大阪 inter.noise 2011
Osaka Japan September 4-7

Impact of dynamic binaural signals on three-dimensional sound reproduction

Tatsuya Hirahara[1], Yuki Sawada[2], and Daisuke Morikawa[3]

[1-3] Toyama Prefectural University

5180 Kurokawa, Imizu, Toyama 939-0398 JAPAN

ABSTRACT

Binaural technology enables the reproduction of the three dimensional (3-D) sound through earphones but requires an acoustically correct sound reproducing system and high-precision individual head-related transfer functions (HRTFs) for spatially distortion-free 3-D sound reproduction. These requirements for the acoustical strictness of binaural systems, however, are based on experimental results of sound localization experimental results using static binaural signals. We used the *TeleHead*, a rotatable dummy head system that quietly and quickly tracks a listener's yaw head movement, to carry out a series of horizontal/median plane sound localization experiments in head-movement condition. Binaural signals were provided using stereo-microphones that only provide interaural time difference (ITD), a simple dummy head without pinnae that provides ITD and interaural level difference (ILD), or personal dummy heads that provide ITD, ILD and spectral cues, *i.e.*, HRTFs. Results showed that the use of dynamic binaural signals, in which a listener's voluntary head movement is reflected as time-varying binaural cues, improved 3-D sound localization. It was also confirmed that the use of the dynamic binaural signals can considerably ease the acoustical requirements in reproducing spatially distortion-free 3-D sound.

Keywords: dynamic binaural signal, 3-D sound localization, head movement, TeleHead

1. INTRODUCTION

The sounds we hear in the real world come from all directions and we have the ability to localize them to avoid danger, hunt prey, or find a mate. By contrast, reproduced sounds somewhat lack 3-D information compared to those we hear in the real world. The sounds played back with earphones or headphones are lateralized between the ears; they sound as if they are inside the head or at the ears. The sounds played back with a 2-channel stereo system are localized between the two speakers. The sounds played back with a 5.1-channel system surround the listener but their localizations are vague.

Modern 3-D sound technologies have been developed to reproduce sound with rich spatial information. Transaural systems control sound pressure at a point, e.g., a listener's ear, with loud-speakers. They, however, have underlying problems in which the inverse filters for cancelling cross-talk are not always stable and the sound source locations must be given. Alternative technologies are multi-channel speaker rendering and wave-field synthesis, which control the sound field around a listener. These technologies can reproduce perceptually good 3-D sound, but they require large-scale equipment and signal processing.

Another 3-D sound technology is a binaural system, which uses earphones instead of loud speakers. The binaural system controls sound pressure at a listener's ear. Unlike a transaural system, the reproduced sound with a binaural system can always be well defined and stable since an earphone generates sound pressure in the ear canal. A binaural system, however, has several acoustical constraint conditions for properly reproducing binaural signals; the head-related transfer functions (HRTFs) must be accurate, the earphones must exhibit free air equivalent coupling (FEC) to the ear, and earphones must have a flat frequency characteristics [1-3]. These are problematic

[1] hirahara@pu-toyama.ac.jp

acoustical constraints. Because an HRTF is highly unique for each person, measuring individual HRTFs accurately is difficult, and there are few FEC earphones available. The actual ear response of an earphone is unique for each ear, and designing an individual equalizer for flattening the earphone actual ear response is also difficult. These constraints may be controllable in a laboratory, but neither accurate HRTFs nor appropriate earphones can practical.

It has been known since 1939 that the head movement can improve localization [4-7]. Our auditory system somehow integrates the temporal changes in interaural time difference (ITD), interaural level difference (ILD), and spectral cues (SCs) that occur with head movement over time. It is often said that we move our head to minimize the interaural differences when we localize a target sound source [8, 9]. This is not necessarily true, because we do NOT always turn our head toward the target sound source in front of us and small head movements help reduce the sound location ambiguities [10].

We present clear evidence that the use of the dynamic binaural signals synchronous to listeners' head movement can considerably ease the acoustical constraints of the binaural system based on a series of sound localization experiments.

2. DYNAMIC BINAURAL SIGNAL

2.1 Cues for Sound Localization

The human auditory system uses several acoustical cues for static-sound localization. One is ITD calculated from lower frequency components below 1.5 kHz, and another is ILD calculated mainly from higher frequency components above 1.5 kHz. The spectral notches appearing in the mid-frequency range, *e.g.*, 4 to 14 kHz, of the amplitude spectrum of the HRTF, are known as spectral cues for the median plane sound localization. These are the well known static cues for sound localization [8]. When a listener moves, he/she moves his/her head, or when the sound source itself moves, all the cues for sound localization change momentarily. The changes in ITD, ILD and spectral cues, *i.e.*, ΔITD, ΔILD, and ΔSC, can be called dynamic cues for sound localization. They are used to localize not only moving sound but also static sound, as the head movement alters physically static sound to perceptually dynamic sound.

2.2 Rotatable Dummy head System: *TeleHead*

The *TeleHead* is a rotatable dummy head system, which tracks a listener's head movement in real time without generating loud mechanical noise [11-14]. Unlike digital binaural systems based on the HRTF convolution to the signals, the *TeleHead* provides perfect dynamic binaural signals. Neither reduction nor simplification of the binaural signals is provided using the *TeleHead*, whereas digital binaural systems do this implicitly. As shown in Figure 1, two small microphones placed at the entrance of outer-ear canals of the dummy head receive actual binaural signals. The binaural signals are just amplified and presented to a listener's ear through headphones.

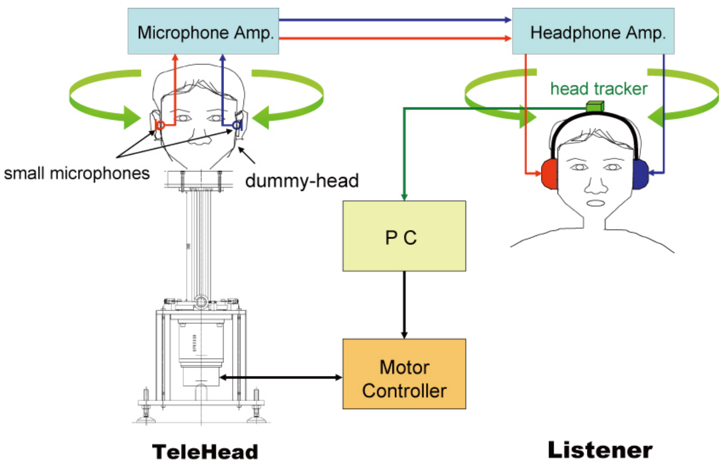

Figure 1 – Schematics of *TeleHead*, which is rotatable dummy head that quietly tracks listener's head movement in real time.

Since there is no digital signal processing stage, factors that deteriorate the binaural signals are transfer characteristics of microphones, microphone amplifiers, headphone amplifiers, and headphones. The transfer characteristics are fairly flat between 20 Hz and 20 kHz, except for headphones. Another factor that deteriorates the binaural signals of the *TeleHead* is the rotational delay caused by the servo loop of the motor controller. The rotational delay is about 100 ms, which is just beyond its perceptual threshold [15].

2.3 Stereo Microphones

Stereo microphones used in this study were two small electret-condenser microphones spaced at 30 cm from each other, which is the geodetic distance between two actual ears (Figure 2 left). The supporting arms of the stereo microphones are made of stainless steel round bars of 5−6 mm ø. The arms are acoustically negligible and do not vibrate mechanically due to rapid rotation.

In the stereo microphones, ITD varied ±900 μs and ILD varied ±3.0 dB with azimuthal angles, shown as red lines in Figure 3. In contrast, ITD was exactly 0 μs and ILD was almost 0 dB for any elevation angle in the median plane.

2.4 Simple Dummy head

The simple dummy head used in this study was an ellipsoidal dummy head made of Styrofoam, originally used for displaying wigs (Figure 2 middle). The dummy head had neither nose nor pinnae. The surface of the dummy head was covered with paper clay to adjust the distance between two small electret-condenser microphones embedded in the lateral side to be 16.5 cm.

In the simple dummy head, ITD varied ±750 μs and ILD varied ±15 dB with azimuthal angles, shown as the blue lines in Figure 3. ITD was almost 0 μs and ILD was almost 0 dB for any elevation angle in the median plane.

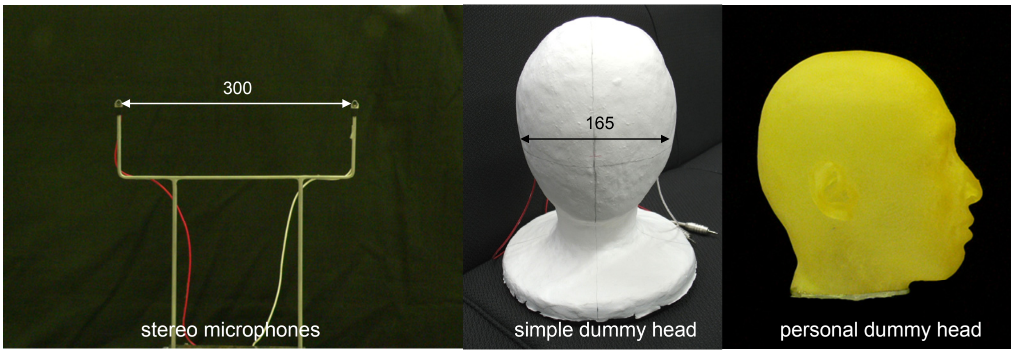

Figure 2 – Photographs of stereo microphones, simple dummy head, and personal dummy head used in experiment.

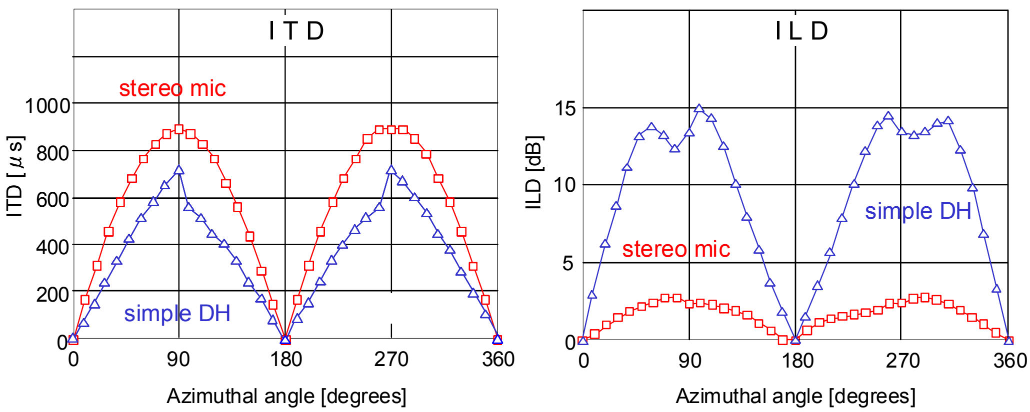

Figure 3 – ITD and ILD variations of stereo microphones (red) and simple dummy head (blue) with azimuthal angles.

2.5 Personal Dummy heads

The personal dummy heads used in this study were made of hard resin dummy heads fabricated using a rapid prototyping system (Figure 2 right). The 3D shapes of two subjects' actual heads were measured using a 1.5-T magnetic resonance imaging (MRI) system. The spatial resolution of the MRI image was 1 × 1 × 1 mm, where the image size was 256 × 256 pixels, field of view was 256 × 256 mm and slice thickness was 1 mm. MRI images in DICOM form were transformed to 5-mm-thick hollow 3D head models in STL format. During the format transformation, the back of the external-ear canal and nostrils were shut, and the head surface was smoothed. Rapid Meister 6000 (Cmet) was used to make the personal dummy heads. The photopolymer materials used in the rapid prototyping system was the TRS-821 epoxy-based resin.

Each personal dummy head and its model head had the same ITD and ILD characteristics and very similar HRTFs. In both types of heads, ITD varied by about ±720 μs and ILD varied by about ±15 dB with azimuthal angles. ITD was almost 0 μs and ILD was almost 0 dB for any elevation angle in the median plane.

3. EXPERIMENTS

3.1 Method

Figure 4 and 5 illustrate the experimental setup. The speaker array consisted of twelve loudspeakers (Vifa, MG10SD0908) placed in a horizontal circle and five loudspeakers placed on the upper-semicircle at intervals of 30 degrees. The diameter of the circles was 1 m and the height of the horizontal speakers was 1.1 m. A subject sat or the *TeleHead* was placed at the center of the speaker array in an experimental room.

Sound stimulus reproduced from a loudspeaker was received by two microphones. The received binaural signals were amplified and transmitted to another experimental room. Finally, the binaural signals were presented to a subject through headphones. The stimulus was 3 seconds white noise generated independently for each trial. Each stimulus was D/A converted (Fs=48 kHz, 24 bits) and presented in random directions at 3-seconds intervals. The sound pressure level of each stimuli produced by the loudspeakers was 80 dB, while that of the binaural stimuli reproduced at the subject's ears was adjusted to 70 dB, which was measured with an IEC60711 coupler. The headphones used were HDA200 (Sennheiser) [16, 17].

The experiment conditions were the combinations of three binaural-signal conditions and two head-movement conditions for the horizontal and median plane conditions. The binaural signals were obtained through the stereo microphones, the simple dummy head, or the personal dummy head. The head-movement conditions were the head-still and the head-movement conditions. Localization experiments using one's ears in both conditions were also conducted as a reference.

The horizontal and median plane localization experiments were carried out separately. For horizontal plane localization, each session consisted of 60 trials; stimulus was presented 5 times in random order from each of the 12 directions. One experiment consisted of 4 sessions, resulting in responses of 20 trials from each of the 12 directions. For median plane localization, each session consisted of 35 trials; stimulus was presented 5 times in random order from each of the 7 directions. One experiment consisted of 4 sessions, resulting in responses of 20 trials from each of the 7 directions.

Five males participated in the experiments involving their own ears, stereo microphone and simple dummy head conditions. For the personal dummy head condition, only two males participated in the experiment because others did not have personal dummy heads made for them. Subjects were asked to localize each stimulus and to note the localized direction from one of the 12 azimuth angles or 7 elevation angles in 30º apart. In the head-still condition, they were asked to keen their heads as still as possible while each stimulus was presented. In the head-movement condition, they were allowed (encouraged) to turn their heads horizontally while each stimulus was presented. The entire experiment was carried out in a particular order; starting with the subjects' own ears, stereo microphone, simple dummy head and finally with personal dummy head. The median plane localization experiments were performed after all of the horizontal localization experiments were completed.

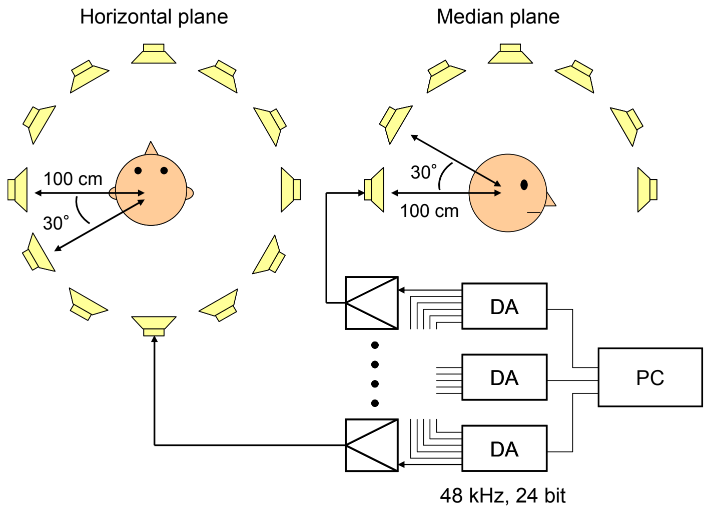

Figure 4 – Schematic of the sound localization experimental system. When a listener localizes stimuli with his ears, he sits at center of the speaker array.

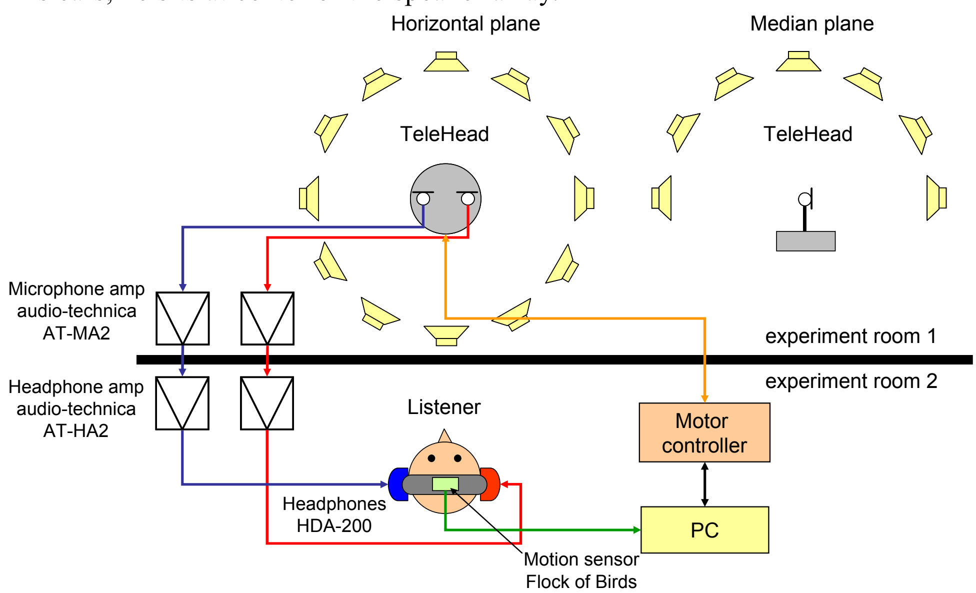

Figure 5 – Schematic of the sound localization experimental system. When a listener localizes stimuli through binaural signals, the *TeleHead* is placed at center of the speaker array and the binaural signals provided using either the stereo microphones, simple dummy head or personal dummy head are presented to the listener sitting in another room via headphones.

3.2 Results

Figure 6 shows the pooled localization results. The left two columns are the horizontal sound localization in the head-still and head-movement conditions. The right two columns are the horizontal sound localization in the head-still and head-movement conditions. In each panel, the areas of blue circles are proportional to in-head localization rates and those of the red circles are proportional to out-of-head localization rates. Responses on the light blue lines are those of front-back confusions.

3.2.1 Subjects' own ears

The panels in the first row of Figure 6 are the pooled localization results with the subjects' own ears. Horizontal plane sound localization was possible either in head-still or head-movement conditions. When the subjects were allowed to move their heads during the stimuli presentation, the correct localization rate increased 3% due to decreasing localization errors to the neighboring angles. This increase is not statistically significant. The median plane sound localization was far from perfect even with the subjects' own ears in the head-still conditions. The localization errors, however, dramatically decreased in the head-movement conditions, yielding a correct localization rate of 90.6 %. The improvement rate was 16%, which is statistically significant ($p<0.01$).

3.2.2 Stereo microphones

The panels in the second row of Figure 6 are the pooled localization results with the stereo microphones. Even horizontal plane sound localization was quite difficult in the head-still condition. Many stimuli were localized right or left and the stimuli presented in front and behind the subject were localized inside of the head with front-back confusion. The localization errors dramatically deceased in the head-movement conditions, yielding a correct out-of-head localization rate of 64.6 %. None of the stimuli were localized inside the head and the front-back confusion decreased 24%. Median plane sound localization was almost impossible in the head-still condition. The correct localization rate was just above the chance level. Horizontal head movement introduced a significant change in median plane sound localization. Although the correct out-of-head localization rate was only 40.4%, median plane sound localization was somewhat possible with the stereo microphones in the head-movement condition.

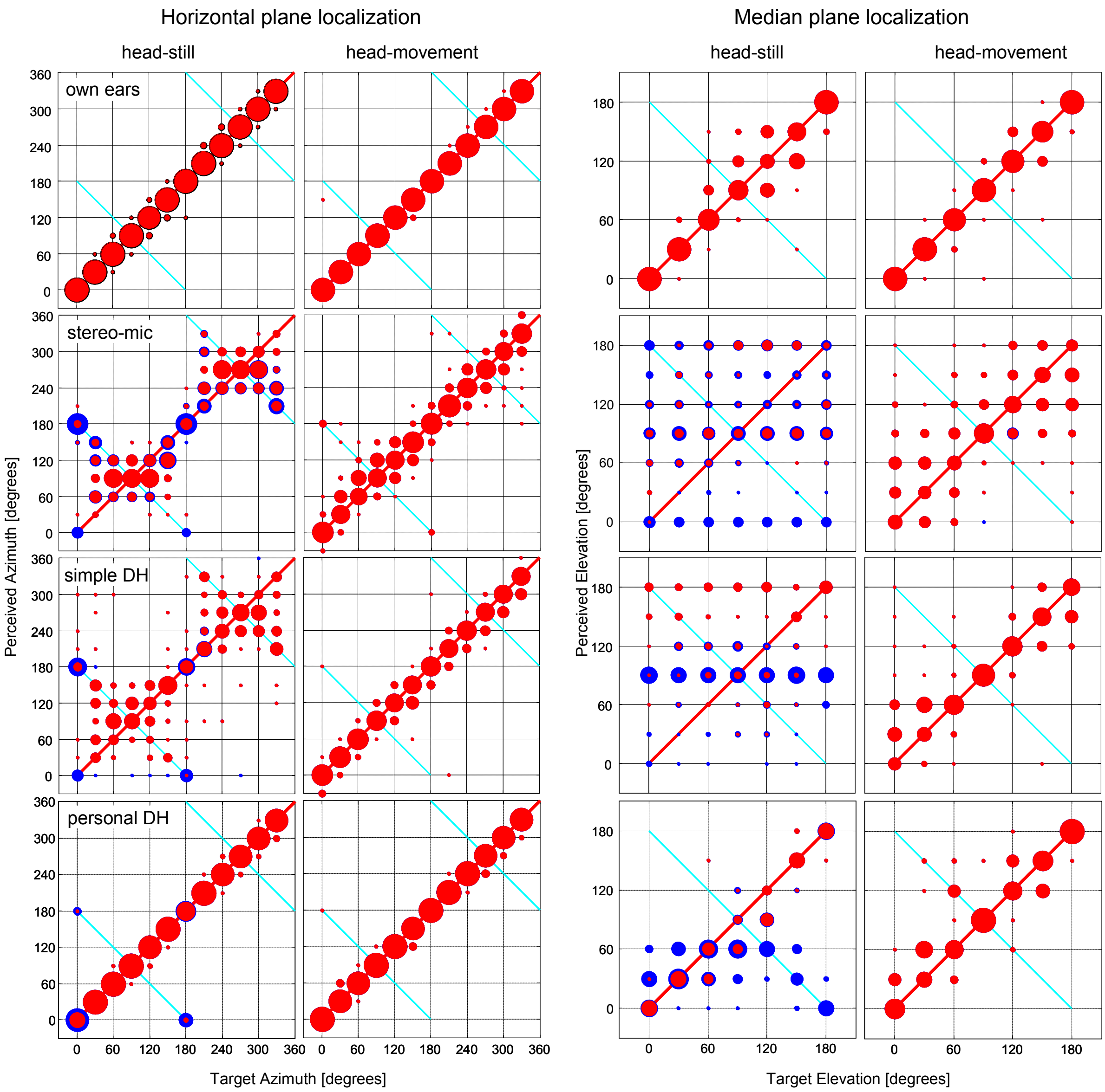

Figure 6 – Pooled horizontal and median plane sound localization results with listeners' own ears, stereo microphones, a simple dummy head, and personal dummy heads in head-still and head-movement conditions. Blue circles represent in-head localizations and red circles represent out-of-head localizations. Area of each circle is proportional to localization rate.

3.2.3 Simple dummy head

The panels in the third row of Figure 6 are the pooled localization results with the simple dummy head. Many stimuli were localized very roughly, and the stimuli presented in front and behind the subject were localized inside the head with front-back confusion. The tendency to localize the stimuli to the right or left was diminished compared to that of the stereo microphones. The correct out-of-head localization rate was 38.3%, which was comparable to that with the stereo microphones. The localization errors dramatically deceased in the head-movement conditions, yielding a correct out-of-head localization rate of 78.3 %, which was 12% better than that with the stereo microphones. Median plane sound localization was almost impossible in the head-still condition. The correct localization rate was just above the chance level. Horizontal head movement improved median plane sound localization. The correct out-of-head localization rate reached 65.7%. All stimuli were localized out of the head modestly well.

3.2.4 Personal dummy heads

The panels in the bottom row of Figure 6 are the pooled localization results with the personal dummy heads. Horizontal plane sound localization was possible in the head-still condition, although some stimuli presented in front of the subject were localized in-head, and some stimuli presented in front and behind the subject were localized with front-back confusions. In the head-movement condition, the in-head localization and front-back confusions disappeared. The correct localization rate was 95%, which is comparable to that with subjects' own ears. In contrast, the median plane sound localization was difficult in the head-still condition. All stimuli tended to be localized in-head, and there were many front-back confusions. In the head-movement condition, such errors disappeared, and 70% of the stimuli were localized correctly. The correct localization rate was similar to that of the simple dummy head, and significantly lower than that with the subjects' own ears.

4. DISCUSSION

When a listener could only use static binaural cues, the horizontal sound localization was very difficult and the median plane sound localization was almost impossible with the binaural signal provided using the stereo microphones or the simple dummy head. Even with those provided using the listener's own dummy head, some sound images were localized in-head. Since the actual-ear responses of the HDA-200 headphones were not equalized and non-FEC, poor sound localization performances might be due to the use of inappropriate headphones.

When a listener could use dynamic binaural cues, the horizontal sounds were localized out-of head almost exactly even with the binaural signal provided using the stereo microphones. Furthermore, median plane sounds were also well localized with the simple dummy head which had no pinnae. Again, the headphones used were the non-equalized and non-FEC.

These results strongly suggest the use of the dynamic binaural signals synchronous to listeners' head movements can considerably ease the acoustical constraints of the binaural reproducing system. Even with non-equalized and non-FEC headphones, both horizontal and median plane sound localizations are possible. Horizontal real-sound localization experiments with own ears [18], non-personal dummy heads [13, 14], scaled dummy heads [19], and incomplete dummy heads [20] in the head-movement condition yielded comparable results. Furthermore, horizontal virtual-sound localization experiments with non-personal HRTFs [21] and with several types of earphones [22, 23] in the head-movement condition also showed similar results. The dynamic binaural system is most likely to be free from the strict acoustical constraints imposed on the static binaural system.

5. CONCLUSIONS

A series of horizontal/median plane sound localization experiments in the head-still and head-movement conditions were conducted with binaural signals provided using stereo microphones, a simple dummy head, listeners' personal dummy heads and listeners' own ears. The headphones used in the experiment were non-FEC headphones HDA200 with which actual ear response was not equalized.

In the head-still conditions, where only static cues were available, sound localizations in the horizontal plane were very difficult and those in the medial plane were almost impossible with the stereo microphones or with the simple dummy head. In contrast, the sound localizations in both planes were possible in the head-movement condition, where dynamic binaural cues were available.

These results suggest that the strict acoustical constraints imposed on the static binaural system can be eased by the use of the dynamic binaural signals synchronous to listeners' head movement.

ACKNOWLEDGEMENTS

The authors thank Mr. Daisuke Yoshisaki for his support in conducting experiments with personal dummy heads. Part of this work was supported by KAKENHI (22300061).

REFERENCES

[1] H. Møller, "Fundamentals of binaural technology," Applied Acoustics, 36(3-4), 171-218 (1992).

[2] H. Møller, D. Hammershøi, C. Jensen and M. Sørensen, "Transfer Characteristics of Headphones Measured on Human Ears," J. Audio. Eng. Soc., 43(4), 203-217 (1995).

[3] H. Møller, "Binaural Technique: Do We Need Individual Recordings?" J. Audio Eng. Soc., 44(6), 451-469 (1996).

[4] H. Wallach, "On sound localization," J. Acoust. Soc. Am., 10, 270-274, 1939.

[5] S. Perrett and W. Noble, "The contribution of head motion cues to localization of low-pass noise," Perception & Psychophysics, 59(7), 1018-1026 (1997).

[6] F. Wightman and D. Kistler, "Resolution of front-back ambiguity in spatial hearing by listener and source movement," J. Acoust. Soc. Am., 105(5), pp.2841-2853 (1999).

[7] D. Begault, E. Wenzel and A. Lee, "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).

[8] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT press, Cambridge, MA, 1996).

[9] Y. Iwaya, Y. Suzuki and D. Kimura, "Effects of head movement on front-back error in sound localization," Acoustical Science and Technology, 24(5), 322-324 (2003).

[10] Y. Toyoda, D. Morikawa, and T. Hirahara, "Head movements during horizontal sound localization," Proc. Inter-Noise 2011 (2011).

[11] I. Toshima, H. Uematsu, and T. Hirahara, "A steerable dummy head that tracks three-dimensional head movement: TeleHead," Acoustical Science and Technology, 24(5), 327-329 (2003).

[12] I. Toshima, H. Uematsu, S. Aoki, and T. Hirahara, "A steerable dummy head that tracks human head movement in real time: TeleHead," J. Acoust. Soc. Japan, 61(4), 197-207 (2005). (in Japanese)

[13] I. Toshima, S. Aoki, and T. Hirahara, "Advanced version of a steerable dummy head that tracks human head movement in real time: TeleHead II" J. Acoust. Soc. Japan, 62(3), 244-254 (2006). (in Japanese)

[14] I. Toshima, S. Aoki, and T. Hirahara, "Sound Localization Using an Acoustical Telepresence Robot: TeleHead II," Presence, 17(4), 365-375 (2008).

[15] S. Yairi, "A Study on System Latency of Virtual Auditory Display Responsive to Head Movement," Ph.D. dissertation, Tohoku University (2006). (in Japanese)

[16] T. Hirahara, "Physical characteristics of headphones used in psychophysical experiments," Acoustical Science and Technology, 25(4), 276-285 (2004).

[17] T. Hirahara, "Discrepancies between actual-ear and artificial-ear responses," Proc. 20th International Congress on Acoustics, 474.1-474.4 (2010).

[18] D. Morikawa, Y. Toyoda, and T. Hirahara, "Impact of head movement on sound localization with band-limited noise," Proc. Inter-Noise 2011 (2011).

[19] T. Hirahara, H. Sagara, and M. Otani, "Sound localization with scaled dummy heads on a TeleHead," Proc. 19th International Congress on Acoustics, PPA-05-013, 1-4 (2007).

[20] T. Hirahara, "Sound localization using an incomplete dummy head, of which rear panel is opened," Proc. Autumn Meeting of the Acoustical Society of Japan, 463-464 (2005). (in Japanese)

[21] T. Hirahara, M. Otani, and I. Toshima, "Issues on HRTF Measurement and Binaural Reproduction," Fundamentals Review, 2(4), 68-85 (2009). (in Japanese)

[22] H. Aoyama, M. Otani, and T. Hirahara, "Sound localization with insert and intra-concha earphones," Proc. Autumn Meeting of the Acoustical Society of Japan, 515-516 (2007). (in Japanese).

[23] T. Hirahara, H. Aoyama, and M. Otani, "Effects of actual-ear response compensation on sound localization of dynamic binaural signals," Proc. Autumn Meeting of the Acoustical Society of Japan, 511-512 (2008). (in Japanese)