Generation of the Individual Head-Related Transfer Functions in the Upper Median Plane Based on the Anthropometry of the Listener's Pinnae

Kazuhiro Iida
Department of Advanced Media
Chiba Institute of Technology
Narashino, Chiba, Japan
kazuhiro.iida@it-chiba.ac.jp

Hikaru Shimazaki and Masato Oota
Graduate School
Chiba Institute of Technology
Narashino, Chiba, Japan

Abstract— **In order to address the individual differences in the head-related transfer functions (HRTFs) of different listeners, the individualization of the HRTF has been investigated. In the present study, multiple regression analyses were carried out as objective variables of the amplitude level of each discrete frequency of the early HRTFs in the upper median plane and as explanatory variables of fourteen anthropometric parameters of the pinnae. The results showed the potential of generation of the individual HRTFs from the listener's anthropometry of the pinnae without any HRTF database.**

Keywords— head-related transfer function; individualization; generation; pinna; anthropometry; the median plane

I. Introduction

There exist remarkable individual differences in head-related transfer functions (HRTFs). This is a serious problem, which prevents acoustic virtual reality from coming into widespread practical use. Recent studies [1, 2] showed that a parametric HRTF recomposed using only the two lowest-frequency notches and the two lowest-frequency peaks provides almost the same localization accuracy as the corresponding measured HRTF. Some methods to estimate the frequency of these notches and peaks from the anthropometry of the listener's pinna have been proposed [3–5]. However, estimation of the level and the band width of the notches and peaks has not succeeded.

The present study proposed a novel method (GAP: Generation of the Amplitude spectrum from the Pinna anthropometry), in which the amplitude level of each discrete frequency of the individual HRTF is generated from the anthropometry of the listener's pinnae without any HRTF database.

II. Pre -procrssing of HRTFs

The notches and peaks in the HRTFs are generated in the pinna. The effect of the pinnae is considered to be included in the early part of the head-related impulse response (HRIR), because the response from the pinna arrives at the entrance of the ear canal earlier than that from the torso. Iida and Oota [6] reported that the early HRIR, of which duration is 1 ms, truncated by a temporal window includes information of the outline of the spectral notches and peaks and provides approximately the same vertical angle of a sound image as those of the full-length HRIR, in the upper median plane.

In the present study, the HRIRs of 27 Japanese adult subjects (54 ears) were measured for seven vertical angles in the upper median plane (0° to 180° in 30° steps) in an anechoic chamber. The sampling frequency was 48 kHz. The vertical angle is defined as the angle measured from the front direction, with 0° indicating the front and 180° indicating the rear. The early HRTFs were obtained using the algorithm, as follows:

(1) Detect the sample for which the absolute amplitude of the HRIR is maximum, S_{max}. (2) Clip the HRIR using a four-term, 96-point Blackman-Harris window, adjusting the temporal center of the window to S_{max}. (3) The amplitude spectra of early HRTFs were obtained by the FFT with 512 samples.

Fig. 1 shows an example of the amplitude spectra of the early HRTF and the usual full-length (5.3 ms) HRTF. The early HRTF retains the outline of the spectral notches and peaks.

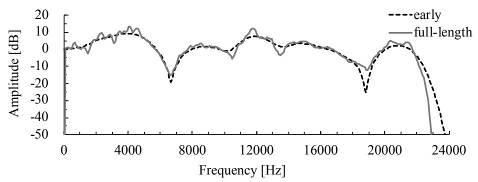

Fig. 1. Example of the amplitude spectra of the early HRTF (broken line) and the usual full-length HRTF (solid line).

III. Anthropometric parameters of the pinnae

Fourteen anthropometric parameters of the pinnae to be analysed (x_1 through x_{14}) were adopted. As shown in Fig. 2, two-dimensional coordinates of p_1 through p_{12} were obtained. The origin of the coordinate (p_0) was set at the entrance of the ear canal. C_1, C_2, and C_3 were the inner border of the helix, the antihelix, and he outer border of the concha, respectively. x_1 through x_{12} were the lengths from p_0 to p_1 through p_{12}, respectively. x_{13} was the tilt of the pinna, and x_{14} was the depth of the concha. x_1 through x_{13} were measured from a photograph and x_{14} was measured using a vernier caliper.

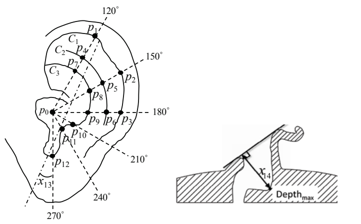

Fig. 2. Fourteen anthropometric parameters of the pinna.

IV. ESTIMATION OF HRTFS BY MULTIPLE REGRESSION ANALYSIS

Multiple regression analyses were carried out as objective variables of the amplitude level of each discrete frequency (in 93.75 Hz steps) of the early HRTFs and as explanatory variables of fourteen anthropometric parameters of the pinnae, as Eq. (1):

$$y(s, \beta, f) = \sum_{i=1}^{14} a_i(\beta, f) x_i(s) + \mathrm{b}(\beta, f) \quad (1)$$

where y, s, β, and f denote the amplitude level of the early HRTF, the subject, the vertical angle, and the discrete frequency, respectively. a, b, and x denote the regression coefficients, a constant, and the anthropometric parameters, respectively. We adopted the combinations of parameters for which the squared multiple correlation coefficient adjusted for the degrees of freedom were the highest.

Examples of the amplitude spectra of the estimated HRTFs and the measured early HRTFs are shown in Fig. 3. The estimated HRTFs have similar prominent notches and peaks to those of the measured early HRTFs. However, the estimation error was observed in the level of notches.

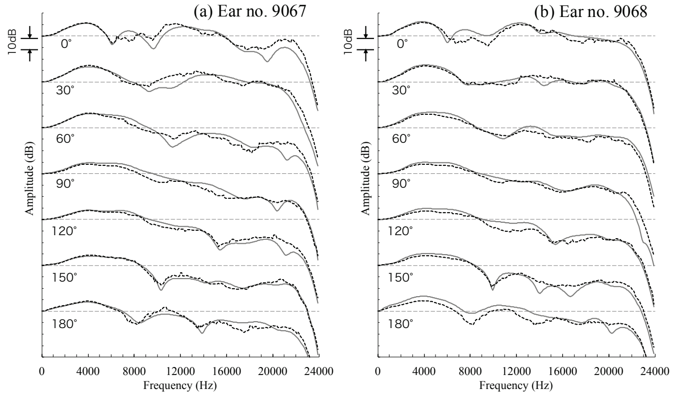

Fig. 3. Examples of the amplitude spectra of the estimated HRTFs (broken line) and the measured early HRTFs (solid line) in the upper median plane.

The multiple correlation coefficients averaged over frequency (93.75 Hz to 19,968.75 Hz) for seven vertical angles were 0.53 to 0.61. The spectral distortion (SD) between the estimated HRTFs and the measured early HRTFs (Eq. (2)) averaged over 54 ears were 3.2 to 4.2 dB.

$$SD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[20\log_{10}\left|\frac{HRTF_e(f_i)}{HRTF_m(f_i)}\right|\right]^2} \quad (2)$$

where $HRTF_e$, $HRTF_m$, and f denote the estimated HRTF, the measured early HRTF, and the discrete frequency, respectively.

V. GENERATION OF HRTFS OF NAIVE SUBJECTS

The validity of the generation of the amplitude spectra of naive subjects were examined. Multiple regression analyses were carried out using the early HRTFs and the pinna anthropometry of 24 subjects (48 ears) of 27 subjects. Then, the amplitude spectra of the three naive subjects (six ears), who were not involved in the multiple regression analysis, were generated using the pinna anthropometry and the partial regression coefficients.

Examples of the amplitude spectra of the generated HRTFs of a naive subject are shown in Fig. 4. The generated HRTFs have similar prominent notches and peaks to those of the measured early HRTFs. However, the generation accuracy decreased compared with the estimation accuracy shown in Fig. 3. The SD between the generated HRTFs and the measured early HRTFs averaged over six ears were 4.2 to 6.4 dB for seven vertical angles.

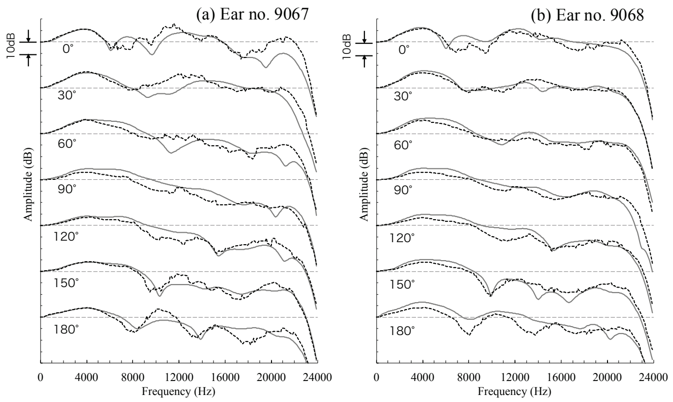

Fig. 4. Examples of the amplitude spectra of the generated HRTFs (broken line) and the measured early HRTFs (solid line) of a naive subject.

The validation of the median plane localization performance of the generated individual HRTFs for the naive subjects remains an issue to be addressed.

REFERENCES

[1] 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. 835-850, 2007.

[2] 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. 239-247, 2018.

[3] K. Iida, Y. Ishii, 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, pp. 317-333, 2014.

[4] S. Spagnol and F. Avanzini, “Frequency estimation of the first pinna notch in head–related transfer functions with a linear anthropometric model” Proc. the 18th Int. Conference on Digital Audio Effects (DAFx-15), Trondheim, Norway, 2015.

[5] K. Iida, M. Oota, and H. Shimazaki, “Estimation of the spectral notch frequency of the individual early head-related transfer function in the upper median plane based on the anthropometry of the listener's pinnae” Conference paper, the International Conference on Spatial Reproduction, Audio Eng. Soc., Tokyo, Japan, 2018.

[6] K. Iida and M. Oota, “Median plane sound localization using early head-related impulse response” Appl. Acoust., vol. 139, pp. 14-23, 2018.