INTERNATIONAL WORKSHOP
ON THE PRINCIPLES AND APPLICATIONS OF SPATIAL HEARING
IWPASH
NOVEMBER 11-13, 2009, ZAO, MIYAGI, JAPAN

# An approach to individualization of head-related transfer functions based on the spectral cues for sound localization

Kazuhiro IIDA*
*Chiba Institute of Technology*

## 1 Introduction

It is well known that the accurate sound image localization is accomplished when the listener's own HRTFs are reproduced [1]. The localization error and the reduction in reality often occur when other's HRTFs are used, due to the difference in the shape and size of the head and the pinnae. One of the methods to settle this problem is to provide the appropriate HRTFs to a listener extracted from the HRTF database.

The proper measure of individual difference in HRTF is necessary in order to provide the appropriate HRTFs to a listener. It is, however, hard to describe the individual difference in HRTFs because the characteristics of HRTFs are complex and the individual difference is also complex (Fig.1). The author proposes to uses the individual difference in the cues for localization, instead of the amplitude spectrum. In this study, spectral cues for sound localization, which the author has reported, are introduced at first, and then, a method to provide the appropriate HRTFs to a listener is proposed.

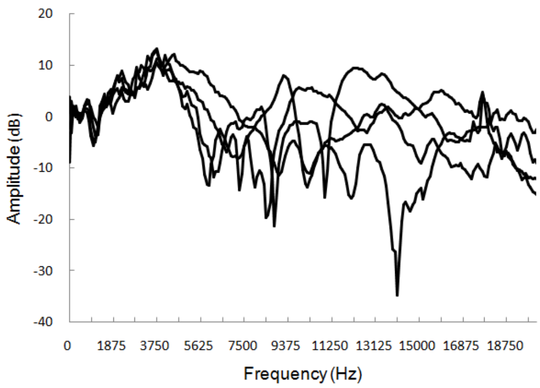


Fig.1 Measured HRTFs of 4 subjects for front direction.

## 2 Spectral cues for sound localization

A number of studies have revealed that spectral distortions caused by pinnae in the high-frequency range above approximately 5 kHz act as cues for elevation perception [2-6]. Hebrank and Wright [2] carried out experiments with filtered noise and reported that spectral cues of median plane localization exist between 4 and 16 kHz, front cues are a one-octave notch having a lower cutoff frequency between 4 and 8 kHz and increased energy above 13 kHz, an overhead cue is a 1/4-octave peak between 7 and 9 kHz, and a behind cue is a small peak between 10 and 12 kHz with a decrease in energy above and below the peak. Moore *et al*. [3] measured the thresholds of various spectral peaks and notches. They showed that the spectral peaks and notches that Hebrank and Wright regarded as cues of median plane localization are detectable by listeners and that thresholds for detecting changes in the position of a sound source 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. Butler and Belendiuk [4] showed that the prominent notch in the frequency response curve moved toward lower frequencies as the sound source moved from above to below the aural axis in the frontal half of the median plane. Raykar *et al*. [5] noted that deep spectral notches attributed to the pinna, which are prominent features observed in the head-related impulse response (HRIR), are important for elevation perception. 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.

Iida *et al*. [6] introduced a parametric HRTF, which is recomposed of all or some of the spectral peaks and notches extracted from the measured HRTF. 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

* E-mail address: kazuhiro.iida@it-chiba.ac.jp

sound source elevation [7], 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 (Fig.2). 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.

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 48kHz, 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 $HRTF_w$, respectively. Thus, the 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 $HRTF_w$. Fig.3 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.

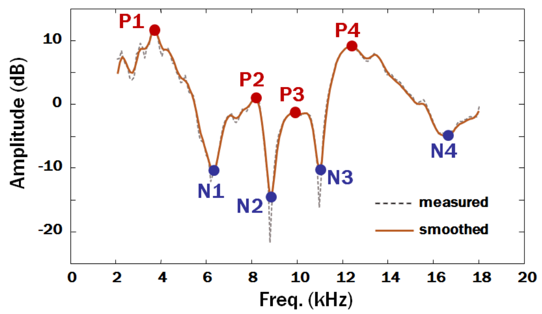


Fig.2 Examples of extracted spectral peaks and notches from measured HRTF.

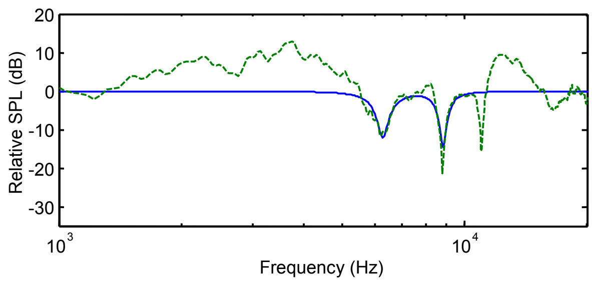


Fig.3 An example of parametric HRTF.
Dashed line: measured HRTF, solid line: parametric HRTF recomposed of N1 and N2

Then, they carried out localization tests in the median plane using the parametric HRTFs. The results revealed that the parametric HRTF recomposed of the first and second notches (N1 and N2) and the first peak (P1), in order of frequency above 4 kHz, provides almost the same localization accuracy as the measured HRTFs (Fig.4). Observations of the spectral peaks and notches indicate that the frequencies of N1 and N2 change markedly with changes in the source elevation, whereas the frequency of P1 is independent of the source elevation (Fig.5). Thus, Iida *et al.* concluded that N1 and N2 could be regarded as spectral cues and that the human auditory system could use P1 as a reference for analyzing N1 and N2.

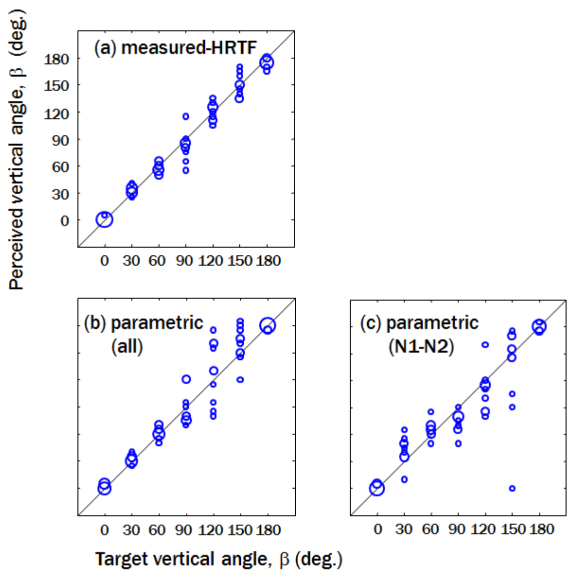


Fig.4 Responses of a subject to stimuli of measured HRTFs and parametric HRTFs in the median plane.

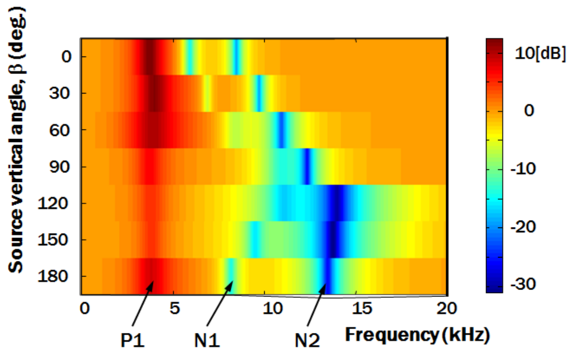


Fig.5 Distribution of frequencies of N1, N2, and P1 in the upper median plane

## 3 NFD (Notch Frequency Difference) as a measure for individual difference of HRTFs

Some appropriate measures are necessary in order to extract the HRTFs which provide accurate localization to a listener from the HRTF database. SD (Spectral Distortion) defined by Eq. (3) has been used as a conventional measure to evaluate the individual difference of HRTFs. It would, however, not be proper to use SD because it calculates the difference in all the range of the spectral components.

$$SD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[20\log_{10}\frac{|HRTF_j(f_i)|}{|HRTF_k(f_i)|}\right]^2} \quad [\mathrm{dB}] \qquad (3)$$

As mentioned above, N1 and N2 could be regarded as spectral cues. Therefore, distribution of the N1 and N2 frequencies among many subjects is observed. Fig. 6 shows the distribution of N1 and N2 frequency of 50 subjects' HRTF for front direction. N1 frequency ranges from 5.5 kHz to 10 kHz, and N2 frequency 7 kHz to 12.5 kHz.

Then, the author proposes NFD (Notch Frequency Distance) as a measure for the individual difference of HRTFs. NFD is defined as Eqs. (4-6).

$$NFD1_{j,k} = \log_2\{f_{N1}(HRTF_j) / f_{N1}(HRTF_k)\} \text{ [oct]} \qquad (4)$$

$$NFD2_{j,k} = \log_2\{f_{N2}(HRTF_j) / f_{N2}(HRTF_k)\} \text{ [oct]} \qquad (5)$$

$$NFD_{j,k} = |\, NFD1_{j,k} \,| + |\, NFD2_{j,k} \,| \quad \text{[oct.]}, \qquad (6)$$

where, $f_{N1}$ and $f_{N2}$ denote the frequency of N1 and N2, respectively.

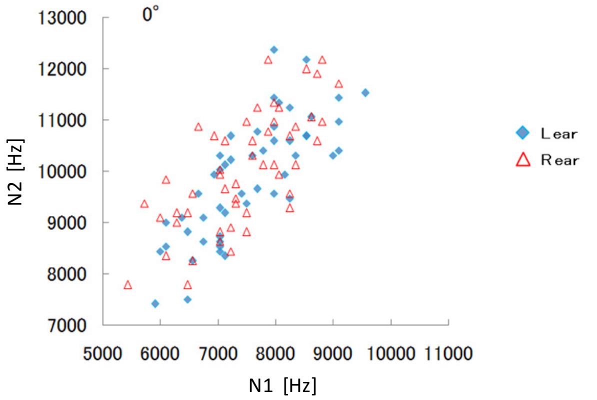


Fig. 6 Distribution of N1 and N2 frequency of 50 subjects' HRTF for front direction.

# 4 An approach to individualization of HRTFs on the basis of NFD

The distribution range of N1 and N2 frequencies is divided by the values of jnd of N1 and N2 frequencies for localization. The jnd is between 0.1oct. and 0.2oct. Fig. 7 shows the extracted N1 and N2 frequencies by dividing the distribution range of them for the front direction by jnd of N1 and N2 frequencies (0.1 oct.). As for the front direction, at which front-back localization error occur frequently due to the individual difference, 38 parametric HRTFs were extracted. It could be considered that this is a required minimum amount of HRTFs for searching the appropriate HRTFs for each listener. It takes only 3 minutes to search them by the localization test.

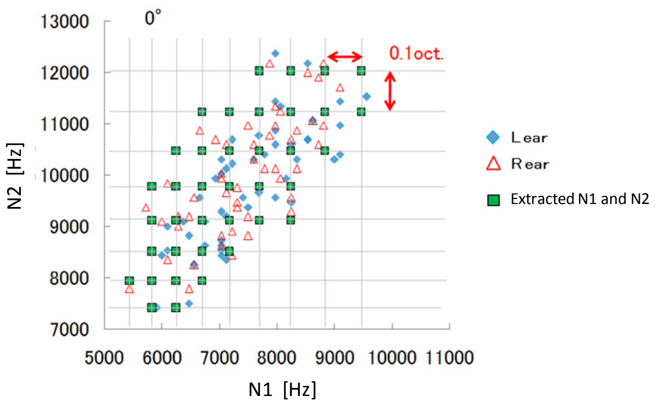


Fig. 7 Extracted N1 and N2 frequency by dividing the distribution range of them by jnd of N1 and N2 frequencies (0.1 oct.)

**References**


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

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

[3] Moore BCJ, Oldfield R, Dooley GJ "Detection and discrimination of peaks and notches at 1 and 8 kHz," J. Acoust. Soc. Am. 85: 820-836 (1989).

[4] Butle A, Belendiuk K "Spectral cues utilizes in the localization of sound in the median sagittal plane," J. Acoust. Soc. Am. 61:1264-1269 (1977).

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

[6] Iida K, Itoh M, Itagaki A, and Morimoto M "Median plane localization using parametric model of the head-related transfer function based on spectral cues," Applied Acoustics, 68: 835-850 (2007).

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