Proceedings of 20th International Congress on Acoustics, ICA 2010

23–27 August 2010, Sydney, Australia

Influence of band-limited flattening of HRTFs on sound localization performance

Kanji Watanabe (1), Ryosuke Kodama (2), Sojun Sato (1), and Shouichi Takane(1)

(1) Faculty of Systems Science and Technology, Akita Prefectural University, Japan
(2) Graduate School of Systems Science and Technology, Akita Prefectural University, Japan

PACS: 43.60.–c, 43.66.–x

ABSTRACT

Data reduction of head-related transfer functions (HRTFs) serves for effective implementation of their synthesis from computational point of view. In order to investigate the ability of data reduction, the following two points are taken into account: 1) The low frequency characteristics of HRTFs has little dependence on sound source direction, 2)it can be found out from the frequency resolution of the auditory system that the detailed spectral peaks and dips of the HRTFs in high frequency region do not fully contribute to the sound localization. In this paper, the low frequency characteristics of HRTFs for both ears were flattened below a certain boundary frequency (called lower boundary frequency), and the HRTF on the contralateral side was flattened above a certain boundary frequency (called higher boundary frequency). Those flattening processes were simultaneously applied retaining the interaural level and time differences to the original ones. A localization test with the sound source in horizontal plane was carried out, and the results showed that the flattening especially affects the front-back confusions, and the flattening of HRTFs does not significantly affect the sound localization performance when the lower and the higher boundary frequency are 1 and 4 kHz, respectively.

INTRODUCTION

When sound travels from the source to the ear canal, its spectrum is transformed by reflection and diffraction around the external ear, head and torso. This transformation is known as the head-related transfer function (HRTF)[1]. In the free field, the HRTF is defined as the relation of sound pressure at a point of measurement in the ear canal to the sound pressure that would be measured at a point corresponding to the center of the head while the subject is not present[1]. The HRTF contains some features varying according to the source direction and distance. Changes in those features can be regarded as cues for sound localization[2–8]. This means that various physical cues for sound localization are included comprehensively in HRTFs. Therefore, synthesis of a set of HRTFs is an important technique for virtual auditory displays[1, 9] which could present any virtual sound environments to any listeners. Most of such systems present sound virtual images generated by convolution of a certain sound signal and HRTFs[9]. To synthesize moving sound images coincident with movements of sound sources or a listener's head, a convolution must be processed nearly in real time. Therefore, data reduction of HRTFs is useful for effective implementation of their synthesis from a computational perspective. As a matter of course, it is necessary that the reduction method does not influence sound localization performance.

In general, HRTFs have a directional dependence. However, there might be a frequency region which does not influence sound localization if that region vary a little for sound source directions. For example, the magnitude characteristics of HRTFs qualitatively stay flat in low frequency region. On the other hand, the magnitude characteristics in high frequency region complexly vary according to source directions. The frequency resolution of the human auditory system is generally rough in that region[10]. Therefore, the detailed spectral form of the HRTFs is unlikely to be perceived. Those might enable some reduction of the HRTF data. Some investigations of data reduction methods have been reported [11, 12]. Most of those methods, however, are applied equally to both right-ear and left-ear HRTFs. The frequency characteristics of HRTFs and temporal waveforms of head-related impulse responses (HRIRs) generally show remarkable differences between those measured at right and left ears. Therefore, the data reduction might be applied in the most suitable way for each of them. This concept might engender more effective and flexible methods than conventional ones. Binaural cues play a principal role in auditory localization. Therefore, some kind of data reduction method for the HRTFs only on the contralateral side might be possible if the HRTFs on the ipsilateral side and the binaural cues are accurately reproduced in high frequency region. The HRTFs on the contralateral side have a complicated spectral form, with peaks and dips, and with low energy compared to those on the ipsilateral side. It remains unclear, however, whether the peaks and dips involved in the contralateral–side HRTFs are required as localization cues. Then, it is necessary to investigate how localization performance is influenced when data reduction is applied to a part of frequency region.

In this paper, a subjective evaluation is conducted to investigate whether data reduction without influence of localization performance is possible. The magnitude characteristics of the HRTFs are flattened partially in frequency band. Two experiments are described in following sections; at first, influence of HRTFs flattened in low frequency region is investigate, and the second, influence of HRTFs flattened not only in low but high frequency region. It is remarkable that the flattening in high frequency region is applied only to the contralateral–side HRTFs.

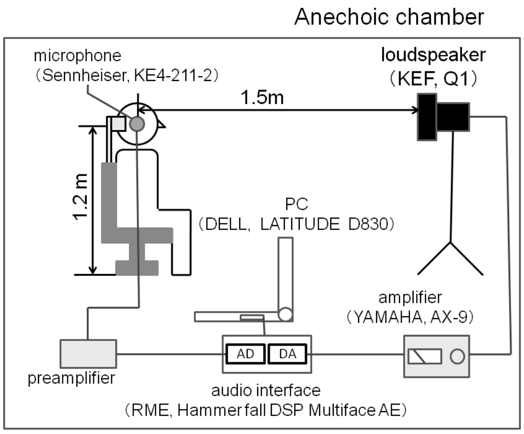
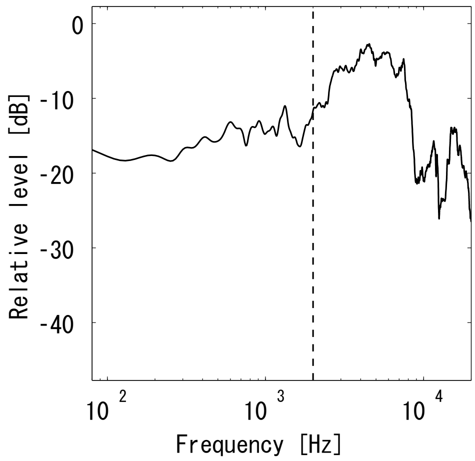
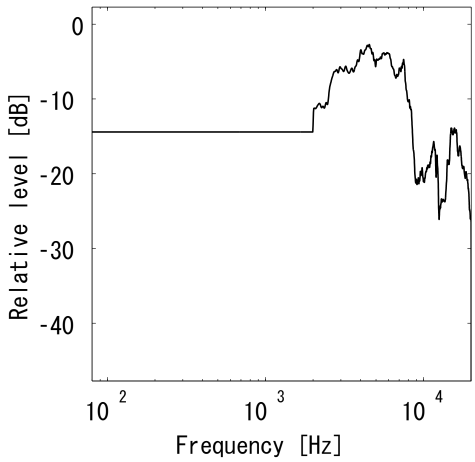


Figure 1: HRTF measurement system



(a) original HRTF



(b) flattened HRTF

Figure 2: Example of the flattening method of the HRTFs in low frequency region (source azimuth: 60°, right ear, boundary frequency: 2 kHz).

FLATTENING METHOD OF HRTFS

HRTF measurement

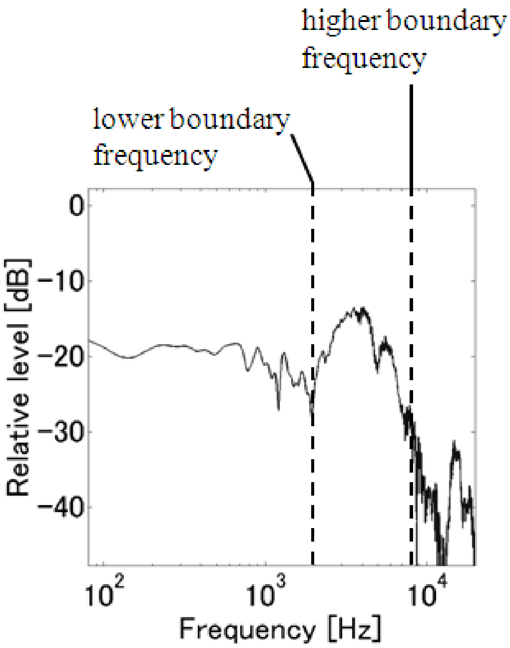
A set of HRTFs for each subject was measured by using TSP (time stretched pulse)[13] with a sampling frequency of 48 kHz. Fig. 1 shows the measurement system. A subject was seated in an anechoic chamber. A loudspeaker (Q1; KEF) was mounted at the position of 1.5 m from the center of the subject's interaural axis. The subject's ear canals were blocked[14], and miniature microphones (KE4-211-2; Sennheiser) were used to pick up the TSP signals. A headrest was mounted on the chair to constrain the subject's head. The chair was rotated so as to measure HRTFs in any azimuths. Anti-alias filtering with cut-off frequency of 20 kHz were applied to each measured HRTFs.

Method of flattening HRTFs in low frequency region

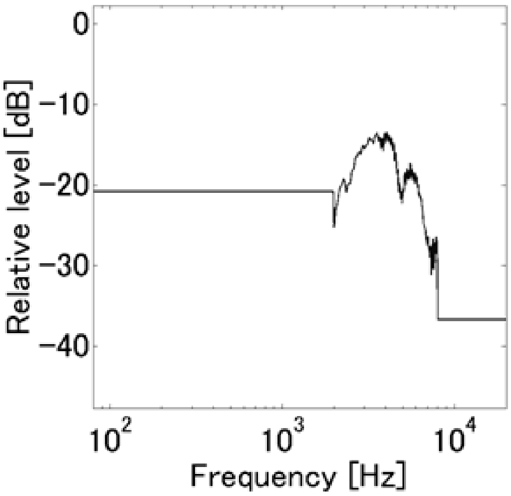
In general, sound waves arrive at both ears with little level difference if they have long wavelengths (low frequency) compared to the size of listeners' head and torso. Therefore, incident sounds have almost the same level in spite of source directions in that frequency region. In other words, the magnitude characteristics of the HRTFs in low frequency region are nearly independent of source directions. This might contribute to some reduction of the HRTF data. In this study, the magnitude characteristics are flattened in the frequency region lower than a certain frequency. This frequency is referred to a "boundary frequency" in this paper. The level of the flattened region is set so as to retain the interaural level difference (ILD) calculated in the overall frequency region. The interaural time difference (ITD) is also maintained by adjusting each time delay included in the processed HRIR to that in the original one. That processing does not accurately retain the frequency characteristics of ILD and ITD to those in the original one; the processing is made so as to retain the binaural cues as much as possible. An example of the method is presented in Fig. 2. Fig. 2(a) shows the frequency characteristic of a certain HRTF as an original. Fig. 2(b) shows that the spectral form in the lower frequency region than the boundary frequency, 2 kHz, is flattened. The relative level in that region is determined to be equal to the average level of the original HRTF in the frequency region corresponding to a flattened one. Consequently, the overall ILD is retained in all directions.

Method of flattening HRTFs in low and high frequency region

In this paper, the influence of flattening HRTFs not only in low but also high frequency region is investigated as well. As mentioned in the previous section, flattening HRTFs in high frequency region might be possible only on contralateral side. Then, the magnitude characteristics of the HRTFs on the contralateral side are flattened in the frequency region higher than a certain frequency. This frequency is referred to a "higher boundary frequency" in this paper. On the other hand, the boundary frequency which is defined as flattening in low frequency region is referred to a "lower boundary frequency." This boundary frequency is the same as one being defined in the previous section. The level of the flattened region is set so as to retain the ILD calculated in the overall frequency region in the same manner as flattening in low frequency region. The ITD is also maintained. An example of the method is presented in Fig. 3. Fig. 3(a) shows the frequency characteristics of a certain contralateral–side HRTF as an original. Fig. 3(b) shows that the spectral form in the lower frequency region than the lower boundary frequency , 2 kHz, and the higher frequency region than the higher boundary frequency, 8 kHz, is flattened.



(a) original HRTF



(b) flattened HRTF

Figure 3: Example of the flattening method applied to contralateral–side HRTFs in low and high frequency region (source azimuth: 60°, lower boundary frequency: 2 kHz, higher boundary frequency: 8 kHz).

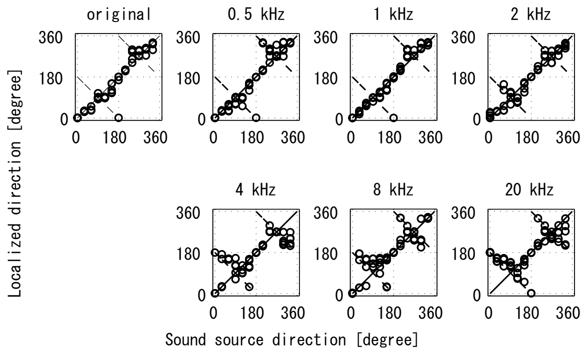
EXPERIMENT I: INFLUENCE OF FLATTENING HRTFS IN LOW FREQUENCY REGION

Outline of the experiment

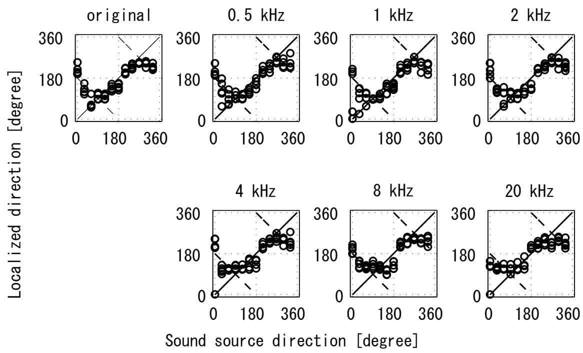
A localization test was conducted to evaluate the influence of flattening the HRTFs in low frequency region. Subjects seated in an anechoic chamber listened to a stimulus with headphones (HD-650; Sennheiser Electronic GmbH and Co. KG). Each stimulus is a sound signal convolved with the subject's own HRTFs or HRTFs with the HRTF flattened below the boundary frequency. The subjects were asked to judge the perceived direction in a horizontal plane.

Table 1: Experimental conditions

subjects	four males and one female
source direction [degree]	0–330 (interval of 30°)
stimulus	pink noise with duration of 1 s
useful frequency region [kHz]	0.08–20
boundary frequency [kHz]	0 (original), 0.5, 1, 2, 4, 8, 20



(a) subject 1



(b) subject 5

Figure 4: Scatterplots of simulated directions versus a subject's perceived directions.

Conditions

Table 1 shows the experimental conditions. Four males and one female subjects with normal hearing acuity participated in the experiment. A set of HRTFs was measured for each subject. In this experiment, the source of 0° is defined as front, and the others are defined in a clockwise rotation. The interval of source directions in which 0° was included was 30° in the horizontal plane. Therefore, the number of source directions was 12. Each source signal was pink noise with duration of 1 s convolved with an HRTF or a flattened HRTF. The conditions of boundary frequency is 0, 0.5, 1, 2, 4, 8, and 20 kHz. The boundary frequency of 20 kHz corresponds to flattening in all frequency region, and that of 0 kHz corresponds to no flattening (referred to "original"). Each stimulus was evaluated five times in random order.

Results

Four of the five subjects showed the similar localization performance. Only subject 5 showed frequently front–back confusion in all conditions including "original". The results for two of the subjects in each condition are presented in Figs. 4(a) and 4(b). The abscissa is the simulated direction and the ordinate is the subject's perceived direction. The boundary frequencies are shown at the top of each panel. In condition "0 kHz (original)", the responses of subject 1 are shown near the diagonal

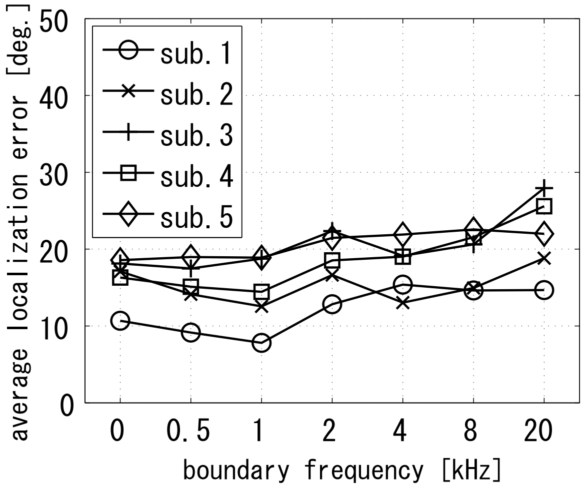


Figure 5: Average localization error for each subject.

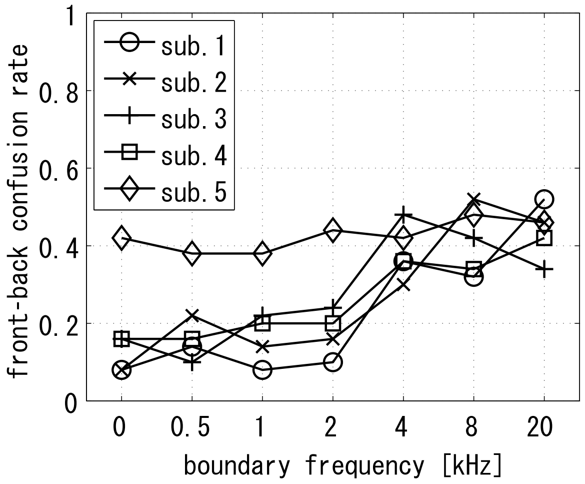


Figure 6: Front–back confusion rates for each subject.

line, while those of subject 5 show that the subject tended to perceive a sound image presented in the frontal direction as being in the rear one, i.e. front–back confusion occurred. Therefore, the tendency of localization is different among subjects. Fig. 4(a) shows that the subject's answers are enormously scattered when boundary frequencies are higher than 4 kHz, i.e. his localization performance degrades.

To evaluate the results of localization for the flattened HRTFs, localization errors and front–back confusion rates are analyzed. In this paper, localization error is defined as the absolute value of the difference between a source direction and a perceived direction from which front–back confusion is extracted in advance. Errors are calculated for all source directions. On the other hand, front–back confusion rates are calculated for all source directions except 90° and 270°. Fig. 5 shows localization errors for each subject as averaged for source directions. The abscissa is the condition of the flattened HRTF; the ordinate is the average localization error. The effects of the flattening, source directions, and subjects were analyzed using analysis of variance (ANOVA). The main effect of the flattened HRTFs was insignificant. The main effect of the source directions was also insignificant. On the other hand, Fig. 6 shows front–back confusion rates for each subject. The abscissa is the condition of the flattened HRTF; the ordinate is the front–

Table 2: Results of the LSD test. The asterisk denotes the boundary frequency conditions in which front–back confusion rate is significantly different from that obtained in condition "original" for each source direction.

source direction [deg.]	boundary frequency [kHz]						
	0	0.5	1	2	4	8	20
0							*
30						*	*
60				*	*	*	*
300					*	*	
330					*		*

Table 3: Experimental condition of boundary frequency [kHz]

lower	0	0.5	1	2	4
higher	4	8	20		

back confusion rates for all source directions except 90° and 270°. As a result of ANOVA of the effects of the flattened HRTF and source directions, the interaction was significant ($F(54,280) = 10.37, p < 0.05$). Because the simple main effect was significant for the flattening when source directions were frontal (0, 30, 60, 300, and 330°), the LSD test was applied for a multiple comparison. The significance for the condition "0 kHz (original)" was shown in Table. 2.

Those results show that the influence of the flattening is different for source directions. At least, the spectral forms of the HRTFs in the frequency region higher than 2 kHz are necessary for front–back discrimination for any source directions.

EXPERIMENT II: INFLUENCE OF FLATTENING HRTFS IN LOW AND HIGH FREQUENCY REGION

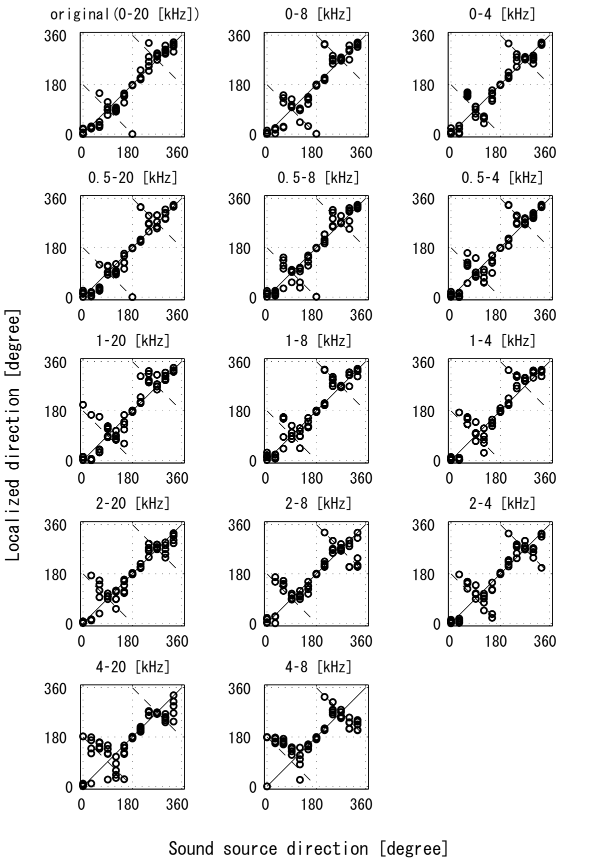
Outline of the experiment

From the results in the previous section, it might be possible to flatten HRTFs in low frequency region without influence of localization. In this section, flattening HRTFs in high frequency region is investigated.

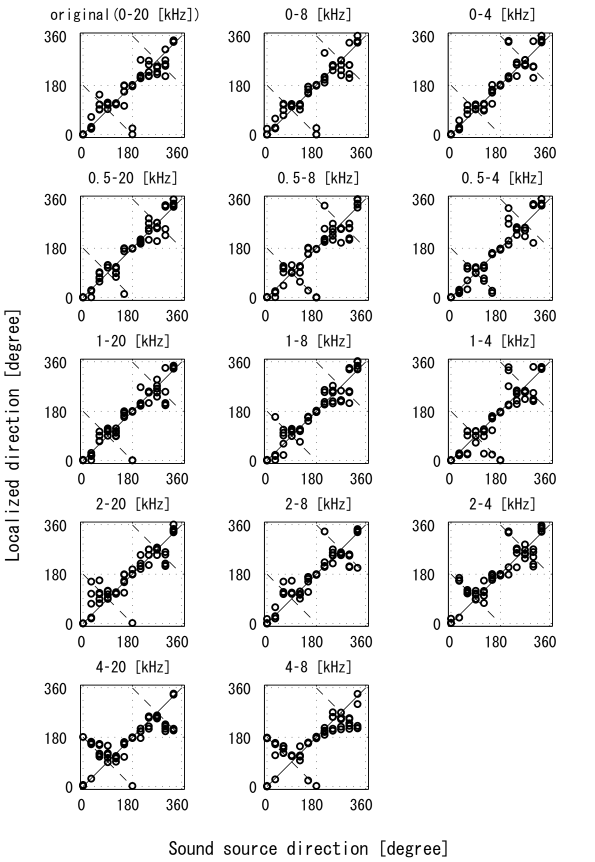
A localization test which is a similar way done in previous section was conducted. The HRTFs on the ipsilateral side were flattened in low frequency region as shown in Fig. 2, while those on the contralateral side were flattened in lower and higher frequency region as shown in Fig. 3.

Conditions

Four males and one female subjects who also participated in the Experiment I participated in the experiment. In this experiment, the conditions of source directions were the same as the Experiment I. Consequently, the contralateral side for right ear corresponds to the source directions, which are larger than 180°, while that for left ear corresponds to those which are smaller than 180°. Flattening was not applied to the HRTFs for two source directions of 0° and 180° because the contralateral side could not be defined. Each source signal was the same used in the Experiment I (pink noise with duration of 1 s). Each stimulus was evaluated five times in random order. Table 3 shows the conditions of the lower and higher boundary frequencies. All combination of lower and higher boundary frequencies was examined except the case of a condition that both frequencies are 4 kHz, i.e. flattening was applied to all frequency region.



(a) subject 1



(b) subject 2

Figure 7: Scatterplots of simulated directions versus a subject's perceived directions. Examples of results of subjects 1 and 2.

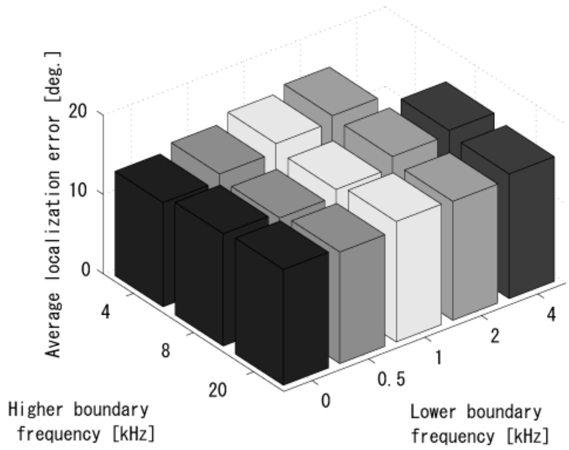


Figure 8: Average localization error.

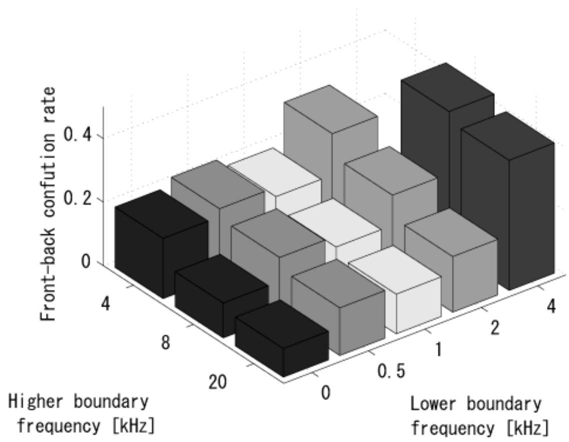


Figure 9: Average front–back confusion rate.

Results

Fig. 7 shows the results of all boundary frequency conditions for subjects 1 and 2 as examples. The abscissa is the simulated direction and the ordinate is the subject's perceived direction. The combination of boundary frequencies (lower–higher) are shown at the top of each panel. If the bandwidth of flattening is narrow, such as conditions of "original" or "1–8 kHz," localization performance is relatively well as the responses of all subjects are shown near the diagonal line. On the other hand, if the bandwidth is become large, such as conditions of "2–4 kHz" or "4–8 kHz," front–back confusions are observed.

Fig. 8 shows the localization error for each boundary frequency conditions averaged for subjects and source directions. Three axes are lower boundary frequency, higher boundary frequency, and localization error, respectively. The combination of a lower boundary frequency of 0 kHz and a higher boundary frequency of 20 kHz corresponds to the condition "original", that is, the HRTFs without flattening were used. Note that the condition "4–4 kHz" was not examined. The effects of combinations of boundary frequencies, source directions, and subjects were analyzed by applying ANOVA. The main effect was insignificant for each factor.

Fig. 9 shows front–back confusion rates averaged for all subjects. Three axes are lower boundary frequency, higher boundary frequency, and front–back confusion rate, respectively. The front–back confusion rates were calculated for all source directions except 90° and 270°. As a result of ANOVA of the effects

Table 4: Results of the LSD test. The asterisk denotes the boundary frequency conditions which is significant difference from "original" for each source direction.

lower boundary frequency [kHz]		2		4	
higher boundary frequency [kHz]		8	4	20	8
	30		*	*	*
source	60	*		*	*
direction	300	*	*	*	*
	330	*	*	*	*

of the flattened HRTF and source directions, the interaction was significant ($F(117,420) = 3.00, p < 0.01$). The analysis of the simple main effect showed that the factor of combinations of boundary frequencies was significant when source directions were 30, 60, 300, and 330°. The LSD test as a multiple comparison was conducted for each source direction. The asterisk in Table 4 denote the combinations of boundary frequencies which is significant for the condition "original."

The results as well as those in previous section show that the influence of the flattening is different for source directions. In the case of lower boundary frequencies, flattening HRTF in frequency region of 1 kHz or lower does not significantly influence localization performance. In the case of higher boundary frequencies, the flattening HRTFs on the contralateral side in frequency region of higher than 4 kHz does not significantly influence if lower boundary frequency is 1 kHz or lower.

GENERAL DISCUSSION

Figs. 5 and 8 present the flattening influence of horizontal localization if the front–back confusion is ignored. It is considered that the spectral forms of the HRTFs do not significantly influence the localization performance. The results might imply that subjects were able to use the ILDs and ITDs as cues for sound localization because they were retained in the flattened HRTFs.

From the results shown in Figs. 6 and 9, front–back confusion rates tend to increase if frequency region to which flattening applied is wide. That is, flattening influence localization performance. Spectral structure of HRTFs is generally known to be important as a cue of front–back discrimination[15, 16]. The results however might imply that it is not necessary to synthesize the spectral structure of HRTFs accurately in all frequency region. The spectral form of the HRTFs in frequency region lower than 1 kHz might enable to be flattened regardless of that in high frequency region because similar results of significance on lower boundary frequency are obtained by statistic analysis shown in Tables 2 and 4. The results of the Experiment II suggest that the flattening of the contralateral-side HRTFs in the frequency region higher than 4 kHz does not influence sound localization when the flattening applied to a low frequency region does not influence.

Those results show that the data reduction of HRTFs might be possible without degradation on localization performance.

CONCLUSIONS

This paper described an investigation on possibility of the data reduction of the HRTFs without influence upon sound localization performance. Two localization tests were conducted using HRTFs partially flattened in frequency characteristics. From the results, it can be concluded as follows:

1. Ignoring front–back confusion, the spectral form of the HRTFs does not influence horizontal localization.
2. Considering front–back confusion, a frequency region of 1 kHz or higher on the ipsilateral side and that between 1 and 4 kHz on the contralateral side are necessary.

Results of our experiment show that data reduction of the ipsilateral- and contralateral–side HRTFs at frequencies lower than 1 kHz and only contralateral–side HRTFs at frequencies higher than 4 kHz is at least possible without a significant effect on localization performance.

Although it remains unclear the influence of flattening for distance localization because subjects was asked to answer only perceived directions. Moreover, the influence of sound quality by transforming frequency spectrum is not taken into account. Further consideration is required in those points.

ACKNOWLEDGEMENT

This work was partially supported by Grant-in-Aid for Young Scientists (B) (21700140).

REFERENCES

[1] J. Blauert. *Spatial Hearing*. The MIT Press, Cambridge, MA, 1983.

[2] F. Asano, Y. Suzuki, and T. Sone. Role of spectral cues in median plane localization. *J. Acoust. Soc. Am.*, 88:159–168, 1990.

[3] A. W. Bronkhorst. Localization of real and virtual sound sources. *J. Acoust. Soc. Am.*, 98:2542–2553, 1995.

[4] J. Kawaura, Y. Suzuki, F. Asano, and T. Sone. Sound localization in headphone reproduction by simulating transfer functions from the sound source to the external ear. *J. Acoust. Soc. Jpn. (E)*, 12:203–216, 1991.

[5] W. L. Martens. Perceptual evaluation of filters controlling source direction: Customized and generalized HRTFs for binaural synthesis. *Acoust. Sci. & Tech.*, 24:220–232, 2003.

[6] M. Morimoto and Y. Ando. On the simulation of sound localization. *J. Acoust. Soc. Jpn. (E)*, 1:167–174, 1980.

[7] M. Morimoto and H. Aokata. Localization cues of sound sources in the upper hemisphere. *J. Acoust. Soc. Jpn. (E)*, 5:165–173, 1984.

[8] M. Morimoto and K. Nomachi. Binaural disparity cues in median-plane localization. *J. Acoust. Soc. Jpn. (E)*, 3:99–103, 1982.

[9] D. R. Begault. *3-D sound for virtual reality and multimedia*. AP Professional, Cambridge, 1994.

[10] B. C. J. Moore. Frequency difference limens for short-duration tones. *J. Acoust. Soc. Am.*, 54(3):610–619, 1973.

[11] A. Kulkarni and H. S. Colburn. Infinite-impulse-response models of the head-related transfer function. *J. Acoust. Soc. Am.*, 115(4):1714–1728, 2004.

[12] M. A. Senova, K. I. Mcanally, and R. L. Martin. Localization of virtual sound as a function of head-related impulse response duration. *J. Audio Eng. Soc.*, 50:57–66, 2002.

[13] Y. Suzuki, F. Asano, H. Y. Kim, and T. Sone. An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses. *J. Acoust. Soc. Am.*, 97:1119–1123, 1995.

[14] H. Møller. Fundamentals of binaural technology. *Applied Acoustics*, 36:172–218, 1991.

[15] J. Hebrank and D. Wright. Spectral cues used in the localization of sound sources on the median plane. *J. Acoust. Soc. Am.*, 56(6):1829–1834, 1974.

[16] M. Morimoto. Role of two ears in upper hemisphere localization. *J. Acoust. Soc. Am.*, 105(2):1163, 1999.