



ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Psychological and Physiological Acoustics
Session 3pPP: Multimodal Influences on Auditory Spatial Perception

3pPP5. Impact of dynamic binaural signal associated with listener's voluntary movement in auditory spatial perception

Tatsuya Hirahara*, Daisuke Yoshisaki and Daisuke Morikawa

*Corresponding author's address: Toyama Prefectural University, Imizu, 939-0398, Toyama, Japan, hirahara@pu-toyama.ac.jp

The effect of listener's voluntary movement on the horizontal sound localization was investigated using a binaural recording/reproduction system with TeleHead, a steerable dummy head. Stimuli were static binaural signals recorded with a still dummy-head in head-still condition, dynamic binaural signals recorded with a dummy-head that followed precise or modified listener's head rotation, dynamic binaural signals produced by steering-wheel rotation with listener's hands in head-still condition, and dynamic binaural signals produced by an experimenter in head-still condition. For the static binaural signals, some were localized within the head and the front-back errors often occurred. For the dynamic binaural signals, none of them was localized within the head, and the front-back confusions seldom occurred. Sound images of the dynamic binaural stimuli produced by head rotation were localized out-of-head, while those produced by the steering-wheel rotation or by an experimenter were moving around the listener's head. Listeners could judge the orientation of each stimulus more correctly with dynamic binaural signals produced by listener's head or steering-wheel rotation than with static binaural signals and with dynamic binaural signals produced by an experimenter. Results suggest that the dynamic binaural signal associated with listener's voluntary movement plays a crucial role in sound localization.

Published by the Acoustical Society of America through the American Institute of Physics

INTRODUCTION

We know that two binaural cues, interaural time difference (ITD) and interaural level difference (ILD), and one monaural cue, spectral cue (SC), are the major cues for sound localization [1]. We also know that temporal variation of these cues provided by head movement greatly contributes to sound localization [2–8]. Since temporal variation of ITD, ILD and SC are available, head movement during sound localization decreases front-back confusions and improves the accuracy of sound localization both in horizontal and median-plane sound localization. In addition, head movement diminishes the strict acoustic constraints required for the binaural reproduction system and head-related transfer functions (HRTFs).

Møller *et al.* showed that the headphones should have flat frequency responses and free air equivalent coupling (FEC) to the ears [9]. However, there are few headphones that fulfill these requirements. HRTFs should be measured with the listener's own head, because the shapes of head and pinna are highly individual [10]. However, few listeners have their own HRTFs precisely measured. Accordingly, when we listen to reproduced binaural signals, we often perceive distorted three-dimensional (3D) sound images: frontal sounds are apt to be localized in-head, front-back confusions occur, elevations as well as distance of sound images are uncertain, among other distortions. However, the use of dynamic binaural signals that have temporal variation of sound localization cues resolves these issues on binaural reproduction. When dynamic binaural signals associated with the listener's head movement are presented, sound images are localized at the correct positions even when using non-flat-frequency response and non-FEC headphones [11], non individual HRTFs [11], scaled dummy heads [12], or simple stereo microphones [13], as well as when using narrow-band-limited sounds [14]. Namely, *Move, they shall be localized.*

One of the unanswered questions on the use of dynamic binaural signals is how accurately the listener's head movement must be associated with them. If they have to be tightly linked together, the use of such signals would be limited. If loose association between them could cause the same effect, we would have more choices than ever before to utilize the dynamic binaural reproduction system. Another question is whether the dynamic binaural signal has to be associated with listener's *head* movement or not. In this study, in order to answer to these questions, the effect on the horizontal sound localization of the listener's voluntary head and hand movements was investigated using a real time binaural recording/reproduction system, a steerable dummy head named *TeleHead* [8, 15].

EXPERIMENT

Experimental System

All stimuli used in the experiments were binaural sounds recorded with a dummy head and reproduced with headphones in real time. The dummy head accurately reproduces a real head of an adult male by using a rapid prototyping system. The 3D shape of the real head was measured by a 1.5 T magnetic resonance imaging (MRI) system. The spatial resolution of the MRI image was $1 \times 1 \times 1$ mm, where the image size was 256×256 pixels, the field of view was 256×256 mm, and the slice thickness was 1 mm. The MRI images in DICOM format were transformed to 5-mm-thick hollow 3D head models in STL format. During the format transformation, the back of the outer-ear canals and nostrils were shut, and the head surface was smoothed. TRS-821 epoxy-based resin was used as the photopolymer material in the rapid prototyping system. The dummy-head model was not the real head of the listeners who participated in the experiments but that of another adult male.

The dummy head was put on a steerable dummy-head rotation system *TeleHead IV*, which tracks the listener's head yaw rotation quickly and quietly (Figure 1). The listener's head movement was tracked by a motion sensor (Ascension Technology, Flock of Birds) fixed at the top of the head. The motion delay of *TeleHead IV* was 120 ms, and the operational noise superimposed on the line signal was less than 30 dB SPL from 100 Hz to 20 kHz.

The dummy head put on *TeleHead IV* was placed in an experimental room that had an A-weighted noise floor level of 22 dB and the reverberation time of approximately 50 ms. Twelve loudspeakers (Vifa, MG10SD0908) were circularly placed around *TeleHead IV* at intervals of 30° in the horizontal plane. The distance from the loudspeakers to the dummy-head center was 1 m.

Small electret condenser microphones (SONY, ECM77B) embedded in earplugs made of silicon impression material were placed at the vicinity of the left and right outer-ear canal entrances of the dummy head. The left- and right-microphone output signals were amplified 35 dB by microphone preamplifiers (Earthworks, 1021) and headphone amplifiers (audio-technica, AT-HA20), then reproduced with closed-type dynamic headphones (Sennheiser, HDA200). The gain of the left- and right-channel headphone amplifiers was adjusted so that the sound

pressure level of the headphone reproduced sound, measured with an IEC60711 coupler built in HATS (Brüel & Kjær, 4128C), was that of the 1 kHz tone emitted from the frontal loudspeaker, measured with a free field microphone (Brüel & Kjær, 4169) in the experimental room.

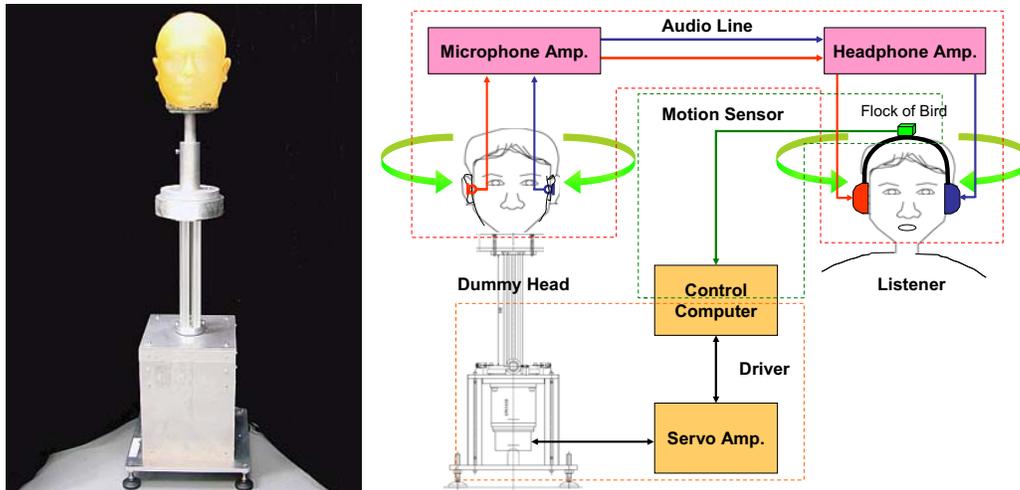


FIGURE 1. Photograph of *TeleHead IV* (left) and schematic diagram of *TeleHead* (right)

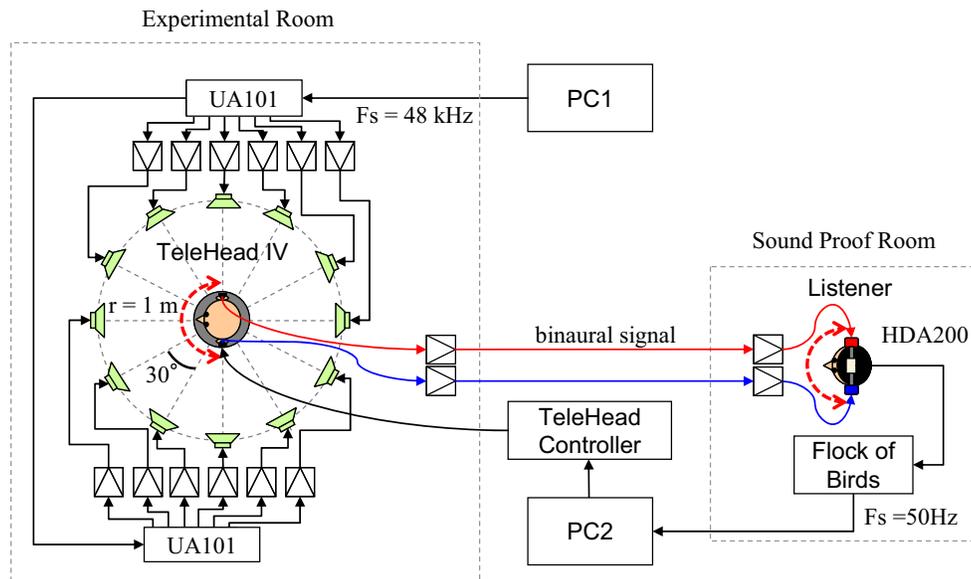


FIGURE 2. Block diagram of the experiment system

Stimuli

Broadband noise, *i.e.*, Gaussian-distributed random noise, was used as the sound stimulus emitted from the speakers. It was generated on a PC with a sampling frequency at 48 kHz, D/A converted by USB audio interface units (Roland, UA-101), amplified by an amplifier (Bose, 1705 II), then fed to the loudspeakers. The binaural signals recorded with the dummy head were transferred to the headphones worn by a listener in real time, without any digital signal processing. The listener thus listened to the sound stimuli through the dummy head's ears, as if he

sat at the center of the speaker array in the experiment room (Figure 2). The sound pressure level of the sound stimuli in the experimental room was set at 70 dB; thus, the binaural stimulus reproduced by headphones was presented to the listener at 70 dB.

Procedure

Sound localization experiments were conducted for the following five conditions separately. Each experiment consisted of four sessions, and every session consisted of 60 trials; broadband noise was randomly presented five times from each of the 12 azimuthal angles. In total, responses were obtained on 20 trials from each of the 12 orientation. Listeners were asked to judge from where the sound was presented and to indicate, on an answer sheet, which of the 12 orientations as well as whether the sound image was in-head.

The first condition was the static binaural condition. The stimuli were binaural signals recorded with the still dummy head by turning off the tracking switch of the *TeleHead*. The listener was instructed to keep his head still during each stimulus presentation. That is, the listener localized the *static* binaural signals keeping his head still.

The second condition was the head-controlled dynamic binaural condition. The stimuli were binaural signals recorded with the rotating dummy head, rotation of which was synchronized with the listener's head movement. The listener was allowed to rotate his head while each stimulus was being presented. That is, the listener localized the *dynamic* binaural signals associated with his voluntary head movement.

The third condition was the hand-controlled dynamic binaural condition. The stimuli were binaural signals recorded with the dummy head, rotation of which was synchronized to a steering wheel rotated by the listener's hands. The motion sensor was placed on the steering wheel. The listener was instructed to keep his head still while each stimulus was being presented. That is, the listener localized the *dynamic* binaural signals associated with his voluntary hand movement keeping his head still.

The fourth condition was the experimenter-controlled dynamic binaural condition. The stimuli were binaural signals recorded with the dummy head, rotation of which was synchronized with the pre-recorded head rotation of an experimenter who listened to the same stimuli in the head-controlled dynamic binaural condition. The listener was instructed to keep his head still while each stimulus is being presented. That is, the listener localized the *dynamic* binaural signals according to the other's head movement keeping his own head still.

The fifth condition was the slowed-down head-controlled dynamic binaural condition. The stimuli were binaural signals recorded with the rotating dummy head, rotation of which was synchronized with the slowed-down listener's head movement. The rotation speed of the dummy head was slowed down to 10% or 50% of that of the listener's head. Thus, the rotation angle of the dummy head was 10% or 50% of that of the listener's head. The listener was, of course, allowed to rotate his head while each stimulus was being presented. That is, the listener localized the slowed-down *dynamic* binaural signals associated with his voluntary head movement.

Six male listeners participated in the experiments for the first four conditions. Four different male listeners participated in the experiment for the fifth condition, as well as for the first and second conditions. All listeners were in their twenties with normal hearing abilities.

RESULTS

Figure 3 shows the pooled results of four conditions for six listeners. In each panel, the area of the blue-filled circles is proportional to the in-head localization rate, while that of the red-filled circles is proportional to the out-of-head localization rate.

In the static binaural condition, most stimuli were localized out-of-head, whereas a number of sound images of the stimuli presented at the front position were localized in-head. Some front-back confusions and near-miss orientation-judgment errors to adjacent orientations occurred. Accordingly, the mean correct-sound-source orientation-judgment rate, \overline{P}_c , was only 53%.

In the head-controlled dynamic binaural condition, listeners perceived stationary sound images of the stimuli in the spatial coordinate system. Since the dummy head rotated synchronously with the listener's head rotation, in terms of dynamics, the binaural sounds presented to a listener's ears via headphones were the same dynamic binaural sounds that arrived at the ears directly from the loudspeaker. As a result, almost all in-head localization and many front-back confusions disappeared; however, near-miss orientation-judgment errors remained, resulting in a \overline{P}_c value of 81%.

In the hand-controlled dynamic binaural condition, listeners perceived moving sound images around the head. Binaural signals presented to the listener's ears were dynamic since the dummy head was synchronously rotated with the listener's handling of the steering wheel. Keeping his head still altered the perception that sounds were moving around his head. Using those moving sound images, listeners could judge from which orientation each stimulus was presented, because they knew that the sound images moved according to their handling of the steering wheel. As a result, front-back confusion in this condition was as low as that in the head-controlled dynamic binaural condition, resulting in a \bar{P}_c value of 85%.

In the experimenter-controlled dynamic binaural condition, listeners perceived moving sound images around the head. Binaural signals presented to the listener's ears were dynamic, but the dummy-head rotation was controlled by an experimenter rather than by the listener himself. Thus, listeners thought that the stimuli themselves were moving sound sources. In this condition, sound images were not localized in the head, front-back confusions seldom occurred, and \bar{P}_c was 69%.

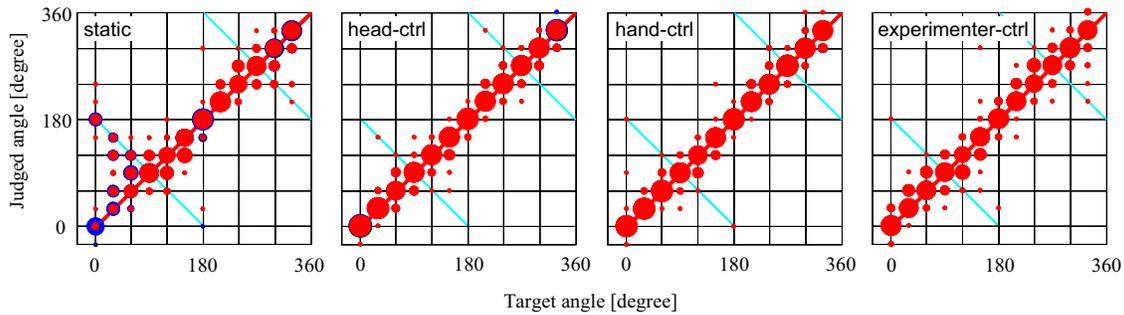


FIGURE 3. Pooled results for six listeners of the sound localization experiment for four conditions.

Figure 4 shows the mean correct-sound-source orientation-judgment rates, \bar{P}_c , and standard deviations (left); front-back error rates and standard deviations (middle); and near-miss orientation-judgment error rates and standard deviations (right) for the six listeners for each condition. \bar{P}_c of the static binaural condition is significantly lower ($p < 0.005$) than that of the head- and hand-controlled dynamic binaural conditions. These differences resulted from the decrease in the front-back errors. \bar{P}_c of the experimenter-controlled dynamic binaural condition is significantly lower ($p < 0.05$) than that of the hand-controlled dynamic binaural condition, due to a decrease in near-miss orientation-judgment errors. By contrast, no significant difference was found between either the static and the experimenter-controlled dynamic binaural conditions or the head- and hand-controlled dynamic binaural conditions.

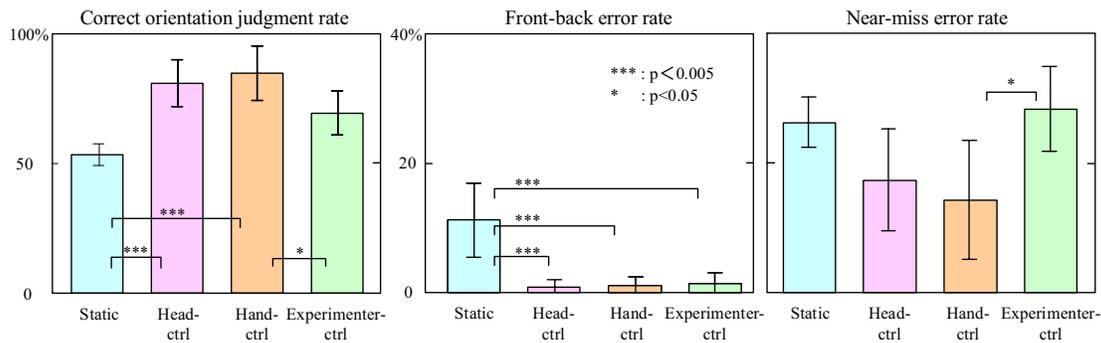


FIGURE 4. Mean correct sound source orientation judgment rates \bar{P}_c (left); mean front-back error rates (middle); and mean near-miss orientation judgment error rates (right), with standard deviations for the six listeners for each condition.

In the slowed-down head-controlled dynamic binaural condition, listeners perceived stationary sound images. Figure 5 shows the pooled results of the sound localization experiment for the fifth conditions for four listeners. The left panel shows the results of the static binaural condition. The middle two panels are those for the head-controlled dynamic binaural condition with slowed-down synchronization speed at 10% and 50% of the listeners head movement. The right panel shows the results of the normal head-controlled dynamic binaural condition.

The pooled result for the static and head-controlled dynamic binaural conditions was similar to that of the previous experiment. In the slowed-down synchronization speed conditions, listeners perceived stationary sound images of the stimuli in the spatial coordinate system; the front-back localization error decreased, but some in-head localizations remained.

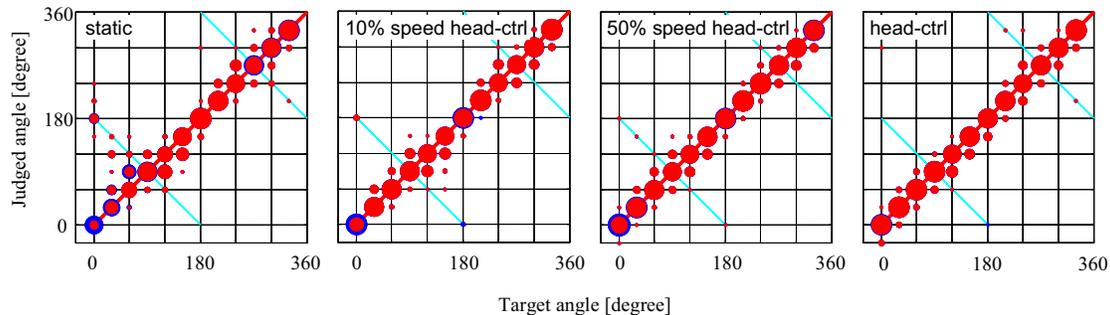


FIGURE 5. Pooled results of the sound localization experiment; static binaural (left); dynamic binaural condition with slowed-down synchronization speed to the listener's head movement (middle two); and dynamic binaural conditions for four listeners.

Figure 6 shows \bar{P}_c s and standard deviations for the four listeners for each condition. \bar{P}_c of the static binaural condition was 60%, which is higher than that in Figure 4 because of the listener differences. \bar{P}_c of the dynamic binaural condition was 80%, which was almost the same as that in the previous experiment. The difference of \bar{P}_c 's between the static and dynamic binaural conditions was highly significant ($p < 0.005$). \bar{P}_c of the slowed-down head-controlled dynamic binaural condition was 78% for 50% speed and 72% for 10% speed. Both these values were also significantly higher than those for the static binaural condition. The differences resulted from decreases in in-head localization and front-back confusion because of head rotation, regardless of the synchronization speed. Near-miss orientation-judgment errors, however, did not differ among the four conditions.

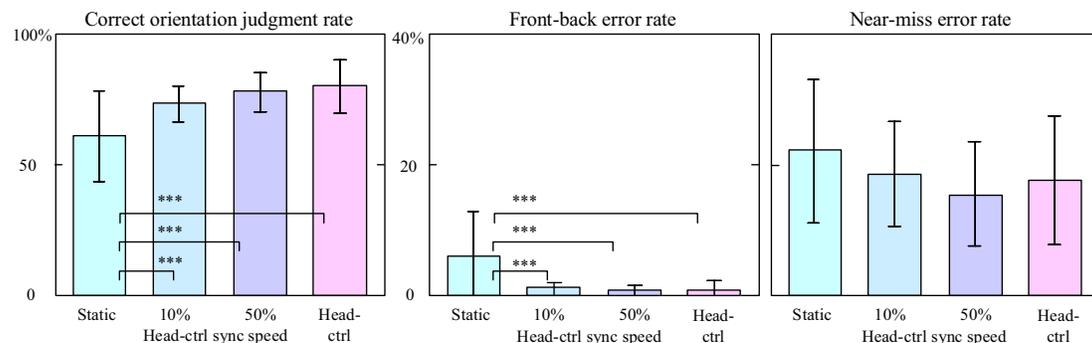


FIGURE 6. Mean correct sound source orientation judgment rates \bar{P}_c (left); mean front-back error rates (middle); and mean near-miss orientation judgment error rates (right), with standard deviations for the four listeners for each condition.

DISCUSSION

Localization performance for the static binaural condition was poor, as the brain has to *calculate* sound-source orientation from only the auditory information, which involves static binaural cues. Part of this poor performance can be attributed to the use of another dummy head as well as the use of non-ideal headphones for binaural reproduction systems. The frequency response of the headphones used in the experiments was not flat and they are not FEC to the ears. The listeners had normal sound localization abilities because \overline{P}_c was 98% when they listened to the sound stimuli directly with their real heads and ears sitting at the center of the speaker array.

The result for the head-controlled dynamic binaural condition reconfirms the impact of the head movement on sound localization. That is, the listeners can judge sound source orientation more accurately with dynamic binaural signals associated with the listener's voluntary movement than with static binaural signals. The brain can use consistent auditory, neck somatic/kinetic, and equilibrium information to *calculate* the sound source orientation.

The result for the slowed down head-controlled dynamic binaural condition indicates that the listener can judge sound source orientations more accurately with these dynamic binaural signals than with static binaural signals, even when the synchronization speed was slowed down to 10% of the head movement. This result suggests that the association between the dynamic binaural signal and a listener's head movement is necessary, but sound localization does not necessarily require a very accurate association. In this case, the brain can use less consistent auditory, neck somatic/kinetic, and equilibrium information to calculate sound-source orientation compared to the fully synchronized case.

The result for the hand-controlled dynamic binaural condition indicates that the listener can judge the orientation of the sound source fairly well from moving sound images when the dynamic binaural signals are synchronized to the listener's voluntary rotation of the steering wheel. In this case, the brain can use the auditory as well hand somatic/kinetic information, but neck somatic/kinetic and equilibrium information were unavailable as the listener kept his head still. The result implies that the dynamic binaural signals need not necessarily be associated with the listener's *head* movement to perform sound localization. The listener's voluntary *hand* movement makes sound localization easier, even though the sound image itself moves around his head.

By contrast, the result for the experimenter-controlled dynamic binaural condition indicates that the listener cannot judge sound-source orientations accurately. This situation is similar to that of the static binaural condition. As the dynamic binaural signals were not synchronized to a listener's voluntary movement and the head was kept still, the brain could only use the auditory information. Although the auditory information involves the dynamic cues for sound localization, the listener could not judge the sound-source orientation accurately without neck somatic/kinetic and equilibrium information. Auditory information not associated with listener's voluntary movement must be less informative for the brain. Since it is natural in daily life for us to listen to sounds by moving our head and body, the brain must have developed under such an environment. The more consistent are the multimodal source of information, the more informative they are for informing the brain about events in the world.

SUMMARY

The effect of a listener's voluntary movement on horizontal sound localization was investigated. Listeners could judge the orientation of each sound source more accurately with dynamic binaural signals associated with the listener's voluntary head or steering-wheel rotation than with static or dynamic binaural signals not associated with the listener's movement. In addition, dynamic binaural signals produced by the listener's slowed-down voluntary head rotation also made the sound-source orientation judgment more accurate. These results suggest that the dynamic binaural signals associated with a listener's voluntary movement play a crucial role in sound localization.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Number 22300061.

REFERENCES

1. J. Blauert, *Spatial Hearing* (The MIT Press, Cambridge, Mass., 1997).
2. H. Wallach; "The role of head movements and vestibular and visual cues in sound localization," J. Exp. Psychol. **27**, 339–368 (1940).

3. M. Wenzel; "Localization in virtual acoustic displays," *Presence*, **1**, 80-107 (1992).
4. S. Perrett, W. Noble; "The contribution of head motion cues to localization of low-pass noise," *Perception and Psychophysics* **59**, 1018-1026 (1997).
5. F. Wightman, D. Kistler; "Resolution of front-back ambiguity in spatial hearing by listener and source movement," *J. Acoust. Soc. Am.* **105(5)**, 2841-2835 (1999).
6. D. Begault, E. Wenzel, M. Anderson; "Direct comparison of the impact of head-tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," *J. Audio Eng. Soc.* **49(10)**, 904-916 (2001).
7. Y. Iwaya, Y. Suzuki, D. Kimura; "Effects of head movement on front-back error in sound localization," *Acoust. Sci. & Tech.* **24(5)**, 322-324 (2003).
8. I. Toshima, S. Aoki, T. Hirahara; "Sound Localization Using an Acoustical Telepresence Robot: TeleHead II," *Presence* **17(4)**, 365-375 (2008).
9. H. Møller, D. Hammershøi, C. Jensen, M. Sørensen; "Transfer Characteristics of Headphones Measured on Human Ears," *J. Audio Eng. Soc.* **43(4)**, 203-217 (1995).
10. H. Møller, M. Sørensen, C. Jensen, D. Hammershøi; "Binaural Technique: Do We Need Individual Recordings?" *J. Audio Eng. Soc.* **44(6)**, 451-469 (1996).
11. T. Hirahara, M. Otani, I. Toshima; "Issues on HRTF Measurement and Binaural Reproduction," *Fundamentals Review* **2(4)**, 68-85 (2009).
12. T. Hirahara, H. Sagara, M. Otani; "Sound localization with scaled dummy-heads on a TeleHead," *Proc. 19th ICA Madrid*, PPA-05-013, 1-4 (2007).
13. T. Hirahara, Y. Sawada, D. Morikawa; "Impact of dynamic binaural signals on three-dimensional sound reproduction," *Proc. Inter-Noise 2011*, SS20-Mon4-18-1-6 (2011).
14. D. Morikawa, T. Hirahara; "Effect of head rotation on horizontal and median sound localization of band-limited noise," *Acoust. Sci. & Tech.* **34(1)**, 56-58 (2013).
15. I. Toshima, H. Uematsu, T. Hirahara; "A steerable dummy head that tracks three-dimensional head movement: TeleHead," *Acoust. Sci. & Tech.* **24 (5)**, 327-329 (2003).