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

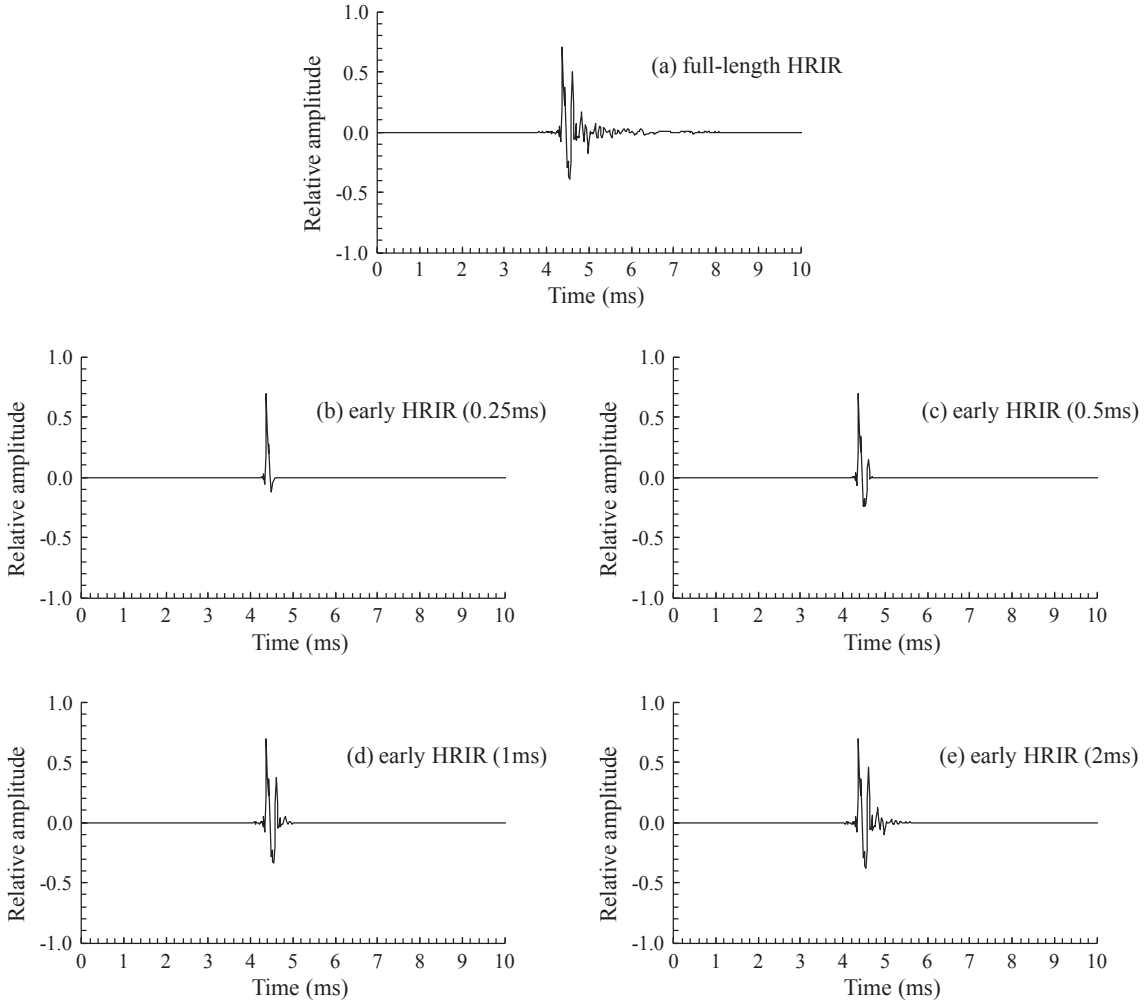

Fig. 1. Examples of a full-length HRIR and the early HRIRs for the front direction of subject HRM.

vertical localization performance of the early HRIRs in the upper median plane.

3.1. Method of experiment 1

A laptop computer (Panasonic CF-R8), an audio interface (RME Fireface 400), an amplifier (Marantz PM4001), open-air headphones (AKG K1000), earplug-type microphones, and an A/D converter (Roland M-10MX) were used for the localization tests. The localization tests were conducted in a quiet soundproof room. The working area of the room was 4.6 m (width) by 5.8 m (depth) by 2.8 m (height). The background A-weighted sound pressure level (SPL) was 20 dB.

Each subject's own full-length HRIRs and early HRIRs of four lengths (0.25, 0.5, 1, and 2 ms), which were generated using the algorithm described in Section 2.2, were used. The target vertical angles were seven directions, in steps of 30°, in the upper median plane. The five subjects (HRM, ISI, OOT, TKH, and SKW) participated in the sound localization tests. All of the subjects self-reported normal hearing sensitivity.

In order to reproduce the sound pressure at the eardrum for the open-ear-canal condition, P, the signal at the entrance of the blocked ear canal ($S \times HRTF$) was processed with a compensation filter, G, and presented through headphones [14], as follows:

$$P(\omega) = S(\omega) \times HRTF(\omega) \times G(\omega), \tag{2}$$

$$G(\omega) = \frac{1}{M(\omega) \times PTF(\omega)} \times \frac{Z_{ear\ canal}(\omega) + Z_{headphone(\omega)}}{Z_{earcanal}(\omega) + Z_{radiation}(\omega)}, \tag{3}$$

$$\triangleq \frac{1}{M(\omega) \times PTF(\omega)} \times PDR(\omega), \tag{4}$$

where S denotes the sound source, $M \times PTF$ is the electroacoustic transfer function from the headphones to the earplug-type microphones, $Z_{ear\ canal}$ and $Z_{headphone}$ denote the impedances of the ear canal and headphones, respectively, and $Z_{radiation}$ is the free-air radiation impedance as observed from the ear canal. The second term on the right-hand side of Eq. (3) is referred to as the pressure division ratio (PDR). Møller *et al.* defined free-air equivalent coupling to the ear (FEC) headphones as headphones for which the PDR reduces to unity. K1000 headphones (AKG), which are regarded as FEC headphones, were used in the localization tests.

The compensation of $M \times PTF$ was processed from 200 Hz to 17 kHz by the following procedure. The subjects sat at the center of the soundproof room. 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, in the same manner as in the HRIR measurements described in Section 2.1. The subjects wore the headphones, and maximum-length sequence signals (48 kHz sampling, 12th-order, and no repetitions) were emitted through the headphones. The signals were received by the earplug-type microphones, and $M \times PTF$ was obtained. The earplug-type microphones were then removed without displacing the headphones because the pinnae of the subject were not enclosed by the headphones. The typical peak-to-peak range of the transfer functions between the headphones and the earplug-type microphones from 200 Hz to 17 kHz was approximately 20 dB. This was

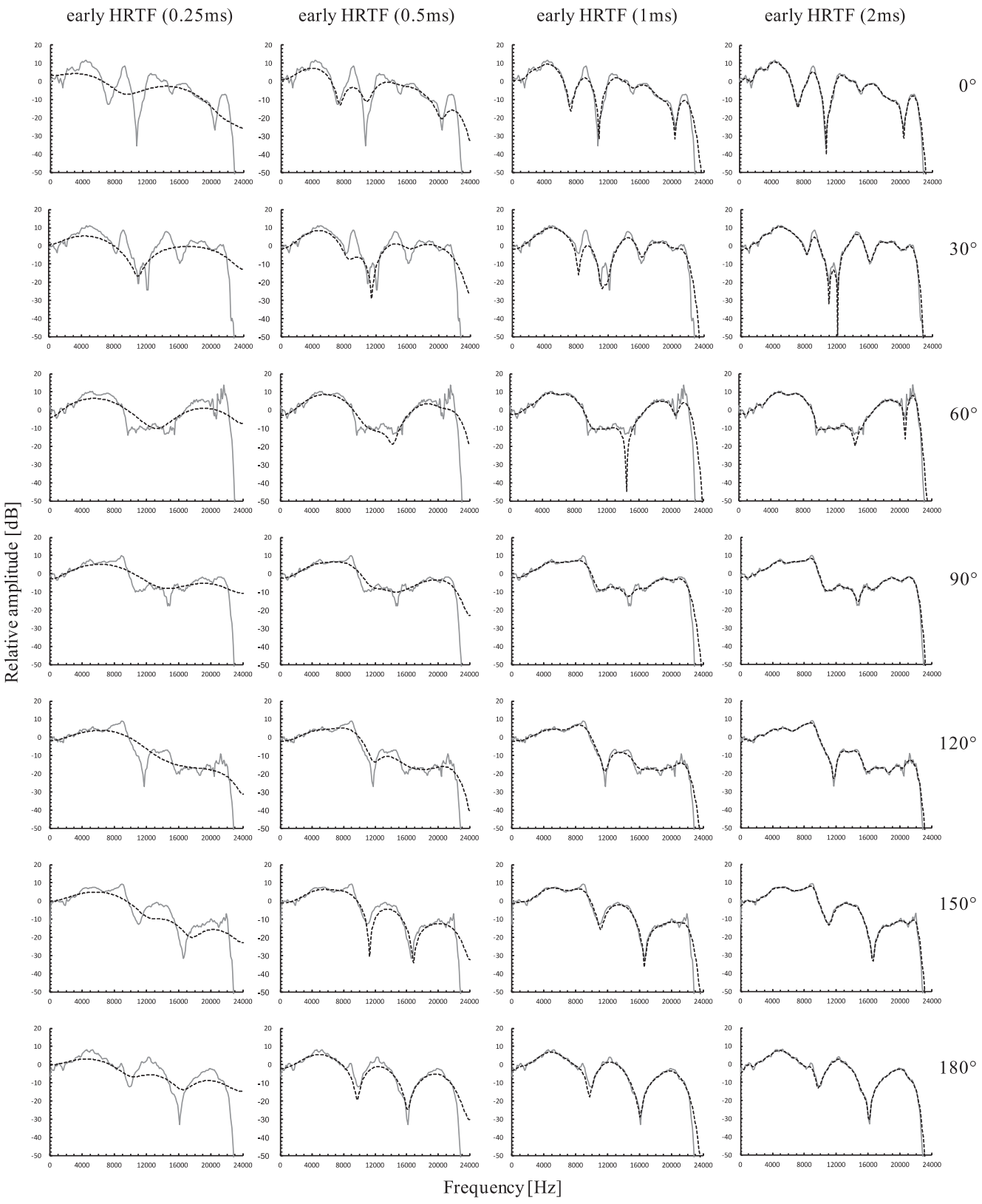

Fig. 2. Examples of the amplitude spectra of the full-length HRTFs and the early HRTFs for the seven vertical angles for subject HRM. The solid lines and the broken lines indicate the full-length HRTFs and the early HRTFs, respectively.

reduced to 3 dB by the compensation filter, *G*.

The source signal was a wideband Gaussian white noise from 200 Hz to 17 kHz. Stimuli were delivered at 63 dB SPL at the entrance of each ear. The duration of the stimuli was 1.2 s, including the rise and fall times, each of which was 0.1 s.

The mapping method was adopted as a response method in order to respond on a continuous scale rather than selecting between distinct locations. A circle and a horizontal arrow through the center of the

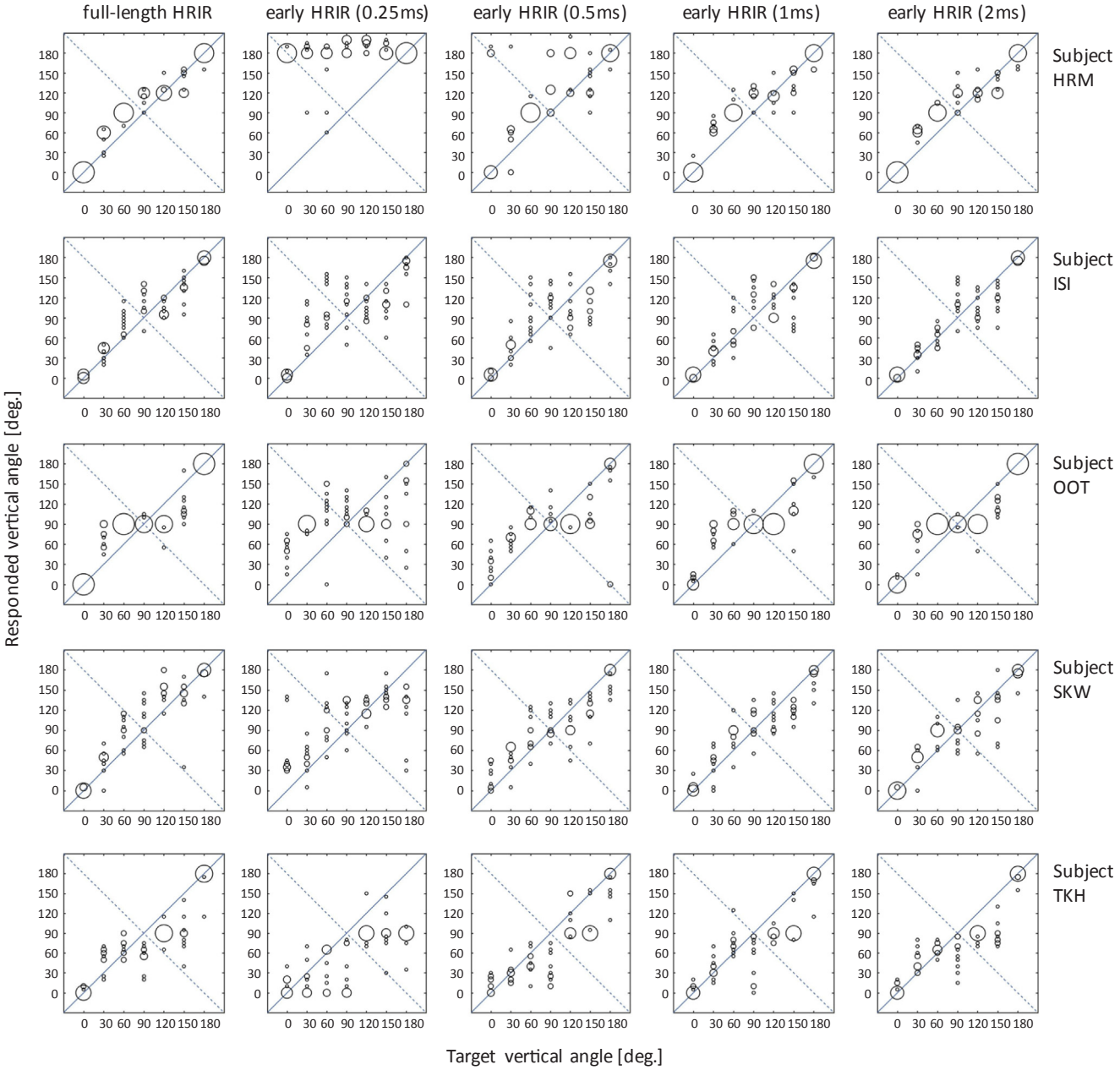

Fig. 3. Responses to the full-length HRIRs and the early HRIRs. The ordinate represents the responded vertical angle, and the abscissa represents the target vertical angle. The diameter of each circle is proportional to the number of responses with a resolution of 5°.

circle, which indicated the median plane and the front-back axis, respectively, were displayed on the screen of a laptop computer. The subject's task was to click on the perceived vertical angle on the circle shown on the computer display using a mouse. Each subject was also instructed to check the box on the display when he/she perceived a sound image inside his/her head.

In one test block, 35 stimuli (five durations, seven directions) were randomized and presented to a subject. The duration of one block was approximately 7 min. Ten test blocks were performed by each subject. Therefore, each subject responded to each stimulus 10 times. The tests were carried out using a double-blind method.

3.2. Individual responses

Fig. 3 shows the responses to the subject's own full-length HRIRs and early HRIRs for the five subjects. The ordinate represents the responded vertical angle, and the abscissa represents the target vertical angle. The diameter of each circle is proportional to the number of responses with a resolution of 5°.

For the full-length HRIR, most of the responses were distributed around the target vertical angles. However, the responses of subject OOT were distributed around 90° for the target vertical angles of 60° and 120°. The responses of subject TKH were distributed around 90° for the target vertical angle of 150°.

For the early HRIR of 0.25 ms, the distribution of most of the responses differed from that for the full-length HRIR. Most of the responses of subject HRM were distributed rearward, whereas those of ISI, OOT, and SKW were distributed widely. Most of the responses of TKH were distributed from forward to upward. On the other hand, the distribution of the responses of the four subjects other than HRM was approximately the same as that for the full-length HRIR for the target vertical angle of 120°. The distribution of the responses was also approximately the same as that for the full-length HRIR for HRM at 180°, ISI at 0°, SKW at 90° and 150°, and TKH at 150°.

For the early HRIR of 0.5 ms, the distribution of most of the responses also differed from that for the full-length HRIR. The responses of subject HRM were distributed around both the target vertical angle and 180° at the target vertical angles of 0°, 90°, and 120°. The responses for ISI were distributed widely for the target vertical angles of 60° to 120°. The responses of OOT were distributed from forward to upward for the target vertical angle of 0°, and were distributed around both the target vertical angle and 0° at the target vertical angle of 180°. The responses of SKW were distributed from forward to diagonally upward for the target vertical angle of 0°. On the other hand, the distribution of the responses was also approximately the same as that for the full-length HRIR for HRM at 60° and 180°, ISI at 0° and 90°, OOT at 30° and 120°, SKW at 30° to 90°, and TKH at 90°. Thus, the distribution of most of the responses for the early HRIRs of 0.25 and 0.5 ms differed from that for the full-length HRIR. However, for the target vertical angles around 120°, the distribution of the early HRIRs of 0.25 and 0.5 ms tended to be approximately the same as that for the full-length HRIR.

For the early HRIRs of 1 and 2 ms, the distribution of the responses was approximately the same as that for the full-length HRIR for each subject and target vertical angle.

3.3. Mean vertical localization error

The mean vertical localization error for each subject, HRIR, and target vertical angle was calculated (Table 1). The mean vertical localization error is defined as the absolute difference between the responded and target vertical angles averaged over all of the responses. Then, Tukey's multiple comparison test was performed in order to determine whether a difference in the mean vertical localization error between the full-length HRIRs and each of the early HRIRs is statistically significant (Table 2).

Table 1
Mean vertical localization error for each subject, HRIR, and target vertical angle [°].

Subject	HRIR	Target vertical angle [°]						
		0	30	60	90	120	150	180
HRM	Full-length	0.5	24.2	27.7	25.5	4.5	16.5	3.1
	0.25 ms	178.3	144.9	98.5	100.0	73.4	35.3	0.7
	0.5 ms	70.9	43.3	32.3	40.8	39.3	24.6	3.3
	1 ms	2.8	37.8	35.4	28.6	9.7	17.6	5.8
	2 ms	0.5	30.7	32.8	26.6	7.1	20.2	4.7
ISI	Full-length	2.8	11.8	23.5	29.3	16.4	18.4	3.1
	0.25 ms	3.3	45.8	51.7	32.3	18.3	41.0	21.6
	0.5 ms	5.2	18.2	36.9	30.6	31.2	43.1	9.6
	1 ms	3.3	14.3	22.6	36.0	18.9	38.7	2.8
	2 ms	3.3	11.6	11.8	31.0	22.8	31.6	3.4
OOT	Full-length	0.6	40.3	30.5	3.1	34.1	37.8	0.4
	0.25 ms	50.1	57.7	60.1	20.4	24.3	54.8	59.1
	0.5 ms	29.2	37.0	39.3	9.6	30.3	43.5	40.3
	1 ms	5.1	45.3	34.1	2.5	30.2	34.5	2.2
	2 ms	3.0	41.9	30.4	2.5	34.3	28.9	0.5
SKW	Full-length	1.9	20.3	33.3	23.5	30.2	20.3	6.3
	0.25 ms	55.5	23.3	46.7	28.3	11.8	13.3	63.3
	0.5 ms	20.7	24.5	28.9	15.3	28.8	28.9	14.4
	1 ms	4.4	20.8	25.6	19.8	20.1	31.0	11.4
	2 ms	1.7	22.9	27.3	15.4	22.2	33.8	6.0
TKH	Full-length	2.7	25.4	12.7	35.1	30.2	61.7	7.2
	0.25 ms	11.4	21.6	30.1	60.5	33.2	60.9	96.7
	0.5 ms	14.1	9.4	19.4	54.0	25.7	48.3	16.9
	1 ms	4.7	12.3	18.6	40.1	32.6	49.8	10.2
	2 ms	5.7	20.6	8.3	33.2	34.4	61.6	3.6
All subjects	Full-length	1.7	24.4	25.5	23.3	23.1	30.9	4.0
	0.25 ms	59.7	58.7	57.4	48.3	32.2	41.1	48.3
	0.5 ms	28.0	26.5	31.4	30.1	31.1	37.7	16.9
	1 ms	4.1	26.1	27.3	25.4	22.3	34.3	6.5
	2 ms	2.8	25.5	22.1	21.7	24.2	35.2	3.6

Table 2
Results of Tukey's multiple comparison test between the full-length HRIRs and the early HRIRs for mean vertical localization error.

Subject	Early HRIR	Target vertical angle [°]						
		0	30	60	90	120	150	180
HRM	0.25 ms	**	**	**	**	**	*	
	0.5 ms	**				**		
	1 ms							
	2 ms							
ISI	0.25 ms		**					*
	0.5 ms	*						
	1 ms							
	2 ms							
OOT	0.25 ms	**	*	**	*			*
	0.5 ms	**						
	1 ms							
	2 ms							
SKW	0.25 ms	**						**
	0.5 ms							
	1 ms							
	2 ms							
TKH	0.25 ms							**
	0.5 ms							
	1 ms							
	2 ms							
All subjects	0.25 ms	**	**	**	**			**
	0.5 ms	**						
	1 ms							
	2 ms							

** : $p < 0.01$.
* : $p < 0.05$.

For the full-length HRIR, the mean vertical localization error tended to be small at the target vertical angles near the horizontal plane (0° and 180°) and increased with the vertical angle, as reported by Carlile *et al.* [2] and Majdak *et al.* [10].

For the early HRIR of 0.25 ms, the mean vertical localization errors for most subjects and target vertical angles were larger than those for the full-length HRIR. However, the mean vertical localization errors were approximately the same as those for the full-length HRIR at 120° for the four subjects other than HRM. The mean vertical localization error was significantly larger than that of the full-length HRIR at one to six target vertical angles for each of the subjects. In particular, at the target vertical angles of 0° and 180°, a significant difference was observed for three and four of the five subjects, respectively. For the responses of all of the subjects, a significant difference was observed at the target vertical angles of 0–90° and 180° ($p < 0.01$).

For the early HRIR of 0.5 ms, the mean vertical localization errors for most subjects and target vertical angles were larger than those for the full-length HRIR. In particular, the mean vertical localization errors for the four subjects other than subject ISI was large at the target angle of 0°. However, the mean vertical localization errors were approximately the same as those of the full-length HRIR at 120° for subjects OOT, SKW, and TKH. For subjects HRM, ISI, and OOT, the mean vertical localization error was significantly larger than that for the full-length HRIR at one or two target vertical angles, including 0°. For the responses of all of the subjects, a significant difference was observed at the target vertical angle of 0° ($p < 0.01$).

For the early HRIRs of 1 and 2 ms, the mean vertical localization errors were approximately the same as those for the full-length HRIR for most subjects and target vertical angles. No statistically significant difference was observed at any target vertical angle or for any subject between the full-length HRIR and the early HRIR of 1 ms or between the full-length HRIR and the early HRIR of 2 ms.

Thus, the durations of the early HRIR, the mean vertical localization error of which was approximately the same as that for full-length HRIR

Table 3
Ratio of front-back confusion for each HRIR and target vertical angle.

Subject	HRIR	Target vertical angle [°]						
		0	30	60	90	120	150	180
All subjects	Full-length	0.00	0.02	0.40	–	0.20	0.14	0.00
	0.25 ms	0.24	0.32	0.58	–	0.24	0.24	0.26
	0.5 ms	0.08	0.02	0.40	–	0.28	0.26	0.04
	1 ms	0.00	0.06	0.38	–	0.34	0.16	0.00
	2 ms	0.00	0.02	0.40	–	0.38	0.20	0.00

for all seven vertical angles in the upper median plane, were 1 and 2 ms. On the other hand, the errors of the early HRIRs of 0.25 and 0.5 ms were larger than that for the full-length HRIR for most of the target vertical angles and subjects. These results approximately agree with those of Senova *et al.* [19], indicating that the mean localization error of the truncated HRIRs averaged over 354 sound source directions (in 10° azimuth steps and 10° elevation steps ranging from −40° to 70°) was not disrupted dramatically until the HRIR duration was reduced to 0.64 or 0.32 ms. However, our results suggest that the minimum required duration of the HRIR depends on the direction of a sound source. The errors of the early HRIRs of 0.25 and 0.5 ms were approximately the same as that for the full-length HRIR at the target vertical angle of 120° for the subjects other than HRM.

3.4. Ratio of front-back confusion

Table 3 shows the ratio of front-back confusion for each HRIR and target vertical angle. The ratio of front-back confusion is defined as the ratio of the responses for which the subjects localized a sound image in the quadrant opposite that of the target direction in the upper median plane. Statistical tests (chi-square test) were performed to verify whether the difference in the ratio of front-back confusion between the full-length HRIRs and each of the early HRIRs was statistically significant (Table 4).

For the full-length HRIR, the ratio of front-back confusion at the target vertical angles of 0° and 180° was 0%. At 60° and 120°, the ratios were 40% and 20%, respectively.

For the early HRIR of 0.25 ms, the ratio was higher than that for the full-length HRIR at all of the target vertical angles. The ratios at 0° and 180° were 24% and 26%, respectively. At the other four target vertical angles, the ratios ranged from 24% to 58%. A significant difference was observed at the target vertical angles of 30° ($p < 0.01$), 60° ($p < 0.05$), and 150° ($p < 0.05$). Note that the statistical tests could not be performed at the target vertical angles of 0° and 180° because the ratio for the full-length HRIR was 0%.

For the early HRIR of 0.5 ms, the ratios at 0° and 180° were 8% and 4%, respectively. The ratios were higher than those for the full-length HRIRs at the target vertical angles of 120° and 150°. A significant difference was observed at the target vertical angle of 150° ($p < 0.05$).

For the early HRIR of 1 ms, the ratio at 0° and 180° was 0%. However, the ratio was higher than that for the full-length HRIR at the target vertical angles of 30°, 120°, and 150°. A significant difference was observed at the target vertical angles of 30° and 120° ($p < 0.05$).

For the early HRIR of 2 ms, the ratio at 0° and 180° was 0%. However, the ratio was higher than that for the full-length HRIR at the target vertical angles of 120° and 150°. A significant difference was observed at the target vertical angle of 120° ($p < 0.01$).

Table 4
Results of chi-square tests between the full-length HRIRs and the early HRIRs for ratio of front-back confusion.

Subject	Early HRIR	Target vertical angle [°]						
		0	30	60	90	120	150	180
All subjects	0.25 ms	–	**	*	–		*	–
	0.5 ms	–			–		*	–
	1 ms	–	*		–	*		–
	2 ms	–			–	**		–

** : $p < 0.01$.
* : $p < 0.05$.

Table 5
Ratio of inside-of-head localization for each HRIR and target vertical angle.

Subject	HRIR	Target vertical angle [°]						
		0	30	60	90	120	150	180
All subjects	Full-length	0.00	0.00	0.00	0.00	0.12	0.06	0.00
	0.25 ms	0.06	0.16	0.10	0.06	0.08	0.14	0.18
	0.5 ms	0.06	0.04	0.02	0.08	0.02	0.04	0.04
	1 ms	0.02	0.00	0.00	0.02	0.04	0.02	0.00
	2 ms	0.02	0.00	0.00	0.02	0.02	0.00	0.00

3.5. Ratio of inside-of-head localization

Table 5 shows the ratio of inside-of-head localization for each HRIR and target vertical angle. For the full-length HRIR, the inside-of-head localization was never reported at the target vertical angles of 0–90° and 180°. For the early HRIRs of 0.25 and 0.5 ms, inside-of-head localization was reported at all of the target vertical angles. For the early HRTFs of 1 and 2 ms, the ratio was approximately the same as that for the full-length HRIR.

4. Experiment 2: Distance of a sound image for the early HRIRs

Experiment 1 suggested that the vertical angle of a sound image for the early HRIRs of 1 and 2 ms was approximately the same as that for the full-length HRIR. The purpose of experiment 2 is to examine whether the distance of a sound image for each of the early HRIRs of 1 and 2 ms can be considered to be the same as that for the full-length HRIR.

4.1. Method of experiment 2

The subject's full-length HRIR and early HRIRs of 1 and 2 ms were used. The target vertical angles were 0°, 90°, and 180°. The experiment system, sound source, and subjects were the same as those used in experiment 1.

The distance of a sound image was evaluated using the Scheffe's paired comparison method modified by Ura [18]. Paired stimuli were presented to a subject, who then responded with the degree of the distance of a sound image of the stimuli presented later compared to that presented previously, responding with −2.0 for a very near distance, 0 for the same distance, and +2 for a very far distance.

A total of 72 ($_9P_2$) paired stimuli were prepared from nine (three durations and three vertical angles) stimuli. In one test block, 36 paired-stimuli were randomized and presented to a subject. Each subject was presented with eleven test blocks, including one test block for practice. The responses to the practice block were excluded from the analysis. Therefore, five responses to each paired-stimulus for each subject were analyzed. The tests were carried out using a double-blind method.

4.2. Results of experiment 2

Fig. 4 shows the scale value of the distance of a sound image obtained using the responses of all the subjects. The error bars denote standard errors. The distance of a sound image was affected by the target vertical angle rather than the duration of the HRIR. The distance was long for the front direction and short for the rear direction. The

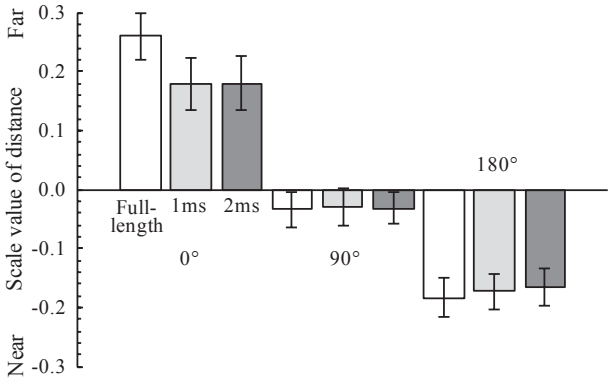

Fig. 4. Scale value of the distance of a sound image. The error bars denote standard errors.

distance for the early HRIRs was short compared to that for the full-length HRIR at the front direction. On the other hand, the distance for the early HRIRs was long compared to that for the full-length HRIR at the above and rear directions. However, the differences of scale value were within the standard error.

Since there exist two possible factors, the duration and the vertical angle of the HRIR, which may affect the distance of a sound image, we performed a two-way repeated measures analysis of variance. The results revealed that no significant effects remain for the duration of the HRIR, and significant effects ($F = 86.9, p < 0.01$) remain for the target vertical angle. No significant interaction effects were observed. Tukey's multiple comparison test revealed the existence of significant differences ($p < 0.01$) between 0° and 90°, between 0° and 180°, and between 90° and 180° for each duration. However, the significance level between 90° and 180° for the durations of 1 and 2 ms was 5%.

These results suggest that the early HRTFs of 1 and 2 ms provide approximately the same sound image distance as that for the full-length HRIRs for the target vertical angles of 0°, 90°, and 180°. However, the reason for the existence of the significant difference in the sound image

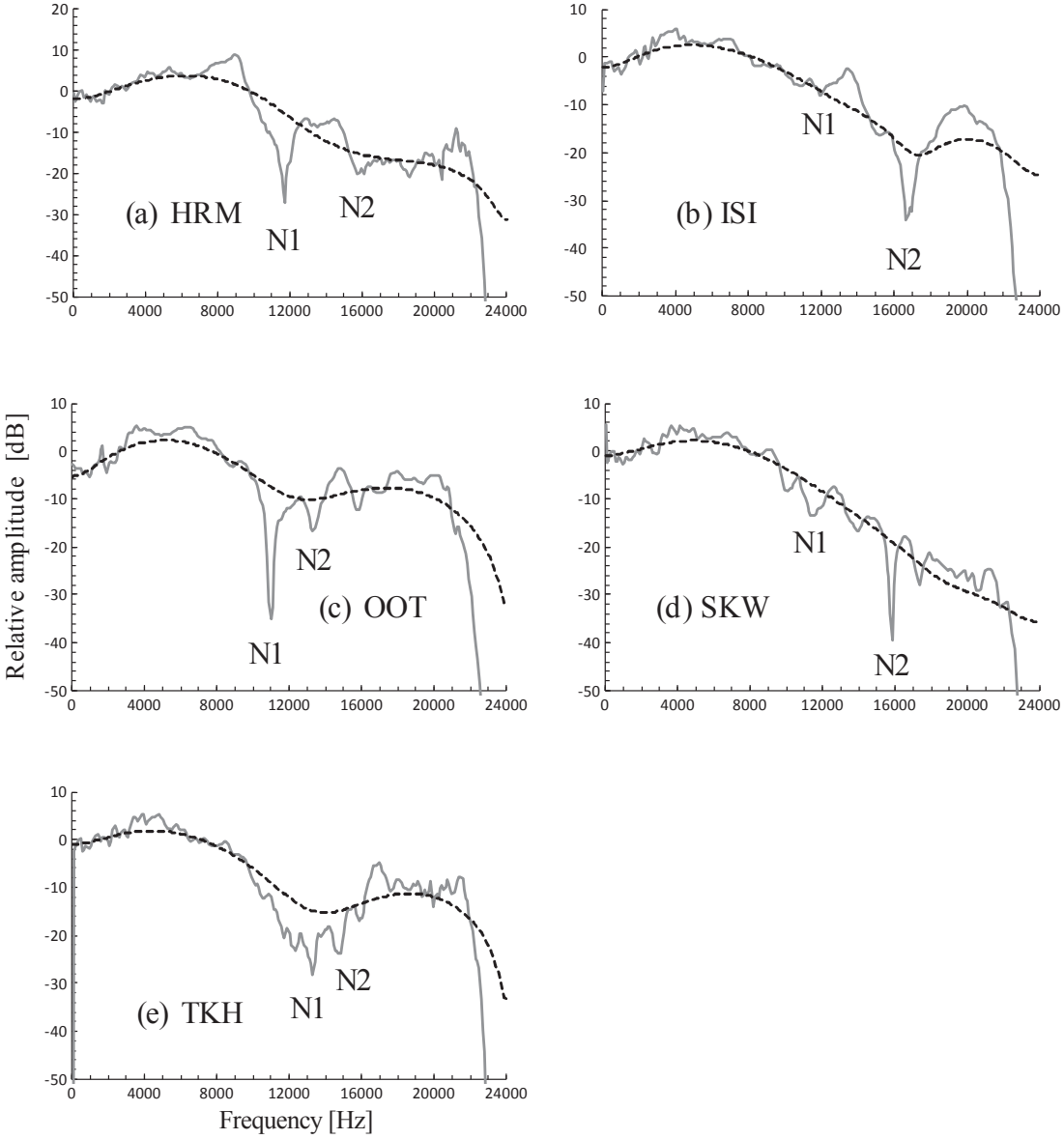

Fig. 5. Amplitude spectra of the full-length HRTFs and the early HRTFs (0.25 ms) for the vertical angle of 120°. The solid lines and the broken lines indicate the full-length HRTFs and the early HRTF of 0.25 ms, respectively.

distance among the target vertical angles is unclear. This is a problem to be solved in the near future.

5. Discussion

Let us next discuss the relation between the duration of the early HRIR and the performance of vertical localization.

5.1. Discussion 1: Reason for high vertical localization performance for the early HRIRs of 1 and 2 ms

In experiment 1, the early HRIRs of 1 and 2 ms provided approximately the same vertical localization performance as the full-length HRIR, while there existed a significant difference in the mean vertical localization error between the full-length HRIR and each of the early HRIRs of 0.25 and 0.5 ms. Let us next discuss the reason for the high vertical localization performance in the early HRIRs of 1 and 2 ms.

At the target vertical angle of 0° for subject HRM, for example, most of the responses for the early HRIR of 0.25 ms were distributed around 180°, and the responses for 0.5 ms tended to shift rearward, as shown in Fig. 3. The responses for 1 and 2 ms were approximately the same as those for the full-length HRIR.

In the amplitude spectra of the early HRIRs of 0.25 and 0.5 ms, as shown in Fig. 2, there existed no remarkable notches, which were observed in the full-length HRIR. On the other hand, for the early HRTF of 1 ms, the outline of the notches and peaks was approximately the same as that for the full-length HRTF, whereas the fine structure differed from that of the full-length HRTF. The spectrum of the early HRTF of 2 ms coincided with that of full-length HRTF to the last detail.

Thus, the cues for the vertical perception were included in the early HRTFs of 1 and 2 ms, whereas they were not fully included in the early HRTFs of 0.25 and 0.5 ms. This might explain the high vertical localization performance for 1 and 2 ms.

5.2. Discussion 2: Reason for high vertical localization performance at 120° even for the early HRIR of 0.25 ms

At the target vertical angle of 120°, the vertical localization performance of the early HRIR of 0.25 ms was approximately the same as that of the full-length HRIR for the four subjects other than HRM.

In a previous study, the notches were reported to be shallow for the upper direction [8]. Fig. 5 shows the amplitude spectra at 120° for the full-length HRIR and the early HRIR of 0.25 ms for all of the subjects.

For the full-length HRIR (shown by the solid line), N1 was shallow, and N2 was located between 16 kHz and 17 kHz for subjects ISI and SKW. Since the frequency range of the sound source was 200 Hz–17 kHz, no remarkable notches appeared in the stimuli for these two subjects. For subject OOT, N1 was very narrow, and N2 was shallow. For TKH, N1 and N2 were bound together into one broad valley. On the other hand, N1 and N2 were remarkable in the amplitude spectrum of subject HRM, the localization performance of which was low at 120°.

For the early HRTF of 0.25 ms, no remarkable notches were observed for any of the subjects. However, the outline of the spectrum of 0.25 ms was approximately the same as that of the full-length HRTF, except for the notches. As a result, the spectrum of the early HRTF of 0.25 ms was approximately the same as that of the full-length HRTF for the four subjects, the full-length HRTF of which does not include the remarkable notches. This might be a reason for the high vertical localization performance at 120°, even for the early HRTF of 0.25 ms, except for subject HRM.

5.3. Discussion 3: Use of an early HRIR for individualization of the HRIR

It is widely accepted that the HRIR of other listeners often causes front-back confusion [15,23] due to individual differences in the HRIR. This is a serious problem, which prevents acoustic virtual reality from coming into widespread practical use.

In order to obtain the individualized HRTF, previous studies proposed a method by which to estimate the frequency of the notches and peaks in the listener's HRTF of the front direction from the anthropometry of the listener's pinna [7,21,12,13]. However, estimation of the notches and peaks of directions other than the front direction has not succeeded. One reason for this failure might be that the usual HRIR includes not only the response of the pinna but also the responses of the head and torso. The responses of the head and torso are assumed to decrease the correlation between the anthropometry of the listener's pinna and the frequency of the notches and peaks of the listener's HRTF, because the notches and peaks are generated not by the head and torso but just by the pinna [22,20].

Using the early HRIR, which is mainly composed of the response of the pinna, instead of the usual HRIR, the estimation of the notches and peaks based on the anthropometry of the listener's pinna is expected to be achieved in the entire median plane.

6. Conclusions

In the present study, we measured the HRIRs for seven target vertical angles in the upper median plane (0–180° in 30° steps) for five subjects and generated the 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. The results suggested the following:

(1) For the early HRTFs of 0.25 and 0.5 ms, one to three notches were observed. However, these notches 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, whereas the fine structure differed from that of the full-length HRTF. The spectrum of the early HRTF of 2 ms coincided with that of the full-length HRTF to the last detail.
(2) For the early HRIRs of 0.25 and 0.5 ms, the mean vertical localization errors for most subjects and target vertical angles were larger than those for the full-length HRIRs. The significant difference in the mean vertical localization error between the full-length HRIR and the early HRIR of 0.25 ms was observed at the target vertical angles of 0–90° and 180° ($p < 0.01$). Between the full-length HRIR and the early HRIR of 0.5 ms, a significant difference was observed at 0° ($p < 0.01$). On the other hand, the mean vertical localization errors for the early HRIRs of 1 and 2 ms were approximately the same as that for the full-length HRIR. No significant difference in the mean vertical localization error was observed between the full-length HRIR and each of the early HRIRs of 1 and 2 ms at all the target vertical angles.
(3) The early HRTFs of 1 and 2 ms provided approximately the same sound image distance as that for the full-length HRIR. No significant difference was observed between the full-length HRIR and each of the early HRIRs of 1 and 2 ms.

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.

Acknowledgement

The present study was supported in part by MEXT KAKENHI through a Grant-in-Aid for Scientific Research (A) (Grant Number 15H01790).

References

[1] Asano F, Suzuki Y, Sone T. Role of spectral cues in median plane localization. J Acoust Soc Am 1990;88:159–68.
[2] Carlile S, Leong P, Hyams S. The nature and distribution of errors in sound localization by human listeners. Hear Res 1997;114:179–96.
[3] Butler A, Belendiuk K. Spectral cues utilized in the localization of sound in the median sagittal plane. J Acoust Soc Am 1977;61:1264–9.
[4] Gardner B, Gardner S. Problem of localization in the median plane: effect of pinna cavity occlusion. J Acoust Soc Am 1973;53:400–8.
[5] Iida K, Yairi M, Morimoto M. Role of pinna cavities in median plane localization. Procee 16th Int Congress Acoustics. 1998. p. 845–6.
[6] 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.
[7] 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;136:317–33.
[8] 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.
[9] Kulkarni A, Colburn HS. Role of spectral detail in sound- source localization. Nature 1998;396:747–9.
[10] Majdak P, Goupell MJ, Laback B. 3-D localization of virtual sound sources: effects of visual environment, pointing method, and training. Atten Percept Psychophys 2010;72:454–69.
[11] Mehrgardt S, Mellert V. Transformation characteristics of the external human ear. J Acoust Soc Am 1977;61:1567–76.
[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;137:690–701.
[13] Mokhtari P, Takemoto H, Nishimura R, Kato H. Vertical normal modes of human ears: individual variation and frequency estimation from pinna anthropometry. J Acoust Soc Am 2016;140:814–31.
[14] Møller H, Hammershøi D, Jensen CJ, Sørensen MF. Transfer characteristics of headphones measured on human ears. J Audio Eng Soc 1995;43:203–17.
[15] Morimoto M, Ando Y. On the simulation of sound localization. J Acoust Soc Jpn (E) 1980;1:167–74.
[16] Morimoto M, Aokata H. Localization cues of sound sources in the upper hemisphere. J Acoust Soc Jpn (E) 1984;5:165–73.
[17] Musicant A, Butler R. The influence of pinnae-based spectral cues on sound localization. J Acoust Soc Am 1984;75:1195–200.
[18] Nagasawa S. Improvement of the Scheffe's method for paired comparisons. Kansei Eng Int 2002;3:47–56.
[19] Senova MA, McAnally KI, Martin RL. Localization of virtual sound as a function of head-related impulse response duration. J Audio Eng Soc 2002;50:57–66.
[20] Shaw EAG, Teranishi R. Sound pressure generated in an external-ear replica and real human ears by a nearby point source. J Acoust Soc Am 1968;4:240–9.
[21] Spagnol S, Avanzini F. Frequency estimation of the first pinna notch in head–related transfer functions with a linear anthropometric model. In: Proceedings of the 18th Int. Conference on Digital Audio Effects (DAFx-15), Trondheim, Norway, Nov 30–Dec 3 (2015).
[22] Takemoto H, Mokhtari P, Kato H, Nishimura R, Iida K. Mechanism for generating peaks and notches of head-related transfer functions in the median plane. J Acoust Soc Am 2012;132:3832–41.
[23] Wenzel EM, Arruda M, Kistler DJ, Wightman FL. Localization using non-individualized head-related transfer functions. J Acoust Soc Am 1993;94:111–23.