

ABS-1018

## Generation of the individualized HRTFs in the upper hemisphere using Parametric Notch-Peak HRTF model

Kazuhiro IIDA<sup>1</sup>; Fuka NAKAMURA<sup>2</sup>

<sup>1</sup> Chiba Institute of Technology, Japan

<sup>2</sup> Chiba Institute of Technology, Japan

### ABSTRACT

In order to realize three-dimensional sound image localization and acoustic VR, methods for generating head-related transfer functions (HRTFs) for individual listeners have been actively studied. In the present paper, we propose a novel method that generates the individualized HRTFs for an arbitrary direction in the upper hemisphere using the Parametric Notch-Peak HRTF model (PNP model) for the zenith direction, which can continuously change the parameters of notches and peaks corresponding to individual differences in HRTFs. Results of the sound image localization tests showed that the individualized HRTFs provided accurate sound image localization in the horizontal plane and the transverse plane. In the median plane, the individualized HRTFs provided accurate sound image localization for the front, zenith, and rear directions, however, there is still room for improvement in localization accuracy for the vertical angles of 30°, 60°, and 120°. In addition, we have developed a toolkit that generates individualized HRTFs based on the proposed method.

Keywords: Head-Related Transfer Function, Individualization, Sound Image Localization

### 1. INTRODUCTION

Methods for obtaining individualized HRTFs for an unknown listener that do not require acoustical measurements can be roughly divided into the following two approaches:

- (1) Select a suitable HRTF from an HRTF database.
- (2) Generate an individual HRTF from the listener's pinna shape.

In approach (1), the larger the database, the higher the probability that HRTFs that are suitable for the listener can be selected. However, the time and effort required for the selection process, such as listening tests, increases as the number of HRTFs included in the database increases. In order to solve this problem, a method by which to reduce the total number of listening tests has been studied [1-3].

In approach (2), two methods have been proposed for generation of the amplitude spectra of the individual HRTFs. One is a method that decomposes the amplitude spectrum of the HRTF into several principal components and synthesizes the HRTF using some of the components with weighting coefficients [4,5]. The weighting coefficients depend on both the listener and the direction of a sound source. They have been estimated based on the anthropometry of the listener's pinnae using multiple regression analysis [6] or using a deep neural network [7]. However, the estimation of the weighting coefficients for an unknown listener has not been successful.

Another method for HRTF individualization estimates the prominent spectral peaks and notches in the individual HRTFs. The minimum HRTF components, which provide approximately the same localization performance as the measured HRTFs, were demonstrated to be the two lowest-frequency notches (N1 and N2) and the two lowest-frequency peaks (P1 and P2) above 4 kHz [8,9].

Each notch and peak (hereinafter referred to as N/P) can be determined by three parameters: center frequency, level, and Q factor by using a peaking filter. Therefore, the generation of individual HRTFs results in the problem of how to set the N/P parameters for each listener and for each direction.

<sup>1</sup> kazuhiro.iida@it-chiba.ac.jp

Of these parameters, the frequencies of the N/P for a sound source at the front direction were reported to be estimated based on the anthropometry of the listener’s pinnae using multiple regression analysis [10-12]. However, estimation of the level and Q factor has not been successful.

In the present paper, we propose a novel method that generates the individualized HRTFs for an arbitrary direction in the upper hemisphere using the Parametric Notch-Peak HRTF model (PNP model) for the zenith direction, which can continuously change the parameters of notches and peaks corresponding to individual differences in HRTFs.

## 2. GENERATION OF INDIVIDUALIZED HRTFS IN THE UPPER HEMISPHERE

### 2.1 PNP HRTF models for the front, zenith, and rear directions

We consider adaptation of the parameters of an HRTF, which is constructed with two notches (N1 and N2) and two peaks (P1 and P2), to a listener. This is an optimization problem in 12 dimensions (three parameters multiplied by four N/Ps) and is not easy to adjust by a listener him/herself.

In order to solve this problem, we have proposed the PNP model [13], which reduces the number of independent parameters, and which can change the parameters continuously to correspond to individual differences in HRTFs by a listener. The frequency of each N/P is expressed by regression equations with the N2 frequency as an independent variable. For the level and Q factor of each N/P, the value averaged over 20 HRTFs was used as the common constant value among listeners.

### 2.2 Generation of individualized HRTF for the zenith direction

Here, we describe the four steps for individualization of the HRTF for the zenith direction. A listener adjusts the N2 frequency while listening through headphones. The user interface of the PNP model is shown in Figure 1.

Step 1: Use the “±500 Hz”, “±100 Hz”, or “±10 Hz” buttons or the slider to adjust the N2 frequency so that a sound image is localized in the zenith direction.

Step 2: The “Move” button provides a stimulus with ITD and ILD added. Use this button to confirm that the sound image moves from the left side to the right side through the zenith.

Step 3: Adjust the P1 level with the “+1 dB” and “-1 dB” buttons. For the P1 level, the following tendencies were observed [13]. The vertical angle of a sound image rises as the P1 level increases or decreases to the horizontal plane as the P1 level decreases. However, if the P1 level is lowered too much, a sound image is localized close to the head or inside the head.

Step 4: The “save” button outputs the individualized HRTF information as head-related impulse responses (HRIRs), together with the parameters of four peaking filters.



Figure 1 GUI of the PNP HRTF model for the zenith direction.

### 2.3 Generation of individualized HRTFs for the front and rear directions

#### 2.3.1 Estimation of the N2 frequencies of the front and rear directions

In order to estimate the N2 frequencies of the front and rear directions, regression analyses were performed with the N2 frequency of the zenith direction as an independent variable and the N2 frequencies of the front and rear directions as dependent variables, using HRTFs for 20 ears. The N2 frequencies of the front and rear directions tend to increase as the N2 frequency of the zenith direction increases. The correlation coefficients between N2 frequency of the zenith and the front and between

N2 frequency of the zenith and the rear were 0.71 and 0.65, respectively.

Figure 2 shows the relation between the measured and the estimated N2 frequencies of the front and rear directions. Broken lines denote the just noticeable differences (JNDs) of the N2 frequency. Here, the JND is regarded to be 0.15 octaves because the JND was reported to range from 0.1 to 0.2 octaves [14]. The residual errors for almost of the ears were under the JND, however, the residual errors for one and two ears exceeded the JND for the front and rear directions, respectively. The mean absolute residual errors for the front and rear directions were 0.06 and 0.07 octaves, respectively.

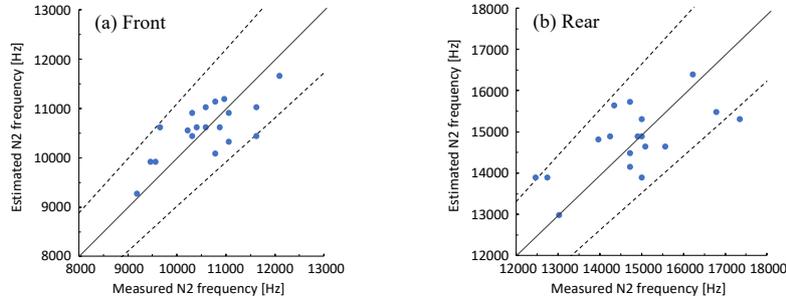


Figure 2 Relationship between the measured and the estimated N2 frequencies of the front and rear directions. Broken lines denote the JNDs of the N2 frequency.

### 2.3.2 Estimation of other N/P parameters

Frequencies of N1, P1 and P2 for the front and rear directions were estimated by PNP models for the front and rear directions using the N2 frequency obtained in Sec. 2.3.1. As mentioned above, for the level and Q factor of each N/P, the value averaged over 20 HRTFs was used as the common constant value among listeners.

## 2.4 Generation of individualized HRTFs in the median plane

The frequencies of N1 and N2 strongly depend on the vertical angle of the sound source (Figure 3) [15]. The N1 frequency increases with increasing vertical angle of the sound source from the front direction to the above direction and then decreases toward rear direction. The N2 frequency increases with increasing vertical angle from the front direction to above direction, whereas the range of the change in frequency between the above direction and rear direction is small. On the other hand, the frequencies of P1 and P2 are almost constant, independent of the vertical angle.

Based on these findings, the parameters of each N/P for the arbitrary directions in the median plane were obtained by interpolation using the parameters of N/Ps of the front, zenith, and rear directions (Figure 4).

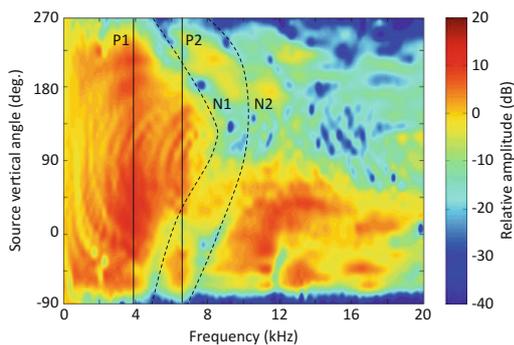


Figure 3 Relationship between vertical angle of a sound source and frequencies of N/Ps [2].

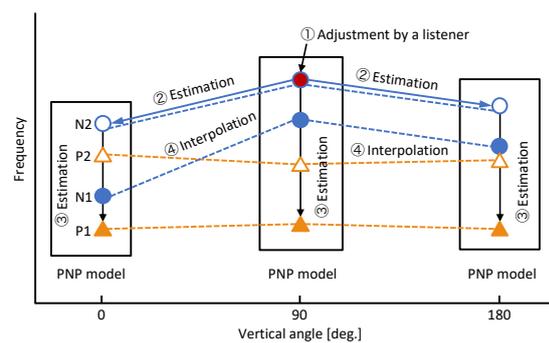


Figure 4 Schematic diagram for estimation of N/P frequencies of individualized HRTFs in the median plane.

## 2.5 Generation of individualized HRTFs in the upper hemisphere

Previous studies have shown that sound image localization in an arbitrary three-dimensional direction can be achieved by adding the interaural difference cues to the spectral cues in the median plane [16]. Based on these findings, the individualized HRTF for an arbitrary direction in the upper hemisphere were obtained by adding the interaural time difference (ITD) and the interaural level difference (ILD) to the individual HRTF in the median plane (Figure 5).

ITDs in the front half of the horizontal plane in four directions (lateral angles  $\alpha$ : 0, 30, 60, and 90°) were obtained from the measured HRIRs of 18 Japanese adult subjects. Since an approximately linear relationship was observed between the ITD and the lateral angle of a sound image, the ITD was obtained by (1).

$$ITD = 0.0078\alpha \text{ [ms]} \quad (1)$$

The ILD varies with both the incident azimuth angle and the frequency of a sound. The results of the experiments on the relationship between the ILD and the lateral localization for a wide-band noise showed that an approximately linear relationship was observed between the ILD and the lateral angle of a sound image [17]. Then, the ILD was obtained by (2).

$$ILD = 9 \times \alpha / 90 \text{ [dB]} \quad (2)$$

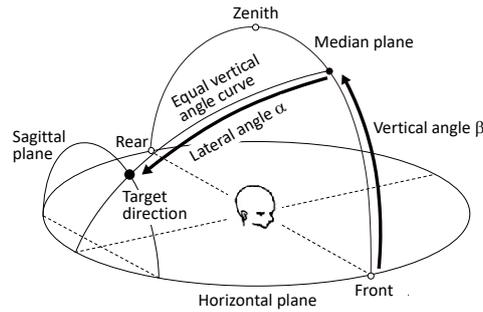


Figure 5 Sound image localization for arbitrary directions in the upper hemisphere using the individualized HRTFs in the median plane and interaural differences.

### 3. LOCALIZATION PERFORMANCE OF INDIVIDUALIZED HRTFS

Localization tests were carried out using the individualized HRTFs generated by the PNP model. The target vertical angles were seven directions (30° steps) in the right half of the horizontal plane, seven directions in the upper median plane, and seven directions in the upper transverse plane.

The source signal was wideband white noise. The stimuli, which were obtained by convolution of the sound source and the individualized HRTFs, were presented to the subjects through free-air equivalent coupling to the ear headphones (beyerdynamic DT990 PRO) [18]. No compensation of the headphone transfer functions was performed. The mapping method was adopted as a response method. Two subjects participated in the sound localization tests.

#### 3.1 Localization in the horizontal plane

Figure 6 shows the responded azimuth angles to the individualized HRTFs in the horizontal plane. For both subjects, the responses were distributed around the target azimuth angles. For subject 2, however, the responses distributed around 30°-120° for the target azimuth angle of 60°.

Figure 7 shows the responded elevation angles. For subject 1, the responses were distributed around the target elevation angle (0°) for the target azimuth angles of 90°-180°. However, the responses shifted slightly upward for the target azimuth angles of 0°-60°. For subject 2, the responses were distributed around the target elevation angle.

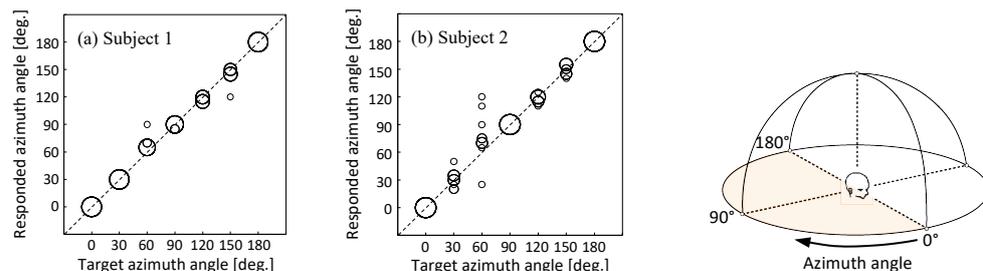


Figure 6 Responded azimuth angle to the individualized HRTFs in the horizontal plane.

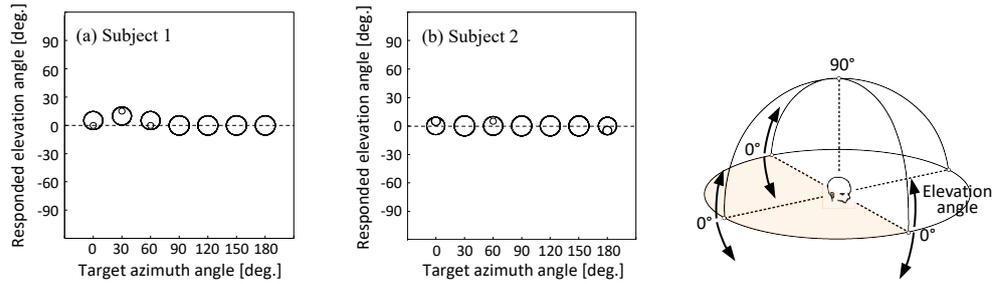


Figure 7 Responded elevation angle to the individualized HRTFs in the horizontal plane.

### 3.2 Localization in the median plane

Figure 8 shows the responded vertical angles to the individualized HRTFs in the median plane. For subject 1, the responses were distributed around the target vertical angles for the target azimuth angles of  $0^\circ$ ,  $90^\circ$ ,  $150^\circ$ , and  $180^\circ$ . However, the responses tended to shift to upward for target vertical angles of  $30^\circ$  and  $60^\circ$ . For subject 2, the responses were distributed around the target vertical angles for the target azimuth angles of  $0^\circ$ ,  $90^\circ$ , and  $180^\circ$ . However, the responses were widely distributed for the target vertical angles of  $30^\circ$ ,  $60^\circ$ , and  $120^\circ$ .

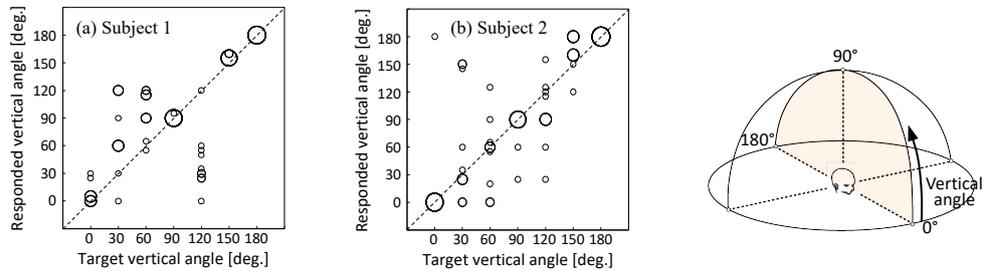


Figure 8 Responded vertical angle to the individualized HRTFs in the median plane.

### 3.3 Localization in the transverse plane

Figure 9 shows the responded lateral angles to the individualized HRTFs in the transverse plane. For both subjects, the responses distributed around the target lateral angles. For subject 2, however, the responses shifted to the median plane for the target lateral angle of  $-60^\circ$ .

Figure 10 shows the responded vertical angles to the individualized HRTFs in the transverse plane. For both subjects, most of the responses were distributed around the target vertical angle ( $90^\circ$ ).

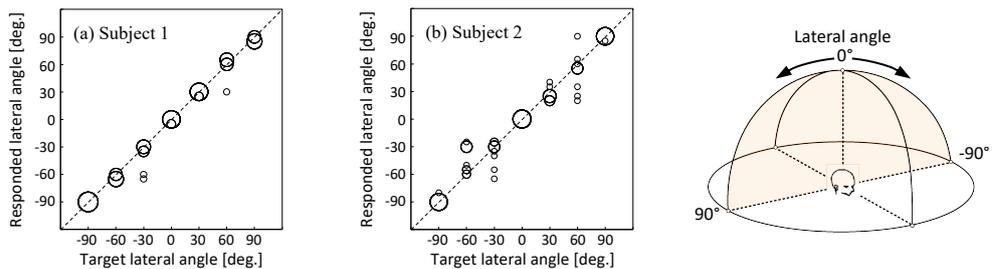


Figure 9 Responded lateral angle to the individualized HRTFs in the transverse plane.

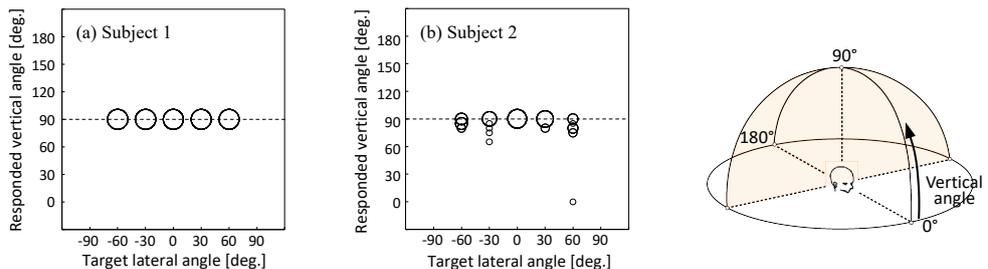


Figure 10 Responded vertical angle to the individualized HRTFs in the transverse plane.

## 4. DISCUSSIONS

For subject 1, the responses to the individualized HRTF for the target vertical angle of  $120^\circ$  in the median plane distributed around  $30^\circ$  (Figure 8(a)). Here, we discuss a possible reason for this.

Figure 11 shows the amplitude spectra of the individualized HRTF for the vertical angle of  $120^\circ$  and the measured subjects own HRTFs for  $120^\circ$  and  $30^\circ$ . This figure indicates that the N2 frequency of the individualized HRTF for  $120^\circ$  was approximately the same as that of the measured own HRTF for  $120^\circ$ . However, the N1 frequency of the individualized HRTF for  $120^\circ$  did not coincide with that of the measured own HRTF for  $120^\circ$ . In fact, the N1 frequency of the individualized HRTF for  $120^\circ$  coincided with that of the measured own HRTF for  $30^\circ$ . This is considered to be a reason for the response distribution around  $30^\circ$  for the individualized HRTF for  $120^\circ$ .

In general, N1 frequency for the vertical angle of  $120^\circ$  tends to be higher than that for  $90^\circ$ , as shown in Figure 3. However, the proposed individualization method did not reflect this tendency because the proposed method estimates the N1 frequency for  $120^\circ$  by linear interpolation between N1 frequencies of  $90^\circ$  and  $180^\circ$  (Figure 4). A study for an improved method, which reflects the detail of the vertical angle dependence of notch frequencies, is underway.

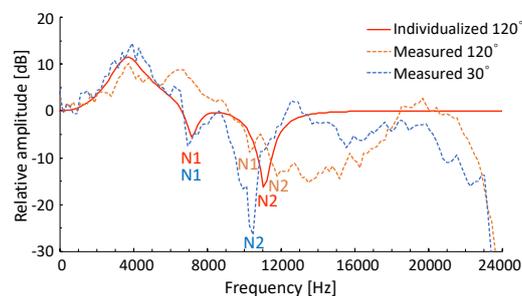


Figure 11 Comparison of the notch frequencies among the individualized HRTF ( $120^\circ$ ) and the measured subject's own HRTFs ( $120^\circ$  and  $30^\circ$ ).

## 5. MULTI-TRACK 3D AUDIO RENDERING TOOLKIT (orion)

We have developed the orion 48-track 3D audio rendering toolkit, which generates individualized HRIRs in arbitrary directions in the upper hemisphere using the PNP HRTF model and convolves the individualized HRIRs with the sound source signal using MATLAB<sup>®</sup> (Figure 12).

An outline of the rendering process is as follows:

- 1) Generate the individualized HRTFs for the zenith direction using the PNP model.
- 2) Obtain the regression equations of N/P parameters in the median plane as the explanation variable of the vertical angle, using the parameters of individualized HRTFs for the zenith direction.
- 3) Set the azimuth angle  $\phi$  and elevation angle  $\theta$ , and relative sound pressure level for each sound source. The azimuth and elevation angles are converted into the lateral angle  $\alpha$  and vertical angle  $\beta$ , respectively.
- 4) Calculate the N/P parameters for the vertical angle of each sound source using the regression equations obtained in 2). The parameters are set in the peaking filter to generate the individualized HRIRs of the vertical angle.
- 5) Add ITD and ILD to the individualized HRIRs corresponding to the lateral angle of each sound source.
- 6) Play the sound created by convolution of each sound source and each individualized HRIR.

## 6. MULTI-TRACK 3D AUDIO CONTENTS

Multi-track 3D audio contents were created using the orion 3D rendering toolkit. Till Eulenspiegel's Merry Pranks by Richard Strauss was rendered. The setting of the instruments for it is shown in Figure 13. The listener sits at the origin of the plan and elevation view. Individualized HRTFs for a female were prepared. Anyone can listen to the 3D rendered audio contents from the following URL: <https://youtu.be/ZTDOhZDkek4>.

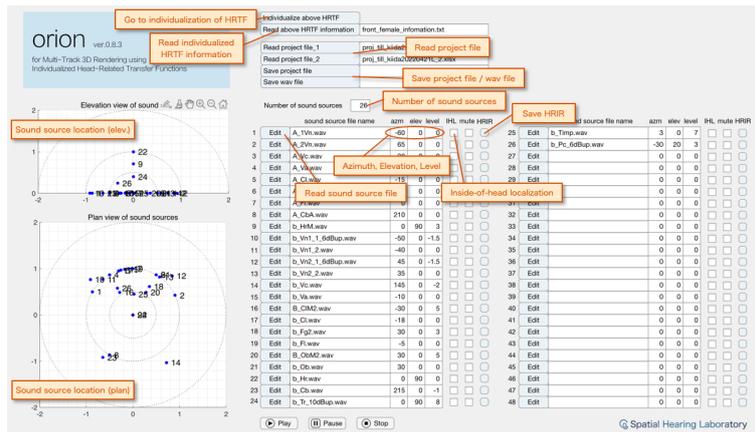


Figure 12 GUI of the orion multi-track 3D rendering using individualized HRTFs.

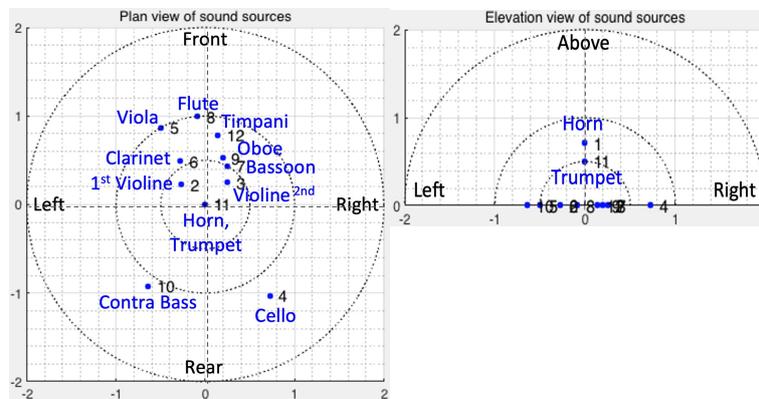


Figure 13 Setting of the instruments for 3D rendering of *Till Eulenspiegel's Merry Pranks* by Richard Strauss.

## 7. FINAL REMARKS

In the present paper, we proposed a method that estimates the notch frequencies for the front and rear directions from the notch frequencies individualized in the zenith direction using the PNP HRTF model and generates individualized HRTFs for an arbitrary direction in the upper hemisphere.

Results of the sound image localization tests showed that the individualized HRTFs generated by the PNP model provide accurate sound image localization in the horizontal plane and the transverse plane. In the median plane, the individualized HRTFs provide accurate sound image localization for the front, zenith, and rear directions, however, there is still room for improvement in localization accuracy for the vertical angles of  $30^\circ$ ,  $60^\circ$ , and  $120^\circ$ . A study for an improved method, which reflects the detail of the vertical angle dependence of notch frequencies, is underway.

The orion 48-track 3D audio rendering toolkit, which generates individualized HRTFs in arbitrary directions in the upper hemisphere using the PNP model, was developed and multi-track 3D audio contents were created using the orion. For trial use of the orion, contact the authors by email.

## ACKNOWLEDGEMENTS

A part of the present study was supported in part by MEXT KAKENHI through a Grant-in-Aid for Scientific Research (C) (Grant Number 19K12068).

## REFERENCES

1. Seeber BU, Fastl H. Subjective selection of non-individual head-related transfer functions. Proc 2003 ICAD; 6-9 July 2003; Boston, USA 2003.

2. Iwaya Y. Individualization of head-related transfer functions with tournament-style listening test: listening with other's ears. *Acoust Sci Technol*. 2006;27(6):340-3.
3. Iida K. *Head-Related Transfer Function and Acoustic Virtual Reality*. Springer; 2019. p. 76-82.
4. Kistler DJ, Wightman FL. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *J Acoust Soc Am*. 1992;91(3):1637-47.
5. Middlebrooks JC, Green DM. Observations on a principal components analysis of head-related transfer functions. *J Acoust Soc Am*. 1992;92(1):597-9.
6. Bomhardt R, Braren H, Fels J. Individualization of head-related transfer functions using principal component analysis and anthropometric dimensions. *Proc meetings on acoustics*; 28 November - 02 December 2016; Honolulu, USA 2016.
7. Miccini R, Spagnol S. HRTF Individualization using Deep Learning. In *IEEE 5th VR Workshop on Sonic Interactions in Virtual Environments*; 2019.
8. Iida K, Itoh M, Itagaki A, Morimoto M. Median plane localization using parametric model of the head-related transfer function based on spectral cues. *Appl Acoust*. 2007;68:835-50.
9. Iida K, Ishii Y. Effects of adding a spectral peak generated by the second pinna resonance to a parametric model of head-related transfer functions on upper median plane sound localization. *Appl Acoust*. 2018;129:239-47.
10. Iida K, Ishii Y, Nishioka S. Personalization of head-related transfer functions in the median plane based on the anthropometry of the listener's pinnae. *J Acoust Soc Am*. 2014;136(1):317-33.
11. Spagnol S, Avanzini F. Frequency estimation of the first pinna notch in head-related transfer functions with a linear anthropometric model. *Proc DAFX-15*; Trondheim, Norway 2015.
12. Mokhtari P, Takemoto H, Nishimura R, Kato H. Frequency and amplitude estimation of the first peak of head-related transfer functions from individual pinna anthropometry. *J Acoust Soc Am*. 2015;137(2):690-701.
13. Iida K, Aizaki T, Kikuchi T. Toolkit for individualization of head-related transfer functions using parametric notch-peak model. *Appl. Acoust*. 2022;189:108610. <https://doi.org/10.1016/j.apacoust.2021.108610>.
14. Iida K. *Head-Related Transfer Function and Acoustic Virtual Reality*. Springer; 2019. p. 61-62.
15. Iida K. *Head-Related Transfer Function and Acoustic Virtual Reality*. Springer; 2019. p. 36-38.
16. Morimoto M, Iida K, Itoh M. Upper hemisphere sound localization using head-related transfer functions in the median plane and interaural differences. *Acoust Sci Technol*. 2003;24:267-75.
17. Blauert J. *Spatial hearing revised edition*. The MIT Press; 1997. p. 155-164.
18. Møller H. Fundamentals of binaural technology. *Appl Acoust*. 1992;36(3-4):171-218.