Journal of Information Hiding and Multimedia Signal Processing ©2017 ISSN 2073-4212
Ubiquitous International Volume **8**, Number **6**, November **2017**

Relationship between Localization of Spatially Segregated Sound and Variation in Interaural Level and Time Differences due to Head Rotation

Daisuke Morikawa

Graduate School of Advanced Science and Technology, Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa, Japan
Current address: Faculty of Engineering, Toyama Prefectural University, 5180 Kurokawa, Imizu, Toyama, Japan
dmorikawa@pu-toyama.ac.jp

Received February 2017; revised June 2017

Abstract. *The role of variations in interaural level difference (ILD) and interaural time difference (ITD) due to head rotation in the localization of spatially segregated sounds was investigated by performing listening tests. The participants were asked to distinguish between two sources of white noise having various ILDs/ITDs under head rotation. Under the ILD condition, the segregation rate reached 80% when the ILD between the two sources, at an angular difference of 36°, corresponded to different sides, i.e., left and right hemisphere. However, the sound image was integrated into one when the sources corresponded to the same side. Under the ITD condition, two or three images were perceived regardless of the ITDs. This was because when only one source was used, it was perceived as separate lower- and higher-frequency images. In the experiments using low- and high-pass noises, the lower-frequency image contained frequency components lower than 2.0 kHz and the higher-frequency image contained frequency component higher than 1.7 kHz.*
Keywords: Interaural level difference; Interaural time difference; Spatially segregated sound; Sound localization; Head rotation

1. **Introduction.** We can often hear what people are saying, even in noisy and crowded surroundings. In particular, we have the ability to segregate the sound streams, i.e., separate the target stream from others and group them into a stream. This ability is called the cocktail party effect and has been known since it was first reported by Cherry [1]. Previous researches have investigated the effects of acoustic features of sound, such as the sound direction, timbre, and temporal structure, on this process. It is well known that the interaural level difference (ILD) and interaural time difference (ITD) are cues for sound localization [2]. However, it is not clear as to which acoustic features are the important cues when sound is segregated. Some researchers have reported that differences in direction are particularly important for the cocktail party effect. In addition, it is easy to hear the signals when the signals and maskers have different interaural cues or are spatially separated. These are also referred as to the binaural masking level difference [3] and spatial release from masking [4, 5]. Moreover, interaural cues are used in sound source localization algorithms based on the cocktail party effect [6]. However, paired stimuli (a signal and masker) have different types of timbre such as pairs of voice and random noise.

The participants were able to perceive the two sound images, which had different ILDs or ITDs, even when the two sounds had the same timbre [7]. Meanwhile, it is also

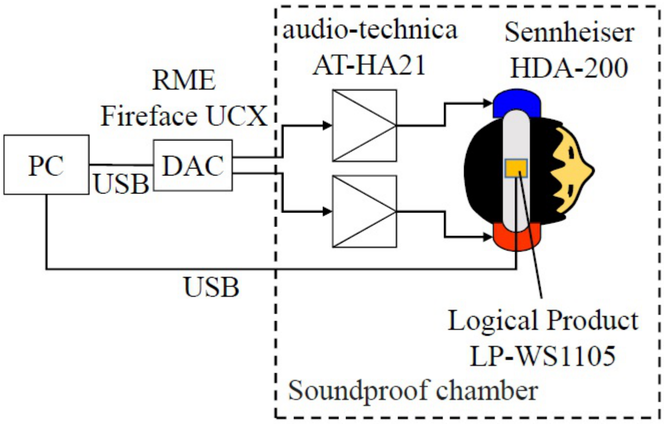

FIGURE 1. Experimental system.

known that the temporal variation of ILD and ITD significantly contributes to sound localization [8]. However, the effectiveness of the temporal variation of ILD and ITD due to head rotation on sound segregation is still not clear.

This paper presents the results of a comprehensive study on understanding the effect of ILD and ITD, caused by head rotation, on sound segregation.

2. Experimentals.

2.1. Experimental system. Figure 1 outlines the experimental system, which consisted of a Windows-based personal computer (PC), digital-to-analog converters (DACs) (RME, Fireface UCX), a headphone amplifier (audio-technica, AT-HA21), headphones (Sennheiser, HDA-200), and a motion sensor (Logical Product, LP–WS1105). The motion sensor was connected to the PC via a USB interface, and it was fastened to the band of the headphones. The sampling frequency of the DACs was 192 kHz and that of the motion sensor was 1 kHz. The apparent sampling frequency of the motion sensor was 200 Hz, because, the sensor sends 5 samples simultaneously. The angular resolution capability was 1°. The experiment was carried out in a soundproof chamber. The background A-weighted sound pressure level of the room was less than 21 dB.

2.2. Stimuli. Two uncorrelated white noise sources (WN1, WN2) were used. The duration of the white noise was 3 s. A 30-ms linear taper window was applied at the beginning and end of the white noise sources.

The PC received the angle of the motion sensor, and switched the ILD or the ITD in response to this angle, in real time. Binaural signals were synthesized from each of the white noise sources. ILD or ITD were generated using the overlap-add method similar to Otani's method [9]. The ILD or the ITD were switched frame by frame. The frames of the left and right channel stimuli are given by Eq. (1) and (2), respectively.

$$S_l = A_l(\theta_1 + \theta_s) \otimes \mathrm{WN1} + A_l(\theta_2 + \theta_s) \otimes \mathrm{WN2} \quad (1)$$

$$S_r = A_r(\theta_1 + \theta_s) \otimes \mathrm{WN1} + A_r(\theta_2 + \theta_s) \otimes \mathrm{WN2} \quad (2)$$

where A_l and A_r are the impulse responses, which are estimated from the delta function δ(t), WN1 and WN2 are the one frame segments of WN1 and WN2, respectively, θ_1 and θ_2 are the configuration angles of WN1 and WN2, respectively, θ_s is the head-rotation angle, and $\otimes$ denotes the convolution. One frame consisted of 4092 points (about 21 ms). The initial value of θ_s was 0° in all the stimuli. The sound pressure level was 70 dB when $\theta_1 + \theta_s$ and $\theta_2 + \theta_s$ were 0°. Figure 2 shows the process of $A_{l|r}(\theta_{1|2} + \theta_s) \otimes \mathrm{WN}$.

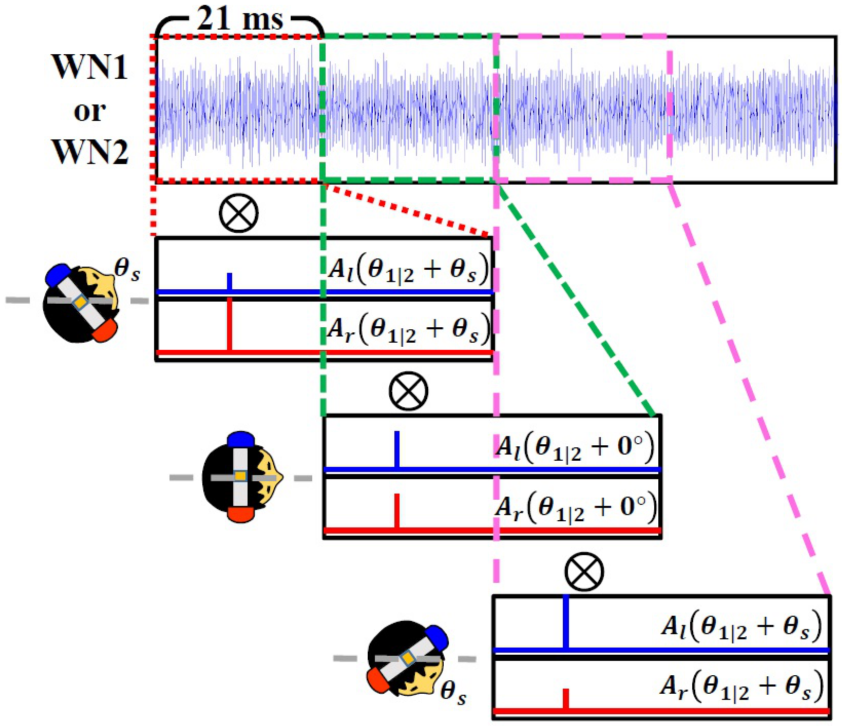

FIGURE 2. Synthesized method.

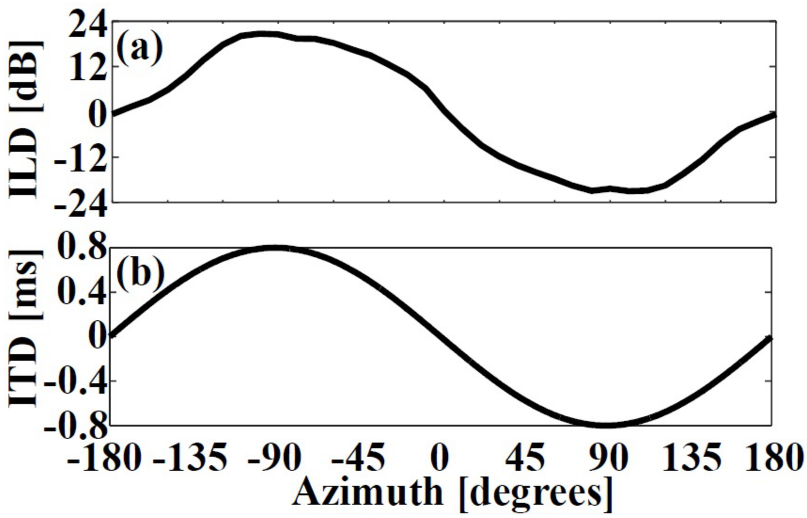

FIGURE 3. Interaural difference (a) the mean of the measured ILDs, (b) ITD model of sphere

Under the ILD condition, the ILD was modified by the amplitudes of A_l and A_r, based on the mean measured ILD. The mean ILD was calculated from the measured head-related transfer functions (HRTFs) [7]. Figure 3(a) plots the mean ILD.

Under the ITD condition, the ITD was modified by the delays of A_l and A_r, based on the ITD model. The ITD model of sphere for each azimuth, proposed by Kuhn [10], was used for the modification. These ITDs are given by Eq.(3), where r is the radius of the sphere, c is the speed of sound, and θ is the stimulus angle.

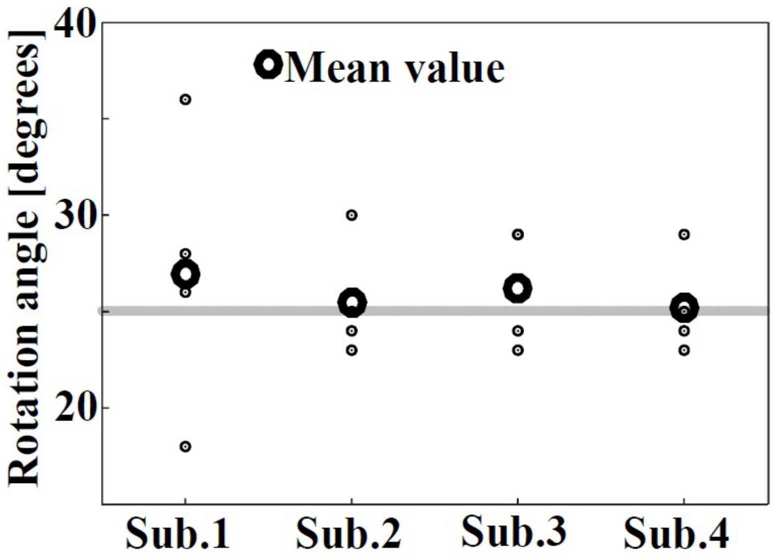

FIGURE 4. Rotation angle of participants when they stop their head.

$$\mathrm{ITD}(\theta) = 3\frac{r}{c}\sin(\theta) \tag{3}$$

Figure 3(b) plots the ITD model when r is 90 mm and c is 340 m/s.

3. ILD Experiments.

3.1. Preliminary experiment. The aim of this preliminary experiment was to find out how the temporal variations of ILD were perceived. The participants were instructed to freely rotate their head in the horizontal direction and were asked to answer the number of perceived sound images, as either one or two. All the participants perceived two separate sounds under a head-still condition [7]. However, when the participants rotated their head considerably, the sound images were sometimes perceived as one sound image. In a case of small θ_1 and θ_2 (at least $|25°|$), particularly, the sound images were always perceived as one sound image.

3.2. Experiment 1. In this experiment, the rotation angle at which there was alternate switching between the segregation and integration of the sound images was investigated. The values of θ_1 and θ_2 were configured as 25° and –25°, respectively. The participants were instructed to turn their heads left or right and stop when the two sound images became one sound image. Four participants with normal hearing and having 23 to 27 years of age participated in this experiment.

Figure 4 plots the rotation angle $|\theta_s|$ when the participants stopped turning their heads. $|\theta_s|$ was about 25°. At this angle, either of the sound sources corresponded to being right in front of the participant. In other words, the angles of A_l and A_r were $|50°|$ and 0°.

3.3. Experiment 2. In this experiment, the differences in head-rotation and head-still in the sound segregation by ILD, when the ILDs corresponded to different sides, was investigated. Here, θ_1 and θ_2 were ±10, 12, 14, 16, 18, 20, 22, 24, 26, and 0°, where 0° was the single sound image stimulus. The experiments included four sessions, and each session consisted of 60 trials. The stimuli were presented in a random order from 10 types of stimuli. Since all the experiments consisted of four sessions, 24 answers were obtained

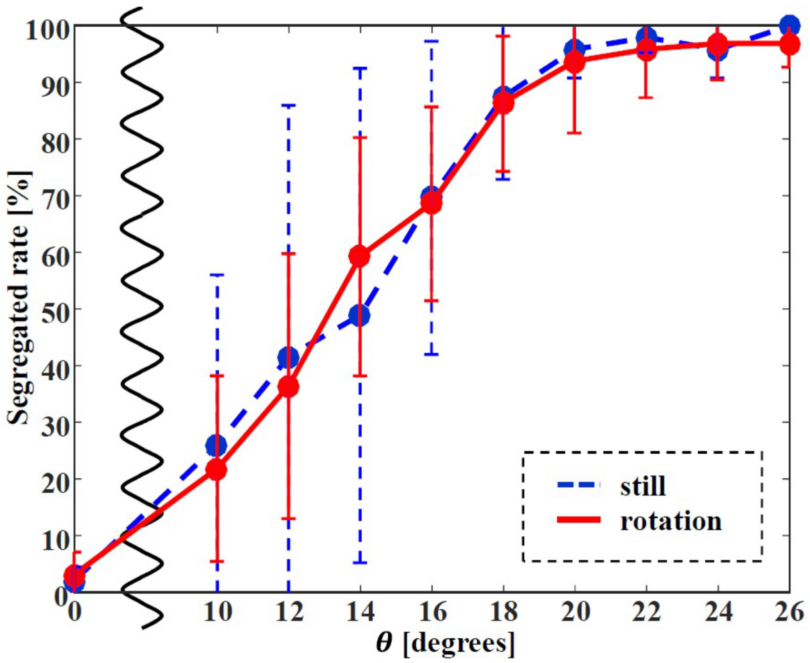

Figure 5. Sound segregation rate under the ILD condition.

for each type of stimuli. The inter-stimulus-interval of each stimuli was 3 s. The same participants who had participated in Experiment 1 participated in this experiment also.

Under head-rotation condition, the participants were instructed to freely rotate their heads in a horizontal direction while the stimulus was being reproduced. Here, no physical restrictions were imposed on their head rotation. In head-still condition, they were instructed to hold their head still while the stimulus was being reproduced. When the motion sensor was turned off, θ_s was always 0°. They were asked to answer the number of the perceived sound images, and to mark either one or two on an answer sheet. They were also asked to answer the number of perceived sound images when their heads were towards the front, and if the number changed when their heads rotated.

Figure 5 plots the rate of marking two (i.e., segregation rate) for each condition by the four participants. Here, θ_1 is the positive theta, and θ_2 is the negative theta. The mean segregation rate reached 80% when $|\theta_1|$ and $|\theta_2|$ were over 18°. This tendency was similar to that in the previous work for head-still condition [7]. In that study, the mean segregation rate reached 80% when $|\theta_1|$ was over 30° and $|\theta_2|$ was 0°. Therefore, an angular difference of over 30° to 36°, between A_l and A_r, was necessary for sound segregation.

There was no significant difference in the segregation rates between head-rotation and head-still conditions. All the participants reported that judging the segregation under head-rotation condition was easier than under head-still condition. The participants were separated into two groups. Two participants rarely gave one sound image as their response, while the other two rarely gave two sound images as their response.

3.4. **Discussion 1.** In Experiment 1, the sound images were integrated when either of the sound sources corresponded to being right in front of the participant. The two sound images were perceived separately when the sound sources corresponded to different sides. This means that the two sound images were integrated into one, when the ILDs corresponded to the same side.

In Experiment 2, two sound images were perceived when the ILD between the two sources had an angular difference of 36°. There was no significant difference between head-rotation and head-still conditions. However, the participants easily judged the segregation by rotating their heads. This means that the various ILDs had some effect on sound segregation.

4. ITD Experiments.

4.1. **Preliminary experiment.** The aim of this preliminary experiment was to find out how the temporal variations of ITD were perceived. The participants were instructed to freely rotate their head in a horizontal direction and were asked to answer the number of the perceived sound images, as either one or two. Two or three sound images were perceived in this condition. In other words, the sound image was not perceived as one in this condition. Therefore, the participants could not answer when they perceived three sound images, and the experiment was discontinued. However, all the participants reported the following results. One lower-frequency sound image and one higher-frequency sound image were perceived, when two sound images were perceived. Two lower-frequency sound images and one higher-frequency sound image were perceived, when three sound images were perceived. The lower-frequency sound images were localized at angles corresponding to θ_1 and θ_2, which were synchronous with the participants' head rotation, and the higher-frequency sound image was localized at the center of the head. In the head-still condition [7], the sound image was not segregated. Therefore, this segregation is considered to be caused by head rotation. In order to discuss the cause of this phenomenon, it is necessary to clarify the components of the lower-frequency sound image and higher-frequency sound image.

4.2. **Experiment 3.** In this experiment, the components of the lower-frequency sound image and higher-frequency sound image were investigated using low-pass and high-pass filtered noises, respectively.

The experiments were conducted using low-pass and high-pass filtered noises, which were reproduced from the convolution of WN1 and the filter. WN2 and θ_2 did not affect S_l and S_r. Therefore, the frames of the left and right channel stimuli are given by Eq. (4) and (5), respectively.

$$S_l = A_l(\theta_1 + \theta_s) \otimes \mathrm{WN1} \otimes \mathrm{filter} \tag{4}$$

$$S_r = A_r(\theta_1 + \theta_s) \otimes \mathrm{WN1} \otimes \mathrm{filter} \tag{5}$$

Here, θ_1 was 0°, and the ITD was modified by the delays of A_l and A_r.

The cut-off frequencies of the low-pass and high-pass filters were 1.0, 1.2, 1.4, 1.7, 2.0, 2.4, 2.8, 3.4, and 4.0 kHz. The noise conditions (e.g., low-pass filtered condition, and high-pass filtered condition) were composed of the 9 filtered-noises and WN1. Four participants with normal hearing and 23 to 30 years of age participated in this experiment. Two of these participants had participated in the ILD experiments, while the other two did not participate in the ILD experiments. The procedures were same as that for the head-rotation condition of Experiment 2.

Figures 6 and 7 plot the mean segregating rate under each condition for the four participants. Fig. 6 indicates that the mean segregation rate increased with cut-off frequency. The result of the two-way repeated-measures analysis of variance of the mean segregation rates, with the factors being the cut-off frequencies and participants, showed a significant difference between these two factors [$F(9, 3) = 2.26$, $p < 0.005$]. The simple main effect of the mean segregation rate for each cut-off frequency was significant [$F(9, 3) = 10.08$, $p < 0.001$]. Table 1 shows the result of the post hoc analysis using Tukey's honestly significant difference test (HSD). When the cut-off frequency was lower than 2.0 kHz, the mean segregation rates were smaller than the mean segregation rates for higher cut-off frequencies, and they were not significantly affected by the cut-off frequency. However, there was no significant difference lower than 2.0 kHz too, since the variances are large. The mean segregation rates for low-pass noises with cut-off frequencies over 2.4 kHz were significantly higher than that of the 1.0 kHz low-pass noise.

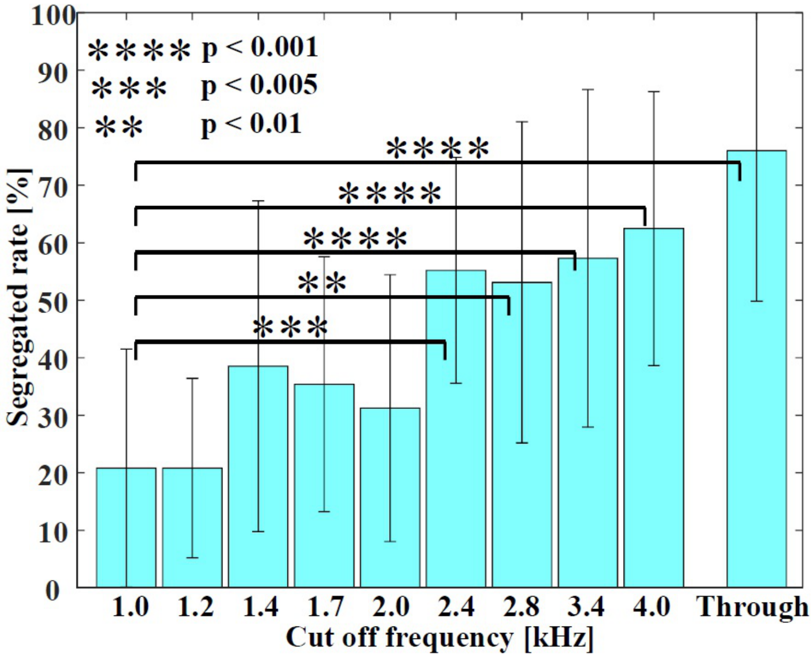

FIGURE 6. Sound segregation rate under low-pass condition.

TABLE 1. Results of HSD under low-pass condition (****: $p < 0.001$, ***: $p < 0.005$, **: $P < 0.01$, *: $p < 0.05$)

	1.0	1.2	1.4	1.7	2.0	2.4	2.8	3.4	4.0	Through
1.0						***	**	****	****	****
1.2						***	**	****	****	****
1.4										****
1.7									*	****
2.0									**	****
2.4	***	***								
2.8	**		**							
3.4	****	****								
4.0	****	****		*	**					
Through	****	****	****	****	****					

Fig. 7 shows that the mean segregation rate dropped with cut-off frequency. The result of the two-way repeated-measures analysis of variance of the mean segregation rates, with the factors being the cut-off frequencies and participants, showed a significant difference between these two factors [$F(9,\ 3) = 2.75$, $p < 0.001$]. The simple main effect of the mean segregation rate for each cut-off frequency was significant [$F(9,\ 3) = 72.46$, $p < 0.001$]. Table 2 shows the result of the post hoc analysis using Tukey's HSD. When the

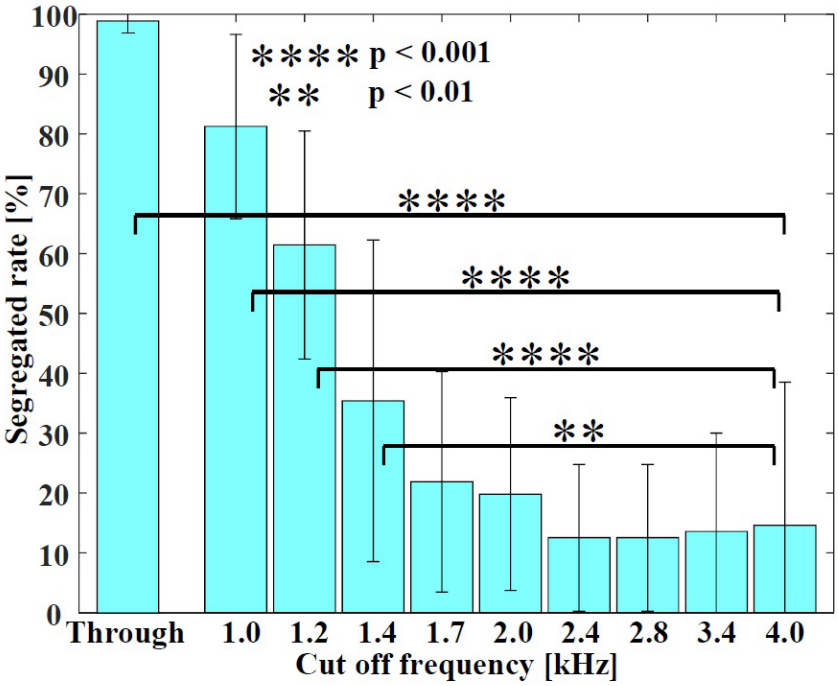

FIGURE 7. Sound segregation rate under high-pass condition.

TABLE 2. Results of HSD under high-pass condition (****: $p < 0.001$, ***: $p < 0.005$, **: $P < 0.01$, *: $p < 0.05$)

	Through	1.0	1.2	1.4	1.7	2.0	2.4	2.8	3.4	4.0
Through		*	****	****	****	****	****	****	****	****
1.0	*		*	****	****	****	****	****	****	****
1.2	****	*		****	****	****	****	****	****	****
1.4	****	****	****				***	***	***	**
1.7	****	****	****							
2.0	****	****	****							
2.4	****	****	****	***						
2.8	****	****	****	***						
3.4	****	****	****	***						
4.0	****	****	****	**						

cut-off frequency was over 1.7 kHz, the mean segregation rates were smaller than the mean segregation rates for lower cut-off frequencies, and they did not show a significant variation with cut-off frequency. The mean segregation rates for high-pass noises with cut-off frequencies under 1.4 kHz were significantly higher than that of the 4.0 kHz high-pass noise.

Figure 8 shows the typical rotation angles of the participants in each session. Each line represents the rotation angle of a participant in each trial. All the participants rotated their head in all the trials, and the mean maximum rotated angle for all the trials was 45°. The head rotation trends were classified into three patterns. In the first pattern, both the head rotation speeds and head rotation directions were almost the same during one session (8(a)). In the second pattern, the head rotation speeds were almost the same, but the head rotation direction was reversed many times (8(b)). In the third pattern, both the

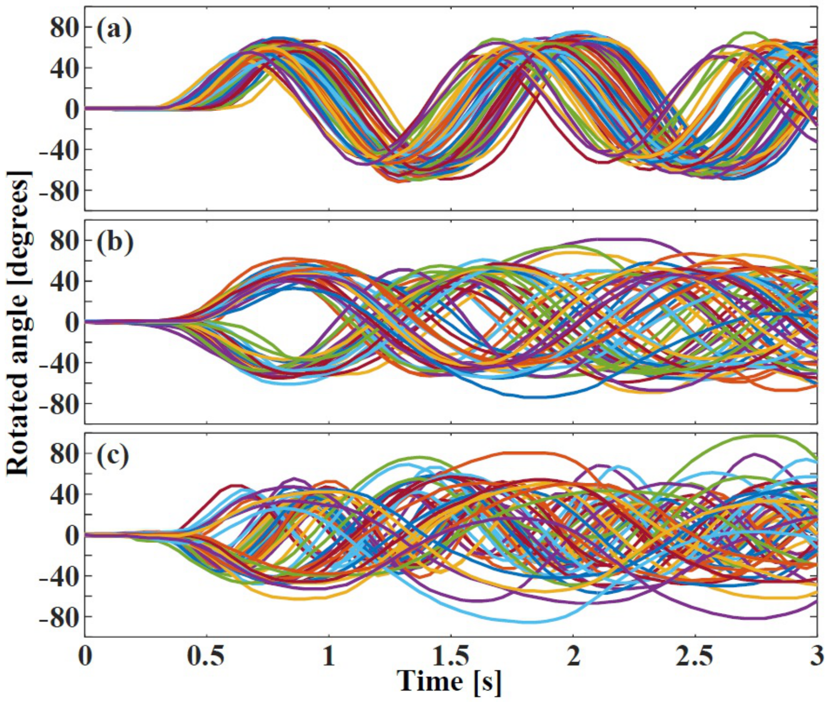

FIGURE 8. Example for typical head rotations.

head rotation speed and head rotation direction were changed randomly (8(c)). However, the results are not affected by the head rotation trends. Merely, the sound images were segregated only if the participants rotated their heads

4.3. **Discussion 2.** In Experiment 3, the lower-frequency sound image contained frequency components lower than 2.0 kHz and the higher-frequency sound image contained frequency components higher than 1.7 kHz. Therefore, the ITD was mainly calculated from frequencies lower than 1.5 kHz, and the ILD was calculated from frequencies higher than 1.5 kHz [2]. In addition, the lower-frequency sound image was localized from the ITD, and the higher-frequency sound image was localized from other cues. It is assumed that the 1.5 – 2.0 kHz components were masked or integrated by the lower-frequency sound image. The higher frequency sound image is not synchronized with the participants' head rotation.

5. **General discussion.** In the ILD condition, the two sound images were integrated into one, when the ILDs corresponded to the same side. This integration does not happen in the modification of ITD condition. These results suggest that in the sound localization strategy, the ITD is more dominant in the perception of sound image segregation, especially on the same side. In other words, the ILD is more dominant in the perception of sound image integration on the same side.

In ITD condition, the lower-frequency sound images were synchronized with the participants' head rotation, and the higher-frequency sound image was localized at the center of the head. This segregation does not happen in the modification of ILD condition. The lower-frequency sound images were localized from the ITD and the higher-frequency sound images were localized from the ILD under the ITD condition. On the other hand, the lower-frequency sound images were not localized from the ITD under the ILD condition. It is assumed that the ILD was preferentially used for sound localization. However, this result was obtained from white noise sources, i.e., the component of the higher-frequency sound image was larger than that of the lower-frequency sound image. Therefore, in order to explain this phenomenon clearly, it is necessary to use various band-limited noises for this experiment.

The knowledge gained from this study can be useful for clarifying the auditory perception ability. However, in this study, white noise was used and the same ILD or ITD were used for all the frequencies. The difference in ILD or ITD for each frequency, and the effect of timbre for sound image segregation and integration, will be the subject of our future works.

6. **Conclusions.** This paper presented the results of a comprehensive study on understanding the effect of ILD and ITD, caused by head rotation, on sound segregation. Sound segregation experiments with two white noise sources, under various ILD and ITD conditions, were conducted under the head-rotated condition. The participants were asked to distinguish between two sources of white noise. Under the ILD condition, the segregation rate reached 80% when the ILD between the two white noise sources, at an angular difference of 36°, corresponded to different sides, i.e., one in the left hemisphere and the other in the right hemisphere. However, the sound image was integrated into one, when the sources corresponded to the same side. Under the ITD condition, two or three images were perceived regardless of the ITDs. This was because, when only one white noise source was used, it was perceived as separate high and low frequency sound images. In the experiments using low- and high-pass noises, the lower-frequency sound image contained frequency components lower than 2.0 kHz and the higher-frequency sound image contained frequency components higher than 1.7 kHz. In this study, white noise was used and the same ILD or ITD were used for all the frequencies. The difference in ILD or ITD for each frequency, and the effect of timbre for sound image segregation and integration, will be the subject of our future works.

Acknowledgment. Part of this work was supported by KAKENHI (15K16022).

REFERENCES

[1] E. C. Cherry, Some experiments on the recognition of speech with one and two ears, *J. Acoust. Soc. Am.*, vol. 25, no. 5, pp. 975–979, 1953.
[2] J. Blauert, *Spatial Hearing*, The MIT press, Cambridge, MA., 1997.
[3] B. C. J. Moore, *An Introduction to the Psychology of Hearing: Sixth Edition*, pp. 271–275, Brill, Leiden, The netherlands, 2012.
[4] M. Ebata, T. Sone, T. Nimura, Improvement of hearing ability by directional information, *J. Acoust. Soc. Am.*, vol. 43, no. 2, pp. 289–297, 1968.
[5] N. Kuroda, J. Li, Y. Iwaya, M. Unoki, M. Akagi, Effects of spatial cues on detectability of alarm signals in noisy environments, *Principles and Applications of Spatial Hearing*, World Scientific, Singapore, pp. 484–493, 2011.
[6] M. Bodden, Modeling human sound-source localization and the cocktail-party-effect, *Acta Acoustica*, vol. 1, pp. 43–55, 1993.
[7] D. Morikawa, Effect of interaural difference for localization of spatially segregated sound, *Proc. IIH-MSP 2014*, pp. 602–605, Kita-kyusyu, Japan, 2014.
[8] H. Wallach, The role of head movements and vestibular and vis-ual cues in sound localization, *J. Exp. Psychol.*, vol. 27, no. 4, pp. 339–368, 1940.
[9] M. Otani, T. Hirahara, Auditory Artifacts due to Switching Head-Related Transfer Functions of a Dynamic Virtual Auditory Display, *IEICE TRANS. FUNDAMENTALS*, vol. E91-A, no. 6, pp. 1320–1328, 2008.
[10] G. F. Kuhn, Model for the interaural time differences in the azimuthal plane, *J. Acoust. Soc. Am.*, vol. 62, no. 1, pp. 157–167, 1977.