Applied Acoustics 139 (2018) 14–23

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Median plane sound localization using early head-related impulse response

Kazuhiro Iida*, Masato Oota

Faculty of Advanced Engineering, Chiba Institute of Technology, 2-17-1 Tsudanuma, Narashino, Chiba 275-0016, Japan

ARTICLE INFO

Keywords:
Head-related impulse response
Head-related transfer function
Localization
Median plane
Distance

ABSTRACT

Previous studies reported that the outline of the spectral notches and peaks in the head-related transfer function (HRTF) in the frequency range above 5 kHz plays an important role in the perception of the vertical angle of a sound image. Moreover, the notches and peaks were reported to be generated in the pinna. These findings imply that the important information for the vertical localization is mostly included in the early part of the head-related impulse response (HRIR), because the response from the pinna arrives at the receiving point (the entrance of the ear canal) earlier than that from the torso. However, the duration of the HRIR required for accurate median plane localization is unclear. In the present study, we measured the HRIRs for seven target vertical angles in the upper median plane (0° to 180° in 30° steps) for five subjects and generated early HRIRs, the durations of which were 0.25, 0.5, 1, and 2 ms. We analyzed the amplitude spectra of the early HRIRs and performed two psycho-acoustical tests with regard to the vertical angle and the distance of a sound image. The results suggested that (1) the outline of the amplitude spectra of the early HRIRs of 1 and 2 ms was approximately the same as that of the full-length HRIR, whereas the outline of the amplitude spectra of the early HRIRs of 0.25 and 0.5 ms differed from that of the full-length HRIR, and (2) no statistically significant difference in the mean vertical localization error or in the scale value of the perceived sound image distance was observed between the full-length HRIR and each of the early HRIRs of 1 and 2 ms at any target vertical angle. These results suggest that the early HRIR of 1 ms includes information of the outline of the spectral notches and peaks with respect to physical aspect. Moreover, the results suggest that the early HRIR of 1 ms provides approximately the same vertical angle and distance of a sound image as the full-length HRIR in the upper median plane with respect to perceptual aspect.

1. Introduction

It is widely known that the spectral notches and peaks in the human head-related transfer functions (HRTFs) in the frequency range above 5 kHz contribute to the perception of the vertical angle of a sound image [3,11,17].

The outlines of the amplitude spectra were found to be more important than their fine structures. Asano *et al.* [1] conducted localization tests in the median plane using the HRTFs, the amplitude spectra of which were smoothed by an auto-regressive moving-average model, and reported that, rather than the information in small peaks and notches, the information in macroscopic patterns is utilized for the perception of the elevation angle. Kulkarni and Colburn [9] carried out localization tests in the horizontal plane (azimuth of −45°, 0°, 45°, and 180°). The magnitude spectra of the subjects' own HRTFs were systematically smoothed in seven levels. The results indicate that the fine spectral structure is relatively unimportant for sound localization, as compared to the outline of the notches and peaks. They also reported that the elevation of a sound image shifted upward under the extreme smoothing condition.

Moreover, among the numerous notches and peaks in the HRTF, some specific notches and peaks have been reported to play an important role in vertical localization. A parametric HRTF model, which is recomposed of all or some of the spectral notches and peaks extracted from a listener's own HRTF, has been proposed [6]. The notches and peaks are labeled in order of frequency (e.g., N1, P1, N2, P2, N3, P3, and so on), and each of the notches and peaks is expressed parametrically in terms of frequency, level, and sharpness. Sound localization tests in the upper median plane, using the parametric HRTFs of various combinations of notches and peaks, demonstrated that the minimum components, which provide approximately the same localization performance as the listener's own HRTFs, were the two lowest-frequency notches (N1 and N2) and the two lowest-frequency peaks (P1 and P2) [8].

The above findings suggest that the outline of N1, N2, P1, and P2 is important for the median plane localization.

* Corresponding author.
E-mail address: kazuhiro.iida@it-chiba.ac.jp (K. Iida).

https://doi.org/10.1016/j.apacoust.2018.03.027
Received 10 December 2017; Received in revised form 24 February 2018; Accepted 22 March 2018

The notches and peaks are reported to be generated in the pinna. When the cavities of the pinna are occluded by clay, the notches and peaks vanish [5], and the front-back confusion of a sound image increases [4,17,5]. At the notch frequency, the anti-nodes are generated at the cymba of concha and the triangular fossa, and a node is generated at the cavity of concha [22]. Peaks are considered to be generated by the resonances of the pinna [20].

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 receiving point (the entrance of the ear canal) earlier than that from the torso. Therefore, the information of the outline of N1, N2, P1, and P2 is supposed to be included in the early part of HRIR (hereinafter early HRIR).

Senova *et al.* [19] compared the localization performance of the HRIR of reduced duration truncated by a rectangular temporal window with that of the actual sound source and reported that localization performance was not disrupted dramatically until the HRIR duration was reduced to 0.64 or 0.32 ms. However, Senova *et al.* evaluated the mean localization error averaged over 354 sound source directions (in 10° azimuth steps and 10° elevation steps ranging from −40° to 70°). The minimum required duration of the HRIR is supposed to depend on the direction of the sound source because the HRTF, the spectral notches of which are deep, may require more time to form the spectral shape than the HRTF, the spectral notches of which are shallow. Thus, the required minimum duration of the median plane HRIR for accurate vertical angle perception remains unclear.

Moreover, it is unclear whether the early part of the HRIR, which provides the same vertical angle perception as the usual HRIR, also provides the same distance perception as the usual HRIR.

The purpose of the present study is to clarify the minimum required duration of the early HRIR of the median plane, for which the perception performance of vertical angle and distance of a sound image is the same as that of the usual HRIR. We measured HRIRs for seven target angles in the upper median plane for five subjects and generated four early HRIRs, the durations of which were 0.25, 0.5, 1, and 2 ms. Then, we analyzed the amplitude spectra of the early HRIRs and performed two psycho-acoustical tests with regard to the vertical angle and the distance of a sound image.

2. Early HRIRs

2.1. HRIR acquisition

The HRIRs of four male subjects (HRM, ISI, OOT, and TKH) and one female subject (SKW), 22–24 years of age, were measured for seven vertical angles in the upper median plane (0–180° in 30° steps) in an anechoic chamber. The vertical angle, which ranges from 0° to 360°, is defined as the angle measured from the front direction in the median plane, with 0° indicating the front, 90° indicating above, and 180° indicating the rear [16].

The test signal was presented in 30° steps by a loudspeaker having a diameter of 80 mm (FOSTEX FE83E) located in the upper median plane. The distance from the loudspeakers to the center of the subject's head was 1.2 m. The test signal was a swept sine wave, the duration and the sampling frequency of which were 2^{18} samples and 48 kHz, respectively. Earplug-type microphones [7] were used to sense the test signals at the entrances of the ear canals of the subject. The earplug-type microphones were placed into the ear canals of the subjects. The diaphragms of the microphones were located at the entrances of the ear canals. This condition is referred to as the blocked-entrances condition [20]. The HRTF was obtained as

$$HRTF_{l,r}(\omega) = G_{l,r}(\omega)/F(\omega) \quad (1)$$

where $F(\omega)$ is the Fourier transform of the impulse response, $f(t)$, measured at the point corresponding to the center of the subject's head in the anechoic chamber without a subject, and $G_{l,r}(\omega)$ is the Fourier transform of the impulse response, $g_{l,r}(t)$, measured at the entrance of the ear canal of the subject with the earplug-type microphones. Both $f(t)$ and $g(t)$ were 512 samples long. The HRIR was obtained by inverse fast Fourier transform (FFT) of the HRTF.

2.2. Generation of early HRIRs

For each subject, ear, and target vertical angle, early HRIRs were generated 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, 2N-point Blackman-Harris window, adjusting the temporal center of the window to S_{max}, using four values of N (12, 24, 48, and 96 samples). These samples correspond to 0.25, 0.5, 1, and 2 ms, respectively.
(3) For comparison, the full-length HRIR was obtained by clipping the HRIR using a rectangular window. The end point of the window was the first zero-cross point after 256 samples from S_{max}.

Fig. 1 shows examples of a full-length HRIR and the early HRIRs of subject HRM for the front direction (vertical angle of 0°). The response of the full-length HRIR was converged within 4 ms from S_{max}. The early HRIR of 0.25 ms includes only the positive part of the first response. The early HRIR of 0.5 ms includes the positive and negative parts of the first response. However, the absolute value of the negative part was decreased by the temporal window. The early HRIR of 1 ms includes the responses until positive part of the second response, which was, however, decreased by the temporal window. The early HRIR of 2 ms includes most of the responses, except for the fine responses of later part.

Fig. 2 shows the amplitude spectra of the full-length HRTFs and early HRTFs of subject HRM for the seven vertical angles. The amplitude spectra were obtained by the FFT with 512 samples, the frequency resolution of which was 93.75 Hz. The full-length HRTFs (solid lines) include several notches and peaks for all seven vertical angles. The notches were deep at the vertical angles near the horizontal plane (0° and 180°) and shallow at the upper direction. The frequency of the notches was highly dependent on the vertical angle, whereas the frequency of the peaks was approximately constant regardless of the vertical angle, as reported by Iida *et al.* [7].

For the early HRTF of 0.25 ms, one notch and two peaks were observed at the target vertical angles of 0–120°, and two notches and three peaks were observed at vertical angles of 150° and 180°. The frequencies of the notches and peaks did not coincide with those of full-length HRTFs. Most of the notches were shallow, and the peak level was low.

For the early HRTF of 0.5 ms, one to three notches were observed. The frequencies of the notches at the target vertical angles of 0°, 150°, and 180° were approximately the same as those of the full-length HRTF (the frequency difference was within 281.75 Hz). However, the levels of the notches did not coincide with those of the full-length HRTF. Three or four peaks were observed, and the lowest-frequency peaks coincided with those of the full-length HRTF. However, most of the other peaks did not coincide with those of the full-length HRTF.

For the early HRTF of 1 ms, the outline of the notches and peaks was approximately the same as that of the full-length HRTF, while the fine structure differed from that of the full-length HRTF. The sharp notch at 60° was not found for the other subjects, and the reason for its existence is unclear.

The spectrum of the early HRTF of 2 ms coincided with that of the full-length HRTF to the last detail.

3. Experiment 1: Vertical angle of a sound image for the early HRIRs

Psycho-acoustical tests were carried out in order to examine the