Sound Localization of Dynamic Binaural Signals Provided Using a Pinna-Less Dummy Head or a Stereo Microphone

Tatsuya HIRAHARA1,*, Yuki SAWADA1 and Daisuke MORIKAWA2

1Faculty of Engineering, Toyama Prefectural University, Imizu 939-0398, Japan
2School of Information Science, Japan Advanced Institute of Science and Technology, Nomi 923-1211, Japan

Received January 30, 2015; final version accepted April 9, 2015

This study compared the horizontal and median plane sound localization performances of binaural signals provided using a pinna-less dummy head or a stereo microphone that turns in synchronization with a listener’s head yaw rotation in the head-still and head-movement tasks. Results show that the sound localization performances in the head-movement tasks are significantly higher than those in the head-still tasks in both the horizontal and median plane. The dynamic binaural cues synchronized with a listener’s head yaw rotation dissolve the distance ambiguities, front-to-back ambiguities and elevation ambiguities, yielding better sound localization performances.

KEYWORDS: dynamic binaural signal, sound localization, stereo microphone, dummy head

1. Introduction

We evolved sound localization ability to avoid danger, hunt prey, or find a mate. Electrically reproduced sounds, however, usually have insufficient spatial information compared with 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 two-channel stereo system are localized between the two loudspeakers. The sounds played back with a 5.1-channel system surround the listener but their localizations are vague.

A number of 3-D sound technologies have been developed to reproduce sound with rich spatial information. One of them is binaural technology that controls sound pressures at a listener’s right and left ears with earphones [1–3]. The principle of binaural technology is simple: when a system can reproduce the same sound pressures at a listener’s eardrums via earphones as would have been produced there by a real sound, the listener should not be able to distinguish between the reproduced sound and the real sound.

It is however not exactly that simple, because the binaural system produces accurate sound pressures at a listener’s eardrums only with accurate head-related transfer functions (HRTFs) of the listener and acoustically appropriate earphones. HRTFs are highly unique for each person; no standard HRTFs exist and the use of non-personal HRTFs is inappropriate. Further, personal HRTFs are difficult to measure accurately. Acoustical requirements on earphones are also problematic: the earphones must exhibit free-air equivalent coupling (FEC) to the ear, and earphones must have flat frequency characteristics [4, 5]. However, few FEC earphones are available. The actual ear response of an earphone is unique for each ear and changes with the slightest of change in its usage conditions [6]. It is hence not easy to design an individual equalizer for flattening the earphone’s actual ear response. These acoustical constraints may be controllable in a laboratory, but neither accurate personal HRTFs nor acoustically appropriate earphones can be used for practical purposes.

All the acoustical requirements described above are quite right for the static binaural system but not necessarily right for the dynamic binaural system, since the head movement can help sound localization [7–16]. Namely, Move, then it shall be localized. The auditory system uses the temporal changes in interaural time difference (ITD), interaural level difference (ILD), and spectral cues caused by head movement to reduce ambiguities in sound localization.

This study clarifies whether using the dynamic binaural signals synchronized with listeners’ head movement improves sound localization performance or not when the signals are provided using a pinna-less dummy head or a stereo microphone that has non-personal HRTFs and reproduced with acoustically inappropriate earphones.

2010 Mathematics Subject Classification: Primary 92F05, Secondary 94A99.
This work was partly supported by JSPS KAKENHI Grant Number 22300061 and 25330203.
*Corresponding author. E-mail: hirahara@pu-toyama.ac.jp
2. Method

2.1 Steerable dummy head system: TeleHead

We used the TeleHead IV [17] to record ‘binaural’ signals with a pinna-less dummy head or a stereo microphone. The schematic diagram of the TeleHead IV is shown in Figure 1. The TeleHead IV is a one-degree-of freedom steering dummy head system, which tracks a listener’s head yaw rotation within ±85° quickly and quietly. Two small microphones placed at the entrance of outer-ear canals of the dummy head receive actual binaural signals synchronized with a listener’s head yaw rotation. The binaural signals are just amplified and reproduced at a listener’s ears through earphones [18–21].

The TeleHead provides perfect dynamic binaural signals synchronized with a listener’s head yaw rotation unlike digital binaural systems based on the HRTF convolution to the signals. Since no digital audio signal processing is used in the TeleHead, recorded dynamic binaural signals are reproduced as they are. There are no modifications, simplifications, or deteriorations of binaural signals due to digital signal processing, such as the use of the finite impulse response length and artifact noise addition due to switching HRTFs [22].

Several factors deteriorate the binaural signals recorded by the TeleHead. One is transfer characteristics of microphones, microphone amplifiers, headphone amplifiers, and earphones. They are fairly flat in the audible range, except for earphones. Another factor that deteriorates the binaural signals is the rotational delay caused by the servo loop of the motor controller. The rotational delay of the TeleHead IV is about 120 ms, which is just beyond its perceptual threshold and is thus perceptible [23]. The noise floor level \(L_A\) of the calibrated earphone output is less than 24 dB, which is compatible with that of the experimental room. When the sound pressure level of a white noise stimulus is 70 dB \(\text{per se}\), the operating noise of the TeleHead IV is masked, and thus not audible.

![Fig. 1. Schematic diagram of the TeleHead IV.](image)

2.2 Pinna-less dummy head and stereo microphone

While the TeleHead IV uses a personal dummy head in the original case, a pinna-less dummy head or a stereo microphone was used as substitute for the personal dummy head in the experiment (Figure 2). A pinna-less dummy head provides sufficient ITD and ILD but no spectral cues in the higher frequency regions. A stereo microphone provides sufficient ITD and small ILD but no spectral cues.

The pinna-less dummy head used in the experiment was an ellipsoidal dummy head made of Styrofoam, originally used for displaying wigs. The dummy head had neither a nose nor pinnae. Its surface was covered with paper clay to adjust the distance between two small electret-condenser microphones embedded in the lateral sides to be 16.5 cm.

With the pinna-less dummy head, [ITD] varied from 0 to 750 μs and [ILD] varied from 0 to 15 dB with azimuthal angles, shown as open triangles in Figure 3. The ranges of the [ITD] and [ILD] are within the ranges of those in adult males. [ITD] and [ILD] were less than 50 μs and 0.5 dB, respectively, for any elevation angle in the median plane.

The stereo microphone used in the experiment were two small omnidirectional electret-condenser microphones spaced 300 mm apart, which is the geodetic distance between two actual ears (Figure 2b). The supporting arms of the stereo microphone are made of stainless steel rods of 5–6 mm φ. The arms do not vibrate mechanically due to rapid rotation.
With the stereo microphone, $|\text{ITD}|$ varied from 0 to 900 μs and $|\text{ILD}|$ varied from 0 to 3.0 dB with azimuthal angles, shown as filled squares in Figure 3. The range of the $|\text{ITD}|$ is within the range of those in adult males. ITD was 0 μs and $|\text{ILD}|$ was less than 0.3 dB for any elevation angle in the median plane.

### 2.3 Sound localization experiment

Figure 4 illustrates the experimental setup. The speaker array consisted of 12 loudspeakers (Vifa, MG10SD0908) placed in a horizontal circle and five loudspeakers placed on the upper-semicircle at intervals of 30 degrees. The circles were 1 m in diameter, and the horizontal speakers were 1.1 m high. A listener sat or the TeleHead IV was placed at the center of the speaker array in an experimental room, which had a noise floor level $L_A$ of 24 dB and a $RT_{500}$ of 50 ms.

Sound stimulus emanated from a loudspeaker was received by two electret condenser microphones (WM-62PC, Panasonic). The received ‘binaural’ signals were amplified, transmitted to another experimental room, and then presented to a listener through earphones. The stimulus was three-second white noise generated independently for each trial. Each stimulus was D/A converted ($F_s = 48$ kHz, 24 bits) with a UA101 (Roland), amplified with a 1075II (Bose), and presented in a random direction at three-second intervals. The sound pressure level of the stimulus reproduced by each loudspeaker was set at 70 dB at the center of the speaker array. The sound pressure level of earphone-reproduced ‘binaural’ signals, which were received by each ‘head’ placed towards the front (0°), for the stimulus emanated from the front loudspeaker was set at 70 dB with an IEC60711 coupler by adjusting the gain of an earphone amplifier. The earphones used were HDA200 (Sennheiser) [6]. Their actual-ear responses were not equalized. The microphone and earphone amplifiers used were AT-MA2s (audio-technica) and AT-HA20s (audio-technica).

The experiment tasks were the combinations of three ‘head’ tasks, two head-movement tasks, and two plane tasks. The three ‘head’ tasks were the actual head task, pinna-less dummy head task and the stereo microphone task. In the actual head task, a listener sat at the center of the speaker array and localized sound stimuli directly with his own head. In the pinna-less dummy head and the stereo microphone tasks, a listener localized reproduced binaural signals provided using the dummy head and the stereo microphone, respectively. The two head-movement tasks were the head-still and the head-movement tasks. In the head-still task, the TeleHead was powered off; a listener localized static binaural signals. In the head-movement task, the TeleHead was powered on and tracked a listener’s head rotation; a listener localized dynamic binaural signals. The two plane tasks were the horizontal and median plane tasks.

The horizontal and median plane localization tasks were carried out separately. In the horizontal plane task, each session consisted of 60 trials; stimulus was presented five times in random order from each of the 12 directions. One experiment consisted of four sessions, resulting in responses of 20 trials from each of the 12 directions. In the median
plane task, each session consisted of 35 trials; stimulus was presented five times in random order from each of the seven directions. One experiment consisted of four sessions, resulting in responses of 20 trials from each of the seven directions.

Five males of their twenties who have normal hearing participated in the experiments. Listeners were asked to localize each stimulus and to note the localized direction from one of the 12 azimuth angles or seven elevation angles 30° apart as well as whether the sound image was in-head or out-of-head on an answer sheet. In the head-still task, they were asked to keep their heads as still as possible while each stimulus was presented. In the head-movement task, they were allowed and encouraged to turn their heads horizontally while each stimulus was presented. The entire experiment was carried out in a particular order: starting with the actual head task, then the stereo microphone task, and finally the pinna-less dummy head task. In each task, the head-still tasks were carried out prior to the head-movement tasks. The median plane localization experiments were performed after all of the horizontal localization experiments had been completed.

![Diagram](figure4.png)

Fig. 4. Sound localization experiment setups. The pinna-less dummy head or the stereo microphone was put on the TeleHead IV in room 1. It turns quietly in synchronization with a listener’s head yaw rotation in the head-movement task. Binaural signals provided by using them are reproduced to a listener in room 2 via earphones.

### 3. Results

Pooled localization results of three ‘head’ tasks for five listeners are summarized in Figures 5, 6, and 7. In each figure, panels in the upper and lower columns are the results of the horizontal and median plane tasks. Panels in the right and left rows are those of the head-still and head-movement tasks. In each panel, the area of the black-filled circles is proportional to the out-of-head localization rate, while that of the gray-filled circles is proportional to the in-head localization rate. Correct responses are on the diagonally right up line and responses on the dashed diagonally right down lines are those of front-back confusion.

Figure 8 compares the mean correct out-of-head-localization rates $\overline{P}_{CS}$, the mean front-back confusion rates $\overline{E}_{FB}$, and the mean neighborhood error rates $\overline{E}_{NH}$ among six tasks in horizontal and median plane tasks. $\overline{E}_{NH}$ is the percentage of mislocalization by which a listener judged a stimuli angel as either +30 or -30 degrees off the presented angle.

Two-way ANOVAs of $\overline{P}_{CS}$ in the horizontal and median plane tasks were performed, considering the head-movement (head-still or head-movement task) and the ‘head’ (actual head, pinna-less dummy head or stereo microphone) as factors. In the horizontal plane task, the interaction between the head-movement and the ‘head’ was significant [$F(2, 24) = 13.91, p < 0.005$]. Simple main effects of the head-movement were found on the pinna-less dummy head [$F(1, 24) = 47.93, p < 0.005$] and stereo microphone [$F(1, 24) = 50.62, p < 0.005$]. Simple main effects of the ‘head’ were found on the head-still task [$F(2, 24) = 75.15, p < 0.005$] and head-movement task [$F(2, 24) = 12.16, p < 0.005$]. In the median plane task, the interaction between the head-movement and the ‘head’ was significant [$F(2, 24) = 15.37, p < 0.005$]. Simple main effects of the head-movement were found on the actual head [$F(1, 24) = 6.62, p < 0.05$], pinna-less dummy head [$F(1, 24) = 106.01, p < 0.005$] and stereo microphone [$F(1, 24) = 57.93, p < 0.005$]. Simple main effects of the ‘head’ were found on the head-still task [$F(2, 24) = 109.68, p < 0.005$] and head-movement task [$F(2, 24) = 34.27, p < 0.005$]. Results of the post-hoc Tukey–Kramer multiple-comparison tests for $\overline{P}_{CS}$ in the horizontal and median plane tasks are shown in the left row of Figure 8.
Two-way ANOVAs of $E_{FB}$s in the horizontal and median plane tasks were performed, considering the head-movement and the ‘head’ as factors. In the horizontal plane task, there was a significant interaction between the head-movement and the ‘head’ $[F(2, 24) = 19.75, p < 0.005]$. Simple main effects of the head-movement were found on the pinna-less dummy head $[F(1, 24) = 65.59, p < 0.005]$ and stereo microphone $[F(1, 24) = 54.56, p < 0.005]$. The simple main effect of the ‘head’ was found on the head-still task $[F(2, 24) = 41.63, p < 0.005]$. In the median plane task, there was significant main effects of the head-movement $[F(1, 24) = 8.84, p < 0.01]$ and the ‘head’ $[F(1, 24) = 3.37, p < 0.05]$, but no interaction between them. Results of the post-hoc Tukey–Kramer multiple-comparison tests for $E_{FB}$s in the horizontal and median plane tasks are shown in the middle row of Figure 8.

Two-way ANOVAs of $E_{NH}$s in the horizontal and median plane tasks were performed, considering the head-movement and the ‘head’ as factors. In the horizontal plane task, there was a significant interaction between the head-movement and the ‘head’ $[F(2, 24) = 19.75, p < 0.005]$. Simple main effects of the head-movement were found on the pinna-less dummy head $[F(1, 24) = 65.59, p < 0.005]$ and stereo microphone $[F(1, 24) = 54.56, p < 0.005]$. The simple main effect of the ‘head’ was found on the head-still task $[F(2, 24) = 41.63, p < 0.005]$. In the median plane task, there was significant main effects of the head-movement $[F(1, 24) = 8.84, p < 0.01]$ and the ‘head’ $[F(1, 24) = 3.37, p < 0.05]$, but no interaction between them. Results of the post-hoc Tukey–Kramer multiple-comparison tests for $E_{NH}$s in the horizontal and median plane tasks are shown in the middle row of Figure 8.
Simple main effects of the head-movement were found on the actual head $F(1, 24) = 7.47, p < 0.05$ and stereo microphone $F(1, 24) = 6.71, p < 0.05$. The simple main effect of the ‘head’ was found on the head-movement task $F(2, 24) = 15.46, p < 0.005$. Results of the post-hoc Tukey–Kramer multiple-comparison tests for $E_{NH}$s in the horizontal and median plane tasks are shown in the right row of Figure 8.

When listeners localized the sound stimuli using their actual heads, all the stimuli presented in the horizontal plane were localized out-of-head, and most were localized correctly either in the head-still or the head-movement tasks. $P_C$ of the head-still and head-movement tasks were 95.1% and 99.5%. The difference between them was not significant ($p < 0.005$). All the stimuli presented in the median plane were also localized out-of-head, but there were many neighborhood errors. $P_C$ of the head-still task was only 76.4%. In contrast, due to significant reduction of the neighborhood errors, $P_C$ of the head-movement task was 90.3%, which was significantly higher than that of the head-still task ($p < 0.005$).
4. Discussion

Binaural cues provided using the pinna-less dummy head in the head-still tasks are static ITD and ILD. The stereo microphone also provides both binaural cues while the static ITD is major. When listeners were only able to use static binaural cues, horizontal plane sound localizations were very difficult and median plane sound localizations were almost impossible. These poor sound localization performances could be due to the use of non-personal HRTFs; ITD and ILD of the stimuli were not exactly those of each listener and the spectral cues were lacking. Another reason for the poor localization performance could be the use of acoustically inappropriate earphones; the HDA200 does not have a flat actual-ear response or FEC characteristics [4, 6].

In contrast, binaural cues available in the head-movement tasks are dynamic ITD and ILD in addition to the static cues. When listeners were able to use these dynamic binaural cues, sounds were localized in both the horizontal and median planes even in the stereo microphone task. Sound localization performance was improved due to significant reduction in in-head localizations and front-back confusions. The effect of head movement was significant in the pinna-less dummy head and the stereo microphone tasks in both planes. The difference in $P_C$ values of the pinna-less dummy head and the stereo microphone tasks in the head-movement task was not significant in the horizontal plane but was in the median plane. The $P_C$ values of the pinna-less dummy head and the stereo microphone tasks in the head-movement task in either plane were significantly lower than those of the actual head tasks ($p < 0.005$). The $P_C$ values of those in the horizontal plane (78.3% and 68.3%) were higher than that in the head-still task (49.0%) and comparable with that in the head-movement task (75.6%) in another experiment [23], in which binaural signals were provided using a personal dummy head of a non-listener and nine listeners who were not the listeners of the current study participated in the experiment.

The HRTFs of the pinna-less dummy head and the stereo microphone are obviously not those of each listener. The earphones used in this experiment were not acoustically appropriate earphones for a binaural reproduction system. Despite these acoustical imperfections of the binaural reproduction system, the sound localization performances were improved in the head-movement task, suggesting that the acoustical imperfections are irrelevant for the dynamic binaural reproduction system.

Even in the pinna-less dummy head task or the stereo microphone task, the dynamic binaural cues played a major role in yielding better sound localization performances. The identities of the dynamic binaural cues are the temporal changes in ITD and ILD ($\Delta$ITD and $\Delta$ILD) synchronized with a listener’s head yaw rotation, which reduced in-head localizations and front-back confusions. It is just conceivable that $\Delta$ITD and $\Delta$ILD reduce in-head localizations and front-back confusion, because they dissolve distance ambiguities and front-to-back ambiguities.

One may find it suspicious that the dynamic binaural cues associated with head yaw rotation help median plane sound localization. The $\Delta$ITD and $\Delta$ILD can also dissolve elevation ambiguities in addition to the distance and front-to-back ambiguities in the median plane. This is because the lower the sound source elevation angle, the larger the azimuthal-angle change of the sound image when a listener rotates his/her head horizontally. When the sound is presented directly above the head, its sound image stays still regardless of the listener’s head yaw rotation angle. As the pinna-less dummy head and the stereo microphone do not provide the spectral cues, which are the major
monaural cues for elevation, listeners must have used the ΔITD and ΔILD as major cues to judge sound image elevation.

5. Conclusion

The sound localization performances in horizontal and median plane are improved significantly by using the dynamic binaural signals synchronized with a listeners’ head yaw rotation when they are provided using the pinna-less dummy head or the stereo microphone. The dynamic binaural cues, i.e. the temporal changes in ITD and ILD synchronized with a listener’s head yaw rotation, are used to dissolve the distance ambiguities, front-to-back ambiguities, and elevation ambiguities. The effects of the dynamic binaural signals are observed without using accurate personal HRTFs and acoustically appropriate earphones, suggesting that the acoustical requirements for the static binaural reproduction system are not very crucial for the dynamic binaural reproduction system.

Acknowledgments

This work was partly supported by JSPS KAKENHI Grant Number 22300061 and 25330203.

REFERENCES

[1] Blauert, J., Spatial Hearing: The Psychophysics of Human Sound Localization, MIT Press, Cambridge, MA (1996).
[2] Wenzel, E., “Localization in virtual acoustic displays,” Presence, 1: 80–107 (1992).
[3] Møller, H., “Fundamentals of binaural technology,” Applied Acoustics, 36(3–4): 171–218 (1992).
[4] Møller, H., Hammershoi, D., Jensen, C., and Sørensen, M., “Transfer characteristics of headphones measured on human ears,” J. Audio. Eng. Soc., 43(4): 203–217 (1995).
[5] Møller, H., “Binaural technique: Do we need individual recordings?” J. Audio Eng. Soc., 44(6): 451–469 (1996).
[6] Hirahara, T., “Physical characteristics of headphones used in psychophysical experiments,” Acoustical Science and Technology, 25(4): 276–285 (2004).
[7] Wallach, H., “On sound localization,” J. Acoust. Soc. Am., 10: 270–274 (1939).
[8] de Boer, K., and Urk, A., “Some particulars of directional hearing,” Philips Tech. Rev., 6: 359–364 (1941).
[9] Thurlow, W., Mangels, J., and Runge, P., “Head movements during sound localization,” J. Acoust. Soc. Am., 42(2): 489–493 (1967).
[10] Thurlow, W., and Runge, P., “Effect of induced head movement in localization of direction of sound,” J. Acoust. Soc. Am., 42(2): 480–488 (1967).
[11] Makous, J., and Middlebrooks, J., “Two-dimensional sound localization by human listeners,” J. Acoust. Soc. Am., 87(5): 2188–2200 (1990).
[12] Kawaura, J., Suzuki, Y., Asano, F., and Sone, T., “Sound localization in headphone reproduction by simulating transfer functions from the sound source to the external ear,” J. Acoust. Soc. Jpn. (E), 12(5): 203–216 (1991).
[13] Perrett, S., and Noble, W., “The contribution of head motion cues to localization of low-pass noise,” Perception & Psychophysics, 59(7): 1018–1026 (1997).
[14] Wightman, F., and Kistler, D., “Resolution of front-back ambiguity in spatial hearing by listener and source movement,” J. Acoust. Soc. Am., 105(5): 2841–2853 (1999).
[15] Begault, D., Wenzel, E., and Lee, A., “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).
[16] Iwaya, Y., Suzuki, Y., and Kimura, D., “Effects of head movement on front-back error in sound localization,” Acoustical Science and Technology, 24(5): 322–324 (2003).
[17] Hirahara, T., Yoshisaki, D., and Morikawa, D., “Impact of dynamic binaural signal associated with listener’s voluntary movement in auditory spatial perception,” POMA, 19(050130): 1–8 (2013).
[18] Toshima, I., Uematsu, H., and Hirahara, T., “A steerable dummy head that tracks three-dimensional head movement: TeleHead,” Acoustical Science and Technology, 24(5): 327–329 (2003).
[19] Toshima, I., Uematsu, H., Aoki, S., and Hirahara, T., “A steerable dummy head that tracks human head movement in real time: TeleHead,” J. Acoust. Soc. Japan, 61(4): 197–207 (2005).
[20] Toshima, I., Aoki, S., and Hirahara, T., “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).
[21] Toshima, I., Aoki, S., and Hirahara, T., “Sound localization using an acoustical telepresence robot: TeleHead II,” Presence, 17(4): 365–375 (2008).
[22] Otani, M., and Hirahara, T., “Auditory artifacts due to switching head-related transfer functions of a dynamic virtual auditory display,” IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences, E91-A(6): 1320–1328 (2008).
[23] Yairi, S., Iwaya, Y., and Suzuki, Y., “Estimation of detection threshold of system latency of virtual auditory display,” Applied Acoustics, 68(8): 851–863 (2007).
[24] Yoshisaki, D., and Hirahara, T., “Effect of head-and-neck somatic information in horizontal sound localization,” IEICE Trans. on Information and Systems, J96-D(3): 630–635 (2013).