Admissible range for individualization of head-related transfer function in median plane

Masato Akagi and Hideki Hisatsune
School of Information Science, Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa 923--1292 Japan
Email: {akagi, hisatsune}@jaist.ac.jp

***Abstract*— Individualization of the head-related transfer function (HRTF) to each listener is important for the listener to localize sound sources accurately. If an inappropriate HRTF is used, the sound localization is incorrect and there is a lower sense of presence. We investigate the admissible ranges for individualization of P1, N1, and N2 frequencies, which are spectral cues of HRTFs to localize sounds, in particular, on the median plane. The admissible ranges for individualization are the valid ranges of the spectral cues on the HRTFs at which each listener can localize presented sounds. We did listening tests to estimate the admissible ranges of P1, N1 and N2. The results suggest that N1 should be accurate; the admissible range is narrow, P1 and N2 are not strict, and also that strict tuning of P1 and N2 is not required. When applying a reconstructed HRTF for a certain listener selected from those of others in the database based on the condition that the N1, N2, and P1 of those HRTFs are limited to the admissible ranges, accuracy of sound localization approaches that applying the most appropriate HRTF for the listener.**

Keywords-sound localization; HRTF; individualization; admissible range

I. INTRODUCTION

The head-related transfer function (HRTF) is the transfer function from a sound source to the head center of the listener. When we convolve HRTF into a sound wave, three-dimensional sounds can be presented to the listeners [1]. However, there is a significant problem: HRTF varies with the shape of the listener's head, body, and auricle. If an HRTF that is inappropriate for the listener is used, the sound localization is incorrect and the sense of presence decreases. To present listeners with a highly accurate sound image in the three-dimensional space, HRTF should be as accurate as possible for each listener. Thus, individualization of HRTF to each listener is an important and challenging task. Although many approaches to individualize HRTF have been reported, for example, measuring each listener's HRTF, selecting the most fitted HRTF from large HRTF databases [2], and synthesizing HRTF from measurements of head shape [3], the measuring equipment, room, and high-performance computing involved means that implementing these methods is costly.

Interaural time difference (ITD) and interaural level difference (ILD) are included in the HRTFs for the listener's two ears, and they are used for cues of sound localization on the horizontal plane. However, to judge the direction of a sound on the median plane, spectral cues also included in the HRTFs are used [4]. Hence, because of the lack of information of ITD and ILD, presentation of a sound to localize on the median plane is more difficult than localization of sound on the horizontal plane. We discuss individualization of spectral cues in the HRTF to localize sounds on the median plane.

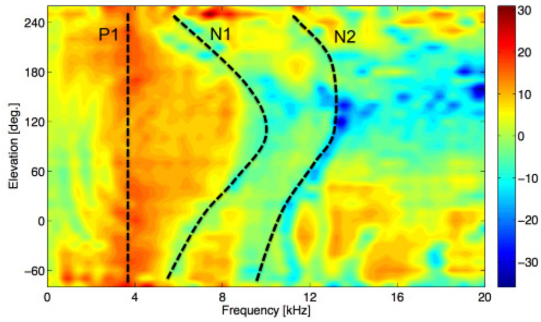


Figure 1. Example of P1, N1, and N2 on median plane

Iida et al. suggested that the sound can be localized on the median plane (elevation perception) using only N1 and N2 notches and the P1 peak in the amplitude spectrum of HRTF [5]. An example of P1, N1, and N2 on the median plane is shown in Fig. 1. This assumption was ensured by one report that describes how removing N1 and N2 by blocking the hollow of the ear of a listener leads to degradation in how accurately elevation is perceived [6]. Another report showed that resonances in the ear involve the origin of peaks and notches [3]. However, there is still less discussion on how accurately an individualized HRTF should be fitted to each listener.

We discuss admissible ranges for P1, N1, and N2, i.e., the ranges for which each listener can localize presented sounds. If we can estimate the admissible ranges for P1, N1, and N2, this can provide new knowledge of individualization of the HRTF on the median plane. To do this, we did listening experiments to select the n-best HRTFs in the database and analyze the variances of P1, N1, and N2 of the selected ones. We compare them with the estimated variances of P1, N1, and N2 of HRTFs in the database. Additionally, we reconstruct a HRTF for a certain listener by selecting from other listeners' HRTFs in the database when the selected HRTFs are in the admissible ranges. Finally, we compare the accuracy of sound localization when using the reconstructed

TABLE I. AVERAGES OF STANDARD DEVIATIONS OF P1, N1 AND N2 (ERB) IN DATABASE

	P1	N1	N2
S.D. (ERB)	0.98	0.91	0.89

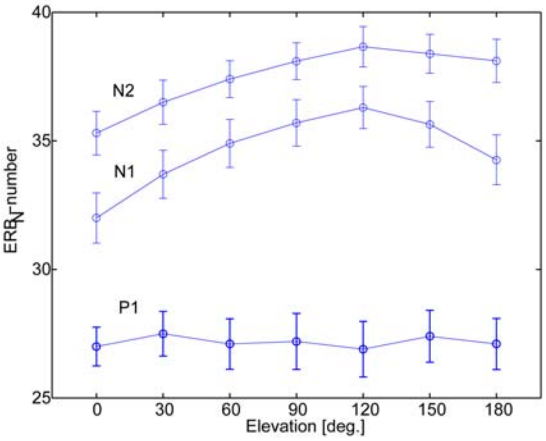


Figure 2. Averages and standard deviations of P1, N1, and N of upper-hemispheric median plane (front: 0 deg to back: 180 deg, 30 deg steps).

HRTF with that using the most appropriate one for the listener.

II. VARIANCES OF P1, N1, AND N2

This section prepares the estimated results of variances of the P1, N1, and N2 in the HRTF database.

A. Analysis method and conditions

To estimate the P1 frequency, we applied the cepstral smoothing to HRTF. Since N1 and N2 are sometimes not clear for some listeners and at some elevations, to estimate N1 and N2 frequencies more precisely, we extracted the initial impulse response (approximately 1-2 [ms] long) and applied FFT to the extracted wave data because of the reducing effects of reflections in the ears [7]. Here, 114 HRTF (of the right ear) data obtained in an anechoic chamber at Tohoku University were used to individualize HRTFs for each listener. Frequencies (Hz) of P1, N1, and N2 were transformed into *ERB_N-number* using Eq. (1) and taking the frequency resolution of human ears into consideration.

$$ERB_N - number = 21.4 \log_{10}\left(\frac{4.37f}{1000} + 1\right) \qquad (1)$$

B. Analyzed results

Averages and standard deviations of the analyzed results of P1, N1, and N2 frequencies on elevation (front: 0 degree to back: 180 degree, 30 degree steps) are shown in Figure 2, and averages of standard deviations on elevations are listed in Table 1.

The figure and table show the following tendencies.

(1) The frequency of P1 is almost constant even when elevation changes.

(2) The frequency of N1 is rising in 0-120 degree then decreasing to 180 degree

(3) The frequency of N2 is rising in almost the same manner as N1 in 0-120 degree then decreasing gradually to 180 degree

(4) Variations of P1, N1, and N2 frequencies are about 1 ERB at every elevation

These results are similar to those of the previous study [8]. This similarity ensures that these analyzed results are valid.

III. ADMISSIBLE RANGES OF P1, N1, AND N2 FREQUENCIES

The results of listening experiments to estimate the admissible ranges of P1, N1, and N2 are described and the results of the listening experiments with the variances estimated in section 2 are compared to discuss how narrow the admissible ranges of P1, N1, and N2 are.

A. Listening experiment

1) Procedure: The listening experiments enabled us to choose the best HRTFs for each subject from the database, by comparing 114 pieces of HRTF data. To choose the best HRTFs accurately, we did the following three experiments.

Experiment I: The subjects selected the n-better ($n > 5$) HRTFs by comparing the sense of sound localization.

Experiment II: The subjects selected the 5 best HRTFs by comparing them extracted in Experiment I. The evaluation procedure is the same as that of Experiment I.

Experiment III: The subjects selected the best HRTF by comparing the 5-best HRTFs.

The experimental system is illustrated in Fig. 3. The listening experiments were carried out in a sound-proof room by using Tucker-Davis Technologies (TDT) System III. The stimuli were presented to each subject through headphones (STAX SR-404).

We used white noise convoluted with the HRTFs in the database as the stimuli. The sampling frequency of the stimuli was 48 [kHz]. Target elevations of the stimuli were seven directions from 0 degree (front) to 180 degree (back) via 90 degree (right above) in 30 degree steps on the median plane. The seven stimuli were presented from front to back in order. Inter stimulus interval of them was 5 [s]. In Experiments I and II, the sound for each elevation was presented in 3 [s]. In Experiment III, the sound for each elevation was presented in 1.5 [s]. In Experiments I and II, the subjects answered how confidently they perceived the direction as indicated at each elevation. In Experiment III, the subjects listened to paired stimuli and answered which one they perceived more accurately.

Eight graduate students who had normal hearing participated in Experiment I. Since three of the eight subjects could not respond stably, five subjects were chosen for Experiments II and III.

2) Results and discussion: About ten HRTFs are selected by each subject, who were able to localize the stimuli produced by the selected HRTFs in Experiment I.

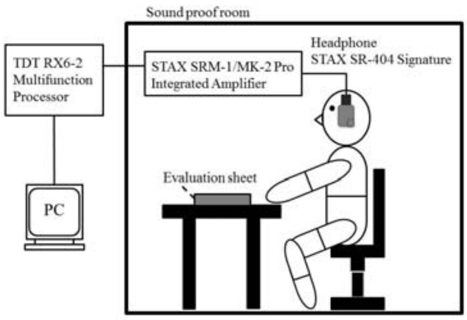


Figure 3. System for experiments.

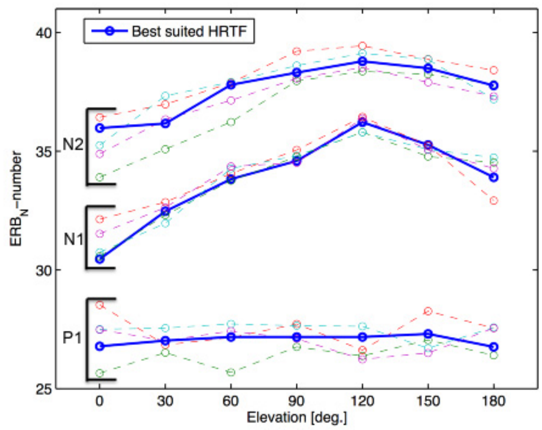


Figure 4. Selected five HRTFs of P1, P2 and N2 (Subject MA): solid line indicates best HRTF for subject.

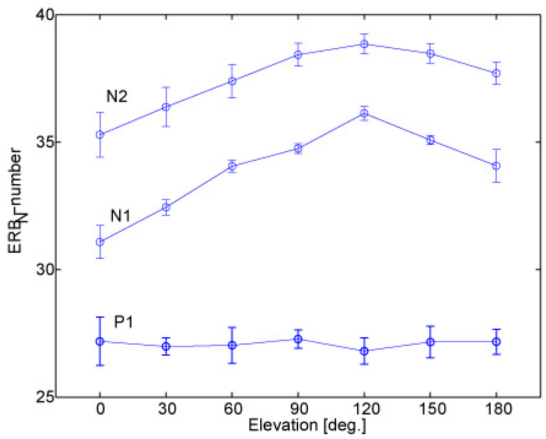


Figure 5. Admissible range of subject MA

According to subjects' introspection, they were unable to localize the sounds or it was difficult for them to do this when using the other (approximately 100) HRTFs. The 5 best HRTFs in those selected in Experiment I were chosen

TABLE II. AVERAGES OF STANDARD DEVIATIONS OF P1, N1, AND N2 (ERB)

	P1	N1	N2
S.D. (Subject MA)	0.57	0.35	0.56
S.D. (Av. among 5 subjects)	0.66	0.41	0.51
S.D. (Same as in Table 1)	0.98	0.91	0.89

for Experiment II. The best HRTF is determined from the 5 best HRTFs in Experiment III. The subjects can easily choose the 5 best and the best HRTFs. The result obtained by the subject MA is shown in Fig. 4, for example. In Fig. 4, the thick solid line indicates the best HRTF and the dotted lines are the 5 best HRTFs of the subject MA.

The averaged standard deviation of the 5 best HRTFs on the elevation is listed at the first row in Table 2. The table shows that standard deviation of the 5 best HRTFs for N1 is small at every elevation and standard deviations of P1 and N2 are not significantly small when we compare them to those of the HRTFs in the database (third row in Table 2). The averages of the standard deviations among the five subjects (second row in Table 2) have the same tendency as that of the subject MA. We assume that the admissible range is $\mu \pm \sigma$ (μ: mean, σ: SD). The admissible range of the subject MA is shown in Fig. 5. The admissible range for N1 is narrow at all elevations, and the admissible ranges for P1 and N2 were not strict. These findings should be considered when individualizing HRTF.

IV. EVALUATION OF ADMISSIBLE RANGE

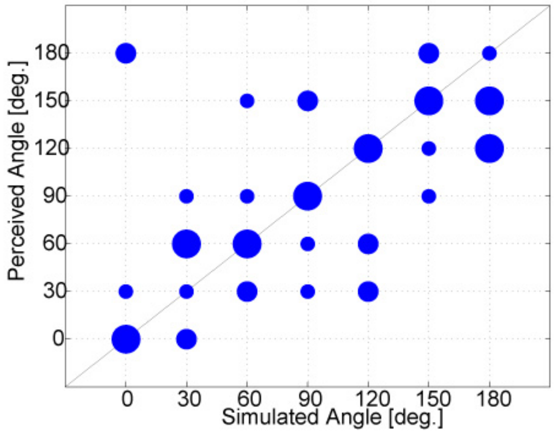
For Experiment IV, to evaluate the estimated admissible ranges, we reconstructed an HRTF for a certain listener in accordance with the admissible ranges and compared the accuracy of sound localization by using the reconstructed HRTF with that using the most appropriate one for the listener.

A. Procedure

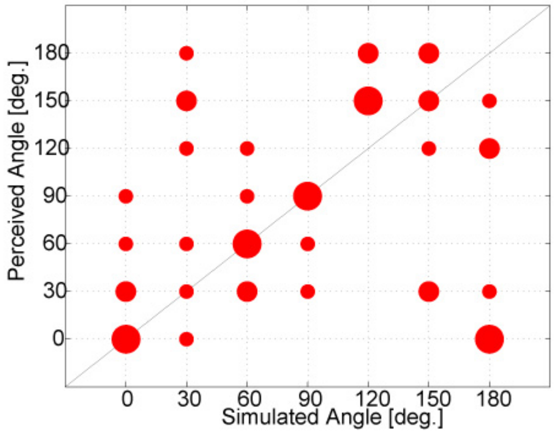
The procedure for Experiment IV is the same as that for Experiments I, II, and III with the exception of the HRTFs used and the subjects' task.

The most appropriate HRTF is the best HRTF for each subject selected in Experiment III and the reconstructed HRTF was designed as follows. An HRTF for each elevation was selected from unselected HRTFs in Experiment I and all selected HRTFs along elevations were combined as a reconstructed HRTF for a certain listener, in which P1, N1, and N2 frequencies of the selected HRTFs are in the admissible ranges of P1, N1, and N2. The reconstructed HRTF is used to produce the sounds presented to each subject. Frequency distances of P1, N1, and N2 between the reconstructed HRTF and the most appropriate one were as follows: P1, 0.10-0.62; N1, 0.05-0.39; and N2, 0.11-0.45 [ERB].

The subjects' task was to state the perceived direction of the randomly presented white noise at an elevation from 0 to 180 degree in 30 degree steps.



(a)



(b)

Figure 6. Performance of subject MA for sound localization, (a) using best HRTF for subject MA, (b) using reconstructed HRTF in accordance with subject MA's admissible ranges

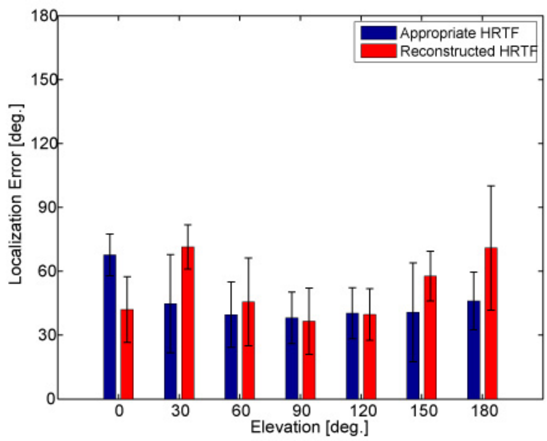


Figure 7. Localization errors using appropriate and reconstructed HRTFs

B. Results and discussion

Responses of one subject (MA) are shown in Fig. 6, and Figure 7 shows errors of perceived directions of five subjects, using the appropriate and reconstructed HRTFs for sound localization done in Experiment IV. Although front-back errors occur more when using the reconstructed HRTF, accuracies of perceived directions are similar among the reconstructed and appropriate HRTFs. This result suggests that localizable sounds can be designed even when using the reconstructed HRTF, if P1, N1, and N2 frequencies of the HRTF can be set in each admissible range. Since the admissible range of N1 is narrow and those of P1 and N2 are not, we should carefully control the N1 frequency.

V. CONCLUSION

We investigated the admissible ranges for P1, N1, and N2 frequencies, which are spectral cues for sound localization on the median plane. Although standard deviations of P1, N1, and N2 frequencies are about 1 ERB, the admissible ranges of them are narrow, in particular, the admissible range of N1 frequency is narrower than others. If P1, N1, and N2 frequencies of an HRTF are in each admissible range, performance of sound localization when using such HRTF is similar to that using the appropriate HRTF. Thus, when individualizing an HRTF to a certain listener, N1 frequency should be taken to be as accurate as possible because the admissible range of N1 frequency is narrow and we should be able to localize sounds in the same way as when using the appropriate HRTF.

ACKNOWLEDGMENT

A part of this work was supported by the JSPS A3 foresight program.

REFERENCES

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

[2] Y. Iwaya, "Individualization of head-related transfer functions with tournament-style listening test: Listening with other's ears," Acoust. Sci. and Tech., 27, 6, 340-343, 2006.

[3] H. Takemoto, P. Mokhtari, H. Kato, R. Nishimura, and K. Iida, "Basic investigation for effects of pinna shapes on head related transfer functions," Proc. Autum Meeting of the Acoustical Society of Japan, 1445-1448, 2009.

[4] J. Blauert, "Spatial Hearing," MIT Press, Cambridge, MA, 1997.

[5] K. Iida, M. Itoh, A. Itagaki, and M. Morimoto, "Median plane localization using parametric model of the head-related transfer function based on spectral cues," Applied Acoustics, 68, 8, 835--850, 2007.

[6] K. Iida, M. Yairi, and M. Morimoto, "Role of pinna cavities in median plane localization," Proc.16th International Congress on Acoustics (Seattle), 103, 5, 845-846, 1998.

[7] K. Iida, N. Gamoh, and Y. Ishii, "A consideration on a method to detect the 1st and 2nd spectral noches in the head-related transfer function," Proc. Fall Meeting of the Acoustical Society of Japan, 473-476, 2010.

[8] Y. Ishii, S. Nishioka, and K. Iida, "Consideration on the individual difference of the pinna shape and the spectral notches on the median plane - Construction of the HRTF database with the quantitative individual difference information -," Proc. Fall Meeting of the Acoustical Society of Japan, 463-466, 2012.