Chapter 3
HRTF and Sound Localization in the Median Plane

Abstract This chapter describes the physical aspects of HRTFs in the median plane and sound image localization by reproducing these HRTFs. Cues for vertical localization, which are included in HRTFs, are also discussed.

3.1 HRTFs in the Median Plane

Figure 3.1 shows the amplitude spectra of the HRTFs for a sound source in the median plane from the front (vertical angle of 0°) to the rear (vertical angle of 180°) in 30° steps. The solid red lines indicate the HRTFs for the right ear, and the blue dotted lines indicate those for the left ear. A difference of 10 dB is indicated on the ordinate. Each HRTF is drawn with a 40-dB shift. The broken line indicates the 0-dB level for each HRTF. Figure 3.1 shows the following:

1) The frequency of the peak around 4 kHz is approximately constant, independent of the vertical angle of the sound source.
2) The frequencies of the notches shift higher as the sound source moves from the front of the subject to above the subject, and reach a maximum at a vertical angle of 120°.
3) The notches are deep for a sound source near the horizontal plane and are shallow for a sound source above the horizontal plane.
4) The HRTFs for the left and the right ears are not identical, even for a sound source located in the median plane. This is due to the asymmetry of the head and pinnae shape.

Thus, the amplitude spectrum of a HRTF varies with the vertical angle of the sound source.

K. Iida, *Head-Related Transfer Function and Acoustic Virtual Reality*,
https://doi.org/10.1007/978-981-13-9745-5_3

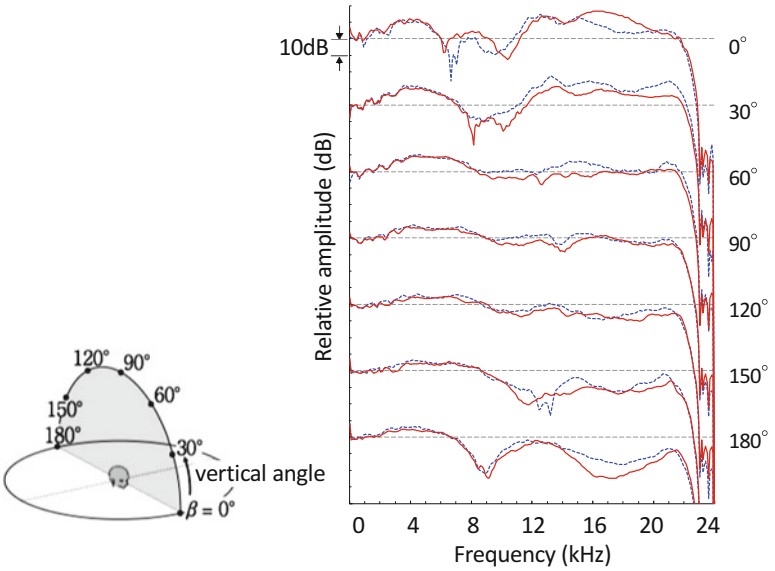

Fig. 3.1 Amplitudes of HRTFs for sound source in upper median plane (0°–180°). The solid red lines and blue dotted lines indicate HRTFs for the right and left ears, respectively

3.2 Sound Localization in the Median Plane

3.2.1 Localization Using listener's Own HRTFs

1. *Reproduction of HRTFs through headphones*

Figure 3.2 shows the results of localization tests in the upper median plane (0°–180°, in 30° steps), in which the listener's own HRTFs were reproduced through headphones using the method described in Sect. 2.2. The subjects were three males and one female, all in their twenties. The sound source was wide-band white noise from 200 Hz to 17 kHz.

Most of the responses are distributed along the solid diagonal line in the figure, whereas the responses for subjects MTZ and YMM are distributed as an inverted s-shaped curve above this diagonal line, which appeared in the responses to the actual sound sources (see Sect. A1.2). These results indicate that the listeners localized a sound image around the target vertical angle when their own HRTFs were reproduced.

2. *Reproduction of HRTFs by a transaural system using two loudspeakers*

Figure 3.3 shows the results of localization tests (Morimoto and Ando 1980) in which the subject's own HRTFs in the upper median plane (0°–180°, in 15° steps) were reproduced at the entrances of the subject's ear canals by a transaural system.

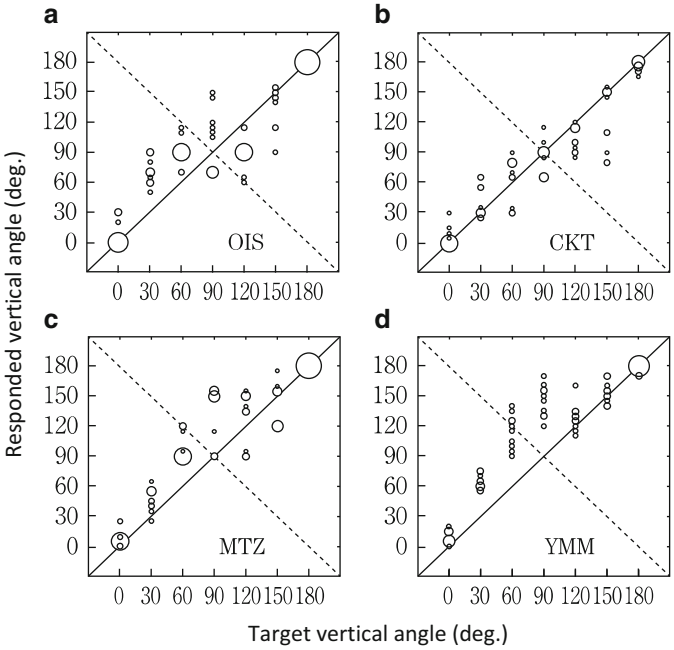

Fig. 3.2 Sound image localization in the upper median plane by reproducing listener's own HRTFs through headphones. The radius of each circle is proportional to the number of responses with a resolution of 5°. (**a**) subject OIS, (**b**) subject CKT, (**c**) subject MTZ, and (**d**) subject YMM

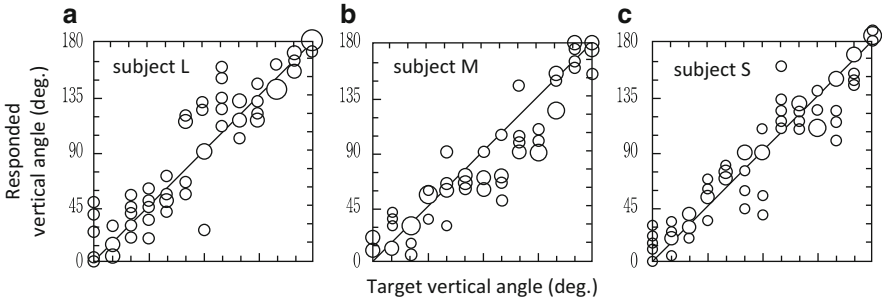

Fig. 3.3 Sound image localization in the upper median plane by reproducing listener's own HRTFs using transaural system. (**a**) subject L, (**b**) subject M, and (**c**) subject S. (Morimoto and Ando 1980)

The subjects were three males (L, M, and S) in their twenties. The sound source was wide-band white noise from 300 Hz to 13.6 kHz.

The responses for each subject are distributed along a diagonal line, although the variances of the responses were larger than those in the horizontal plane. This tendency is observed in localization tests using actual sound sources.

Thus, the direction of a sound image in the median plane can be reproduced by reproducing the listener's own HRTFs using either the headphones or the transaural system.

3.2.2 *Localization Using Others' HRTFs*

1. *Reproduction of HRTFs through headphones*

Figure 3.4 shows the results of localization tests in which others' HRTFs in the upper median plane (0°–180°, in 30° steps) were reproduced through headphones. The subjects are the same as for Fig. 3.2. The HRTFs of a male other than the subject were reproduced. Subject OIS localized a sound image around the target vertical angle. However, the other three subjects localized a sound image at 120°–180° for a target vertical angle of 0°. Moreover, the responses for subject MTZ are widely distributed from forward to rearward for a target vertical angle of 180°. Thus, reproducing other's HRTFs often causes front-back errors.

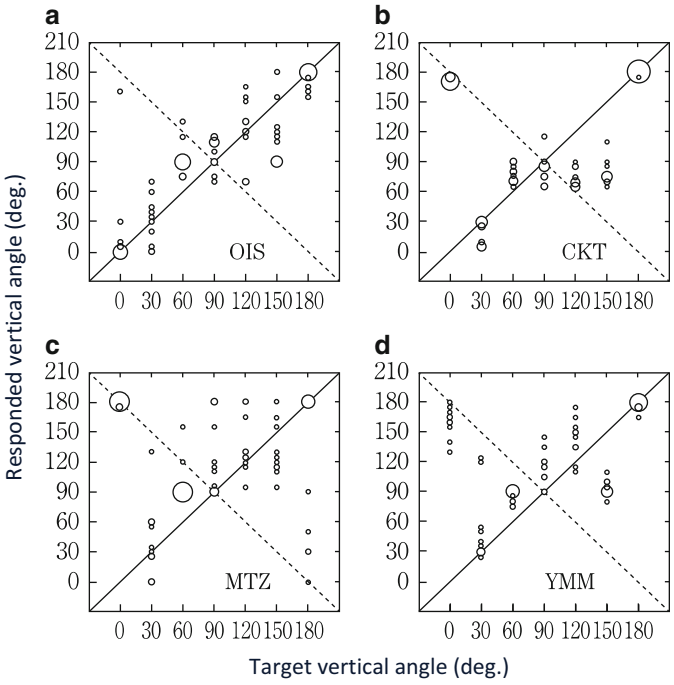

Fig. 3.4 Sound image localization in the upper median plane by reproducing other's HRTFs through headphones. (**a**) subject OIS, (**b**) subject CKT, (**c**) subject MTZ, and (**d**) subject YMM

2. *Reproduction of HRTFs by a transaural system using two loudspeakers*

Figure 3.5 shows the results of localization tests in which others' HRTFs in the upper median plane (0°–180°, in 15° steps) were reproduced by a transaural system (Morimoto and Ando 1980). The experimental conditions are the same as for Fig. 3.3, except for the reproduced HRTFs. Panels (a) and (b) show the responses for subject L to the HRTFs of subjects M and S, respectively, and panels (c) and (d) show the responses for subject S to the HRTFs of subjects L and M, respectively. The designations of the subjects (i.e., subjects S, M, and L) indicate the relative size of the subject's pinnae (i.e., small, medium, and large, respectively). Subject L never localized a sound image at the front for the HRTFs of subjects M and S. Most of the responses for subject L were distributed from 120° to 180°. The responses for subject S shifted upward for the HRTFs of subject L for target vertical angles of 0°–45°. The responses for subject S were widely distributed for the HRTFs of subject M for target vertical angles of 45°–120°.

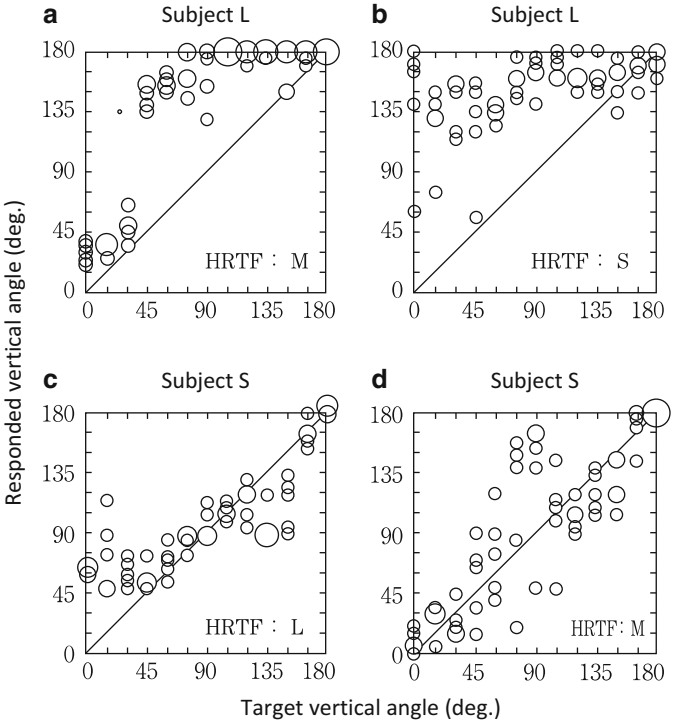

Fig. 3.5 Sound image localization in the upper median plane by reproducing others' HRTFs using transaural system. Panels (**a**) and (**b**) show the responses for subject L to the HRTFs of subjects M and S, respectively, and panels (**c**) and (**d**) show the responses for subject S to the HRTFs of subjects L and M, respectively. (Morimoto and Ando 1980)

3.2.3 *Three Major Errors Regarding Sound Localization in the Median Plane*

Subjects often perceive a sound image in a direction other than the target direction when other's HRTFs are reproduced. The errors fall roughly into the following three categories (Fig. 3.6): (1) front-back confusion, (2) rising of a sound image, and (3) inside-of-head localization (lateralization).

1. **Front-back confusion**

Front-back confusion is a phenomenon in which the target direction and the perceived direction are reversed back to front.

Since the frequencies of the prominent notches (the cues for the front-back direction are described in Sect. 3.3) of other's HRTFs do not often coincide with those of the listener's HRTFs, listeners obtain inadequate information for sound image localization.

Listeners often perceive a sound image to the front for a rear target direction when other's HRTFs are reproduced by a transaural system in which the loudspeakers are visible. This might be explained by visual information affecting the direction of a sound image when the aural information is ambiguous. On the other hand, listeners often perceive a sound image at the rear for a front target direction when other's HRTFs are reproduced through headphones, in which case the loudspeakers are invisible.

2. **Rising of a sound image**

A sound image often rises from the horizontal plane, even though the target direction is on the horizontal plane, when other's HRTFs are reproduced. The reason why sound images rise, but do not fall, is unclear.

3. **Inside-of-head localization (lateralization)**

When diotic sound signals without HRTFs are reproduced at the entrances of the ear canals, listeners perceive a sound image inside their heads. The human auditory system cannot detect vertical angle information in ear-input signals, which does not include HRTFs, and, as a result, the listener perceives a sound image inside his/her head.

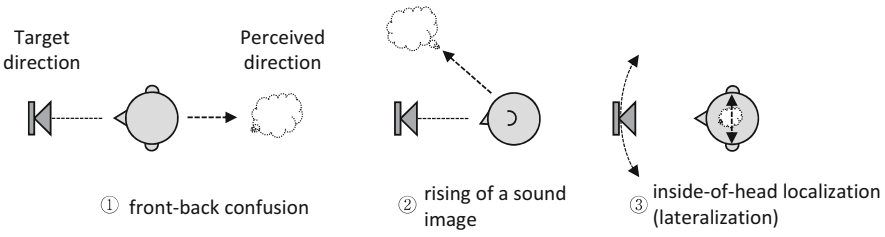

Fig. 3.6 Three major localization errors when other's HRTFs are reproduced

3.3 Cues for Vertical Localization

3.3.1 Overview of Spectral Cues

What are the cues for vertical localization? After the 1970s, a number of studies examined cues for the perception of vertical direction. Consequently, it was found that cues exist in the amplitude spectrum of the HRTF. These are referred to as spectral cues. Moreover, studies to find the specific important part of the amplitude spectrum that acts as a spectral cue have been performed.

These studies revealed that spectral cues exist in the frequency range above 5 kHz. Figure 3.7 shows the effect of the frequency range of the sound source on the accuracy of median plane localization (Morimoto and Saito 1977). For wide-band white noise (a) and noise with a low-pass cut-off frequency of 9.6 kHz (b), the responses are distributed along a diagonal line. For noise with a low-pass cut-off frequency of 4.8 kHz (c), however, the subjects never localized a sound image in the upper direction. For a cut-off frequency of 2.4 kHz (d), front-back errors were observed. On the other hand, for high-pass-filtered noise (e) through (g), the variance of the distribution of the responses tended to be large as the cut-off frequency increased. These results suggest that the frequency components between 5 kHz and 10 kHz are necessary for accurate median plane localization.

Moreover, it has been reported that the frequency components above 16 kHz and below 3.8 kHz do not affect the accuracy of median plane localization (Hebrank and Wright 1974). However, for the sagittal plane far from the median plane, cues for vertical localization were reported to exist below 3 kHz (Algazi et al. 2001).

It is widely known that spectral notches and peaks above 5 kHz contribute to the perception of the vertical angle of a sound image (Hebrank and Wright 1974; Butler and Belendiuk 1977; Mehrgardt and Mellert 1977; Musicant and Butler 1984). The frequency of notches shifts higher as the sound source moves from the front of the subject to above the subject (Butler and Belendiuk 1977; Shaw and Teranishi 1968). The notches are generated in the pinnae (Musicant and Butler 1984; Gardner and Gardner 1973; Lopez–Poveda and Meddis 1996; Iida et al. 1998; Takemoto et al. 2012), and the frequency of the notches depends on the shape of the pinnae as well as the vertical angle (Raykar et al. 2005). Moreover, the outline of the amplitude spectrum has been found to be more important than its fine structure (Asano et al. 1990; Middlebrooks 1992, 1999; Kulkarni and Colburn 1998; Langendijk and Bronkhorst 2002; Macpherson and Sabin 2013). The difference in notch frequency due to the vertical angle has been reported to be detectable by the listener (Moore et al.1989).

3.3.2 Details of Spectral Cues

A parametric HRTF model, recomposed of all or some of the spectral notches and peaks extracted from a listener's own HRTF, taking the peak around 4 kHz as the lower-frequency limit, has been proposed (Iida et al. 2007). These notches and peaks are labeled

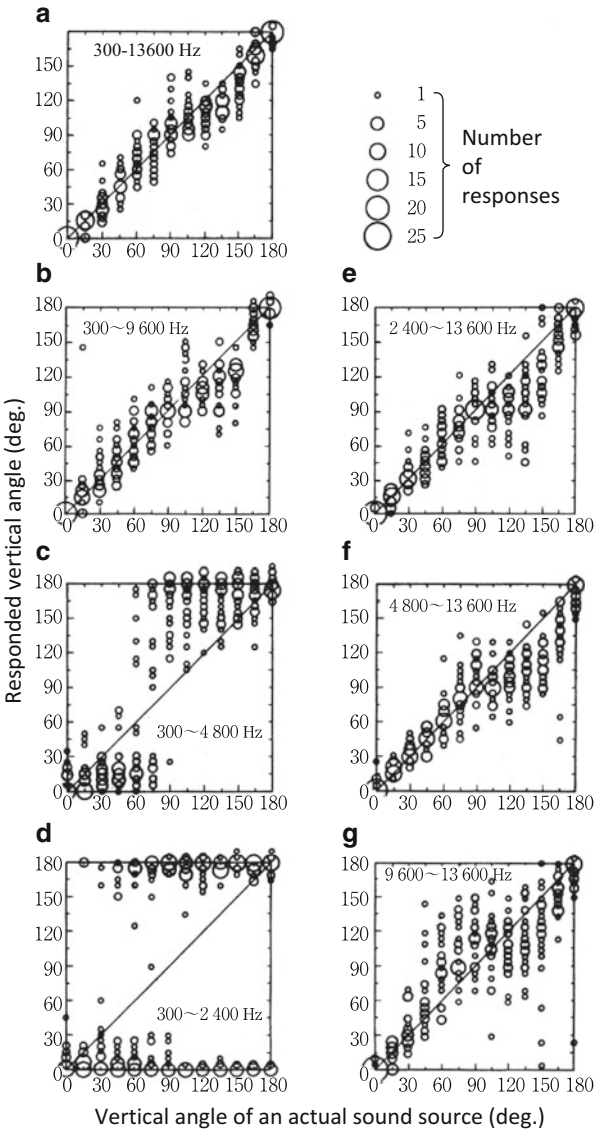

Fig. 3.7 Effects of frequency range of stimuli (white noise) on the median plane localization. (Morimoto and Saito 1977)

in order of frequency (e.g., N1, P1, N2, P2, N3, P3, and so on), and each is expressed parametrically in terms of frequency, level, and sharpness, as shown in Fig. 3.8.

It has been reported that there exist six prominent spectral peaks up to 20 kHz in the HRTFs in the median plane (Shaw 1997; Kahana and Nelson 2006). Figure 3.9 shows a schematic representation of the three lowest-frequency peaks (P1, P2, and P3) and the three lowest-frequency notches (N1, N2, and N3) (Takemoto et al. 2012). The frequencies of the peaks are approximately constant and are thus independent of the vertical angle, whereas the frequencies of the notches are highly dependent on the vertical angle. This vertical angle dependency of the notch

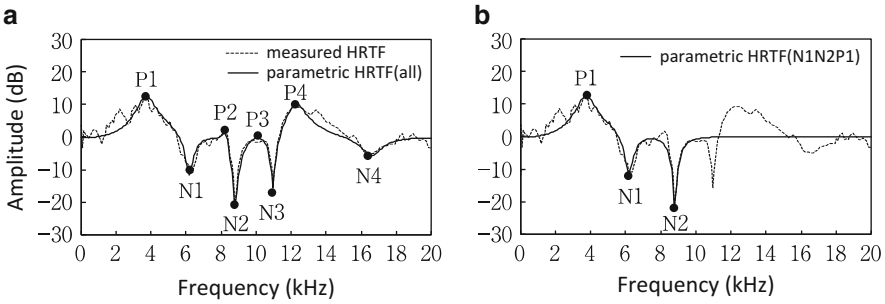

Fig. 3.8 Examples of parametric HRTFs. (**a**) Recomposed of all notches and peaks, (**b**) Recomposed of N1, N2, and P1

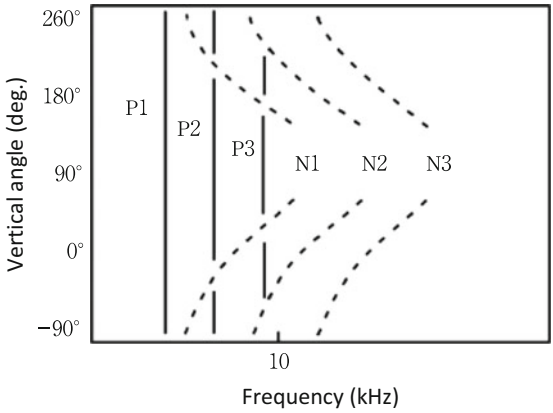

Fig. 3.9 Schematic diagram of notches and peaks in the median plane

frequency is thought to contribute as one of the important cues for vertical localization.

Localization tests and observations of the spectral peaks and notches of the HRTFs in the upper median plane infer the following.

1. **Minimum components of notches and peaks for median plane localization**

We first consider the minimum required number of notches and peaks for accurate median plane localization.

Figure 3.10(a) and (b) show the results of median plane localization tests with parametric HRTFs recomposed of all the notches and peaks extracted from the listener's own HRTFs. The results are seen to be similar to those using the measured HRTFs, as shown in Figs. 3.10(c) and (d) (Iida et al. 2007).

As shown in Fig. 3.11, localization tests were also carried out with parametric HRTFs recomposed of only some of the spectral notches and peaks extracted from the listener's own HRTFs. The results demonstrated that the two lowest-frequency notches (N1 and N2) and the lowest-frequency peak (P1) provided approximately

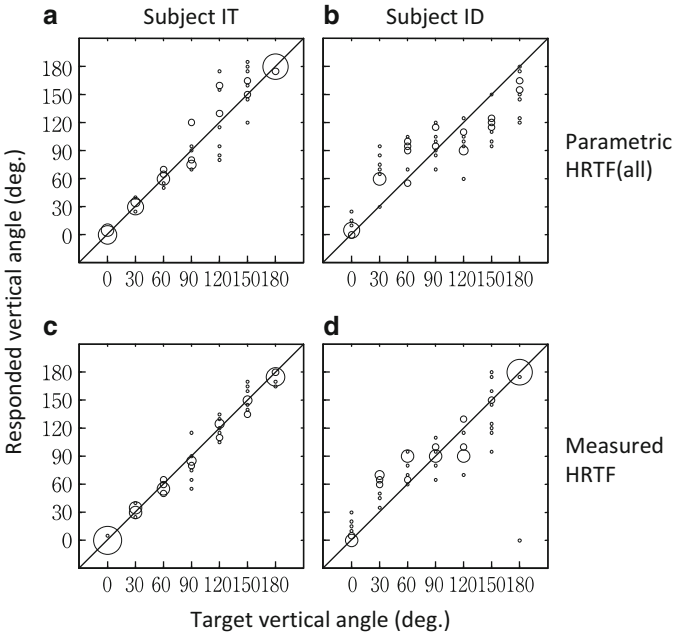

Fig. 3.10 Responses for parametric HRTFs using all notches and peaks and measured HRTFs. (**a**) and (**b**) parametric HRTFs recomposed of all the notches and peaks, (**c**) and (**d**) measured own HRTFs. (Iida et al. 2007)

the same localization performance as the listener's own HRTFs for the front and rear directions (Fig. 3.11(e) to (h)) (Iida et al. 2007).

For the upper directions, however, the localization performance of the parametric HRTF recomposed of N1, N2, and P1 for some of the subjects decreased compared with the subject's own HRTFs (Fig. 3.11(e) to (g)).

Median plane localization tests were then carried out using parametric HRTFs constructed using N1, N2, P1, and P2 (Iida and Ishii 2018). Figure 3.11(i) to (l) show that the localization performance for all subjects was improved at certain target vertical angles by adding P2 to N1N2P1. For subject MKI, the performance at 120° and 150° was improved. For subject OIS, the performance at 90° was improved. For subject OTK, the scatter in the responses observed at 60°, 90°, and 120° for N1N2P1 was not found for N1N2P1P2. The distribution of responses for N1N2P1P2 was approximately the same as that for the measured HRTFs. For subject YSD, the localization performance at 0° was improved. The difference in mean vertical localization errors between N1N2P1P2 and the measured HRTFs became less than 10° for all seven vertical target angles.

These results imply that N1N2P1P2 provides approximately the same vertical localization performance as the measured HRTFs at any of the seven target vertical angles in the upper median plane. In other words, the minimum set of notches and peaks for accurate median plane localization is N1, N2, P1, and P2.

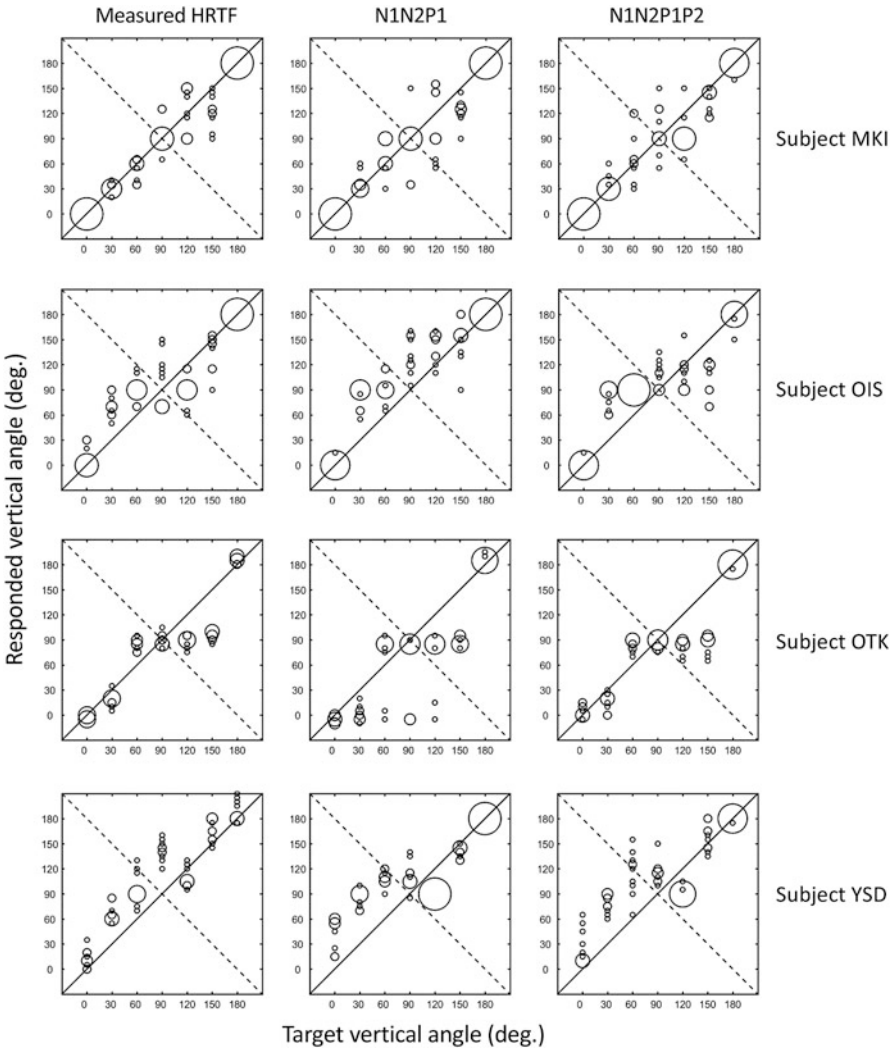

Fig. 3.11 Responses for measured HRTF, parametric HRTF(N1N2P1), and parametric HRTF (N1N2P1P2). (Iida and Ishii 2018)

As a side note, a sound image was not perceived in the upper direction when only P1, P2, or P1P2 were reproduced (Fig. 3.12). These results imply that P1 and P2 were not sufficient in themselves for localization of the upper direction.

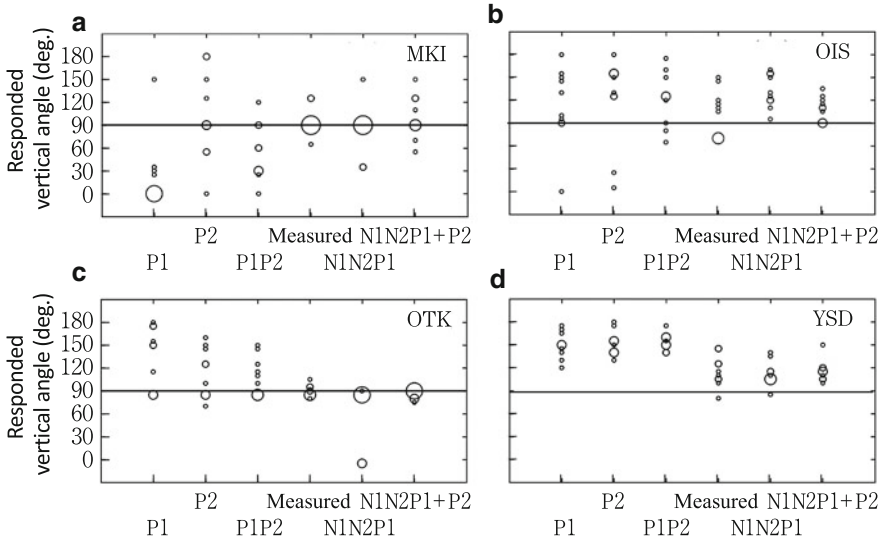

Fig. 3.12 Responses for P1, P2, and P1P2. For comparison, the responses for the measured HRTF, N1N2P1, and N1N2P1P2 in Fig. 3.11 are also shown. (**a**) subject MKI, (**b**) subject OIS, (**c**) subject OTK, and (**d**) subject YSD. (Iida and Ishii 2018)

2. **Vertical angle dependence of notch frequency**

This section examines the relationship between the vertical angle of a sound source and the frequencies of N1 and N2. The frequencies of N1 and N2 strongly depend on the vertical angle of the sound source (Fig. 3.13).

The N1 frequency increases with increasing vertical angle of the sound source from 0° to 120° and then decreases toward 180°. The N2 frequency increases with increasing vertical angle from 0° to 120°, whereas the range of the change in frequency between 120° and 180° is small.

This behavior explains the reason why two notches are necessary for median plane localization. If the notch frequency changed monotonically with the vertical angle of a sound source, then the vertical angle could be determined by extracting only one notch from the ear-input signal. However, since the relationship between the notch frequency and the vertical angle of a sound source is not one-to-one, at least two notches are necessary to determine the vertical angle.

The results of experiments using wide-band white noise with a notch of 8 kHz demonstrated that the subjects were able to detect the notch in the HRTFs and discriminate the difference in frequency of the notch (Moore et al. 1989). These results support the hypothesis that N1 and N2 are the cues for median plane localization.

3. **Vertical angle dependence of peak frequency**

On the other hand, the frequencies of P1 and P2 are almost constant, independent of the vertical angle. Therefore, the physical cue, which depends on the sound source direction is not included in P1 or P2.

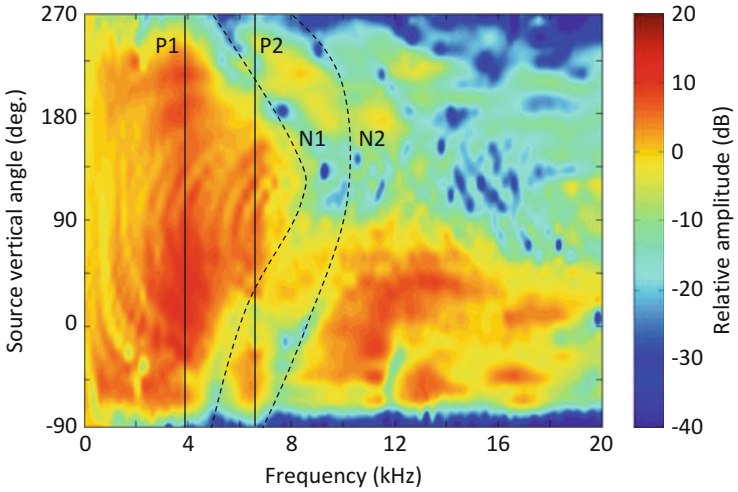

Fig. 3.13 Relationship between vertical angle of sound source and frequencies of N1, N2, P1, and P2

Here, the question is "Why are P1 and P2, the frequencies of which are independent of the vertical angle of a sound source, effective for median plane localization?". The following are possible roles that P1 and P2 play.

One possible interpretation is that the human hearing system uses P1 and P2 as reference information to search for N1 and N2. A listener hears not the HRTFs, but rather the ear-input signals. The ear-input signals are a convolution of the source signal, the spatial (room) impulse response, and the HRIR (see Appendix A.2). Furthermore, the sound pressure level of the ear-input signals varies hour to hour, and the ear-input signals often include background noise.

For extracting N1 and N2 from such ear-input signals, P1 and P2, the frequencies of which are independent of the vertical angle of the sound source, are considered to provide useful reference information to detect N1 and N2 for the human hearing system.

Another possible interpretation is that P1 and P2 emphasize N1 and N2. Figure 3.14 shows the relationship among the frequencies of N1, N2, P1, and P2 and the levels for the vertical angles of 0° (front) and 90° (above). The white and black circles indicate the results for 0° and 90°, respectively.

The two dashed lines indicate the maximum and minimum values of the notch detection threshold for three subjects for the notch having a center frequency of 8 kHz and a band width of 25% of the center frequency (Moore et al. 1989). None of the subjects could detect the notch when its level was higher than −9 dB, whereas they could all detect the notch when its level was less than −20 dB.

For 0°, both N1 and N2 were detectable. For 90°, however, N1 was undetectable, and N2 was detectable for some listeners and undetectable for other listeners. The notch and the peak are located in the order of P1, P2, N1, and N2 from the lower frequency for 90°. Therefore, for the case in which P2 is not reproduced, the contrast

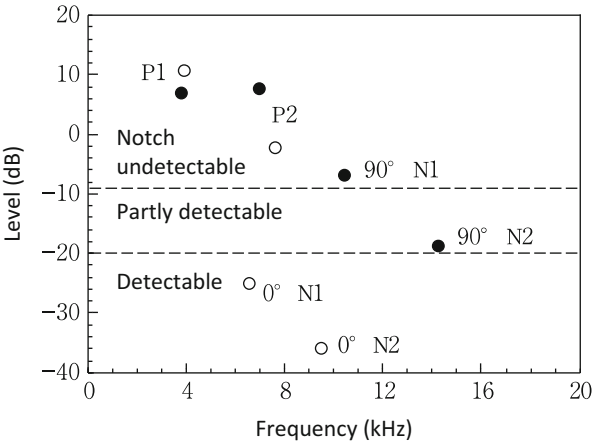

Fig. 3.14 Relationship among frequencies and levels for N1, N2, P1, and P2. Open circles and filled circles indicate the results for 0° and 90°. Broken lines indicate the detection threshold for the notch having a center frequency of 8 kHz and a bandwidth of 25% of the center frequency. (Iida and Ishii 2018)

effects, which emphasize N1, cannot be expected because P1 is at a frequency far from N1. However, for the case in which P2 is reproduced, the relative level of N1 measured from P2 reaches −14.7 dB. At this level, some listeners could detect the notch.

These considerations suggest that P2 could help to improve the accuracy of localization for the upper direction in the median plane by enhancing N1.

3.4 Role of Spectral Information at both Ears in Median Plane Localization

Previous studies clarified that the distortion of the spectral information at one of the ears by occluding the pinna cavities decreases the accuracy of vertical localization (Gardner and Gardner 1973; Morimoto 2001). However, the results of these studies cannot determine whether vertical localization can be accomplished using the spectral information at only a single ear or requires the spectral information at both ears.

The following two hypotheses may explain how the spectral information of the two-ear input signals is processed for vertical angle perception, in other words, the process to extract cues for vertical angle perception from the two ear-input signal spectra.

Hypothesis 1: Spectral cues are extracted from the integrated spectra of the input signals to both ears

The spectra of the two ear-input signals are integrated into one spectrum, and then spectral cues are extracted from the signals and the vertical angle is perceived.

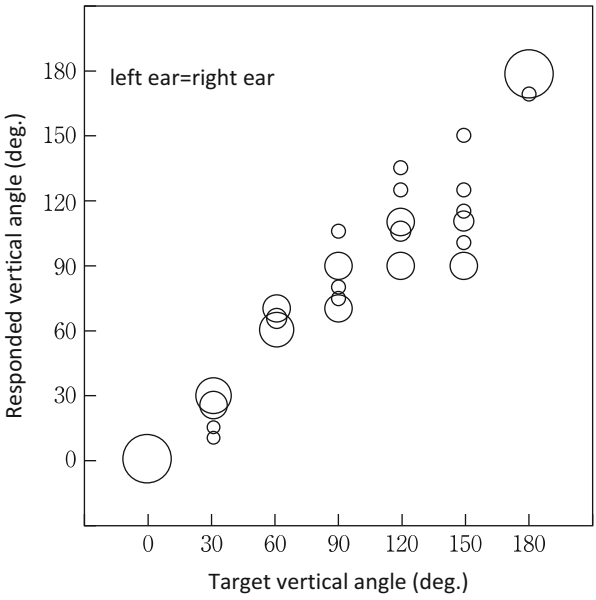

Fig. 3.15 Response when target vertical angles to both ears are identical. (Iida et al. 2018)

Hypothesis 2: Spectral cues are independently extracted from the input signals to each ear

Spectral cues are extracted from the spectrum of input signals to each ear (monaural spectrum) and vertical angle is perceived using these multiple cues.

In order to clarify which hypothesis is valid, median plane localization tests, in which the HRTFs for different target vertical angles were presented to each ear of the listeners, were carried out (Iida et al. 2018).

Figure 3.15 shows the responses for the vertical angle when the HRTFs for identical vertical angles were presented to both ears as usual median plane localization tests. The responses are distributed along a diagonal line, indicating that the subjects perceived the vertical angle of a sound image accurately.

Next, Fig. 3.16 shows the results for the case in which different target vertical angle of HRTFs were provided to the right and left ears. Figure 3.16(a) shows ten responses for the same listener for the case where HRTFs of 0° and 180° were provided to the left and right ears, respectively.

The results showed that the listener either localized a single sound image to the target vertical angle presented to either the left or the right ear, or localized two sound images to both target vertical angles.

Figure 3.16(b) shows the results for the case where HRTFs of 30° and 60° were provided to the left and right ears, respectively. The sound image was perceived at the target vertical angles provided to the left or right ear, and in the intermediate direction only once.

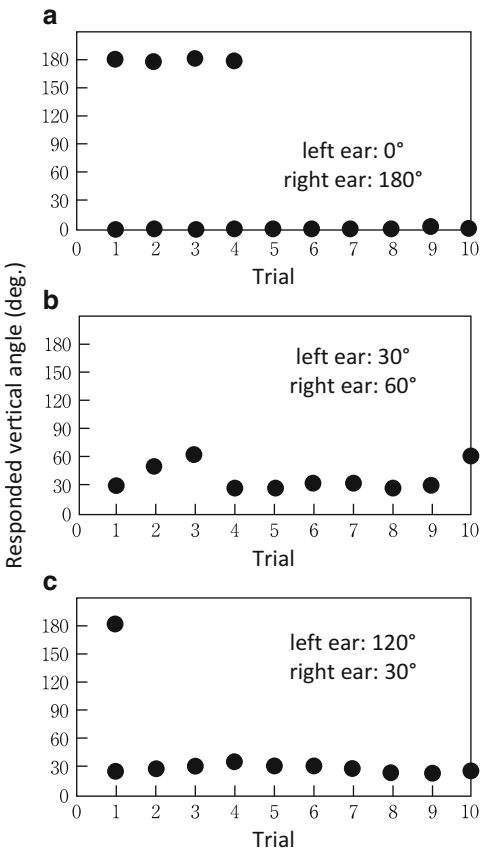

Fig. 3.16 Response when target vertical angles to left and right ears are different

Table 3.1 Mean localization error in the vertical angles for stimuli in which identical target vertical angles were presented to both ears ($\beta_l = \beta_r$) and for stimuli in which different target vertical angles were presented to each ear ($\beta_l \neq \beta_r$). (Iida et al. 2018)

Left ear = Right	Left ear ≠ Right
11.6°	9.3°

Figure 3.16(c) shows the results for the case where HRTFs of 120° and 30° were provided to the left and right ears, respectively. The listener localized most of the sound images to the vertical angle presented to the right ear. However, the listener localized two sound images to both target vertical angles only once.

Table 3.1 shows the mean localization error of the vertical angles for stimuli in which identical target vertical angles were presented to both ears ($\beta_l = \beta_r$) and for stimuli in which different target vertical angles were presented to each ear ($\beta_l \neq \beta_r$). The smaller error for the two target vertical angles was adopted for the stimuli $\beta_l \neq \beta_r$. Both errors were obtained when the subjects localized two sound images. The mean localization error for $\beta_l \neq \beta_r$ was less than that for $\beta_l = \beta_r$. This is probably due to adopting the smaller error in the case of stimuli $\beta_l \neq \beta_r$.

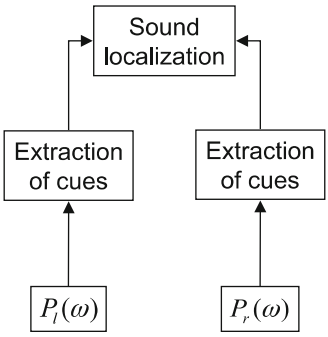

Fig. 3.17 Possible extraction process for spectral cues in human auditory system. The spectral cues are independently extracted from the spectrum of the input signal to each ear, $P_l(\omega)$ and $P_r(\omega)$. (Iida et al. 2018)

A t-test was performed in order to determine whether the difference in the localization error between $\beta_l = \beta_r$ and $\beta_l \neq \beta_r$ was statistically significant. There was no statistically significant difference between the localization errors for $\beta_l \neq \beta_r$ and $\beta_l = \beta_r$.

These results showed that the subjects localized a single sound image to the target vertical angle presented to either the left or right ear or localized two sound images to both target vertical angles, when different target vertical angles were presented to the left and right ears. This implies that a listener perceives the vertical angle of a sound image with the spectral information at only a single ear.

Based on the results, a possible extraction process for spectral cues in the human auditory system is that spectral cues are extracted from the spectrum of the input signal to each ear independently (Fig. 3.17). However, it is not clear which of the ears is dominant in the extraction process. Moreover, what determines this dominance remains unknown.

3.5 Origin of Spectral Cues

We next consider how and where such spectral cues are generated.

3.5.1 Contribution of Pinnae

The HRTFs differ for different sound source directions because of the asymmetry of the pinna, head, and torso in the front-back, up-down, and left-right directions. In particular, the pinnae have the greatest impact on the HRTFs. The pinna is a fold of skin that is supported by cartilage and is 5 to 7 cm in length and 3 to 3.5 cm in width. The pinnae have many cavities and elevated areas, as shown in Fig. 3.18.

Now, which part of the pinnae contributes to the generation of the notches and peaks, and what is the mechanism? A number of studies have examined these problems. Figure 3.19 shows HRTFs calculated by the FDTD method using the

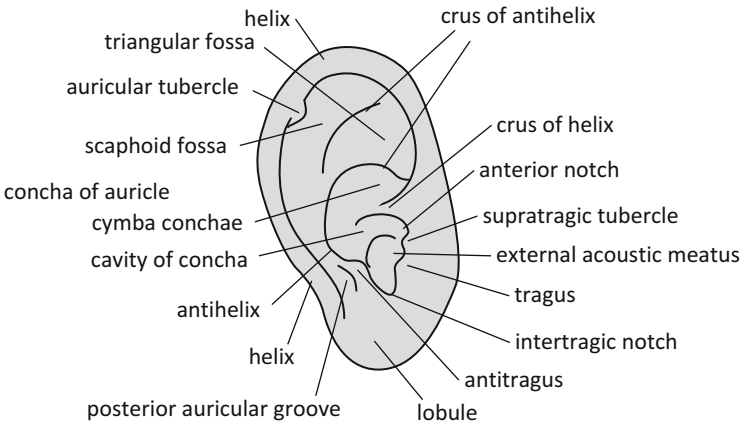

Fig. 3.18 Anthropometry of pinnae

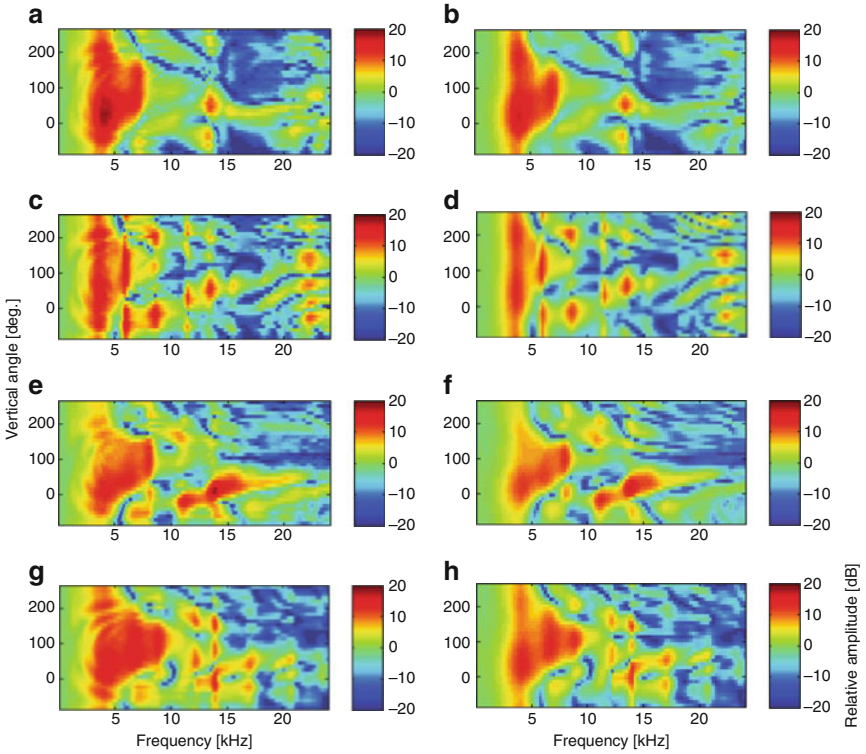

Fig. 3.19 Transfer functions calculated using shape of entire head of subjects M1, M2, F1, and F2 (a, c, e, g) and functions calculated using only shape of pinnae (b, d, f, h). (Takemoto et al. 2012)

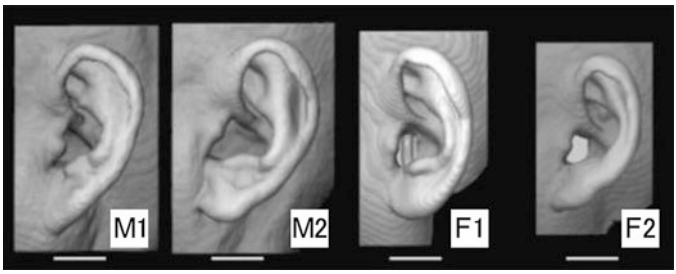

Fig. 3.20 Shape of the pinnae of the four subjects (M1, M2, F1, and F2). The white scale bars indicate 2 cm. (Takemoto et al. 2012)

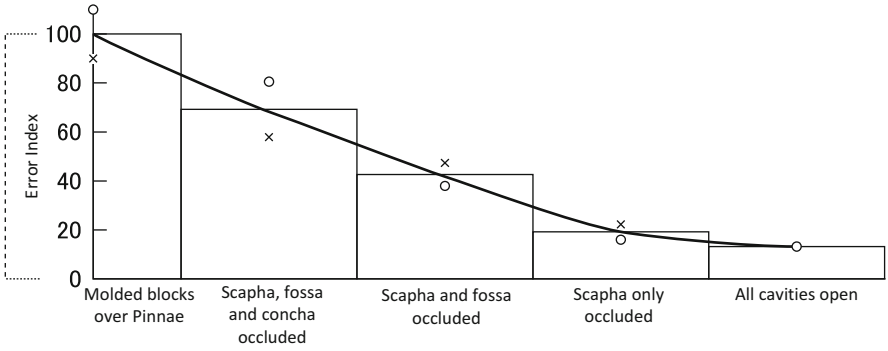

Fig. 3.21 Effects of pinna cavity occlusion on localization in anterior sector of median plane. (Gardner and Gardner 1973)

shape of the entire head of the subjects (a, c, e, g), and HRTFs calculated using only the shape of the pinnae (b, d, f, h) (Takemoto et al. 2012), as shown in Fig. 3.20.

The effects of head shape mainly appear at frequencies lower than 5 kHz. For the case in which the head is considered (a, c, e, g), a region of high sound pressure appears in a concentric pattern with its center at a vertical angle of 90°. This pattern appears to be generated by diffraction of the sound wave around the head. However, no effect of the head on the prominent peaks and notches, namely P1, P2, P3, N1, and N2, is observed. Notches and peaks of the HRTFs in the median plane are determined not by the head, but rather by the pinnae.

We next present experimental results, which verified the effects of the pinnae on the perception of the vertical angle of a sound image. Figure 3.21 shows the error index of the median plane localization for the case in which the three main cavities (triangular fossa, scaphoid fossa, and cavity of concha) of the pinnae were occluded with rubber one by one, while the entrances of the ear canals were open (Gardner and Gardner 1973). The error index increased as the cavities were occluded.

The HRTFs when these three cavities were occluded were also measured (Iida et al. 1998). The amplitude spectra are shown in Fig. 3.22. Figure 3.23 shows the

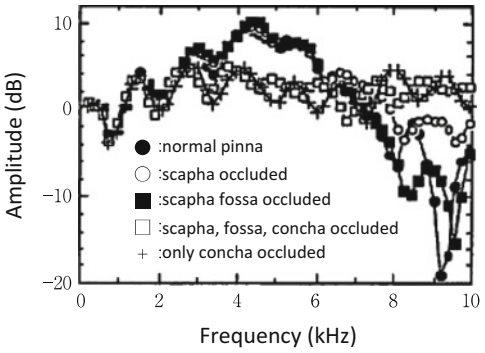

Fig. 3.22 HRTFs of occluded pinnae. (Iida et al. 1998)

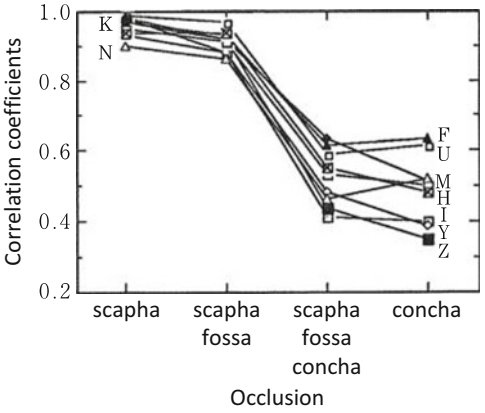

Fig. 3.23 Correlation coefficients between amplitude spectra of HRTFs of occluded pinnae and those of normal pinnae for nine subjects

correlation coefficients between the amplitude spectra of the HRTFs of occluded pinnae and those of the normal pinnae for each of the nine subjects.

Occlusion of the scapha had little effect on the amplitude spectrum, and high correlation coefficients over 0.90 were obtained for all subjects. The notches and peaks for the pinnae, the scapha and fossa of which were occluded, were approximately the same as those for the normal pinnae. The correlation coefficients were over 0.86. For the pinnae, the scapha, fossa, and concha of which were occluded, the notches and the peaks vanished, and the amplitude spectrum was flat. The correlation coefficients were low. In the case that only the concha was occluded, the notches and peaks vanished, and the correlation coefficients were low.

These results suggest that the concha are important for generating the prominent notches and peaks in HRTFs.

Furthermore, sound image localization tests were performed in the upper median plane for seven target vertical angles (0° to 180°, in 30° steps). Table 3.2 shows the number of directions for which a significant difference ($p < 0.01$) was observed in the localization accuracy between the occluded pinnae and the normal pinnae.

Table 3.2 Number of directions for which a significant difference in the localization accuracy was observed compared with the normal pinnae (total of seven directions). (Iida et al. 1998)

	Subject								
Conditions of pinna occlusion	U	M	Z	Y	F	I	K	H	N
Scapha	0	0	1	0	1	0	1	0	0
Scapha+Fossa	0	1	2	1	1	4	3	0	0
Scapha+Fossa+concha	4	4	5	4	4	5	5	0	0
Concha	5	4	3	3	5	4	6	2	1

For the pinnae, the scapha of which was occluded, no significant difference was observed between the normal pinnae and the occluded pinnae in six out of nine subjects for all directions. A significant difference was observed only in one direction for the remainder of the subjects. No tendency for a significant difference for a specific direction was observed.

For the pinnae, the scapha and fossa of which were occluded, a significant difference was observed for six subjects in one to four directions. The subjects perceived a sound image in specific directions (one subject perceived a sound image at around 75° and the other subject at 0°) regardless of the sound source direction).

For the pinnae, the scapha, fossa, and concha of which were occluded, a significant difference was observed for seven subjects in more than four directions. They perceived a sound image in specific directions (one subject perceived a sound image at around 75° and the other subjects perceived a sound image at 0°) regardless of the sound source direction. However, for subjects N and H, no significant difference was observed for any of the sound source directions. The accuracy of localization for the occluded pinnae was approximately the same as that for the normal pinnae. It is not clear which cues subjects N and H used to perceive the vertical angle of a sound image.

For the pinnae, only the concha of which was occluded, a significant difference was observed for seven subjects in more than three directions, as well as the pinnae, the scapha, fossa, and concha of which were occluded. For subjects N and H, a significant difference was observed in one or two directions.

The above results suggest that the cavity of the concha contributes to the generation of the notches and the peaks of HRTFs and significantly affects the perception of the vertical angle of a sound image.

3.5.2 Origin of Peaks

Experiments using a physical pinna model (Shaw 1997) and a simulation using the BEM (Kahana and Nelson 2006) and FDTD methods (Takemoto et al. 2012) revealed that the origin of the peaks is a resonance mode in the pinnae. These studies analyzed the distribution of the antinode of the sound pressure, intensity, and phase difference at the frequencies of the peaks.

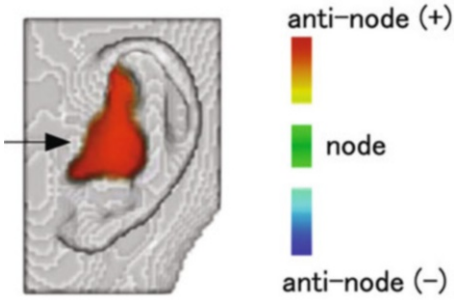

Fig. 3.24 Distribution of sound pressure in pinna at P1 frequency (3.5 kHz) for sound source direction of 0°. (Takemoto et al. 2012)

The results showed that an antinode is generated in the cavity of concha at the peak frequencies. An antinode is generated in the cavity of concha at the P1 frequency. At the P2 frequency, one antinode is generated in the cavity of the concha, and another antinode is generated in the upper cavities of the pinna. At the P3 frequency, one antinode is generated in the cavity of the concha, and two additional antinodes are generated in the upper cavities of the pinna.

Since the antinodes are located in the pinna cavities vertically, these modes are referred to as vertical modes. The origin of each peak is explained in detail below.

1. **First peak (P1)**

The P1 is the first mode generated in the depth direction of the concha. Therefore, the P1 frequency corresponds to the inverse of the wavelength, which is equal to one-fourth the depth of the concha cavity.

Figure 3.24 shows the distribution of sound pressure in the pinna at the P1 frequency (3.5 kHz) for the front direction (0°). An arrow indicates the direction of a sound source. The antinode (+) and antinode (−) indicate high absolute values of the sound pressure, the signs of which are positive and negative, respectively. The node indicates the low absolute value of the sound pressure. The figure shows that an antinode is generated over the entirety of the pinna cavities.

2. **Second peak (P2)**

The origin of P2 is the first mode in the vertical direction in the pinna cavities. This mode is generated along the pinna surface. Two antinodes, the phases of which are reversed, are generated around the entrance of the ear canal and the upper part of the pinna cavities.

Figure 3.25 shows the distribution of sound pressure in the pinna at the P2 frequency (6 kHz) for the upper direction (90°). The antinode of the cavity of concha and that of the cymba conchae and triangular fossa are reversed phase.

Qualitatively, P2 can be described as natural resonance in a rectangular solid room as:

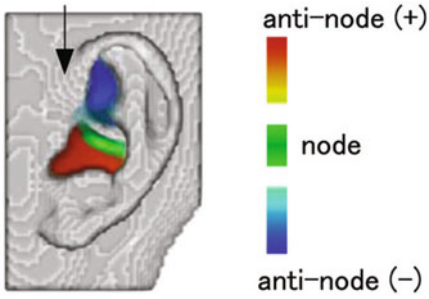

Fig. 3.25 Distribution of sound pressure in pinna at P2 frequency (6 kHz) for sound source direction of 90°. (Takemoto et al. 2012)

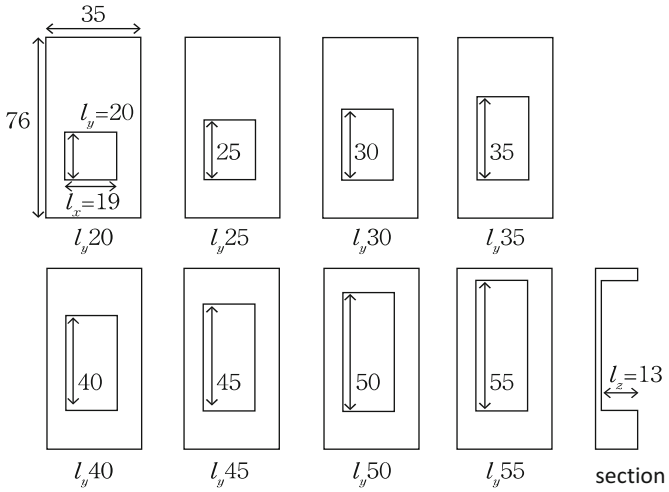

Fig. 3.26 Rectangular solid pinnae model

$$f_n = \frac{c}{2}\sqrt{\left(\frac{n_x}{l_x}\right)^2 + \left(\frac{n_y}{l_y}\right)^2 + \left(\frac{n_z}{l_z}\right)^2}, n_x, n_y, n_z = 0,1,2,\cdots \tag{3.1}$$

where f_n and c denote the natural resonance frequency and the speed of sound, respectively, and l_x, l_y, and l_z indicate the length of each side of the room. Here, P2 corresponds to the first mode for the longest side. (If the longest side is l_x, then, $n_x = 1$, $n_y = n_z = 0$.)

Figure 3.27 shows the measured P2 frequency for the rectangular solid pinnae model shown in Fig. 3.26 and the primary natural resonance frequency calculated using Eq. (3.1). The calculated frequencies show approximately the same behavior

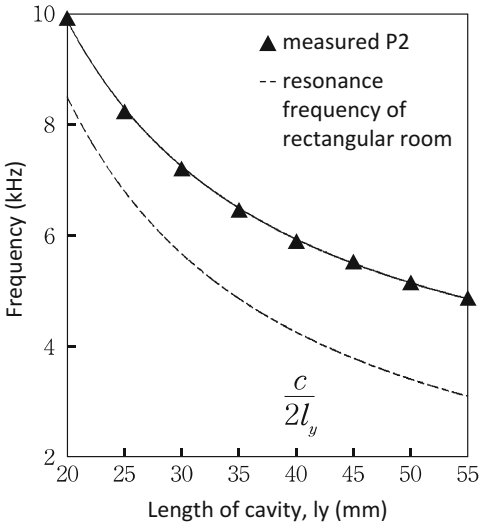

Fig. 3.27 Measured P2 frequency for rectangular solid pinnae model (▲) and calculated primary natural resonance frequency

as that of the measured frequencies. However, the calculated frequencies are lower than the measurement frequencies by approximately 1.5 kHz. The reason for this appears to be that one side of the rectangular solid was open.

3. **Third peak (P3)**

The origin of P3 is the second mode in the vertical direction in the pinna cavities. In this mode, one antinode is generated around the entrance of the ear canal, and two antinodes are generated in the upper cavities of the pinna. Among the two upper antinodes, the antinode close to the entrance of the ear canal has the opposite phase to that of the antinode around the entrance of the ear canal, and the other antinode is in-phase.

Opinion is divided regarding the positions of the two antinodes. It has been reported that the two antinodes are generated in the cymba conchae and the triangular fossa (Shaw 1997), in the cymba conchae and the scaphoid fossa (Kahana and Nelson2006), or in the back side of the cavity of the concha and in the triangular fossa (Takemoto et al. 2012).

The phases and positions of the three antinodes are approximately the same as those of the second natural resonance of a rectangular solid room (Fig. 3.28).

Figure 3.29 shows the distribution of the sound pressure at the P3 frequency (8.25 kHz) for a sound source direction of 120°. The antinode in the concha cavity and the triangular fossa is in-phase, and that at the back side of the concha cavity is reversed phase.

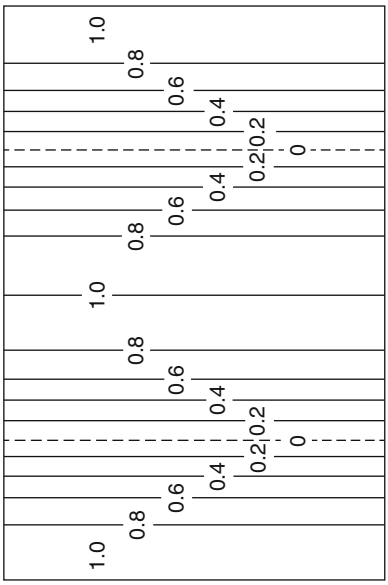

Fig. 3.28 Second natural resonance of rectangular solid room

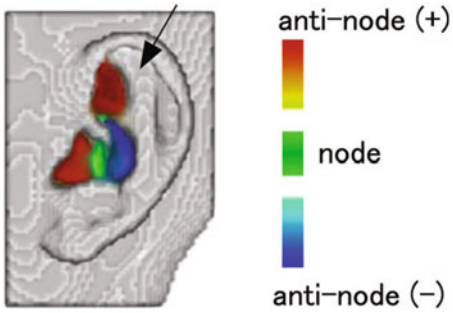

Fig. 3.29 Distribution of sound pressure in pinna at P3 frequency (8.25 kHz) for sound source direction of 120°. (Takemoto et al. 2012)

3.5.3 Origin of Notches

The generation mechanism for the notches is more complicated than that for the peaks, and two hypotheses have been proposed.

One is that the node is generated at the entrance of the ear canal by the interference of the direct wave and the reflected wave from the concha wall (Fig. 3.30) (Raykar et al. 2005). In this hypothesis, the notch frequencies are expressed by the following equation:

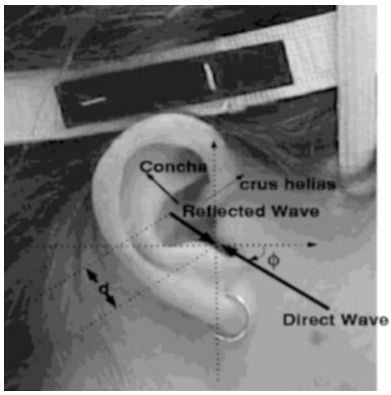

Fig. 3.30 Model of notch generation based on interference of direct wave and reflection from concha wall. (Raykar et al. 2005)

$$f_n(\phi) = \frac{(2n+1)}{2t_d(\phi)}, \quad n = 0,1,2,\cdots \tag{3.2}$$

where f_n is the notch frequency (Hz), t_d is the time difference (s) due to the path length difference between the direct wave and reflected wave, ϕ is the incidence angle of the direct wave, and n is the notch number ($n = 0$ corresponds to the first notch).

However, for a sound source in the upper directions, there exists no reflection point (concha wall). Even for a sound source in the front direction, the notch frequencies calculated by Eq. (3.2) have been reported to not coincide with the measured notch frequencies (Iida et al. 2011).

The other hypothesis is that multiple antinodes with different phases are generated in the pinna, and the node is formed around the entrance of the ear canal (Takemoto et al. 2012). The sound pressure at the entrance of the ear canal becomes a minimum at the N1 frequency.

Figure 3.31 shows the distribution of the sound pressure at the N1 frequencies for six sound source directions. Figure 3.31(a) and (b) show the distribution of the sound pressure for vertical angles of $-30°$ and $0°$, respectively. For these angles, N1 appears at 5.25 kHz and 5.5 kHz, respectively. In either case, antinodes are generated in the triangular fossa and the cymba conchae, and a node is generated in the concha cavity. The location of the antinodes and the nodes varies slightly depending on the vertical angle of the sound source.

Figure 3.31(c) and (d) show the distribution of the sound pressure for vertical angles of $30°$ and $50°$, respectively. For these angles, N1 appears at 6.25 kHz and 7.0 kHz, respectively. In this case, two antinodes, the phases of which are reversed, are generated in the triangular fossa, the cymba conchae, and a part of the concha cavity, and a node is generated in the concha cavity.

Figure 3.31(e) and (f) show the distribution of the sound pressure for vertical angles of $150°$ and $180°$, respectively. For these angles, N1 appears at 8.25 kHz and 6.75 kHz, respectively. In this case, two reverse-phase antinodes are generated in the

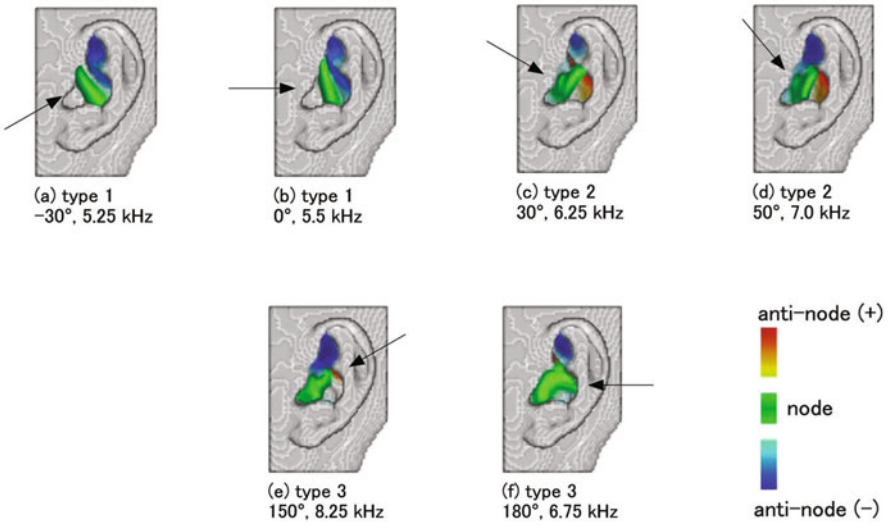

Fig. 3.31 Antinodes and nodes of sound pressure in pinnae. The arrows indicate the direction of the sound source. (Takemoto et al. 2012)

cymba conchae and the triangular fossa, and a node is generated in the region between the two antinodes through the concha cavity.

Numerical calculations on the distribution of the sound pressure around the pinnae, in which a sound source was placed at the entrance of the ear canal, have revealed that nodal lines are generated at specific directions at the N1 frequency. Figure 3.32 shows the sound pressure distributions for the frequencies of (a) 5.75 kHz, (b) 5.95 kHz, and (c) 6.35 kHz. The two straight lines in each figure indicate the vertical angles at which N1 appears. The vertical angles are (a) 19° and 221°, (b) 26° and 216°, and (c) 37° and 186°. These lines coincide with the nodal lines. The vertical angle of the nodal line increases with increasing frequency.

The above observation suggests that two resonances at the same frequency but with opposite phase are generated in the concha cavity and in the area between the cymba conchae and triangular fossa. These resonances then generate the nodal lines. A sound source located on the nodal line gives rise to N1 at the resonance frequency.

3.6 HRTF Learning by Subjects

In order to perceive the direction of a sound image by spectral cues detected from ear input signals, the relationship between the sound source direction and spectral cues must have been acquired by learning.

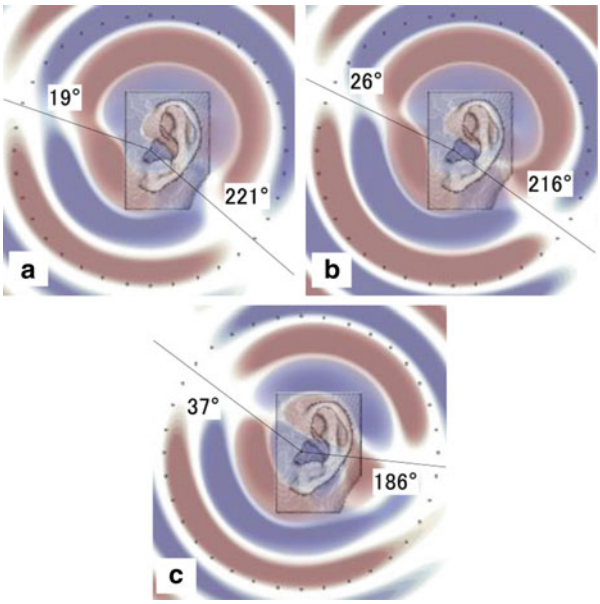

Fig. 3.32 Distribution of instantaneous sound pressure around pinnae. The red and blue regions indicate the local maxima and local minima of the sound pressure, respectively

A study on relearning of the spectral cues by adult subjects has been conducted (Hofman et al. 1998). In the study, the cavities of the pinnae of the adult subjects were occluded with polyester and wax to invalidate the spectral cues the subjects have already acquired. The deterioration of the localization accuracy of vertical directions was confirmed. Then, the subjects were requested to continue daily life under the occluded pinna condition. Three to six weeks were required until the subject was able to achieve accurate sound image localization by relearning spectral cues.

After relearning was accomplished, the fillings were removed, and the pinnae returned to the original state. Interestingly, the subjects localized a sound image accurately both with and without the filling.

This result suggests that the look-up table in the brain, which shows the relationship between the sound source direction and the spectral cues, is not overwritten by relearning, but is newly created.

It is presumed that learning is continuously performed in the developmental stage of the pinnae. The width and length of human pinnae reach their adult size at the age of 3 to 4 years and 9 to 10 years, respectively. Therefore, the learning is considered to be finished in childhood.

3.7 Knowledge of Sound Source

What a listener hears is not HRTFs, but rather ear-input signals. Ear-input signals are expressed, in the frequency domain, by complex multiplication of the spectra of the sound sources, the space transfer functions, and the HRTFs.

As such, the spectra of ear-input signals are not determined by the HRTFs themselves. The question then arises as to whether humans learn not only the spectra of HRTFs, but also the spectra of sound sources. For example, will the sound image localization accuracy differ between sound sources that the subject has heard and sound sources that the subject has never heard? The results of experiments to determine this are shown in Figs. 3.33 and 3.34.

Figure 3.33(a) and (b) show the results of median plane sound image localization tests for the voice of someone with whom the subjects are familiar and for the voice of someone with whom the subjects are not familiar, respectively.

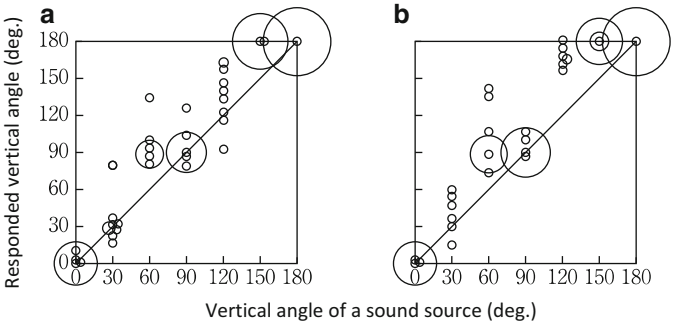

Fig. 3.33 Vertical angle response for (**a**) familiar and (**b**) unfamiliar voices

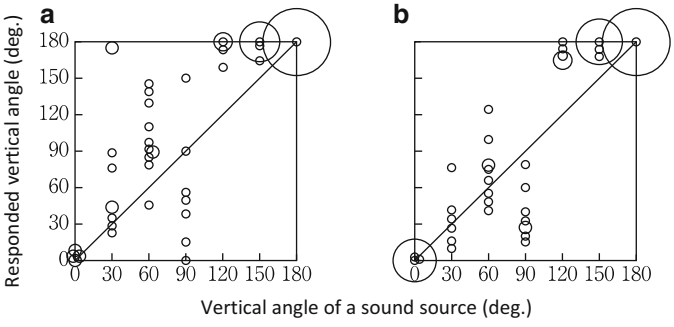

Fig. 3.34 Vertical angle response for (**a**) violin (four seconds) and (**b**) wide-band stationary noise corresponding to average spectrum of violin

Figures 3.34(a) and (b) show the results of median plane sound image localization tests for a violin solo (four seconds) and wide-band stationary noise, the spectrum of which is the same as the average spectrum of the violin solo, respectively.

These results show that the localization accuracy for the unfamiliar sound source is approximately the same as that for the familiar sound source. These results suggest that humans do not learn spectral differences in sound sources.

3.8 Physiological Mechanism of Notch Detection

Does a physiological mechanism to detect notches exist in the human auditory system? Experiments with cats have revealed that the dorsal cochlear nucleus (DCN) distinguishes notches in HRTFs and that Type IV neurons in the DCN extract not the center frequency of the notches, but rather the edges of the high-frequency side of the notches (Reiss and Young 2005).

Furthermore, it has been shown that a function to extract the edges of the high-frequency side is important for vertical angle perception for various kinds of sound sources with different spectra (Baumgartner et al. 2014).

The theory that the edge on the high-frequency side of the notch is important qualitatively supports the statement that an "HRTF of a large pinna tends to be applicable to a listener with small pinnae" in Sect. 2.2.2. Since the notch frequency of the HRTF of a large pinna is lower than that of the small pinna, the edge of the high-frequency side of the notch of the large pinna is included in the notch of the small pinna. On the other hand, the frequency of the edge of the high-frequency side of the notch of the small pinna is higher than that of the large pinna.

3.9 Head Movement

The above sections have discussed vertical perception while keeping the head immobile. Head movement is considered to be another cue for front-back perception. The characteristics of the ear-input signals change according to the head movement of the listener. Suppose the front-back direction of a sound image is uncertain. If the listener turns his/her head to the right and the sound image moves to the left, then the listener judges the sound source to be in the front.

The effects of head movements have already been examined in various sound image localization tests (Perrett and Noble 1997; Kato et al. 2003; Iwaya et al. 2003). Figure 3.35 shows the front-back error rate obtained by sound localization tests, using band-limited noise presented from one of 12 loudspeakers placed in the horizontal plane at an interval of 30° in an anechoic room.

In the figure, the white bars and the black bars indicate the front-back error rates for the restricted head movement and the prompted head movement, respectively. The low-pass noise (LPN) has a cut-off frequency of 1 kHz, and the high-pass noise

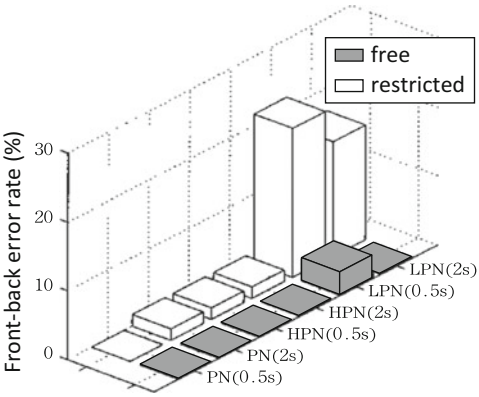

Fig. 3.35 Front-back error rates for restricted head movement (white bars) and prompted head movement (gray bars). (Iwaya et al. 2003)

(HPN) has a cut-off frequency of 3 kHz. Finally, pink noise is denoted as PN. The parentheses indicate the duration of the presentation of the stimuli.

For LPN, in which the spectral cues were lost, the front-back error rate was around 20% for the restricted head-movement condition. On the other hand, the rate was less than several percent when the head movement was prompted.

As such, the accuracy of front-back judgment is improved in the experiment in which the subjects were prompted to move their heads. However, humans do not use head movement as cues for front-back judgment in daily life.

Even barn owls, which identify the direction of prey based on the sound emitted by the prey, do not move their heads to obtain cues for localization. Barn owls do not search for a sound source by changing the direction of their head, but rather determine the position of the sound source before moving their head. They then memorize the position of the sound source and turn their heads in the direction of the prey. Similarly, humans are able to perceive the sound direction without moving their head, and moving the head spontaneously to confirm the sound direction is rather unusual (Nojima et al. 2013).

References

Algazi VR, Avendano C, Duda RO (2001) Elevation localization and head–related transfer function analysis at low frequencies. Acoust Soc Am 109:1110–1122

Asano F, Suzuki Y, Sone T (1990) Role of spectral cues in median plane localization. J Acoust Soc Am 88:159–168

Baumgartner R, Majdak P, Laback B (2014) Modeling sound–source localization in sagittal planes for human listeners. J Acoust Soc Am 136:791–802

Butler A, Belendiuk K (1977) Spectral cues utilizes in the localization of sound in the median sagittal plane. J Acoust Soc Am 61:1264–1269

Gardner MB, Gardner RS (1973) Problem of localization in the median plane: effect of pinnae cavity occlusion. J Acoust Soc Am 53:400–408

Hebrank J, Wright D (1974) Spectral cues used in the localization of sound sources on the median plane. J Acoust Soc Am 56:1829–1834
Hofman PM, Van Riswick JGA, Van Opstal AJ (1998) Relearning sound localization with new ears. Nat Neurosci 1:417–421
Iida K, Yairi M, Morimoto M (1998) Role of pinna cavities in median plane localization. In 16th international congress on acoustics, Seattle, 845–846
Iida K, Itoh M, Itagaki A, Morimoto M (2007) Median plane localization using parametric model of the head–related transfer function based on spectral cues. Appl Acoust 68:835–850
Iida K, Gamoh N, Ishii Y (2011) Contribution of the early part of the head-related impulse response to the formation of two spectral notches of vertical localization cues. Forum Acusticum 2011. Aalborg:2241–2245
Iida K, Ishii Y (2018) 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 129:239–247
Iida K, Itoh M, Morimoto M (2018) Upper median plane localization when head-related transfer functions of different target vertical angles are presented to the left and right ears. Acoust. Sci. & Tech. 39:275–286
Iwaya Y, Suzuki Y, Kimura D (2003) Effects of head movement on front-back error in sound localization. Acoust Sci Tech 24:322–324
Kahana Y, Nelson PA (2006) Numerical modelling of the spatial acoustic response of the human pinna. J Sound Vibration 292:148–178
Kato M, Uematsu H, Kashio M, Hirahara T (2003) The effect of head motion on the accuracy of sound localization. Acoust Sci Tech 24:315–317
Kulkarni A, Colburn HS (1998) Role of spectral detail in sound-source localization. Nature 396:747–749
Langendijk EHA, Bronkhorst AW (2002) Contribution of spectral cues to human sound localization. J Acoust Soc Am 112:1583–1596
Lopez–Poveda EA, Meddis R (1996) A physical model of sound diffraction and reflections in the human concha. J Acoust Soc Am 100:3248–3259
Macpherson EA, Sabin AT (2013) Vertical–plane sound localization with distorted spectral cues. Hear Res 306:76–92
Mehrgardt S, Mellert V (1977) Transformation characteristics of the external human ear. J Acoust Soc Am 61:1567–1576
Middlebrooks JC (1992) Narrow–band sound localization related to external ear acoustics. J Acoust Soc Am 92:2607–2624
Middlebrooks JC (1999) Virtual localization improved by scaling non individualized external–ear transfer functions in frequency. J Acoust Soc Am 106:1493–1510
Moore BCJ, Oldfield SR, Doole GJ (1989) Detection and discrimination of spectral peaks and notches at 1 and 8kHz. J Acoust Soc Am 85:820–836
Morimoto M, Saito A (1977) 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.; H-40-1 (in Japanese)
Morimoto M, Ando Y (1980) On the simulation of sound localization. J Acoust Soc Jpn (E) 1:167–174
Morimoto M (2001) The contribution of two ears to the perception of vertical angle in sagittal planes. J Acoust Soc Am 109:1596–1603
Musicant AD, Butler RA (1984) The influence of pinnae–based spectral cues on sound localization. J Acoust Soc Am 75:1195–1200
Nojima R, Morimoto M, Sato H, Sato H (2013) Do spontaneous head movements occur during sound localization? J Acoust Sci Tech 34:292–295
Perrett S, Noble W (1997) The effect of head rotations on vertical plane sound localization. J Acoust Soc Am 104:2325–2332

Raykar VC, Duraiswami R, Yegnanarayana B (2005) Extracting the frequencies of the pinna spectral notches in measured head related impulse responses. J Acoust Soc Am 118:364–374

Reiss LAJ, Young ED (2005) Spectral edge sensitivity in neural circuits of the dorsal cochlear nucleus. J Neuroscience 25:3680–3691

Shaw EAG, Teranishi R (1968) Sound pressure generated in an external–ear replica and real human ears by a nearby point source. J Acoust Soc Am 44:240–249

Shaw EAG (1997) Acoustical features of the human external ear, binaural and spatial hearing in real and virtual environments. Edited by Gilkey RH, Anderson TR. Lawrence Erlbaum Associates, Mahwah, NJ, pp 25–47

Takemoto H, Mokhtari P, Kato H, Nishimura R, Iida K (2012) Mechanism for generating peaks and notches of head–related transfer functions in the median plane. J Acoust Soc Am 132:3832–3841