are generated in the pinnae [4,6–9], and the frequency of the notches depends on the shape of the pinnae as well as the vertical angle [10]. The difference in notch frequency due to the vertical angle has been reported to be detectable by the listener [11]. Moreover, the envelope of the amplitude spectrum has been found to be more important than its fine structure [12–17].

A parametric HRTF reconstructed using some or all of the notches and peaks extracted from the measured HRTF has been proposed [18]. It was shown that the minimum parametric HRTF components that provide sound image localization accuracy comparable to that for the measured HRTF in any direction in the upper median plane are the two lowest-frequency notches (N1 and N2) and the two lowest-frequency peaks (P1 and P2) above 4 kHz [18,19].

However, front-back discrimination of a sound image can be achieved with only the components below 4 kHz [20]. It has also been reported that smoothing the amplitude spectrum of the HRTF below 2 kHz increases the front-back error [12]. Therefore, it is considered that there exist some cues for front-back discrimination of a sound image below 4 kHz.

The purpose of the present study is to clarify the cues for rear localization at frequencies below 4 kHz. We analyze the HRTFs for 118 ears for the front and rear directions and show that the sound pressure around 1 kHz in the rear HRTF is larger than that in the front HRTF. This band is herein referred to as the boosted band (P0). We then propose a hypothesis for the cues for vertical localization by combining P0 with N1, N2, P1 and P2. Finally, the validity of P0 as a cue for rear localization is evaluated by three psychoacoustic experiments.

2. Controversy between notch-peak model and directional bands theory

Blauert [21] performed sound localization tests in which 1/3-octave

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

https://doi.org/10.1016/j.apacoust.2024.110370
Received 29 October 2023; Received in revised form 30 September 2024; Accepted 24 October 2024
0003-682X/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

band noise was presented randomly from the front, above, and rear in the median plane, and reported that there were bands that were perceived in specific directions, regardless of the direction of the sound source. These bands were referred to as the directional bands. Furthermore, Blauert analyzed HRTFs and reported that the energy in the directional band was larger than in other directions (hereinafter referred to as the boosted band). However, it has been reported that although directional bands occur in narrow-band signals such as 1/3-octave band noise, for a wide-band signal that does not include any notches, even if a directional band is emphasized, the sound image is not perceived in that direction [18,19,22].

It should be noted that the directional band, which could be regarded as a kind of a spectral peak, and the spectral notches are apparently contradictory cues. Since both findings are reproducible, a hypothesis that integrates these two kinds of cues should be constructed.

3. Comparison of front and rear HRTFs below 4 kHz

In order to find the cues for rear localization below 4 kHz, we analyzed the difference between the HRTFs for the front and the rear directions. In this analysis, the vertical angle, which ranged from 0° to 360°, was defined as the angle measured from the front direction in the median plane, with 0° indicating the front, 90° indicating above, and 180° indicating the rear [23].

The HRTFs for 59 Japanese adult subjects (118 ears) were measured in an anechoic chamber for seven vertical angles in the upper median plane (0° to 180° in 30° steps) using earplug-type microphones, which were placed in the ear canals of the subjects [24]. The diaphragms of the microphones were located at the entrances of the ear canals, under blocked canal conditions [5].

Fig. 1 shows an example of the HRTF for a typical subject for the front and the rear directions. At most frequencies, the sound pressure level (SPL) in the rear direction is lower than that in the front direction. However, the SPL in the rear direction is approximately 10 dB higher than that in the front direction at frequencies around 1 kHz and 10 kHz.

In order to verify whether such a tendency is generally observed, we calculated the difference in the SPL between the rear and front using the HRTFs for 118 ears. Each point in Fig. 2 represents the level difference (rear – front) for each ear. Since most of the points were below 0 dB for most frequencies, the SPL in the front was higher than that in the rear. Large individual differences were observed above 4 kHz. For reference, the interaural SPL difference was calculated for each subject. The mean absolute value averaged over 59 subjects was approximately 2 dB at 1 kHz and approximately 7 dB at 10 kHz. This indicates that individual SPL differences are greater than the interaural SPL difference.

On the other hand, as shown in Fig. 1, the SPL in the rear is higher around 1 kHz, and the individual differences are relatively small, as indicated by the blue arrow in Fig. 2. At around 10 kHz, some ears had a comparatively larger rear SPL, whereas others had a comparatively smaller rear SPL, and there were large individual differences.

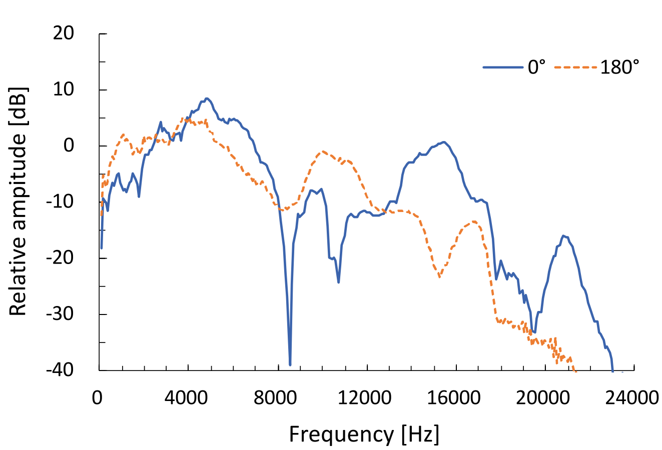

Fig. 1. Examples of HRTF for rear (180°) and front (0°) directions.

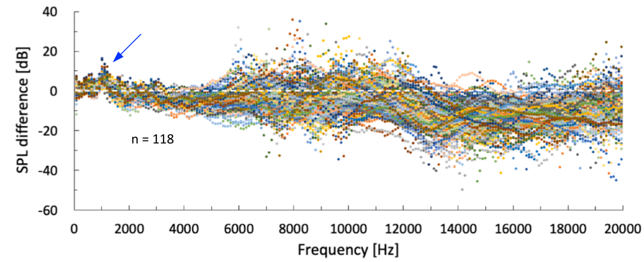

Fig. 2. Difference spectrum, subtracting sound pressure level (SPL) for HRTF for front direction from that for rear direction for 118 ears.

In the present study, we focused on frequencies below 4 kHz, so we calculated the mean SPL difference between the rear and front averaged over 118 ears in the band below 1,500 Hz. As shown in Fig. 3, the maximum mean SPL difference was 5.6 dB at 1,031.25 Hz. The 95 % confidence interval was 0.48 dB.

Fig. 4 shows a normalized histogram of the frequencies at which the difference in SPL between the rear and front HRTFs is a maximum for 118 ears. The frequency resolution was 93.75 Hz. In total, 49 % of the 118 ears had a maximum SPL difference at 1,031.25 Hz, and 86 % had a maximum SPL difference in the 1/3-octave band centered at 1,031.25 Hz.

The above observation suggests that the SPL for the rear HRTF is larger than that for the front HRTF at around 1 kHz, and that individual differences are small.

4. Do directional bands play a role in localization cues for wide-band signals?

4.1. Relationship between vertical angle of sound source and frequencies of N1, N2, P1, and P2 in the median plane

Fig. 5(a) shows an example of a measured HRTF and the corresponding parametric HRTF using all notches and peaks. Fig. 5(b) shows a parametric HRTF reconstructed using only N1, N2, P1, and P2. According to the study in which the model was originally presented [18,19], the parametric HRTF provides a sound image localization accuracy that is approximately the same as that for the measured HRTF in any direction in the upper median plane.

Fig. 6 shows the relationship between the vertical angle of a sound source and the N1, N2, P1, and P2 frequencies in the HRTFs in the median plane. The N1 frequency increases with increasing vertical angle from 0° to 120° and then decreases toward 180°. The N2 frequency

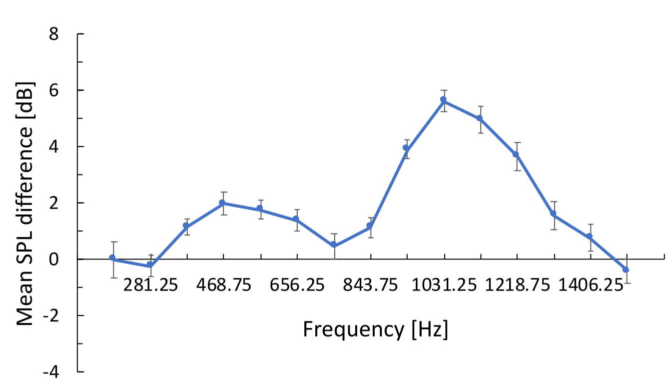

Fig. 3. Mean difference in SPL between rear and front HRTFs for 118 ears. Bars denote the 95% significance range.

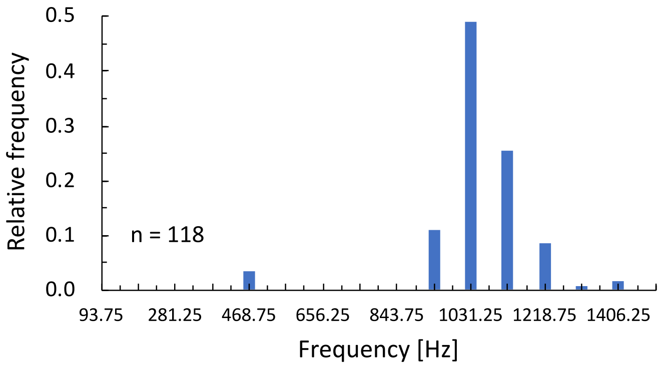

Fig. 4. Normalized histogram of frequencies for which SPL difference between rear and front HRTFs is maximum.

increases with increasing vertical angle from 0° to 120°, but then decreases only slowly for 120° to 180°. Therefore, the vertical angle of a sound source cannot be uniquely determined using either N1 or N2 alone. This explains why two notches are necessary for median plane localization.

On the other hand, the P1 and P2 frequencies are almost independent of the sound source direction. Although various hypotheses have been put forward as to why P1 and P2 are necessary for vertical angle perception [25], none has been firmly established.

Observing Fig. 6 again, it is found that the red color for P1 and P2 is darkest (peak level largest) for the front and above directions, respectively. Thus, although the P1 and P2 frequencies do not depend on the sound source direction, their levels could act as cues for front and above localization, respectively.

4.2. Relationship between mean rear-front difference spectrum and directional bands

The mean SPL difference between the rear and front HRTFs for 118 ears is shown in Fig. 7 (solid orange line). Positive values indicate higher rear SPLs, and negative values indicate higher front SPLs. The regions outlined in orange and green correspond to the rear and front directional bands, respectively. The boosted band for the rear direction at 1 kHz coincides with the rear directional band, and the boosted band for the front direction at 3–5 kHz (shown with negative values) coincides with the front directional band. Therefore, just as P1 and P2 contribute to localization of sound images in the front and above, the presence of P0 is expected to improve the accuracy of localization in the rear.

At 9–14 kHz, it can be seen that the mean SPL difference is not positive, even though this is the rear directional band according to Blauert. As shown in Fig. 2, at around 10 kHz, there are large individual variations in the SPL differences between the rear and front directions. Furthermore, it has been reported that there are individual differences in the directional band, particularly above 9 kHz [27]. This suggests that there are individual differences in the boosted band within this frequency range. This is presumably the reason why the mean SPL difference does not have positive values.

We next compared the directional bands and the P1 and P0 frequencies. The dashed green line in Fig. 7 is the mean P1 frequency averaged over 118 ears and is observed to coincide with the front directional band. Similarly, the dashed orange line is the mean P0 frequency and is seen to coincide with the rear directional band.

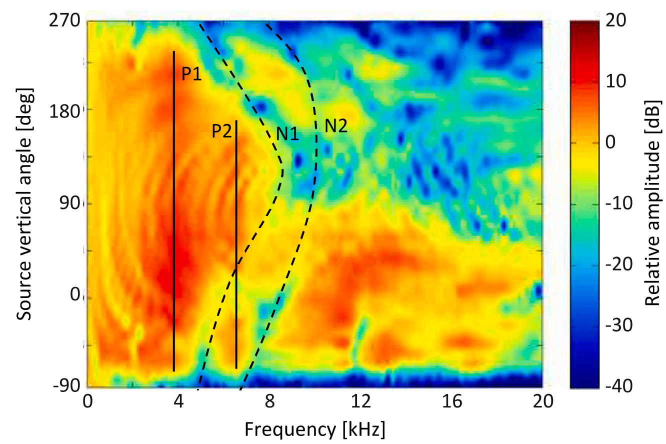

Fig. 6. Relationship between vertical angle of sound source and N1, N2, P1, and P2 frequencies in HRTFs in the median plane.

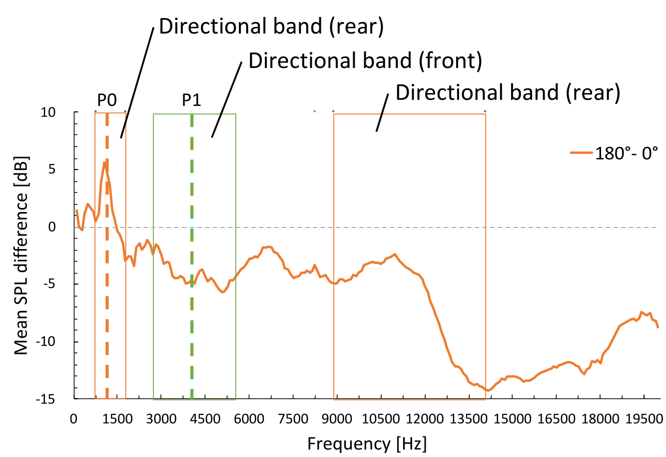

Fig. 7. Relationship between mean SPL difference between rear and front HRTFs for 118 ears (orange solid line) and directional bands for rear (orange outline) and front (green outline). The orange and green dashed lines indicate the mean P0 and P1 frequencies, respectively, averaged over 118 ears.

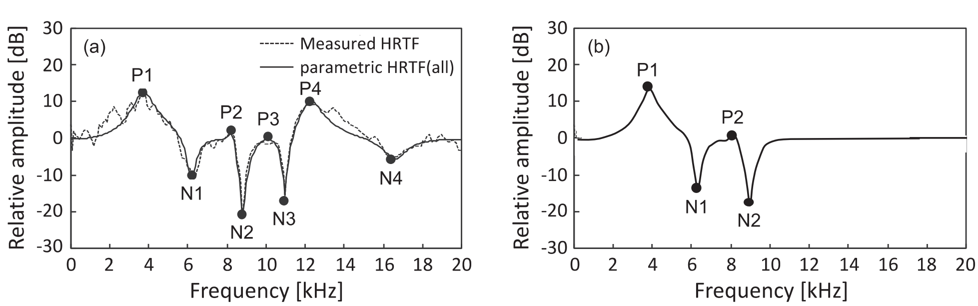

Fig. 5. Examples of parametric HRTFs [26]. (a) measured and parametric HRTFs with all notches and peaks, (b) parametric HRTF reconstructed using N1, N2, P1, and P2.

4.3. Hypothesis combining directional bands with spectral notches and peaks

Based on the above considerations, we propose a hypothesis that the notches in the HRTF are cues for accurate perception of the vertical angle of a sound image, but the presence of P0 combined with the notches acts as a cue to localize a sound image for the rear direction.

This offers a non-contradictory explanation for both the auditory phenomenon due to the directional band for a narrow-band signal and that due to the notches and peaks for a wide-band signal. Based on this hypothesis, P0 is thought to contribute to a narrow-band signal as a directional band and also to a wide-band signal as a cue for the perception of the rear direction together with the notches.

5. Effects of P0 on localization for rear direction

In order to verify the effects of P0 on the localization for the rear direction, three preliminary psychoacoustic experiments were conducted.

5.1. Experiment 1: Sound localization test using stimuli at frequencies below 4 kHz

In Experiment 1, we examined whether P0 acts as a cue for rear localization at frequencies below 4 kHz.

5.1.1. Method of localization tests

The subjects' HRTFs were measured in seven directions in the upper median plane and 12 directions in the horizontal plane more than six months before the experiment. The measurement method was the same as that described in Section 3.

The localization tests were conducted in a soundproof room. The background A-weighted SPL was 19.5 dB. A notebook computer (Apple, MacBook Air), a D/A converter (TEAC, HA-P50SE), and open-air headphones (SONY, MDR-MV1), which are considered to be free-air equivalent coupling (FEC) headphones [28], were used for the localization tests. The headphone transfer functions were not compensated because a preliminary experiment showed no significant differences in the accuracy of sound image localization with and without compensation.

The following stimuli were used at a vertical angle of 180°: (1) a convolution of wide-band white noise and the head-related impulse response (HRIR) for each subject (hereinafter referred to as WB-own), (2) a convolution of low-pass white noise and the HRIR for each subject (LP-own), (3) a convolution of low-pass white noise and the HRTF for each subject following P0 elimination (LP-P0-eliminated). Here, P0-eliminated means that the frequency components of the P0 band for each subject are eliminated. Details of the signal processing LP and P0-eliminated are described below.

The cut-off frequency of the low-pass filter was 4 kHz and the slope was −80 dB per 1/3 octave. P0-eliminated HRTFs were generated by a convolution of the measured HRIR and the P0-eliminated finite impulse response (FIR) filter for each subject. The P0-eliminated FIR filter was generated using the window function method with a Blackman-Harris window. The number of taps for the FIR was 32768. The center frequency of the P0-eliminated filter was set to the P0 frequency for each subject. The bandwidth of the eliminated band was 1/3 octave, and the stopband level was below −80 dB with reference to the passband level. Fig. 8 shows an example of a P0-eliminated HRTF for different frequency ranges.

Each stimulus had a duration of 1 s and was delivered at 62 dB SPL at the entrance of the ear canal. A total of 30 stimuli (three kinds of HRTFs for 180°, ten times) were randomly presented to the subject.

A mapping method was adopted in order to respond on a continuous scale. The subjects' task was to click on the perceived vertical angle of the sound in the median plane shown on a computer display. The subjects were also instructed to check a box on the display when they perceived a sound image inside their heads.

Five subjects (two males and three females, subjects A-E) participated in the sound localization tests. All subjects self-reported normal hearing sensitivity.

5.1.2. Results of localization tests

A. Individual responses.

Fig. 9 shows the individual responses for each subject to the three kinds of stimuli. The diameter of each circle is proportional to the number of responses, and the vertical angle resolution of the circles is 5°. For WB-own, the distribution of the responses is consistent with those of a previous study [29], except for subject E. The responses for subjects A, B, and D were distributed around the target vertical angle of 180°. For subject C, most of the responses were distributed between 130° and 165°. The responses for subject E were distributed in the front half of the upper median plane (0–90°). The responses of subject E for WB-own suggest either a problem with the HRTF measurement for this subject or insufficient auditory localization skills. Regardless of the cause, including this subject's responses in the analysis is unreasonable as this subject is not representative of a qualified participant. Therefore, the responses of subject E were excluded from the analysis.

For LP-own, the distribution of the responses is consistent with those of a previous study [20]. Most of the responses for subjects B, C, and D were distributed around the target vertical angle of 180°. For subject A, the responses were distributed in the rear half of the upper median plane.

For LP-P0-eliminated, the responses were shifted towards the upward direction for subjects A, B, C, and D.

B. Mean elevation angle error.

The mean elevation angle error for each HRTF was calculated, and the results are shown in Table 1. For WB-own, the error was less than 5° for subjects A, B, and D. The error for LP-P0-eliminated was larger than that for LP-own for all subjects. Averaged over four subjects, the error was 18.7° for LP-own and 38.1° for LP-P0-eliminated.

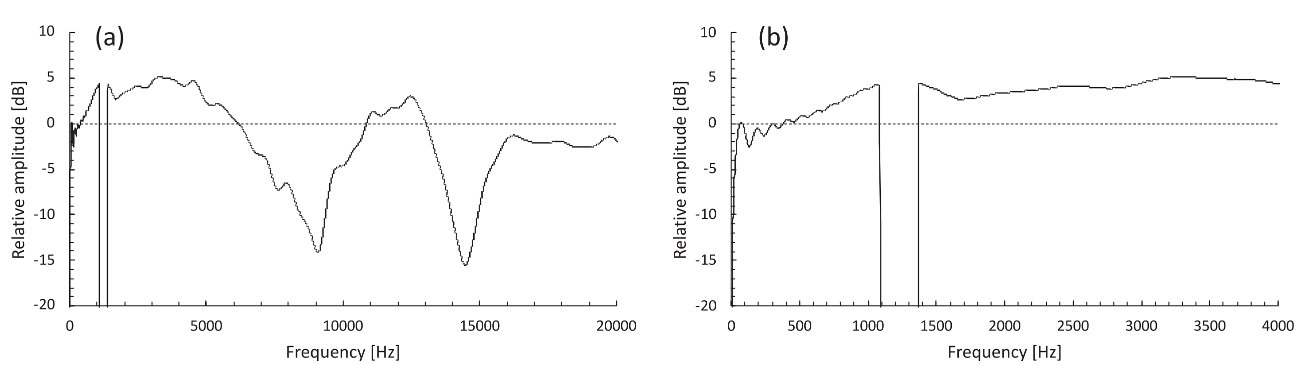

Fig. 8. Example of transfer function obtained by convolution of subject's HRIR at 180° and P0-eliminated filter. (a): 0–20 kHz, (b): 0–4 kHz.

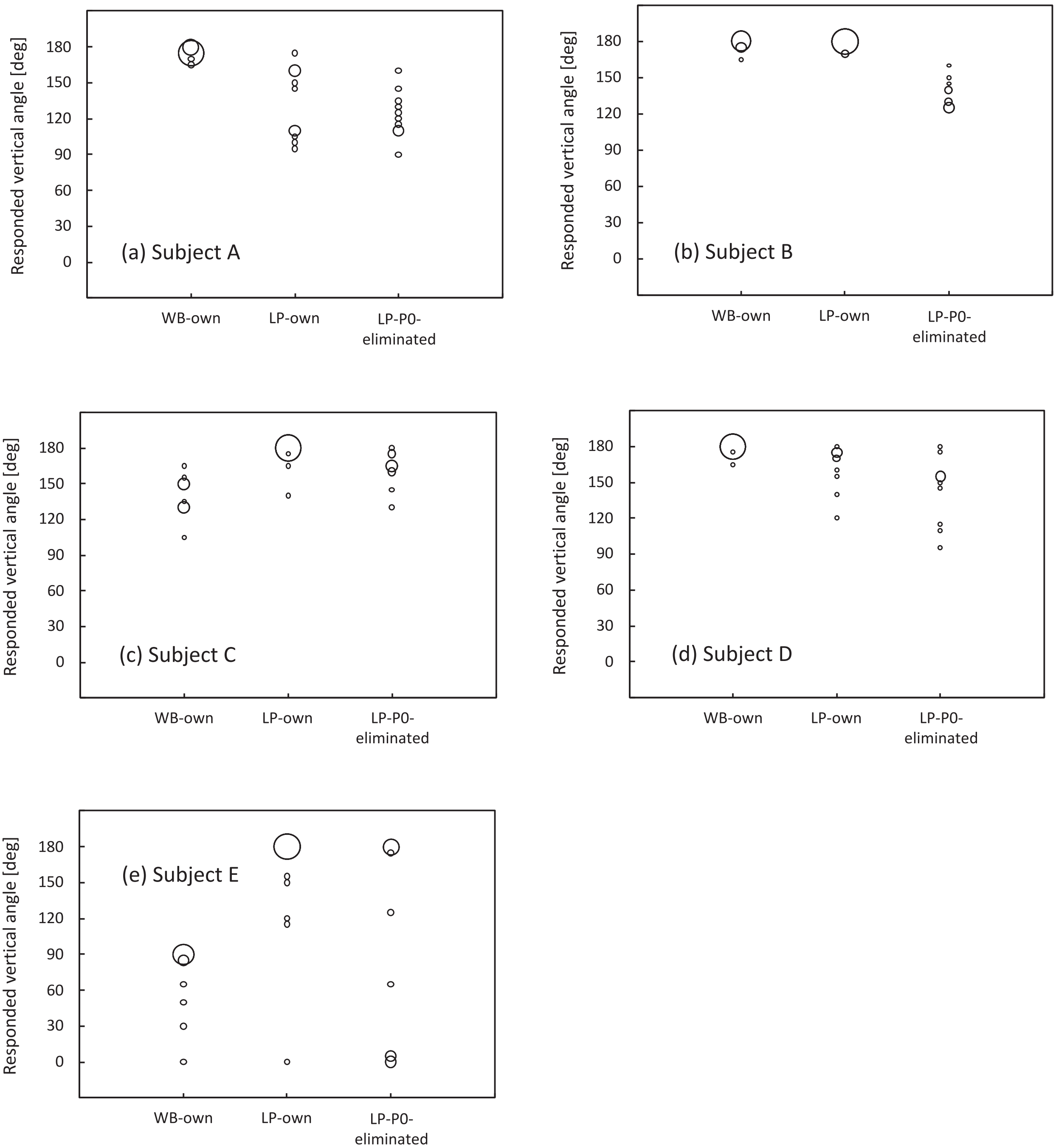

Fig. 9. Individual responses for subjects A–E to WB-own, LP-own, and LP-P0-eliminated stimuli at 180°. The diameter of each circle is proportional to the number of responses. The vertical angle resolution is 5°.

Table 1
Mean elevation angle error [deg].

Subject	WB-own	LP-own	LP-P0-eliminated
A	4.5	48.5	56.1
B	3.5	2.0	42.4
C	39.3	6.2	18.2
D	2.5	18.2	35.7
Ave.	12.4	18.7	38.1

Table 2
Results of Mann-Whitney U test for LP-own and LP-P0-eliminated under 4 kHz. **: $p < 0.01$, *: $p < 0.05$.

Subject			
A	B	C	D
	**	**	

Table 2 shows the results of a Mann-Whitney U test for the LP-own and LP-P0-eliminated stimuli. The mean elevation angle errors for LP-P0-eliminated are significantly larger than those for LP-own for two of the four subjects.

C. Ratio of front-back confusion

No front-back errors occurred for subjects A-D.

D. Ratio of inside-of-head localization

Table 3 shows the ratio of inside-of-head localization. For WB-own,

Table 3
Ratio of inside-of-head localization.

Subject	WB-own	LP-own	LP-P0-eliminated
A	0.10	0.90	1.00
B	0.00	0.10	0.60
C	0.00	0.00	0.10
D	0.00	1.00	1.00
Ave.	0.03	0.50	0.68

inside-of-head localization occurred once for subject A. For LP-own, the ratio of inside-of-head localization was high for subjects A and D. The ratio of inside-of-head localization for LP-P0-eliminated was higher than that for LP-own for subject B, and slightly higher for subjects A and C.

Previous studies using wide-band signals have reported that localization errors when listeners receive inadequate spectral information for vertical localization can be roughly classified into three categories: (1) front-back confusion, (2) rising sound image, and (3) inside-of-head localization [30].

In this experiment, even when the major cues for perception of the vertical angle (N1, N2, P1, and P2) were removed by applying a low-pass filter, the subjects perceived a sound image at the rear for the rear target direction. This is consistent with previous experimental results [20]. The reason why the subjects perceived a sound image at the rear is thought to be that P0 acted as a cue for rear localization. Furthermore, by removing P0, the above-mentioned rising sound image and inside-of-head localization occurred. However, front-back confusion did not occur. In general, if the stimulus includes N1, N2, P1, and P2 for the rear direction of other listeners or dummy heads, and the frequencies of these notches and peaks are closer to those in the subject's front HRTF than rear HRTF, front-back confusion occurs [30]. The reason why front-back confusion did not occur in this experiment is thought to be that there were no cues to localize a sound image for the front direction because N1, N2, P1, and P2 were removed.

These results show that eliminating P0 tends to increase the localization errors for frequencies below 4 kHz and imply that P0 contributes to rear localization for stimuli under 4 kHz.

5.2. *Experiment 2: Sound localization test using wide-band stimuli*

In Experiment 2, we examined whether P0 contributes to rear localization for wide-band signals using the minimum parametric HRTF that provides a sound image localization accuracy comparable to that for the measured HRTF.

5.2.1. *Generation of parametric HRTFs*

The parametric HRTF reconstructed using N1, N2, P1, and P2 (hereinafter referred to as N1N2P1P2) was generated by superposition of the notches and peaks. Parametric HRTF N1N2P1P2 + P0 was created by adding P0 to N1N2P1P2. Each notch and peak was reproduced using a second-order infinite impulse response (IIR) filter. The center frequency, level, and Q factor for each IIR filter were manually adjusted to the notches and peaks in the subject's measured HRTFs using the custom-made software shown in Fig. 10.

Fig. 11 shows examples of a measured HRTF and parametric HRTFs N1N2P1P2 and N1N2P1P2 + P0. By adding P0, the spectrum of N1N2P1P2 + P0 comes close to that of the measured HRTF around 1 kHz.

5.2.2. *Method of localization tests*

Localization tests were conducted in a quiet laboratory room with a background A-weighted SPL of 32.0 dB. A notebook computer (Apple, MacBook Air), a D/A converter (TEAC, HA-P50SE), and open-air headphones (beyerdynamic, DT990 PRO), which are regarded as FEC headphones [28], were used for the tests.

The source signal was wide-band white noise from 200 Hz to 17 kHz. The following HRTFs were used: (1) each subject's measured HRTF for 0°, (2) each subject's measured HRTF for 180°, (3) each subject's N1N2P1P2 for 180°, and (4) each subject's N1N2P1P2 + P0 for 180°. Measured HRTFs for 0° and 180° were included in order to validate the

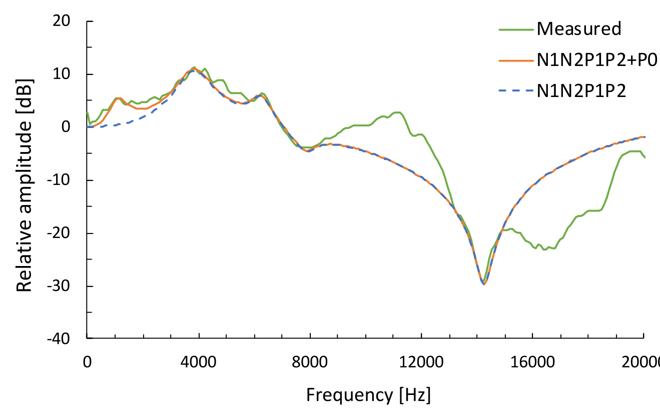

Fig. 11. Examples of measured HRTF and generated parametric HRTFs (N1N2P1P2 and N1N2P1P2 + P0) for rear direction.

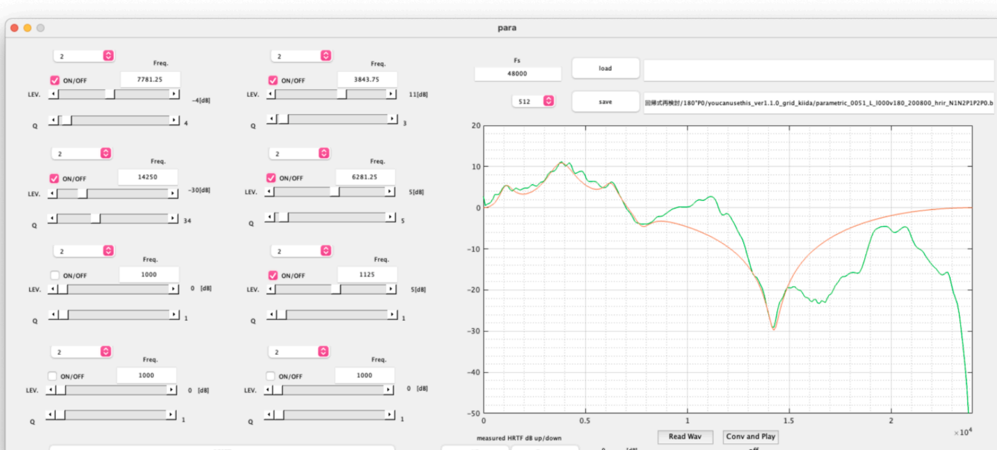

Fig. 10. Custom-made software for generating parametric HRTFs. The left side of the panel shows parameter settings for eight IIR filters, each of which creates a notch or a peak.

experimental system and each subject's localization ability.

Each stimulus was delivered at 63 dB SPL at the entrance of each ear canal. The duration of the stimuli was 1 s. A total of 40 stimuli (four HRTFs, ten times) were randomly presented to a subject. Therefore, each subject responded to each stimulus 10 times. The mapping method described in Section 5.1.1 was again used for the localization tests.

Four subjects (two males and two females, subjects B, F-H) participated in the sound localization tests. All subjects self-reported normal hearing sensitivity.

5.2.3. *Results of localization tests*

A. Individual responses.

Fig. 12 shows the responses to the subject's own measured HRTFs, N1N2P1P2, and N1N2P1P2 + P0 for the four subjects. For the measured HRTFs at 0° and 180°, the responses for subjects B, F, and G are consistent with those in a previous study [29]. These responses were distributed around the target vertical angles. For subject H, the response was to the rear in two out of ten cases for the measured HRTF for target vertical angle of 0° and to the front in one out of ten cases for the measured HRTF for target vertical angle of 180°.

The responses to N1N2P1P2 were shifted upward for subjects B, F, and G. Most of the responses for subject H were distributed between 110° and 180°, and the response was to the front in three out of ten cases.

The responses to N1N2P1P2 + P0 were also distributed around the target vertical angle of 180° for subjects B, F, and G. For subject H, the response was to the front in two out of ten cases.

B. Mean elevation angle error.

Table 4 shows the mean elevation angle error for each HRTF. For the measured HRTFs for 0° and 180°, the error averaged over four subjects was 2.7° and 4.6°, respectively. For 180°, adding P0 to N1N2P1P2 reduced the error for all subjects; averaged over four subjects, it was 15.6° for N1N2P1P2 and 8.8° for N1N2P1P2 + P0. The error for subject H was larger than that for the other three subjects for all HRTFs.

Table 4
Mean elevation angle error [deg].

Subject	0° Measured	180° N1N2P1P2	180° N1N2P1P2 + P0	180° Measured
B	2.0	10.2	4.2	1.9
F	1.6	10.6	4.2	3.2
G	0.4	4.2	2.6	0.4
H	6.8	37.2	24.0	12.7
Ave.	2.7	15.6	8.8	4.6

Table 5 shows the results of Tukey's multiple comparison tests for the measured HRTF, N1N2P1P2, and N1N2P1P2 + P0 for a target direction of 180°. The mean elevation angle error for N1N2P1P2 was significantly larger than those for the measured HRTFs for three of the four subjects. On the other hand, no statistically significant difference was observed between the measured HRTFs and N1N2P1P2 + P0 for any of the subjects. A significant difference was observed between N1N2P1P2 and

Table 5
Results of Tukey's multiple comparison test for measured HRTF and parametric HRTFs for rear direction. **: $p < 0.01$, *: $p < 0.05$.

Subject		Measured	N1N2P1P2	N1N2P1P2 + P0
B	Measured	–		
	N1N2P1P2	**	–	
	N1N2P1P2 + P0	n.s.	*	–
F	Measured	–		
	N1N2P1P2	n.s.	–	
	N1N2P1P2 + P0	n.s.	n.s.	–
G	Measured	–		
	N1N2P1P2	*	–	
	N1N2P1P2 + P0	n.s.	n.s.	–
H	Measured	–		
	N1N2P1P2	*	–	
	N1N2P1P2 + P0	n.s.	n.s.	–

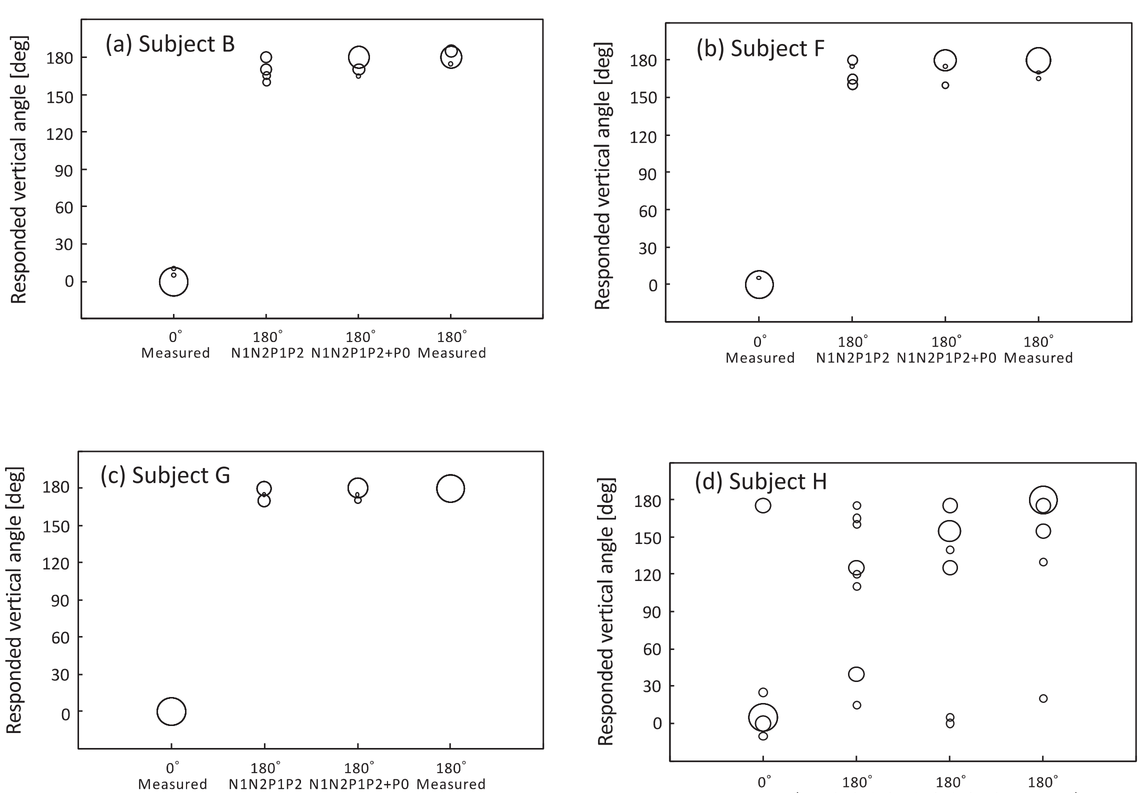

Fig. 12. Responses to measured HRTF, N1N2P1P2, and N1N2P1P2 + P0 for front and rear directions.

N1N2P1P2 + P0 for one subject.

C. Ratio of front-back confusion

Table 6 shows the ratio of front-back confusion. As described above, front-back error occurred only for subject H, and the ratio of front-back confusion for this subject was 0.10–0.30.

D. Ratio of inside-of-head localization

Table 7 shows the ratio of inside-of-head localization. Inside-of-head localization occurred only for subject H, and the ratio for this subject was 0.10–0.30.

These results imply that N1N2P1P2 + P0 provides approximately the same vertical localization performance as the measured HRTFs at 180° for wide-band signals.

5.3. *Experiment 3: Psychoacoustical tests for auditory source width*

It was pointed out in the subject's introspection report that there was a difference in the width of a sound image (auditory source width: ASW) between the stimuli. The ASW is defined as the width of a sound image fused temporally and spatially with the direct sound image [31,32]. In general, the direction of a sound image is clear and easy to localize for the case of a narrow ASW. For a wide ASW, on the other hand, the direction of a sound image is unclear and hard to localize. We conducted experiments to compare the ASW for the measured HRTF, N1N2P1P2, and N1N2P1P2 + P0 using Ura Variation of Scheffe's paired comparison.

5.3.1. *Method of comparison tests*

The comparison tests were conducted in an anechoic room. The experimental apparatus was the same as that used in Experiment 2. The source signal was wide-band white noise from 200 Hz to 17 kHz. The following HRTFs were used: (1) each subject's own measured HRTF for 180°, (2) each subject's N1N2P1P2 for 180°, and (3) each subject's N1N2P1P2 + P0 for 180°. Each stimulus was delivered at 72 dB SPL at the entrance of each ear canal. The duration of the stimuli was 1 s.

Stimuli were presented in pairs. The interval between stimuli was 1 s. Each subject's task was to indicate how narrow or wide the ASW of the subsequently presented stimulus was compared to the previously presented stimulus. Subjects responded by marking with a pen a number line from −2.0 to 2.0 at 0.1 intervals on the response sheet, −2.0 for very narrow, 0.0 for the same, and 2.0 for very wide. There were $3 \times 2 = 6$ stimulus pairs, and each pair was presented to the subjects five times in random order.

Five subjects (two males and three females, subjects A-B, F-H) participated in the tests. All subjects self-reported normal hearing sensitivity.

5.3.2. *Results of comparison tests*

Fig. 13 shows the constellation of ASWs for the measured HRTF, N1N2P1P2, and N1N2P1P2 + P0 for each subject.

The ASW for the measured HRTF was the narrowest for all subjects except subject B. For subject B, the ASW for N1N2P1P2 + P0 was the narrowest. The ASW for N1N2P1P2 was the widest for all subjects.

Compared with the yardstick with a significance level of 5 %, the ASW for N1N2P1P2 + P0 was significantly narrower than that for N1N2P1P2 for all subjects except subject H. There was no significant difference between N1N2P1P2 + P0 and the measured HRTF for subjects A, B, and F. N1N2P1P2 was significantly wider than that for the measured HRTF for all subjects.

Namely, the ASW for N1N2P1P2 tends to be wider than the measured HRTF (sound image was blurred) but adding P0 to N1N2P1P2 makes the ASW closer to that for the measured HRTF.

Table 6
Ratio of front-back confusion.

Subject	0° Measured	180° N1N2P1P2	180° N1N2P1P2 + P0	180° Measured
B	0.00	0.00	0.00	0.00
F	0.00	0.00	0.00	0.00
G	0.00	0.00	0.00	0.00
H	0.20	0.30	0.20	0.10
Ave.	0.05	0.08	0.05	0.03

Table 7
Ratio of inside-of-head localization.

Subject	0° Measured	180° N1N2P1P2	180° N1N2P1P2 + P0	180° Measured
B	0.00	0.00	0.00	0.00
F	0.00	0.00	0.00	0.00
G	0.00	0.00	0.00	0.00
H	0.00	0.30	0.30	0.10
Ave.	0.00	0.08	0.08	0.03

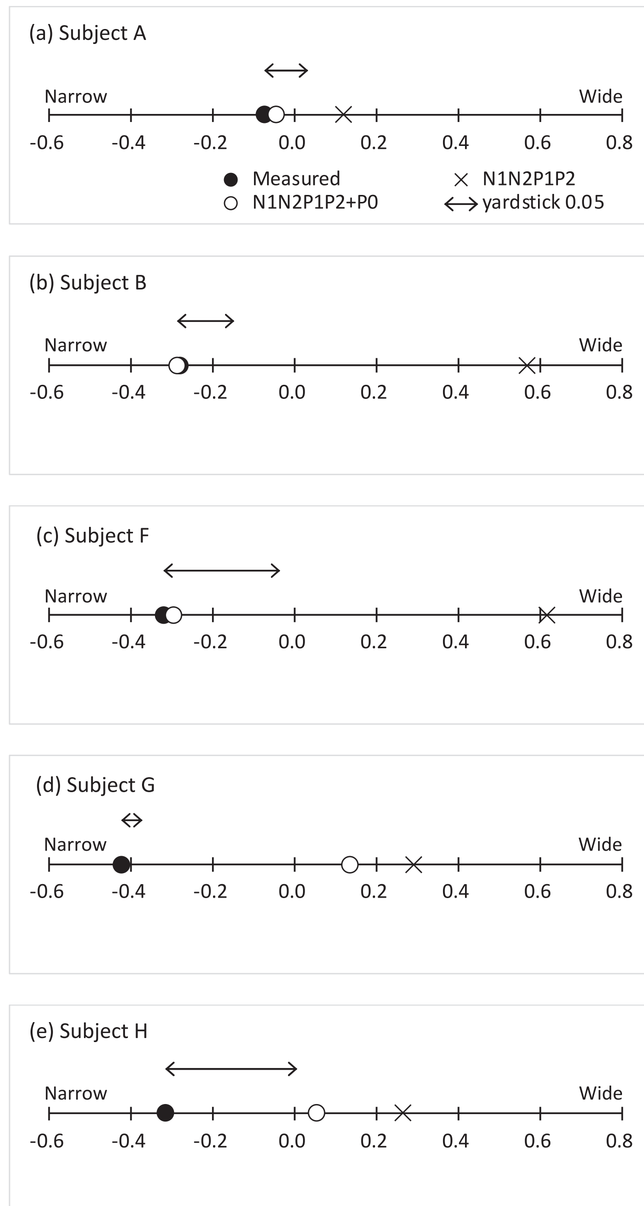

Fig. 13. Auditory source width for measured HRTF (●), N1N2P1P2 (×), and N1N2P1P2 + P0 (○) for rear direction. Arrow indicates yardstick (0.05).

6. Discussion

6.1. Comparison of findings regarding P0 with previous studies

In Section 4.2, we showed that the P0 frequency coincides with the rear directional band. Here, we compare the findings regarding P0 with previous studies. Several studies have examined the effects of low-frequency components of HRTFs on sound localization. Yao *et al*. [33] investigated the contribution of frequency components to median plane localization using the CIPIC database [34]. They selected three frequency bands (400 Hz to 1.2 kHz, 4 to 8 kHz, and 12 to 14 kHz) as the high-F-ratio components. The lowest frequency band (400 Hz to 1.2 kHz) coincides with the P0 frequency. Lee [35] compared front-back symmetrical sound sources below 3 kHz and emphasized the importance of this frequency region for localization.

6.2. Experimental support for proposed hypothesis

Experiment 1 showed that eliminating P0 increased localization errors, and that P0 acted as a cue for rear localization at frequencies below 4 kHz. Furthermore, the results of Experiment 2 indicated that adding P0 to N1N2P1P2 reduced the mean vertical localization error. Finally, Experiment 3 showed that the ASW for N1N2P1P2 was wider (sound image was blurred) than the measured HRTF, but adding P0 to N1N2P1P2 reduced the ASW to almost the same value as that for the measured HRTF.

These results support the hypothesis proposed in Section 4.4 that although the notches in the HRTF are cues for accurate perception of the vertical angle of a sound image, the additional presence of P0 combined with the notches acts as a cue for localization in the rear direction.

6.3. Reason why amplitude of rear HRTF is dominant at 1 kHz

Why does the amplitude of the rear HRTF dominate that of the front HRTF around 1 kHz? In a previous study, a comparison was made between the HRTF calculated by the FDTD method using only the shape of the pinnae, and the HRTF calculated using the entire head shape including the pinnae [9]. In the former case, no effect of the pinnae was observed for any vertical angle at a frequency of 1 kHz. In the latter case, effects of the head were observed, but no difference was found between the front and rear.

Furthermore, from a wavelength point of view, the dominance of the rear HRTF cannot be considered to be the effect of the pinnae. Given the small individual differences in the boosted band around 1 kHz, one possible reason may be the effect of the torso. Further investigation is therefore required in order to clarify the origin of the boosted band.

6.4. Cues for front sound image localization

Thus far, we have focused on rear localization cues at frequencies below 4 kHz, but here we consider front localization cues. We suggest that one such cue could be P1, which is a peak in the minimum parametric HRTF that provides sound image localization accuracy comparable to that for the measured HRTF in any direction in the upper median plane, as described in Section 1. P1 also coincides with the front directional band, as described in Section 4.3.

The mean P1 frequency for the 118 ears in the present study was 4020.0 Hz with a standard deviation of 249.5 Hz. Because P1 is a broad peak with a Q-factor of 1–2, it could act as a cue for front localization even for stimuli below 4 kHz.

7. Conclusions

In the present study, we investigated cues for rear sound image localization at frequencies below 4 kHz. We analyzed the difference between the HRTFs for the front and the rear directions for 118 ears. We then proposed a hypothesis that although the notches in the HRTF are cues for accurate perception of the vertical angle of a sound image, the additional presence of P0 acts as a cue for localizing a sound image for the rear direction. In order to verify the effects of P0 on localization for the rear direction, three preliminary psychoacoustic experiments were conducted.

The results are summarized below.

(1) The sound pressure around 1 kHz in the rear HRTF is larger than that in the front HRTF. This boosted band coincides with Blauert's directional band.
(2) Eliminating P0 tends to increase the localization errors for frequencies below 4 kHz. For wide-band signals, adding P0 to the parametric HRTFs reconstructed using N1N2P1P2 tends to reduce the mean vertical localization error, and makes the ASW closer to that for the measured HRTF.
(3) These preliminary results support the proposed hypothesis, and imply that P0, the boosted band at 1 kHz in the rear HRTF, acts as a cue for rear sound image localization.

Future runs are planned due to the low number of participants in the current study.

CRediT authorship contribution statement

Fuka Nakamura: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Kazuhiro Iida:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

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.

Data availability

The data that has been used is confidential.

References

[1] 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.
[2] Butler A, Belendiuk K. Spectral cues utilizes in the localization of sound in the median sagittal plane. J Acoust Soc Am 1977;61:1264–9.
[3] Mehrgardt S, Mellert V. Transformation characteristics of the external human ear. J Acoust Soc Am 1977;61:1567–76.
[4] Musicant A, Butler R. The influence of pinnae-based spectral cues on sound localization. J Acoust Soc Am 1984;75:1195–200.
[5] 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.
[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] 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.
[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] 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.
[10] 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.
[11] 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.
[12] Asano F, Suzuki Y, Sone T. Role of spectral cues in median plane localization. J Acoust Soc Am 1990;88:159–68.
[13] Middlebrooks JC. Narrow-band sound localization related to external ear acoustics. J Acoust Soc Am 1992;92:2607–24.

[14] Middlebrooks JC. Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. J Acoust Soc Am 1999;106:1493–510.
[15] Kulkarni A, Colburn HS. Role of spectral detail in sound- source localization. Nature 1998;396:747–9.
[16] Langendijk EHA, Bronkhorst AW. Contribution of spectral cues to human sound localization. J Acoust Soc Am 2002;112:1583–96.
[17] Macpherson EA, Sabin AT. Vertical-plane sound localization with distorted spectral cues. Hear Res 2013;306:76–92.
[18] 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.
[19] 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.
[20] Morimoto M, Saito A. On sound localization in the median plane - Effects of frequency range and intensity of stimuli -. Technical Report of Technical Committee of Psychological and Physiological Acoustics. Acoust Soc Jpn 977; H-40-1 (in Japanese).
[21] Blauert J. Sound localization in the median plane. Acust 1969;70(22):205–13.
[22] Iida K. In: Head-Related Transfer Function and Acoustic Virtual Reality. Springer; 2019. P 128.
[23] Morimoto M, Aokata H. Localization cues of sound sources in the upper hemisphere. J Acoust Soc Jpn (E) 1984;5:165–73.
[24] 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(1):317–33.
[25] In IK. Head-Related Transfer Function and Acoustic Virtual Reality. Springer 2019: 36–8.
[26] Iida K. In: Head-Related Transfer Function and Acoustic Virtual Reality. Springer; 2019. P 33.
[27] Itoh M, Iida K, Morimoto M. Individual differences in directional bands. Appl Acoust 2007;68:909–15.
[28] Møller H, Hammershøi D, Jensen CJ, Sørensen MF. Transfer characteristics of headphones measured on human ears. J Audio Eng Soc 1995;43:203–17.
[29] Iida K. In: Head-Related Transfer Function and Acoustic Virtual Reality. Springer; 2019. P 27.
[30] Iida K. In: Head-Related Transfer Function and Acoustic Virtual Reality. Springer; 2019. P 30.
[31] Morimoto M, Iida K, Sakagami K. The role of reflections from behind the listener in spatial impression. Appl Acoust 2001;62:109–24.
[32] Morimoto M, Iida K. A practical evaluation method of auditory source width in concert halls. J Acoust Soc Jpn (E) 1995;16:59–69.
[33] Yao D, Li J, Xia R, Yan Y. The role of spectral cues in vertical plane elevation perception. Acoust Sci and Tech 2020;41:435–8.
[34] Algazi V.R, Duda RO, Thompson DM, Avendano C. The cipic HRTF database. Proc. 2001 IEEE Workshop Applications of Signal Processing to Audio and Acoustics 2001;99–102.
[35] Lee H. Sound Source and Loudspeaker Base Angle Dependency of the Phantom Image Elevation Effect. J Audio Eng Soc 2017;65:733–48.