大阪
inter.noise 2011
Osaka Japan September 4-7

3D sound image control by individualized parametric head-related transfer functions

Kazuhiro IIDA[1] and Yohji ISHII

[1] Chiba Institute of Technology

2-17-1 Tsudanuma, Narashino, Chiba 275-0016 JAPAN

ABSTRACT

It is well known that the listener's own head-related transfer functions (HRTFs) provide accurate 3D sound image localization. However, the HRTFs of other listeners often cause degradation of localization accuracy. Though the 3D auditory display with a head-motion trucker, which provides a dynamic spatial cue to a listener, improves the rate of front-back confusion, this dynamic cue is not enough to accomplish accurate 3D sound image control. In other words, the individualization of HRTFs is necessary for accurate and realistic 3D sound image control. The authors have shown that the frequency of the first and the second spectral notches (N1 and N2) above 4 kHz in HRTFs act an important role as spectral cues for vertical localization. Furthermore, it is indicated that a sound image in any direction in the upper hemisphere can be localized with parametric median-plane HRTFs composed of N1 and N2 combined with frequency-independent interaural time difference. This means that the individualization of HRTFs of all the directions in the upper hemisphere can be replaced by that of N1 and N2 frequency of the HRTFs on the median plane. Thus, this paper describes a 3D sound image control method using individualized parametric HRTFs. In addition, our 3D auditory display system named SIRIUS (Sound Image Reproduction system with Individualized-HRTF, graphical User-interface, and Successive head-movement tracking) is introduced.

Keywords: Sound image control , head-related transfer function, Individualization

1. 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 [1-11]. 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. Shaw and Teranishi [2] 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.* [11] 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. 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. However, it has been unclear which component of HRTF plays an important role of as a spectral cue.

2. CUES FOR MEDIAN PLANE LOCALIZATION

The authors have proposed a parametric HRTF model to clarify the contribution of each spectral peak and notch as a spectral cue for vertical 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 frequency, level, and sharpness. Localization tests were

[1] kazuhiro.iida@it-chiba.ac.jp

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 [12].

2.1 Parametric HRTFs

As mentioned above, the spectral peaks and notches in the frequency range above 5 kHz prominently contribute to the perception of sound source elevation. Therefore, the spectral peaks and notches are extracted from the measured HRTFs regarding the peaks 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 frequency, level, and sharpness. The amplitude of the parametric HRTF is recomposed of all or some of these spectral peaks and notches. Fig.1 shows examples of the parametric HRTFs recomposed of N1 and N2. 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.

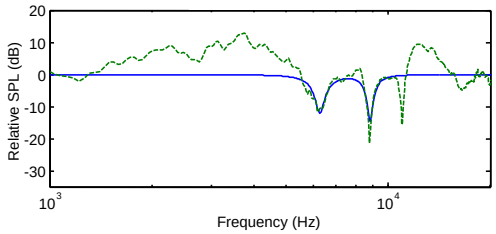


Figure 1 – An example of parametric HRTF. Dashed line: measured HRTF, solid line: parametric HRTF recomposed of N1 and N2

2.2 Method of Sound Localization Tests

Localization tests in the upper median plane were carried out using the subjects' own measured HRTFs and the parametric HRTFs. A notebook computer (Panasonic CF-R3), an audio interface (RME Hammerfall DSP), open-air headphones (AKG K1000), and the ear-microphones [12] 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. 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. (1):

$$P_{l,r}(\omega) = S(\omega) \cdot H_{l,r}(\omega) / C_{l,r}(\omega), \quad (1)$$

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, which were recomposed of all or a part of the spectral peaks and notches, in the upper median plane in 30-degree steps were used. 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 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. A circle and an arrow, which indicated the median and horizontal planes, respectively, were shown on the display of the notebook computer. The subject's task was to plot the perceived elevation on the circle, by clicking a mouse, on the computer display. The subject could hear each stimulus over and over again. However, after he plotted the perceived elevation and moved on to the next stimulus, the subject could not return to the previous stimulus. The order of presentation of stimuli was randomized. The subjects responded ten times for each stimulus.

2.3 Results of the Tests

Figure 2 shows the examples of responses of a subject to the measured and the parametric HRTFs 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. Hereafter, the measured HRTF and parametric HRTF are expressed as the mHRTF and pHRTF, respectively. The subject perceived the elevation of a sound source accurately at all the target elevations for the mHRTF. For the pHRTF(all), which is recomposed of all the spectral peaks and notches, the responses distribute along a diagonal line, and this distribution is practically the same 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. The pHRTF(N1–N2), which is recomposed of N1 and N2, provides almost the same accuracy of elevation perception as the mHRTF at most of the target elevations. However, for some subjects P1 is also necessary for accurate localization in addition to N1 and N2.

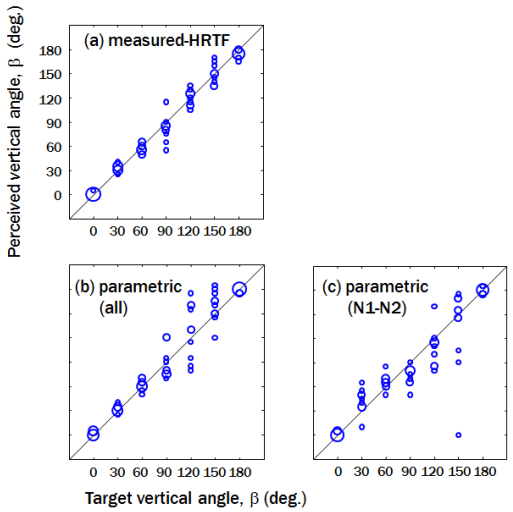


Figure 2 – Examples of responses to stimuli of measured HRTFs and parametric HRTFs in the median plane

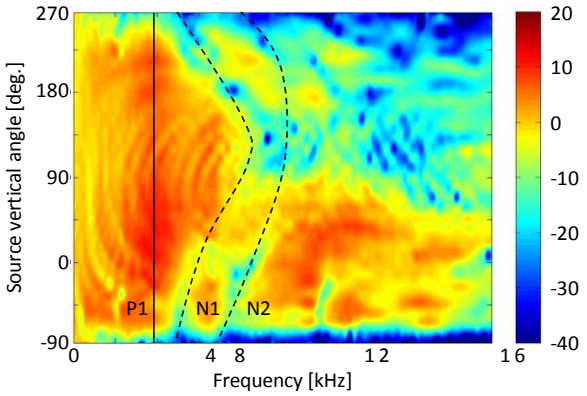


Figure 3 – Distribution of frequencies of N1, N2, and P1 in the median plane

2.4 Discussions

The reason why some spectral peaks and notches markedly contribute to the perception of elevation is discussed. Fig. 3 shows the distribution of the spectral peaks and notches of the measured HRTFs in the median plane. This figure shows that the frequencies of N1 and N2 change remarkably as the elevation of a sound source changes. Since these changes are non-monotonic, neither only N1 nor only N2 can identify the source elevation uniquely. It seems that the pair of N1 and N2 plays an important role as the vertical localization cues. 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. 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.

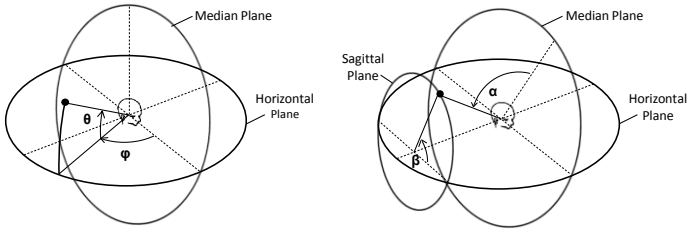
3. 3D SOUND IMAGE CONTROL USING pHRTF AND ITD

As mentioned above, the direction of a sound image is controlled in the median plane by pHRTFs(N1-N2-P1). In this chapter, the authors try to expand the method to arbitrary 3D direction in the upper hemisphere.

Morimoto and Aokata [9] demonstrated that an interaural-polar-axis coordinate system, as shown in Fig. 4, is more suitable for explaining sound localization in any direction in the upper hemisphere than a geodesic coordinate system defined by the azimuth angle ϕ and the elevation angle θ. In an interaural-polar-axis coordinate system, the angle α is the angle between the aural axis and a straight line connecting the sound source with the center of a subject's head, and the angle β is the angle between the horizontal plane and the perpendicular from the sound source to the aural axis, that is, the vertical angle in a plane parallel to the median plane, called the sagittal plane. According to the results of localization tests, Morimoto and Aokata determined that angle α and angle β are independently determined by binaural disparity cues and spectral cues, respectively.

Another localization test [13], using HRTFs measured on the median plane and interaural differences measured on the frontal horizontal plane, showed that the vertical angle α and the lateral angle β of the sound images could be perceived with much the same accuracy as those of real sound sources in the upper hemisphere.

These results infer that the N1 and N2 frequencies are similar among sagittal planes for the same vertical angle β regardless of lateral angle α. The possibility of sound image control in the upper hemisphere using pHRTFs(N1-N2-P1) on the median plane and interaural differences measured on the frontal horizontal plane is discussed in this chapter.



(a) geodesic coordinate system (b) interaural-polar-axis coordinate system

Figure 4 – Definition of the geodesic coordinate system and interaural-polar-axis coordinate system. ϕ: azimuth, θ: elevation, α: lateral angle, and β: vertical angle

3.1 Method of Localization Tests

Localization tests were carried out in an anechoic room. A notebook computer (DELL Vostro 1520), an audio interface (RME Hammerfall DSP), an amplifier (marantz PM4001), an A/D converter (Rolland EDIROL M-10MX), open-air headphones (AKG K1000), and the ear-microphones [12] were used for the localization tests. The source signal was a wide-band white noise from 200 Hz to 20 kHz. The duration of the stimulus was 1.2 s, including the rise and fall times of 0.1 s, respectively. The target directions are 22 in the upper hemisphere. That is, seven vertical angles β ranging from front to rear in 30 degree steps in the three upper sagittal planes ranging from the median plane to right hand sagittal plane in 30 degree steps of lateral angles α, and α=90 degrees.

The subject's own pHRTFs(N1-N2-P1) for the seven vertical angles in the upper median plane were prepared. In addition, interaural time differences (ITD) for each lateral angle α were measured at four lateral angles (α= 0, 30, 60, and 90 degrees) on the right side of the frontal horizontal plane (β=0 degrees). The signals used for the measurements of ITD were obtained by convolving a wide-band white noise with the HRTFs measured at the four lateral angles. The ITD was defined as the time lag at which the interaural cross-correlation of the signals reached a maximum.

For the localization test, 28 directions (seven pHRTFs(N1-N2-P1) × four measured ITD) were simulated. Although the position at α=90 degrees is defined only for lateral angle α, and not for the vertical, the median-plane pHRTFs for all seven vertical angles β were used. The idea here was to examine whether or not all of the responses would be concentrated around the target position at α=90 degrees, despite the variation in HRTFs associated with the seven β angles.

Stimuli $P_{l,r}(\omega)$ were created by Eq. (1). The order of presentation of stimuli was randomized. The subject's task was to plot the perceived azimuth and elevation on the response sheet. Subjects responded ten times for each stimulus. The subjects were two males (ISY and GMU).

3.2 Results of Localization Tests

Figure 5 shows the responses of subject ISY. The circular arcs denote the lateral angle α, and the straight lines from the center denote the vertical angle β. The outermost arc denotes the median plane (α=0 degrees), and the center of the circle denotes the extreme side direction (α=90 degrees). The target α and β are shown in bold lines. The intersection of the two bold lines indicates the target direction. The diameter of the circular plotting symbols is proportional to the number of responses within each cell of a sampling grid with 5 degree resolution.

Broadly speaking, the responses of the lateral angle α are concentrated around the target directions except for the side direction. The responses of the vertical angle β are concentrated around the target directions of forward and rearward, however scattered across the target direction of rearward.

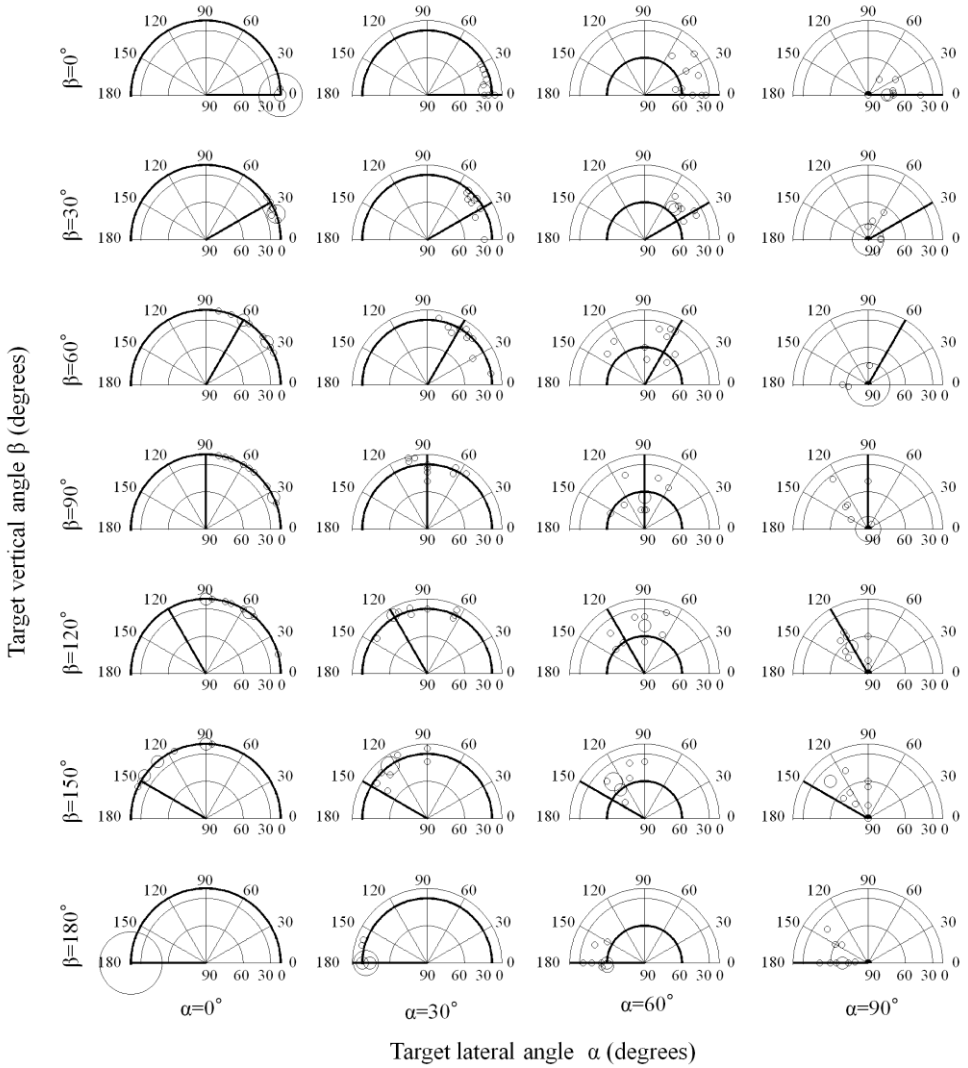


Figure 5 – Responses to the stimuli which simulated parametric HRTF(N1,N2,P1) in the median plane and interaural time differences in the horizontal plane for subject ISY

The statistical test (t-test) in mean localization error between actual sound sources and proposed sound image control method was conducted. The results are shown in Table 1. There are no significant differences at 22 of 28 directions and 26 of 28 directions for subject ISY and GMU, respectively. The localization error of subject ISY for the extreme side direction (α=90 degrees) varied associated with the β angles. These results show that the proposed method provides accurate sound image control for almost of all the direction in the upper hemisphere.

Table 1 – Results of statistical test

(a) Subject ISY

Target lateral angle, α [deg.]	Target vertical angle, β [deg.]						
	0	30	60	90	120	150	180
0					**		
30	*						
60							
90	*				**	**	**

*: p<0.05 **: p<0.01

(b) Subject GMU

Target lateral angle, α [deg.]	Target vertical angle, β [deg.]						
	0	30	60	90	120	150	180
0	**						
30			*				
60							
90							

4. METHODS OF INDIVIDUALIZATION OF HRTFs

It is well known that the HRTFs of other listeners often cause degradation of localization accuracy. Though the 3D auditory display with a head-motion trucker, which provides a dynamic

spatial cue to a listener, improves the rate of front-back confusion, this dynamic cue is not enough to accomplish accurate 3D sound image control. In other words, the individualization of HRTFs is necessary for accurate and realistic 3D sound image control.

Generally speaking, HRTF of all the direction must be individualized for entire 3D sound image control. However, individualization of HRTFs of all the direction could be replaced by that of frequencies of N1 and N2 of the HRTFs in the median plane if the 3D sound image control method described in chapter 3 is used. A method to individualize the frequencies of N1 and N2 of the HRTFs in the median plane is discussed in this chapter. The authors have been researching the individualization of HRTFs in following two approaches;

1) search for the appropriate HRTF for the listener from the minimal HRTF database,

2) estimation of the listener's own HRTF from the shape of the pinnae.

This chapter describes only the first approach due to the limitation of space.

4.1 Individualization Using Minimal Parametric HRTF Database

Searching methods of appropriate HRTFs for a listener using HRTF database have been proposed [14, 15]. However, the previous methods require so much time to find the appropriate HRTFs from a lot of HRTFs in the database. In order to reduce the search time and burden of the listeners, the database, which is composed of the required minimum number of parametric HRTFs, is created by the following procedure.

4.1.1 Individual Difference in N1 and N2 Frequencies of Front Direction

The range of individual differences of N1 and N2 frequencies was obtained for many listeners for the front direction, at which the front-back localization error occurs frequently due to the individual difference. The distribution range of the N1 and N2 frequencies of 50 subjects (100 ears) for the front direction is shown in Fig. 6. One-hundred ears are considered a sufficient number of samples as a subgroup of the population. This figure indicates that the individual differences in N1 and N2 are very large. The N1 frequency ranges from 5.5 kHz to 10 kHz, and the N2 frequency ranges from 7 kHz to 12.5 kHz.

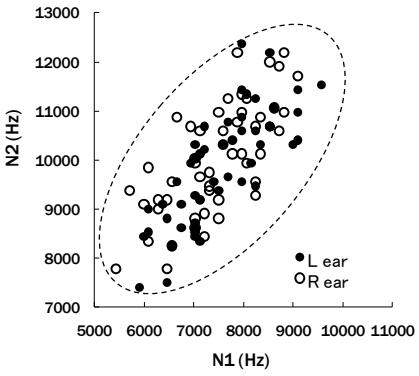


Figure 6 – N1 and N2 frequencies of 50 subjects (100 ears) for the front direction.

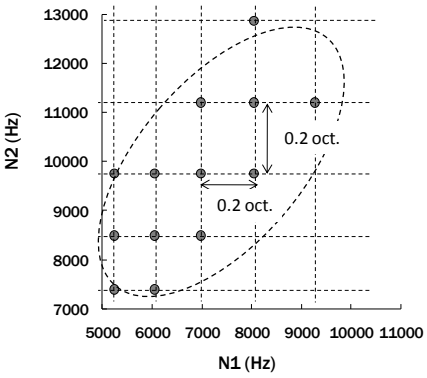


Figure 7 – Extracted pairs of N1and N2; the distribution range is divided by the jnd of *NFD*.

4.1.2 Notch Frequency Distance as a Physical Measure for Individual Difference in HRTF

A physical measure for individual difference of HRTFs, NFD (Notch Frequency Distance) has been proposed. NFD expresses the distance between $HRTF_j$ and $HRTF_k$ in the octave scale, as follows:

$$NFD_1 = \log_2\left\{f_{N1}(HRTF_j) / f_{N1}(HRTF_k)\right\} \quad \text{[oct.]} \tag{2}$$

$$NFD_2 = \log_2\left\{f_{N2}(HRTF_j) / f_{N2}(HRTF_k)\right\} \quad \text{[oct.]} \tag{3}$$

$$NFD = |NFD_1| + |NFD_2| \quad \text{[oct.]} \tag{4}$$

Localization tests were carried out to clarify the just noticeable difference of NFD in vertical localization for front direction. The results show that the jnd is approximately 0.2 octaves both for N1 and N2. In other words, the individual difference in N1 and N2 frequencies within 0.2 octaves is acceptable

4.1.3 Minimal Pairs of N1 and N2 for Parametric HRTF Database

The distribution range of N1 and N2 was divided by the jnd of NFD for frontal localization (0.2 octave), as shown in Fig. 7. Then, pairs of N1 and N2 frequencies on grid points were extracted. Fig. 7 shows 13 extracted pairs of N1 and N2 frequencies. In this way, the minimum database consists of only 13 pHRTFs(N1-N2-P1) was created.

The parametric HRTF, by which a listener localizes a sound image at the front, is selected among 13 pHRTFs as the appropriate one for the front direction. This selection task takes only 1minutes.

4.1.4 Generation of the Individualized Parametric HRTFs in the Median Plane

The behavior of the N1 and N2 frequencies as a function of elevation seems to be common among listeners, even though the frequencies of N1 and N2 of the front direction highly depend on the listener (Fig. 8). The individualized N1 and N2 frequencies in the median plane are obtained by the regression equations (5) and (6), and by using the constant term given by the selected pHRTF mentioned in 4.1.3.

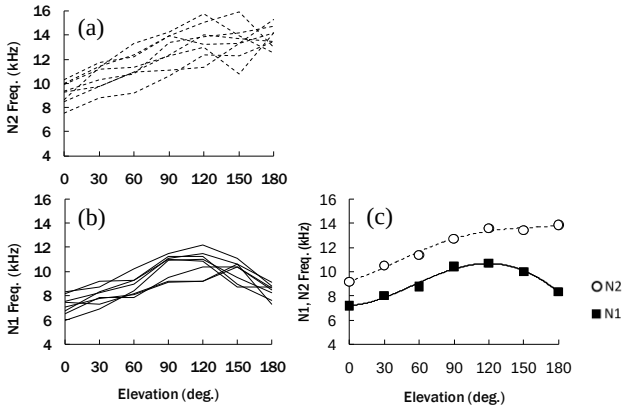


Figure 8 – N1 and N2 frequencies as a function of elevation. (a) measured N1 frequencies of 6 subjects; (b) measured N2 frequencies of 6 subjects; (c) regression curves obtained from the mean values of 6 subjects

$$f_{N1}(\beta) = 1.328\times10^{-10}\times\beta^6 - 1.363\times10^{-7}\times\beta^5 + 5.373\times10^{-5}\times\beta^4 - 9.845\times10^{-3}\times\beta^3 + 7.156\times10^{-1}\times\beta^2 + 2.418\times\beta + 6.930\times10^{3} \quad \text{[Hz]}, \quad (5)$$

$$f_{N2}(\beta) = 4.668\times10^{-10}\times\beta^6 - 4.413\times10^{-7}\times\beta^5 + 1.556\times10^{-4}\times\beta^4 - 2.512\times10^{-2}\times\beta^3 + 1.628\times\beta^2 + 1.421\times10^{1}\times\beta + 9.348\times10^{3} \quad \text{[Hz]}, \quad (6)$$

5. 3D DYNAMIC AUDITORY DISPLAY (SIRIUS)

A 3D dynamic auditory display named SIRIUS (Sound Image Reproduction system with Individualized-HRTF, graphical User-interface and Successive head-movement tracking) has been developed utilizing the abovementioned findings. SIRIUS has very small pHRTF database compared with the previous one. Furthermore, Individualization of HRTFs provides accurate and realistic sound images in arbitrary direction in 3D space. Figure 9 shows the system configuration of SIRIUS.

Figure 10 shows the GUI for the process of individualization of HRTF. A listener is requested to choose the appropriate pHRTF, by which he/she localizes a sound image at front, and to sting a thumbtack. Table 2 shows the specifications of SIRIUS.

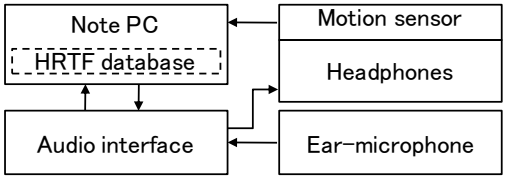


Figure 9 – System configuration of SIRIUS

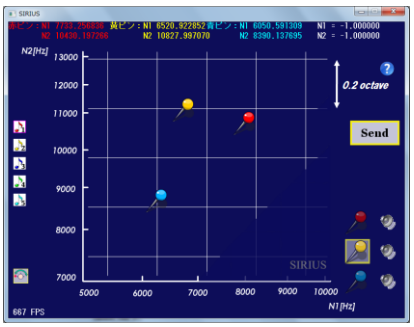


Figure 10 – GUI for individualization of HRTF

Table 2 – Specifications of SIRIUS

software development language		C++, C#, MATLAB
OS		Windows XP, Vista, 7 (32bit)
HDD		＞50MB
head motion sensor		ZMP e-nuvo IMU-Z, USB/Bluetooth
sound image control		measured HRTFs
		measured HRTFs (median plane) ＋ ITD
		parametric HRTFs (median plane) ＋ ITD
individualization of HRTFs		minimum parametric HRTF database
direction control	azimuth (resolution)	0 – 360° （＜1° ）
	elevation (resolution)	±90° （＜1° ）
distance cointrol		based on Binaural SPL
maximum number of sound sources		7
maximum latency		21ms

6. CONCLUTIONS

The authors have shown that the frequency of the first and the second spectral notches (N1 and N2) above 4 kHz in HRTFs act an important role as spectral cues for vertical localization. Furthermore, it is indicated that a sound image in any direction in the upper hemisphere can be localized with parametric median-plane HRTFs composed of N1 and N2 combined with frequency-independent interaural time difference. This means that the individualization of HRTFs in any direction in the upper hemisphere can be replaced by that of N1 and N2 frequency of the HRTFs on the median plane. In addition, a 3D auditory display system named SIRIUS, which localizes a sound image by individualized parametric HRTFs is introduced.

ACKNOWLEDGEMENTS

A part of this work is supported in part by Grant-in-Aid for Scientific Research (A) 22241040.

REFERENCES

[1] K. Roffler, A. Butler, "Factors that influence the localization of sound in the vertical plane", J. Acoust. Soc. Am. 43, 1255-1259 (1968).
[2] E. A. G. Shaw, 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).
[3] J. Blauert, "Sound localization in the median plane", Acustica 22, 205-213 (1969/70).
[4] B. Gardner, S. Gardner, "Problem of localization in the median plane: effect of pinna cavity occlusion", J. Acoust. Soc. Am. 53, 400-408 (1973).
[5] J. Hebrank, D. Wright, "Spectral cues used in the localization of sound sources on the median plane", J. Acoust. Soc. Am. 56, 1829-1834 (1974).
[6] A. Butler, K. Belendiuk, "Spectral cues utilizes in the localization of sound in the median sagittal plane", J. Acoust. Soc. Am. 61, 1264-1269 (1977).
[7] S. Mehrgardt, V. Mellert, "Transformation character-istics of the external human ear", J. Acoust. Soc. Am. 61, 1567-1576 (1977).
[8] A. J. Watkins, "Psychoacoustic aspects of synthesized vertical locale cues", J. Acoust. Soc. Am. 63, 1152–1165 (1978).
[9] M. Morimoto, H. Aokata, "Localization cues of sound sources in the upper hemisphere", J. Acoust. Soc. Jpn (E). 5, 165-173 (1984).
[10] J. C. Middlebrooks, "Narrow-band sound localization related to external ear acoustics", J. Acoust. Soc. Am. 92, 2607-2624 (1992).
[11] K. Iida, M. Yairi, M. Morimoto, "Role of pinna cavities in median plane localization", Proc. 16th Int'l Cong. on Acoust. 1998; 845-846 (1998).
[12] K. Iida, M. Itoh, A. Itagaki, M. Morimoto, "Median plane localization using a parametric model of the head-related transfer function based on spectral cues", Applied Acoustics, 68, 835-850 (2007).
[13] 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)
[14] J. C. Middrebrooks, E. A. Macpherson, and Z. A. Onsan, "Psychological customization of directional transfer functions for virtual sound localization", J. Acoust. Soc. Am. 108, 3088-3091 (2000).
[15] Y. Iwaya, "Individualization of head-related transfer functions with tournament-style listening test: Listening with other's ears," Acoust. Sci. & Tech. 27, 340-343 (2006).