TOOLKIT FOR 3D AUDIO RENDERING USING INDIVIDUALIZED HEAD-RELATED TRANSFER FUNCTIONS IN THE UPPER HEMISPHERE

Kazuhiro IIDA

Spatial Hearing Laboratory,
Department of Advanced Media,
Chiba Institute of Technology
2-17-1 Tsudanuma,
Narashino, Chiba 275-0016, Japan
kazuhiro.iida@it-chiba.ac.jp

Fuka NAKAMURA

Spatial Hearing Laboratory,
Department of Advanced Media,
Chiba Institute of Technology
2-17-1 Tsudanuma,
Narashino, Chiba 275-0016, Japan

ABSTRACT

Although sound image control and sound field reproduction based on head-related transfer functions (HRTFs) have been studied for years, they have not been put to practical use. The main reason for this is that individual differences in HRTFs have not been overcome. In the present paper, the orion 48-track 3D audio rendering toolkit, which generates individualized HRTFs in arbitrary directions in the upper hemisphere using the parametric notch-peak (PNP) HRTF model, is proposed. The results of sound image localization tests using the generated individualized HRTFs show that the individualized HRTFs provide accurate sound image localization in the horizontal plane. However, there is still room for improvement in sound image localization in the upper directions. The present paper is accompanied by binaural audio contents generated by the proposed 3D rendering toolkit using the individualized HRTFs in the upper hemisphere.

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 [1,2].

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 [3,4]. 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 [5] or using a deep neural network [6]. 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 [7,8].

Each notch and peak (hereinafter referred to as N/P) can be determined by three parameters: center frequency, level, and sharpness (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.

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 [9,10,11] or discrimination analysis [12]. However, estimation of the level and Q factor for the individual N/P has not been successful.

In the present paper, we focus on a novel individual HRTF generation method, in which the listener adjusts the HRTF parameters while listening. However, even the minimum configuration of the HRTF, which is constructed with two notches and two peaks, requires optimization of 12 parameters (three parameters multiplied by four N/Ps), which are not easy to adjust. In order to solve this problem, we propose a parametric notch-peak HRTF model (PNP model), which reduces the number of independent parameters and which can change the parameters continuously to correspond to individual differences in HRTFs. Then, we have developed a toolkit that generates individualized HRTFs by having the listener adjust the parameters using the PNP model.

2. GENERATION OF INDIVIDUALIZED HRTFS IN THE UPPER HEMISPHERE

We consider adaptation of the parameters of an HRTF, which is constructed with two notches and two peaks, to a listener. This is an optimization problem in 12 dimensions (three parameters multiplied by four N/Ps) and is not easy to solve. In order to reduce the number of independent parameters, we tried to estimate some parameters from other parameters, or treat these parameters as constants using the PNP model.

2.1. Generation of individualized HRTFs for the front, zenith, and rear directions

The individualized HRTFs for the front, zenith, and rear directions were generated using the PNP model (Fig. 1). The PNP HRTF model is composed of two notches (N1 and N2) and two peaks (P1 and P2). In the PNP model, the frequency

https://doi.org/10.21785/icad2022.034

of each notch and peak (hereinafter referred to as 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 typical HRTFs [2] was used as the common constant value among listeners.

Figure 1: GUI of the PNP model for individualization of HRTFs.

2.2. Generation of individualized HRTFs for the arbitrary directions in the median plane

The frequencies of N1 and N2 strongly depend on the vertical angle of the sound source (Fig. 2) [2]. 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 regression analysis using the parameters of N/Ps of the front, zenith, and rear directions.

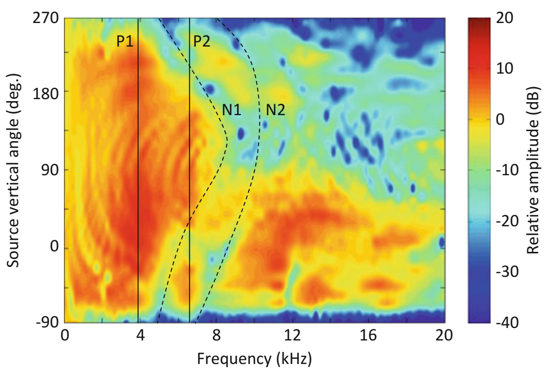

Figure 2: Relationship between vertical angle of a sound source and frequencies of N1, N2, P1, and P2 [2].

2.3. Generation of individualized HRTFs for the arbitrary directions 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 [13]. 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 (Fig. 3).

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

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

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 [14]. The ILD was obtained by (2).

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

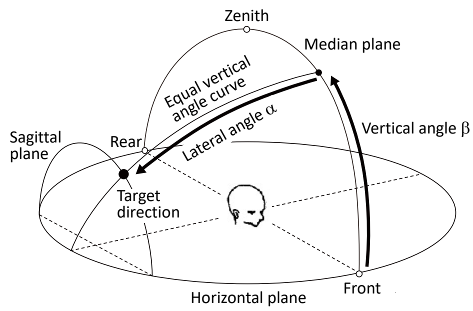

Figure 3: Sound image localization for arbitrary directions in the upper hemisphere using the individualized HRTFs in the median plane and interaural differences.

3. EXAMINATION OF PERFORMANCE OF HRTF INDIVIDUALIZATION

Localization tests were carried out using the subject's own HRTFs and 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 various HRTFs, were presented to the subjects through free-air equivalent coupling to the ear headphones (beyerdynamic DT990 PRO) [15]. 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.

Figure 4 shows a subject's responses to the subject's own HRTFs and the individualized HRTFs in the horizontal plane. For the subject's own HRTF, the responses were distributed around the diagonal line. However, the responses tended to shift to 90°–120° for the target azimuth angle of 60°. For the individualized HRTF, most of the responses were distributed around the diagonal line, although front-back confusion occurred once and twice for target azimuth angles of 0° and 30°, respectively.

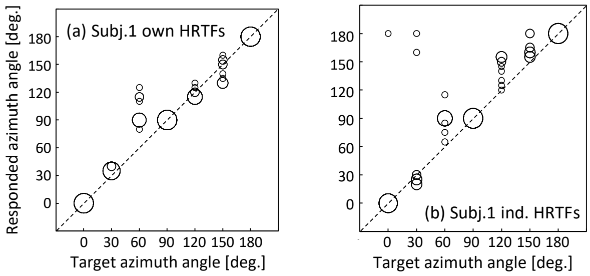

Figure 4: Responded azimuth angle to the subject's own HRTFs and individualized HRTFs in the horizontal plane.

Figure 5 shows a subject's responses to the subject's own HRTFs and the individualized HRTFs in the median plane. For the subject's own HRTF, the responses were distributed around the diagonal line. However, the responses tended to shift to upward (90°) for target vertical angles of 60° and 150°. For the individualized HRTF, the responses were distributed around the target directions for 0° and 180°. However, the responses were distributed near the horizontal plane for target vertical angles of 60° and 120°. For the target vertical angle of 90°, the responses were widely distributed.

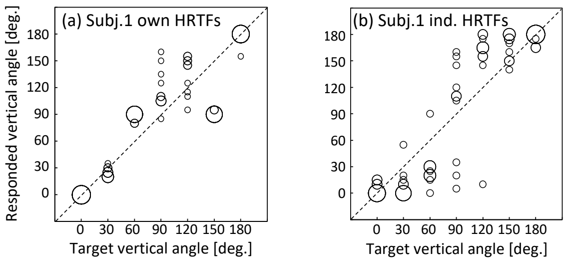

Figure 5: Responded vertical angle to the subject's own HRTFs and individualized HRTFs in the median plane.

Figure 6 shows the responded lateral angles to the subject's own HRTFs (0° and ±90°) and the individualized HRTFs in the transverse plane. The responses distributed around the target lateral angles for both the subject's own HRTFs and the individualized HRTFs. Figure 7 shows the responded vertical angles to the subject's own HRTFs (target lateral angle of 0°) and the individualized HRTFs (target lateral angles of 0°, ±30°, and ±60°) in the transverse plane. Although the target vertical angle was 90°, the responses distributed around 130-170°.

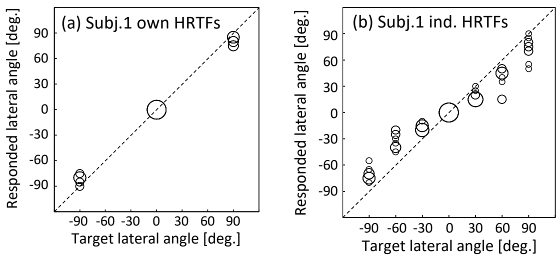

Figure 6: Responded lateral angle to the subject's own HRTFs and individualized HRTFs in the transverse plane.

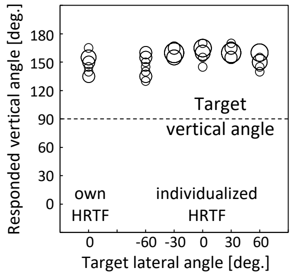

Figure 7: Responded vertical angle to the subject's own HRTFs and individualized HRTFs in the transverse plane.

4. DISCUSSION

The subjects' responses to both the subject's own HRTFs and the individualized HRTFs were distributed around the target directions for the horizontal plan. However, the accuracy of localization was low for the upper target directions in the median plane and the transverse plane. One of possible reasons is that compensation of the headphone transfer functions was not performed.

We measured the transfer functions between the headphones and the entrances of ear canals (HpTFs) using ear-plug type microphones, of which diaphragms were located at the entrances of the ear canals. Peaks and notches were observed in the HpTFs in the frequency range of N1 and N2 of the HRTFs (Fig. 8(a)). This infers that the peaks and notches of the HpTFs affect the perception of the vertical angle of a sound image.

Then, an inverse filter of HpTFs common to subjects was generated. Fig. 8 (b) shows the responded vertical angles to the subject's own HRTFs and the individualized HRTFs using the inverse filter of HpTFs in the transverse plane. The responses distributed around 90-120°. However, there is still room for improvement in sound image localization in the upper directions.

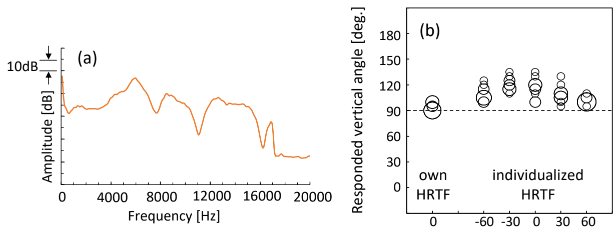

Figure 8: Measured HpTF (a) and responded vertical angle to the subject's own HRTFs and individualized HRTFs in the transverse plane using inverse HpTF (b).

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® (Fig. 9).

An outline of the rendering process is as follows:

1) Generate the individualized HRTFs for the front, zenith, and rear directions 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 front, zenith, and rear directions.

3) Set the azimuth angle ϕ and elevation angle θ, and relative sound pressure level for each sound source. The azimuth and elevation angles are converted into the lateral angle α and vertical angle β, 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.

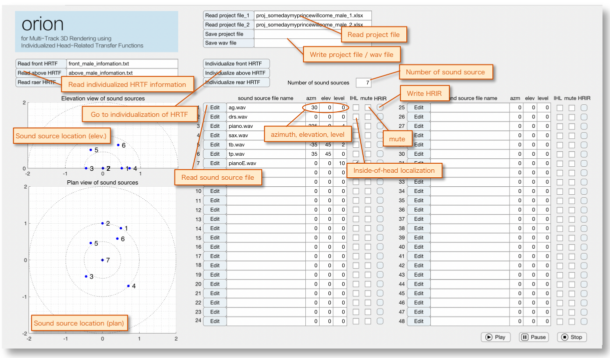

Figure 9: GUI of the orion multi-track 3D rendering using individualized HRTFs.

6. MULTI-TRACK 3D AUDIO CONTENTS

Multi-track 3D audio contents were created using the orion 3D rendering toolkit. Two kinds of music, *Till Eulenspiegel's Merry Pranks* by Richard Strauss and *Symphony No. 7, 1st mov.* by Beethoven were rendered. The setting of the instruments for *Till Eulenspiegel's Merry Pranks* is shown in Fig.10. The listener sits at the origin of the plan and elevation view. Individualized HRTFs for a female and a male were prepared. Anyone can listen to the 3D rendered audio contents from the following URL: https://youtu.be/ZTDOhZDkek4.

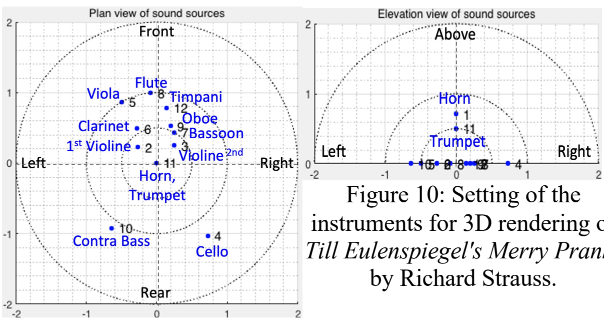

Figure 10: Setting of the instruments for 3D rendering of *Till Eulenspiegel's Merry Pranks* by Richard Strauss.

7. SUMMARY

In the present paper, a toolkit for 3D audio rendering that generates individualized HRTFs in the upper hemisphere using the PNP model has been developed. The results of sound image localization tests showed that the individualized HRTFs provided an accurate sound image localization in the horizontal plane. However, there is still room for improvement in sound image localization in the upper directions.

8. ACKNOWLEDGMENT

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

9. REFERENCES

[1] B. U. Seeber and H. Fastl, “Subjective selection of non-individual head-related transfer functions,” in *Proc. of the 2003 Int. Conf. on Auditory Display*, 2003.

[2] K. Iida, *Head-Related Transfer Function and Acoustic Virtual Reality*, Springer; 2019.

[3] D. J. Kistler and F. L. Wightman, “A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction,” *J. Acoustical Soc. Am.*, vol. 91, no. 3, pp. 1637-47, 1992.

[4] J. C. Middlebrooks and D. M. Green, “Observations on a principal components analysis of head-related transfer functions,” *J. Acoustical Soc. Am.*, vol. 92, no. 1, pp. 597-9, 1992.

[5] R. Bomhardt, H. Braren, and J. Fels, “Individualization of head-related transfer functions using principal component analysis and anthropometric dimensions,” in *Proc. of meetings on acoustics*, 2016.

[6] R. Miccini and S. Spagnol, “HRTF Individualization using Deep Learning,” in *IEEE 5th VR Workshop on Sonic Interactions in Virtual Environments*, Mar. 2020.

[7] 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,” *Appl. Acoust.* vol.68, pp. 835-50, 2007.

[8] K. Iida and Y. Ishii, “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*. vol. 129, pp. 239-47, 2018.

[9] K. Iida, Y. Ishii, and S. Nishioka, “Personalization of head-related transfer functions in the median plane based on the anthropometry of the listener's pinnae,” *J. Acoust. Soc. Am.* vol. 136, no. 1, pp. 317-33, 2014.

[10] S. Spagnol and F. Avanzini, “Frequency estimation of the first pinna notch in head-related transfer functions with a linear anthropometric model,” in *Proc. the 18th Int. conference on digital audio effects (DAFx-15)*, 2015. Trondheim, Norway.

[11] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, “Frequency and amplitude estimation of the first peak of head-related transfer functions from individual pinna anthropometry,” *J. Acoust. Soc. Am.* vol. 137, no. 2, pp. 690-701, 2015.

[12] K. Iida, O. Nishiyama, and T. Aizaki T, “Estimation of the category of notch frequency bins of the individual head-related transfer functions using the anthropometry of the listener's pinnae,” *Appl. Acoust.* 2021;177:107929. https://doi.org/ 10.1016/j.apacoust.2021.107929.

[13] M. Morimoto, K. Iida, and M. Itoh, “Upper hemisphere sound localization using head-related transfer functions in the median plane and interaural differences,” *Acoust. Sci. Technol.* vol. 24, pp. 267-75, 2003.

[14] J. Blauert, *Spatial hearing revised edition*, The MIT Press; 1997.

[15] H. Møller, “Fundamentals of binaural technology,” *Appl. Acoust.* vol. 36 pp. 171-218, 1992.