WESPAC IX 2006
The 9th Western Pacific Acoustics Conference
Seoul, Korea, June 26-28, 2006

A NOVEL HEAD-RELATED TRANSFER FUNCTION MODEL BASED ON SPECTRAL AND INTERAURAL DIFFERENCE CUES

Kazuhiro IIDA, Motokuni ITOH
AV Core Technology Development Center, Matsushita Electric Industrial Co., Ltd.
1006 Kadoma,Osaka 571-8501, Japan
E-mail:iida.kazuhiro@jp.panasonic.com

Atsue ITAGAKI, Masayuki MORIMOTO
Environmental Acoustics Laboratory, Faculty of Engineering,
Kobe University
E-mail: mrmt@kobe-u.ac.jp

ABSTRACT

Authors propose the parametric HRTF, which is recomposed of all or a part of spectral peaks and notches extracted from the measured HRTF, and interaural differences. The results of localization tests in the median plane show that (1) some spectral peaks and notches play an important role in determining the perceived elevation, whereas some peaks and notches do not; (2) listeners perceive a source elevation accurately using the parametric HRTF recomposed of the first and second notches (N1 and N2), and the first peak (P1); (3) the center frequency, level, and sharpness of N1 and N2 change remarkably as the source elevation changes. 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, which does not have a directional dependence, as the reference information to analyze N1 and N2 in ear-input signals.
KEYWORDS: Head-Related Transfer Function, Spectral Cue, Sound Localization

INTRODUCTION

It is generally known that spectral information is a cue for median plane localization. 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. Mehrgardt and Mellert [1] have shown that the spectrum changes systematically in the frequency range above 5 kHz as the elevation of a sound source changes. Hebrank and Wright [2] carried out experiments with band noise and reported the following: spectral cues of median plane localization exist between 4 and 16 kHz; front cues are a 1-octave notch having a lower cutoff frequency between 4 and 8 kHz and increased energy above 13 kHz; an above cue is a 1/4-octave peak between 7 and 9 kHz; a behind cue is a small peak between 10 and 12 kHz with a decrease in energy above and below the peak. Butler and Belendiuk [3] showed almost the same results as those of Hebrank and Wright on the relationship between sound localization and spectral notches in the frontal half of the median plane. Moore *et al*. [4] measured the thresholds of various spectral peaks and notches. They showed that the spectral peaks and notches that Hebrank and Wright regarded as the cues of median plane localization are detectable for listeners, and thresholds for detecting changes in the position of sound sources in the frontal part of the median plane can be accounted for in terms of thresholds for the detection of differences in the center frequency of spectral notches. Shaw and Teranishi [5] reported that a spectral notch changes from 6 kHz to 10 kHz when the elevation of a sound source changes from -45 to 45°. Iida *et al*. [6] carried out localization tests and HRTF measurements 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. Rayker *et al*. [7] noted that one of the prominent features observed in the HRIR and one that has been shown to be important for elevation perception are the deep spectral notches attributed to the pinna. They proposed a method of extracting the frequencies of pinna spectral notches from the measured HRIR, distinguishing them from other confounding features. The extracted notch frequencies are related to the physical dimensions and shape of 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.

In this study, a method of HRTF modeling based on spectral and interaural difference cues is investigated. Authors propose the parametric HRTF, which is recomposed of all or a part of the spectral peaks and notches extracted from the measured HRTF. These spectral peaks and notches are expressed parametrically with center frequency, level, and sharpness. The localization tests in the median plane clarify the accuracy of sound localization with the parametric HRTF.

METHOD OF RECOMPOSING PARAMETRIC HRTFS

Measurements of HRTFs. The subjects' own HRTFs in the upper median plane were measured in an anechoic chamber in 30-degree steps. Ear-microphones [8] 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. This is so called the "blocked entrances condition" [9].

Extraction of spectral peaks and notches. As mentioned above, the spectral peaks and notches in the frequency range above 5 kHz prominently contribute to the perception of sound source elevation. The spectral peaks and notches are extracted from the measured HRTFs regarding the peaks around 4 kHz, which are independent of sound source elevation [5], 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 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):

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

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_1 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 *HRTFw*, 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 *HRTFw*. The parametric HRTF for the contralateral ear includes the time delay corresponding to the interaural time difference (ITD) of the measured HRTF. The ITD was operationally defined as the time lag at which the interaural cross-correlation of the measured head-related impulse responses reached its maximum. An example of a parametric HRTF is shown in Fig.1.

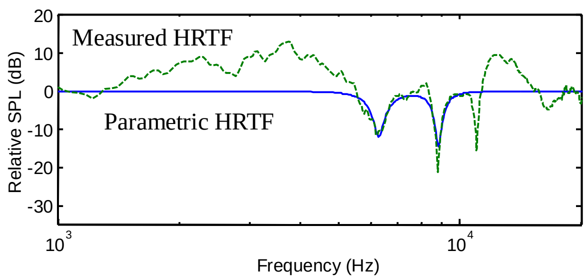

Fig.1 An example of parametric HRTF (subject IT, 0 °).

Dashed line: measured HRTF, solid line: parametric HRTF (N1-N2)

LOCALIZATION TEST I

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

Method. Subjects were two males, 30 and 44 years of age, with a normal hearing sensitivity. They were experienced in this type of localization test. The localization tests were conducted in a quiet listening room. A note PC (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 in 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. Therefore, this is so called the "blocked entrances condition" [9]. 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 $P_{l,r}(\omega)$

were created by Eq. (3):

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

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 parametric HRTFs in the upper median plane in 30-degree steps were used. The parametric HRTFs of all the combinations of all or a part 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.

A stimulus was delivered by hitting a key of the note PC at 60 dB SPL. 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, by clicking a mouse, on the computer display. 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.

Results. Figure 2 shows the distributions of the responses of subject IT for the target elevation of 0°. The diameter of each circle plotted is proportional to the number of responses within five degrees. The ordinate 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 mHRTF, the subject perceived the elevation of a sound source accurately. On the other hand, the variance of responses was large for the stimulus without an HRTF. 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. One peak or notch did not provide 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. The diameter of each circle plotted is proportional to the number of responses within five degrees. 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°.

The localization error was calculated. The error is defined as the mean absolute deviation of the perceived elevation from the target elevation [10]. 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 shows the results of the *t*-test. There was no statistically significant difference between the pHRTF(all) and the mHRTF

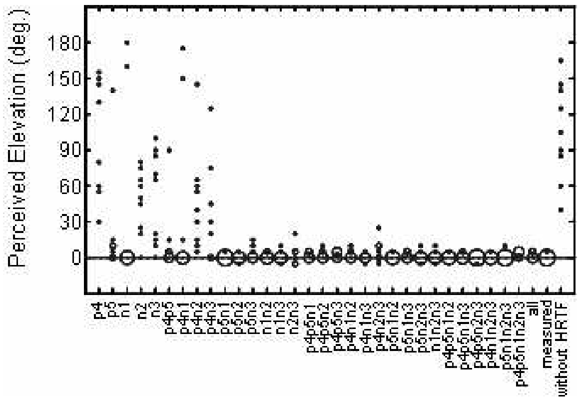

Fig.2 Localization responses of subject IT for the target elevation of 0°. The diameter of the data points indicates the number of responses.

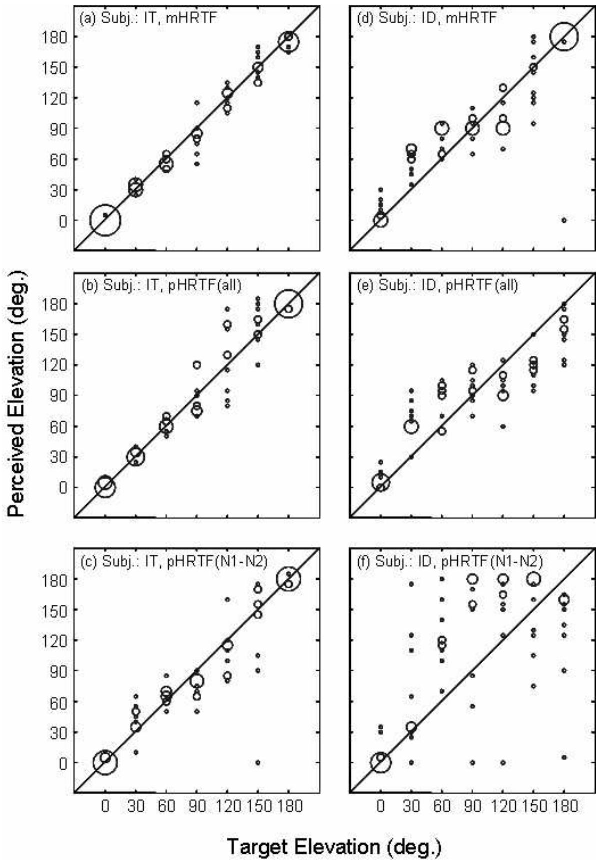

Fig.3 Localization responses to mHRTF, pHRTF(all), and pHRTF(N1-N2)

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.

Table 1 Statistically significant difference in localization error

*between the mHRTFs and the pHRTFs for subjects IT and ID (**: p < 0.01, *: p < 0.05).*

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

LOCALIZATION TEST II

The results of localization test I show that N1 and N2 play an important role in the perception of elevation in the median plane, and that P1, the characteristics of which are independent of source elevation, could be a cue for the perception. The purpose of localization test II is to clarify the effect of P1 on median plane localization.

Method. The apparatus, source signal and target elevation were the same as those used in localization test I. The subject's own mHRTFs and pHRTFs in the upper median plane in 30-degree 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. She was not experienced in this type of localization test.

Results. 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° shifted 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).

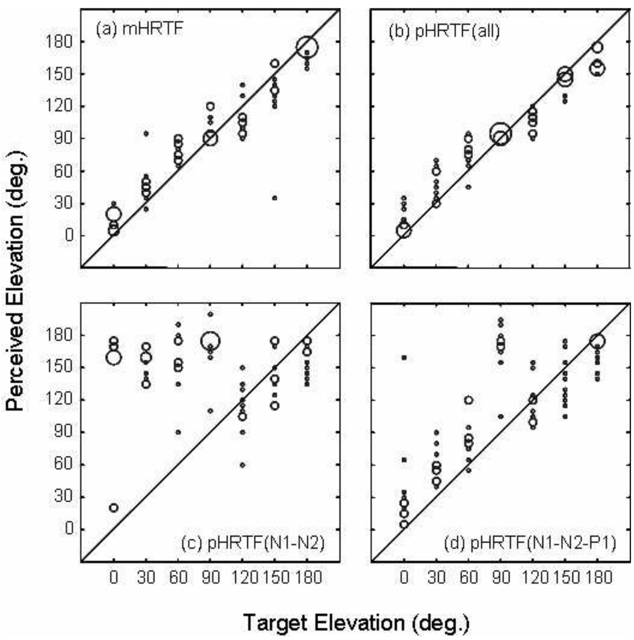

Fig.4 Localization responses to mHRTF, pHRTF(all), pHRTF(N1-N2), and pHRTF(P1-N1-N2).

Table 2 shows 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.

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

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

DISCUSSIONS

The reason why some spectral peaks and notches markedly contribute to the perception of elevation is discussed. The parameters of N1 and N2, i.e., center frequency, level, and sharpness, change more remarkably than those of other spectral peaks and notches, as the elevation of a sound source changes in the upper median plane. It is supposed that the changes in the characteristics of these two spectral notches play an important role in the perception of elevation. This hypothesis is substantiated by other reported experiments. Hebrank and Wright [2] carried out experiments with a narrow-band noise and reported the following: front cues are a 1-octave notch having a lower cutoff frequency between 4 and 8 kHz (corresponding to N1) and increased energy above 13 kHz; a behind cue is a small peak between 10 and 12 kHz with a decrease in energy above and below the peak (corresponding to N1 and N2). The cues reported by Hebrank and Wright are consistent with our results.

P2 corresponds to the above cue, a 1/4-octave peak between 7 and 9 kHz, reported by Hebrank and Wright. Blauert [11] also reported a similar result; that is, 8 kHz is a directional band for the above. For the target elevation of 90°, there was a statistically significant difference between the pHRTF(N1-N2) and the mHRTF in the responses of subject ID. In the responses of subject MK, significant differences were observed between the mHRTF and the pHRTF(N1-N2), and between the mHRTF and the pHRTF(N1-N2-P1) for the target elevation of 90°. These results infer that P2 plays an important role in the above localization.

On the other hand, the parameters of P1 do not depend on the source elevation. According to Shaw and Teranishi [5], 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. 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 [12] 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. A similar consideration was introduced by Asano *et al.* [13]. They proposed the following hypothesis: The boundary of the frequency areas that separate the front-rear judgment cues in the low frequency range and elevation angle judgment cues in the high frequency range exists at around 2-5 kHz, and the perceived elevation angle is related to the ratio of the power in the area above that boundary to that of the area below.

CONCLUSIONS

Authors proposed the parametric HRTF, which is recomposed of all or a part of spectral peaks and notches extracted from the measured HRTF, and interaural differences. These 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 parametric HRTFs. The results show that (1) some spectral peaks and notches play an important role in determining the perceived elevation, whereas some peaks and notches do not; (2) listeners perceive a source elevation accurately using the parametric HRTF recomposed of the first and second notches (N1 and N2), and the first peak (P1); (3) the center frequency, level, and sharpness of N1 and N2 change remarkably as the source elevation changes. 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, which does not have a directional dependence, as the reference information to analyze N1 and N2 in ear-input signals.

REFERENCES

1. S. Mehrgardt and V. Mellert, "Transformation characteristics of the external human ear," *J. Acoust. Soc. Am.*, **61**, 1567-1576 (1977)

2. J. Hebrank and D. Wright, "Spectral cues used in the localization of sound sources on the median plane," *J. Acoust. Soc. Am.*, **56**, 1829-1834 (1974)

3. A. Butler and K. Belendiuk, "Spectral cues utilizes in the localization of sound in the median sagittal plane," *J. Acoust. Soc. Am.*, **61**, 1264-1269 (1977)

4. B. C. J. Moore, R. Oldfield, and G. J. Dooley, "Detection and discrimination of peaks and notches at 1and 8 kHz," *J. Acoust. Soc. Am.*, **85**(2), 820-836 (1989)

5. E. A. G. Shaw and R. Teranishi, "Sound pressure generated in an external-ear replica and real human ears by a nearby point source," *J. Acoust. Soc. Am.*, **44**, 240-249 (1968)

6. K. Iida, M. Yairi, and M. Morimoto, "Role of pinna cavities in median plane localization," Proceedings of 16th International Congress on Acoustics, 845-846 (1998)

7. V. C. Raykar, R. Duraiswami, and B. Yegnanarayana, "Extracting the frequencies of the pinna spectral notches in measured head related impulse responses," *J. Acoust. Soc. Am.*, **118**(1), 364-374 (2005)

8. M. Morimoto, K. Iida, and M. Itoh, "Upper hemisphere sound localization using head-related transfer functions in the median plane and interaural differences," *Acoustical Science and Technology*, **24**(5), 267-275 (2003)

9. D. Hammershøi and H. Møller, "Sound transmission to and within the human ear canal," *J. Acoust. Soc. Am.*, **100**(1), 408-427 (1996)

10. M. Morimoto and Y. Ando, "On the simulation of sound localization," *J. Acoust. Soc. Jpn.(E)*, **1**(3), 167-174 (1980)

11. J. Blauert, " Sound localization in the median plane," *Acustica*, **22**, 205-213 (1969/70)

12. K. Iida, M. Morimoto, "*A priori* knowledge of the sound source spectrum in median plane localization," *J. Acoust. Soc. Am.*, **105**(2) Pt.2, 1391 (1999)

13. F. Asano, Y. Suzuki, and T. Sone, "Role of spectral cues in median plane localization," *J. Acoust. Soc. Am.*, **88**(1), 159-168 (1990).