19th INTERNATIONAL CONGRESS ON ACOUSTICS
MADRID, 2-7 SEPTEMBER 2007

A MODEL OF THE HEAD-RELATED TRANSFER FUNCTION BASED ON SPECTRAL CUES

PACS: 43.66.Qp, 43.66.Pn, 43.66Ba

Iida, Kazuhiro[1]; Itoh, Motokuni[2]; Itagaki, Atsue[3]; Morimoto, Masayuki[3]
[1]Chiba Institute of Technology; Tsudanuma 2-17-1, Narashino, Chiba 275-0016, Japan; kazuhiro.iida@it-chiba.ac.jp
[2]Matsushita Electric Ind. Co., Ltd, Saedo 600, Tsuzuki, Yokohama 224-8539, Japan; itoh.gempo@jp.panasonic.com
[3]Kobe University; Rokko, Nada, Kobe 657-8501, Japan; mrmt@kobe-u.ac.jp

ABSTRACT

In this study, a parametric HRTF model, which is recomposed only of extracted spectral peaks and notches is introduced. To examine this model, localization tests in the median plane were carried out. The results show that the parametric HRTF recomposed using the first and second notches (N1 and N2) and the first peak (P1) provides almost the same localization accuracy as the measured HRTFs. Observations of the spectral peaks and notches indicate that N1 and N2 change remarkably as the source elevation changes, whereas P1 does not depend on the source elevation. In conclusion, N1 and N2 can be regarded as spectral cues, and the hearing system could utilize P1 as the reference information to analyze N1 and N2.

1. INTRODUCTION

Most previous studies showed that spectral distortions caused by pinnae in the high-frequency range approximately above 5 kHz act as cues for median plane localization [1–14]. Shaw and Teranishi [2] reported that a spectral notch changes from 6 to 10 kHz when the elevation of a sound source changes from -45 to 45°. Hebrank and Wright [5] carried out experiments with filtered noise and reported that sound spectra from 4 to 16 kHz are necessary for median plane localization. Butler and Belendiuk [6] showed that the prominent notch in the frequency response curve moved toward the lower frequencies as the sound source moved from above to below the aural axis in the frontal half of the median plane. Mehrgardt and Mellert [7] have shown that the spectrum changes systematically in the frequency range above 5 kHz as the elevation of a sound source changes. Asano *et al*. [11] carried out median plane localization tests with the HRTFs smoothed by ARMA models through headphones. The results of this study indicate that major cues for judgment of elevation angle exist in the high-frequency region above 5 kHz. Iida et al. [13] carried out localization tests and measurements of head-related transfer functions (HRTFs) with the occlusion of the three cavities of pinnae, scapha, fossa, and concha. Then they concluded that spectral cues in median plane localization exist in the high-frequency components above 5 kHz of the transfer function of concha. Raykar et al. [14] noted that one of the prominent features observed in the head-related impulse response (HRIR) and one that has been shown to be important for elevation perception are the deep spectral notches attributed to the pinna. The results of these previous studies imply that spectral peaks and notches due to the transfer function of concha in the frequency range above 5 kHz prominently contribute to the perception of sound source elevation. Furthermore, there might be a potential of HRTF modeling based on the knowledge on spectral cues.

The present paper has two purposes. One is to examine a simulation method for localizing a sound image. Authors propose a parametric HRTF model to simulate vertical sound localization. The parametric HRTF is recomposed only of the spectral peaks and notches extracted from the measured HRTF, and the spectral peaks and notches are expressed parametrically with center frequency, level, and sharpness. The other purpose is to clarify the contribution of each spectral peak and notch as a spectral cue. Localization tests are carried out in the upper median plane using the subjects' own measured HRTFs and the parametric HRTFs with various combinations of spectral peaks and notches.

2. METHOD OF RECOMPOSING PARAMETRIC HRTFS

2.1. Measurements of HRTFs

The subjects' own HRTFs in the upper median plane were measured in an anechoic chamber in 30° steps. The distance from the sound sources to the center of the subject's head was 1.5 m. Ear-microphones [15] were used in the HRTF measurements. The ear-microphones were fabricated using the subject's ear molds. Miniature electret condenser microphones of 5 mm diameter (Panasonic WM64AT102) and silicon resin were put into the ear canals of the ear molds and consolidated. In the HRTF measurements, the ear-microphones were put into the ear canals of the subject. The diaphragms of the microphones were located at the entrances of the ear canals. Therefore, this is so called the ''meatus-blocked condition'' [2], in other words, the ''blocked entrances condition'' [16].

2.2. Extraction of spectral peaks and notches

The spectral peaks and notches are extracted from the measured HRTFs regarding the peak around 4 kHz, which are independent of sound source elevation [2], as a lower frequency limit. Then, labels are put on the peaks and notches in order of frequency (e.g., P1, P2, N1, N2 and so on). The peaks and notches are expressed parametrically with center frequency, level, and sharpness. The amplitude of the parametric HRTF is recomposed of all or some of these spectral peaks and notches. In order to extract the essential spectral peaks and notches, the microscopic fluctuations of the amplitude spectrum of HRTF were eliminated by Eq. (1):

$$HRTF_w(k) = \sum_{n=-n_1}^{n_1} HRTF(k+n)W(n), \qquad (1)$$

where *W*(n) is a Gaussian filter defined by Eq. (2). *k* and *n* denote discrete frequency. The sampling frequency was 48 kHz, and the duration of HRTFs was 512 samples. In this study, *n* and σ were set to be 4 and 1.3, respectively.

$$W(n) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-n^2}{2\sigma^2}} \qquad (2)$$

The spectral peak and notch are defined as the maximal and minimal levels of $HRTF_w$, respectively. Thus, the center frequencies and the levels of the spectral peaks and notches are obtained. The sharpness of the peak and notch is set to be their envelopment fit with that of $HRTF_w$. Figure 1 shows examples of the parametric HRTFs recomposed of N1 and N2 in the median plane. As shown in the figure, the parametric HRTF reproduces all or some of the spectral peaks and notches accurately and has flat spectrum characteristics in other frequency ranges.

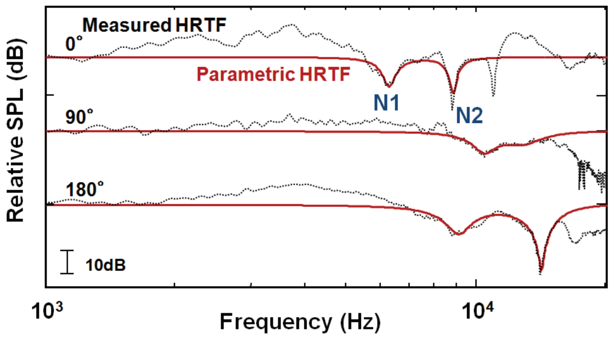

Figure 1. Examples of measured HRTFs (dotted line) and parametric HRTFs recomposed of N1 and N2 (solid line) in the upper median plane.

3. LOCALIZATION TEST I

Localization tests in the upper median plane were carried out using the subjects' own measured HRTFs and the parametric HRTFs.

3.1. Method

The localization tests were conducted in a quiet listening room. The background noise level was below 20 dB(A). A notebook computer (Panasonic CF-R3), an audio interface (RME Hammerfall DSP), open-air headphones (AKG K1000), and the ear-microphones mentioned above were used for the localization tests. The subjects sat at the center of the listening room. The ear-microphones were put into the ear canals of the subject. The diaphragms of the microphones were located at the entrances of the ear canals in the same way as in the HRTFs measurements. Then, the subjects wore the open-air headphones, and the stretched-pulse signals were emitted through them. The signals were received by the ear-microphones, and the transfer functions between the open-air headphones and the ear-microphones were obtained.

Then, the ear-microphones were removed, and stimuli were delivered through the open-air headphones. Stimuli $P_{l,r}(x)$ were created by Eq. (3):

$$P_{l,r}(\omega) = S(\omega) \times H_{l,r}(\omega) / C_{l,r}(\omega), \quad (3)$$

where $S(\omega)$ and $H_{l,r}(\omega)$ denote the source signal and HRTF, respectively. $C_{l,r}(\omega)$ is the transfer function between the open-air headphones and the ear-microphones.

The source signal was a wide-band white noise from 280 Hz to 17 kHz. The measured subjects' own HRTFs and the parametric HRTFs in the upper median plane in 30° steps were used. The parametric HRTFs of all the combinations of all or some of the spectral peaks and notches were prepared. However, P1 was not used in the tests except in the combination consisting of all the peaks and notches, since P1 did not have a directional dependence and seemed to have a small contribution to the perception of elevation. For comparison, stimuli without an HRTF convolution, that is, stimuli with $H_{l,r}(\omega) = 1$, were included in the tests. Subjects were two males, 30 and 44 years old, with a normal hearing sensitivity.

A stimulus was delivered at 60 dB SPL, triggered by hitting a key of the notebook computer. The duration of the stimulus was 1.2 s, including the rise and fall times of 0.1 s, respectively. The subject's task was to plot the perceived elevation on the computer display by clicking a mouse. Each stimulus set contained 100 and 68 different stimuli for subjects IT and ID, respectively. The difference in the number of stimuli was due to the differences in the numbers of spectral peaks and notches between subjects. Ten such sets were prepared for the test. The order of presentation of stimuli depended on the test.

3.2. Results

Figure 2 shows the examples of distributions of the responses of subject IT for the target elevation of 0 and 90°. The ordinate of each panel represents the perceived elevation, and the abscissa, the kind of stimulus. Hereafter, the measured HRTF and parametric HRTF are expressed as the mHRTF and pHRTF, respectively. For the stimuli without an HRTF, the perceived elevation was not accurate, and the variance of responses was large. On the other hand, the subjects perceived the elevation of a sound source accurately for the mHRTF. This means that the method and system of the localization tests are adequate.

For the pHRTF(all), which is the parametric HRTF recomposed of all the spectral peaks and notches, the perceived elevation was as accurate as that for the mHRTF. In other words, the elevation of a sound source can be perceived correctly when the amplitude spectrum of the HRTF is reproduced by the spectrum peaks and notches. For the pHRTF recomposed of only one spectral peak or notch, the variances of the responses were large. Namely, neither one peak nor one notch provides sufficient information for localizing the elevation of a sound source. The accuracy of localization improved as the numbers of peaks and notches increased. Careful observation of the results indicates that the pHRTF recomposed of N1 and N2 provides almost the same accuracy of elevation perception as the mHRTF at most of the target elevations.

Figure 3 shows the responses to the mHRTF, pHRTF(all), and pHRTF(N1–N2) for seven target elevations. The ordinate of each panel represents the perceived elevation, and the abscissa, the target elevation. For the pHRTF(all), the responses distribute along a diagonal line, and this distribution is practically the same as that for the mHRTF. For the pHRTF(N1–N2), the responses distribute along a diagonal line, in the case of subject IT. However, in the case of subject ID, the responses to the stimuli tend to localize in a rear direction for the target elevations of 30, 60, 90, and 120°.

Then, the localization error was calculated. The error is defined as the mean absolute deviation of the perceived elevation from the target elevation. Statistical tests were performed to determine whether a difference in localization error between the mHRTF and the pHRTF is statistically significant or not. Table 1 show the results of the *t*-test. There was no statistically significant difference between the pHRTF(all) and the mHRTF except for subject IT for the target elevation of 120°. Moreover, no statistically significant difference was observed between the pHRTF(N1–N2) and the mHRTF for all the target elevations, except for subject IT for the target elevation of 30° and subject ID for the target elevations of 60, 90, and 120°. This implies that N1 and N2 play an important role as spectral cues.

The pHRTF(all) of the target elevations of 60, 90, and 120° of subject ID consisted of only four peaks and notches, that is, P1, P2, N1, and N2. As mentioned above, for the target elevations of 60, 90, and 120°, statistically significant differences were observed between the pHRTF(N1–N2) and the mHRTF, but not between the pHRTF(all) and the mHRTF. Furthermore, statistically significant differences were observed between the pHRTF(N1–N2–P2) and the mHRTF for the target elevations of 60 and 90°. This implies that P1, which was not used in the tests, could contribute to the perception of elevation.

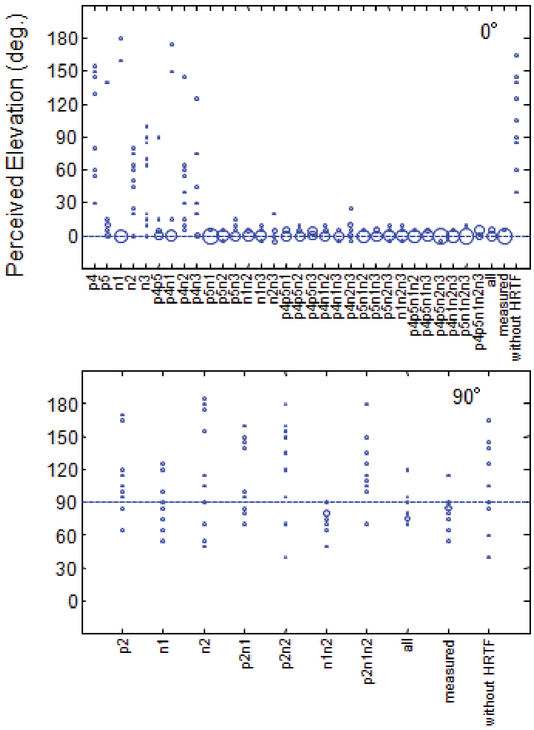

Figure 2. Examples of localization responses of subject IT for the target elevation of 0 and 90°.

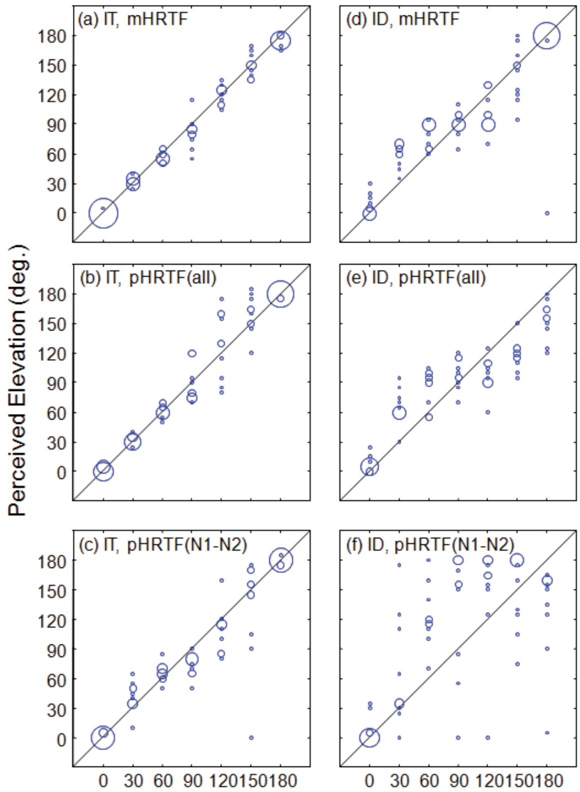

Figure 3. Localization responses to mHRTF, pHRTF(all), and pHRTF(N1-N2).

Table 1 Statistically significant difference in localization error between mHRTFs and pHRTFs for subject IT and ID. * $p < 0.05$, ** $p<0.01$

Subject	pHRTF	Target elevation (degree)						
		0	30	60	90	120	150	180
IT	All	-	-	-	-	**	-	-
	N1-N2	-	**	-	-	-	-	-
ID	All	-	-	-	-	-	-	-
	N1-N2	-	-	**	**	*	-	-
	N1-N2-P2	-	-	**	*	-	-	-

4. LOCALIZATION TEST II

The purpose of localization test II is to clarify the effect of P1 on the localization.

4.1. Method

The apparatus, source signal and target elevation were the same as those used in localization test I. However, test II was carried out in an anechoic chamber. The subject's own mHRTFs and pHRTFs in the upper median plane in 30° steps were used. The number of combinations of spectral peaks and notches of the pHRTF was eight; (1) N1, (2) N2, (3) P1, (4) N1–N2, (5) N1–P1, (6) N2–P1, (7) N1–N2–P1, and (8) all peaks and notches. The number of stimuli was 70. The subject was a female, 22 years of age, with a normal hearing sensitivity.

4.2. Results

For the mHRTF, the subject perceived the elevation of a sound source accurately at all the target elevations. Similar results to localization test I were obtained for the pHRTF(all) and pHRTF recomposed of only one spectral peak or notch. In other words, the perceived elevation was as accurate as that for the mHRTF at all the target elevations for the pHRTF(all), and the variances of the responses were large for the pHRTF recomposed of only one spectral peak or notch at all the target elevations.

Figure 4 shows the responses to the mHRTF, pHRTF(all), pHRTF(N1–N2), and pHRTF(N1–N2–P1) for seven target elevations. For the pHRTF(all), the responses distribute along a diagonal line, and this distribution is practically the same as that for the mHRTF. For the pHRTF(N1–N2), the responses distribute along a diagonal line for the target elevations of 120, 150, and 180°, but the responses for the target elevations of 0, 30, 60, and 90° shift to the rear. For the pHRTF(N1–N2–P1), the responses generally distribute along a diagonal line, except for the target elevation of 90°. The accuracy of elevation localization of the pHRTF(N1–N2–P1) is better than that of the pHRTF(N1–N2).

Table 2 show the localization error and the results of the *t*-test. There was no statistically significant difference between the pHRTF(all) and the mHRTF. Statistically significant differences were observed between the pHRTF(N1–N2) and the mHRTF except for the target elevations of 120 and 150°. Moreover, there was no significant difference between the pHRTF(N1–N2–P1) and the mHRTF, except for the target elevation of 90°. The pHRTF(all) for 90° was recomposed of N1, N2, P1, P2 , and P3. There was no significant difference between the pHRTF(all) and the mHRTF for the target elevation of 90°. This implies that one of P2 and P3 or both of them is required in addition to N1, N2, and P1 to localize the elevation of 90° accurately in this case.

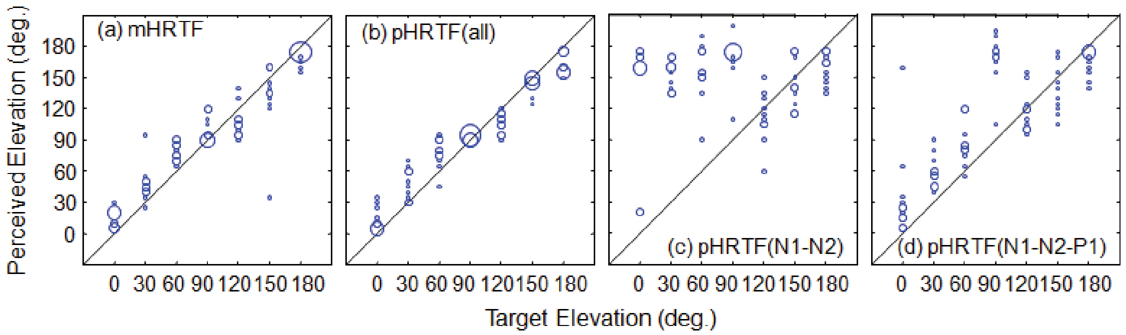

Figure 4. Localization responses to mHRTF, pHRTF(all), pHRTF(N1-N2), and pHRTF(N1-N2-P1).

Table 2 Statistically significant difference in localization error between mHRTFs and pHRTFs for subject MK. * $p < 0.05$, ** $p<0.01$

Subject	pHRTF	Target elevation (degree)						
		0	30	60	90	120	150	180
	All	-	-	-	-	-	-	-
MK	N1-N2	**	**	**	**	-	-	*
	N1-N2-P1	-	-	-	**	-	-	-

5. DISCUSSIONS

The reason why some spectral peaks and notches markedly contribute to the perception of elevation is discussed. Figure 5 shows the center frequency of N1, N2, and P1 of the measured HRTFs of subject IT in the upper median plane in 30° steps. This figure shows that the frequency of N1 and N2 changes remarkably as the elevation of a sound source changes. However, neither only N1 nor only N2 can identify the source elevation uniquely because these changes are non-monotonic. It seems that the pair of N1 and N2 plays an important role as the vertical localization cues.

On the other hand, the frequency of P1 does not depend on the source elevation. According to Shaw and Teranishi [2], the meatus-blocked response shows a broad primary resonance, which contributes almost 10 dB of gain over the 4–6 kHz band, and the response in this region is controlled by a “depth” resonance of the concha.

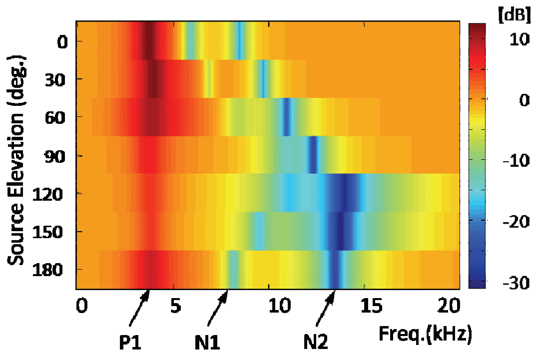

Figure 5. Distribution of N1, N2, and P1 in the upper median plane.

Therefore, the contribution of P1 to the perception of elevation cannot be explained in the same manner as those of N1 and N2. It could be considered that the hearing system of a human being utilizes P1 as the reference information to analyze N1 and N2 in the ear-input signals. Some previous studies support this hypothesis. Iida and Morimoto [17] reported that the hearing system determines the elevation of a sound source using only the spectrum information of the ear input-signals, regardless of *a priori* knowledge on the source signal. The spectrum analysis is easier if a frequency band that has common characteristics independent of incident sound elevation exists.

6. CONCLUSIONS

A method of HRTF modeling based on spectral cues for vertical localization was investigated. Authors proposed a parametric HRTF model for simulating vertical sound localization. The parametric HRTF is recomposed only of the spectral peaks and notches extracted from the measured HRTF, and the characteristics of the spectral peaks and notches are expressed parametrically with center frequency, level, and sharpness. Localization tests were carried out in the upper median plane using the subjects' own measured HRTFs and the parametric HRTFs with various combinations of spectral peaks and notches. The results show that (1) perceived elevation for the parametric HRTF recomposed of all the spectral peaks and notches is as accurate as that for the measured HRTF; (2) some spectral peaks and notches play an important role in determining the perceived elevation, whereas some peaks and notches do not; (3) the parametric HRTF recomposed of the first and second notches (N1 and N2) and the first peak (P1) provides almost the same accuracy of elevation perception as the measured HRTFs.

Observations of the spectral peaks and notches of the HRTFs in the upper median plane indicate that (1) the center frequency, level, and sharpness of N1 and N2 change remarkably as the source elevation changes; (2) whereas, P1 does not depend on the source elevation.

From these results, it is concluded that (1) N1 and N2 can be regarded as spectral cues; (2) the hearing system of a human being could utilize P1 as the reference information to analyze N1 and N2 in ear-input signals.

References:

[1] K. Roffler, A. Butler: Factors that influence the localization of sound in the vertical plane. Journal of the Acoustical Society of America **43** (1968) 1255–1259.

[2] E. A. G. Shaw, R. Teranishi: Sound pressure generated in an external-ear replica and real human ears by a nearby point source. Journal of the Acoustical Society of America **44** (1968) 240–249.

[3] J. Blauert: Sound localization in the median plane. Acustica **22** (1969–1970) 205–213.

[4] B. Gardner, S. Gardner: Problem of localization in the median plane: effect of pinna cavity occlusion. Journal of the Acoustical Society of America **53** (1973) 400–408.

[5] J. Hebrank, D. Wright: Spectral cues used in the localization of sound sources on the median plane. Journal of the Acoustical Society of America **56** (1974) 1829–1834.

[6] A. Butler, K. Belendiuk: Spectral cues utilizes in the localization of sound in the median sagittal plane. Journal of the Acoustical Society of America **61** (1977) 1264–1269.

[7] S. Mehrgardt, V. Mellert: Transformation characteristics of the external human ear. Journal of the Acoustical Society of America **61** (1977) 1567–1576.

[8] A. J. Watkins: Psychoacoustic aspects of synthesized vertical locale cues. Journal of the Acoustical Society of America **63** (1978) 1152–65.

[9] M. Morimoto, H. Aokata: Localization cues of sound sources in the upper hemisphere. Journal of the Acoustical Society of Japan (E) **5** (1984) 165–173.

[10] B. C. J. Moore, R. Oldfield, G. J. Dooley: Detection and discrimination of peaks and notches at 1 and 8 kHz. Journal of the Acoustical Society of America **85** (1989) 820–836.

[11] F. Asano, Y. Suzuki, T. Sone: Role of spectral cues in median plane localization. Journal of the Acoustical Society of America **88** (1990) 159–168.

[12] J. C. Middlebrooks: Narrow-band sound localization related to external ear acoustics. Journal of the Acoustical Society of America **92** (1992) 2607–2624.

[13] K. Iida, M. Yairi, M. Morimoto: Role of pinna cavities in median plane localization. In: Proceedings of 16th International Congress on Acoustics (1998) 845–846.

[14] V. C. Raykar, R. Duraiswami, B. Yegnanarayana: Extracting the frequencies of the pinna spectral notches in measured head related impulse responses. Journal of the Acoustical Society of America **118** (2005) 364–374.

[15] M. Morimoto, K. Iida, M. Itoh: Upper hemisphere sound localization using head-related transfer functions in the median plane and interaural differences. Acoustical Science and Technology **24** (2003) 267–275.

[16] D. Hammershøi, H. Møller: Sound transmission to and within the human ear canal. Journal of the Acoustical Society of America **100** (1996) 408–427.

[17] K. Iida, M. Morimoto: *A priori* knowledge of the sound source spectrum in median plane localization. Journal of the Acoustical Society of America **105** (1999) 1391.