Proceedings of 20th International Congress on Acoustics, ICA 2010

23–27 August 2010, Sydney, Australia

Sound localization in a median plane using an avatar robot "TeleHead" with synchronization of a listener's horizontal head rotation

Yukio Iwaya (1)(2), Yusuke Masuyama (2), Makoto Otani (3), and Yôiti Suzuki (1)(2)

(1) Research Institute of Electrical Communication, Tohoku University, 2–1–1 Katahira, Aoba-ku, Sendai, Miyagi, Japan
(2) Graduate School of Information Sciences, Tohoku University, 6–3–09 Aramaki Aza Aoba, Aoba-ku, Sendai, Miyagi, Japan
(3) Faculty of Engineering, Shinshu University, 4–17–1 Wakasato, Nagano, Nagano, Japan

PACS: 43.66Pn,43.66Qp

ABSTRACT

Demands for realization of communications with a high sense of presence have become great. For such communications, it is important acoustically to capture and transmit comprehensive sound space information of remote places to a local site. The authors developed an avatar robot, which was developed as a simplified version of TeleHead proposed by Toshima et al. The robot's head moves synchronously, following the listener's horizontal head rotation. The authors investigated perceptual sound localization accuracy in the median plane at a remote site. Results show that sound localization accuracy is improved when the robot rotates synchronously with the listener's head rotation. Another sound localization test was conducted to examine the effects of manipulating the ratio of the rotation angle. The ratio between the listener's actual movement and that of the robot varied systematically. Results show that the ratios only slightly affect the accuracy of perceived elevation angles, suggesting large robustness in using cues provided by head rotation.

INTRODUCTION

In the near future, advanced and natural communications with a person in remote places might be well realized using a system with a high sense of presence. For such communications, it is important to capture and transmit comprehensive sound space information of a remote place to a local site. Another important point of interactive communications is to reproduce a sound space so that the synthesized sound field becomes responsive to a listener's movement. Several research results have revealed that the accuracy of sound localization can be improved in a horizontal plane when we allow movement of the head and body in both real [1, 2, 3, 8] and virtual [4, 5, 6, 7] environments.

Several methods have shown great potential to realize responsiveness to listeners' movements in sound reproduction. Wave field synthesis (WFS) [9] and boundary sound control (BoSC) [10] based on Kirchhoff–Helmholtz integral equations permit a listener to change position and execute head movements freely when in the controlled area. Furthermore, Ambisonics technique [11] allows a listener's head movements at and near the sweet spot of listening. However, these methods require many loudspeakers to control sound fields with high accuracy. In recording, numerous microphones are also needed. Sakamoto proposed a novel sound capturing system, SENZI[12]. A spherical microphone array with many microphones installed is used for capturing three-dimensional sound space information. These signals are adequately converted to binaural signals using simple digital signal processing. Because the spherical microphone array is symmetric, the signal processing can be changed according to head movement that is sensed using a position sensor.

An important point related to reproduction of a sound field is to synthesize the sound field so that head-related transfer functions (HRTFs) [13] of a listener are adequately convolved before sounds arrive at the ear drums. In WFS, BoSC, and Ambisonics techniques, HRTFs are naturally convolved at a listening point with the listener's actual figures. In contrast, measurements or numerical estimations of listener's HRTFs are needed in the SENZI system because HRTFs closely depend on a listener's head, body, and ear shapes. However, the measurement of HRTFs for a specific listener requires a huge measurement apparatus, time, and effort [14]. The numerical estimation demands many computation resources[15, 16, 17, 18]. TeleHead [19] in a remote site can move synchronously to the person's various head movements by sensing head movements via a position sensor attached at the person's head. The person can listen to sounds via TeleHead as an avatar if the person wears headphones whose inputs are connected to TeleHead's ears. The head of TeleHead can be exchanged and a listener can use her own personal figure of head as an avatar at the remote site. Because two microphones are installed at ear entrances of the dummy-head at the remote site, her own HRTFs are naturally convolved when the listener uses her own head figure as the dummy-head. Although the listener must prepare such a head figure in advance, the TeleHead appears promising to enable us to sense the whole remote sound space information interactively as an avatar of the listener with the listener's own HRTFs.

Several researchers have reported that head movement during listening in a sound space can enhance the accuracy of sound localization and the reality, or the sense-of-presence, of the perceived sound space in real and virtual environments[1, 2, 3, 4, 7]. Toshima et al. also investigated sound localization accuracy using TeleHead in a horizontal plane and median plane [20]. They used head shapes of two types and discussed the effects of head shapes. They pointed out that the localization accuracy can be improved using TeleHead even when the head shapes of TeleHead are not a listener's own one. They also reported that synchronization of a listener's head movements was important. However, in their experiment, TeleHead could

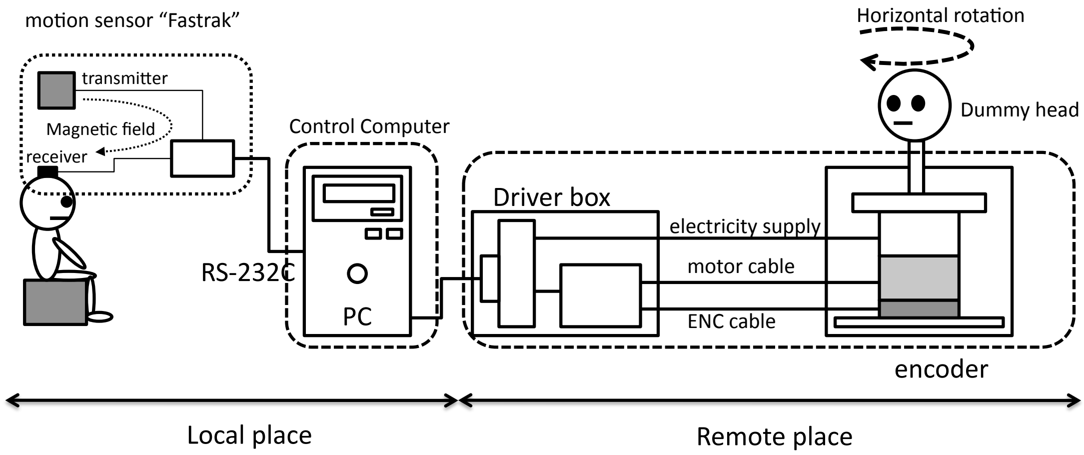


Figure 1: Schematic overview of the current version of TeleHead

move the head in yaw, pitch, and roll axes. The effects of these rotations were not separately described.

It is known that horizontal head rotation plays the most important role in localizing sound [1]. Therefore, we developed a simplified TeleHead whose head moves synchronously: following the listener's horizontal head rotation. Because this TeleHead used in this study does not follow the listener's head rotation in yaw and pitch axes, the effect of the horizontal rotation was specifically studied. It is both interesting and fruitful to investigate the reason why the synchronized movement of the TeleHead only in horizontal rotation would be effective to render the transmitted sound space. Therefore, the authors have intensively investigated sound localization accuracy in such a situation with this simplified TeleHead. First, the perceptual performance of median plane sound localization via this TeleHead was examined. Another sound localization test also for the median plane localization was conducted to examine effects of manipulating the rotation angle. In this test, the rotation of the TeleHead was controlled so that the ratio of its rotation angle to that of the listener's actual horizontal rotation was set to a specific value of 0.5–1.5.

In the following section, the TeleHead construction process is described. Then, experimental methods of the two listening tests and their results as well as some discussions are introduced, respectively, in the successive sections.

AVATAR ROBOT "TELEHEAD"

In this section, the system of simplified TeleHead we constructed is introduced.

System overview

The TeleHead we developed was based on the fourth version of TeleHead developed by Hirahara et al. A three-dimensional position sensor (Fastrak; Polhemus) with six degrees of freedom (6DOF: *x*, *y*, *z*, *yaw*, *pitch*, and *roll*) was attached at the top of the headphones, or the headband, of a listener in a local place. The detected position data were sent to a personal computer using an RS–232C port with a sampling frequency of 120 Hz. The acquired position was processed with a dynamic link library (DLL); adequate displacement and velocity of the head were calculated for controlling a dummy head in a remote place. A servomotor under the dummy head was controlled based on the control information so that the rotation angle of the dummy head in the remote place is synchronized to the listener's. Two microphones were installed to both ears of the dummy head, and acoustic signals toward the ears from whole directions in the remote place could be captured and transmitted to headphones of the listener in the local place. The TeleHead used in Toshima's experiment was capable of moving the head with three DOF (*yaw*, *pitch*, and *roll*). However, the current version is capable of moving the head only in horizontal rotation even when the listener moved his head in three-dimensional rotation. A system overview of the current version of TeleHead is shown in Fig. 1.

Construction of dummy heads using stereolithography

A listener's own HRTFs are naturally convolved in the transmitted sound signals if the head shape of the dummy-head upon the servomotor is identical to that of the listener. Therefore, the authors developed six dummy-heads for all participants in experiments (described later) using the following procedures. Slice images of each listener's head were measured using magnetic resonance imaging (MRI, Marconi Magnex Eclips 1.5 T; Shimadzu Corp.). Then the images were converted to Tagged image file format (tiff) through Digital Imaging and Communication in Medicine (DICOM) format. The skull shapes were extracted from the tiff images and holes in skull data were filled up. Finally, a data file of Standard Triangulated Language format (STL format) was obtained as three-dimensional shape data of the listener's head. The concrete figure of the dummy-head was constructed with a three-dimensional printer based on stereolithography so that the head shape figure was identical to each listener from the STL format file. For example, the actual head photograph, the three-dimensional model based on STL data, and the dummy-head are compared in Fig. 2.

EXPERIMENT I: SOUND LOCALIZATION WITH TELEHEAD IN THE MEDIAN PLANE

Even when a listener moves her head in 6DOF while listening in a sound space, the TeleHead we constructed moves its head only in horizontal rotation. Therefore, the listener feels that the sound space does not move synchronously with the listener's head movement and he might not localize the sound position accurately. In such a situation, the TeleHead cannot function as an avatar in the remote place. Therefore, we investigate sound localization accuracy in a median plane using the TeleHead in a remote place.

Participants

Six males with normal hearing participated in the experiment.

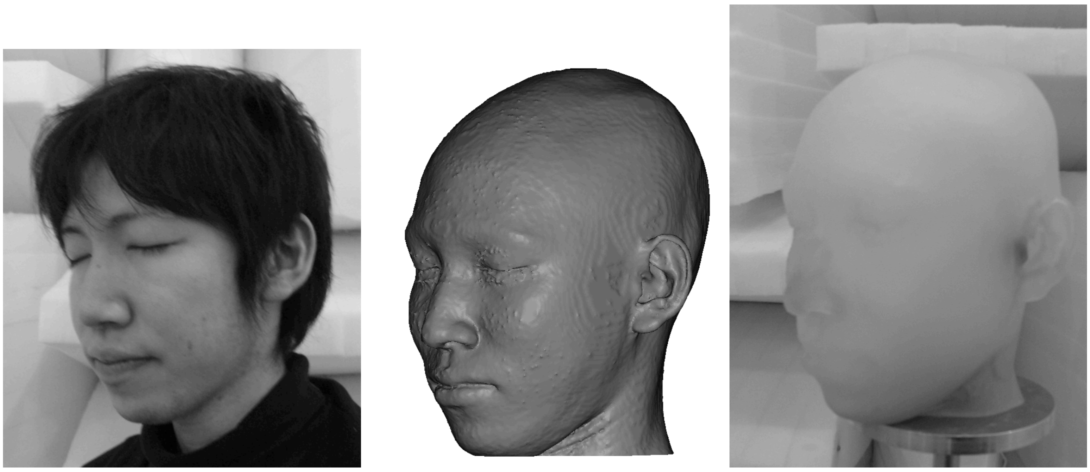


Figure 2: Comparison of an actual head (left), a three-dimensional model head (center), and a dummy-head printed using stereo lithography (right)

Experimental setup

The experimental setup is presented in Fig. 3. An anechoic room and a sound proof room were used, respectively, as a remote place and a local place. The anechoic room was equipped with a vertical-circled speaker array, on which 35 loudspeakers (FE83; Fostex Co.) were installed with an interval of 10 deg. in a median plane (-90 deg. was excluded). The TeleHead was set so that the center of the dummy head was located at the center of the speaker array (Fig. 4). Two small electret condenser microphones (WM-62PC; Panasonic Inc.) were installed at ear-canal entrances of both ears. Signals of microphone output were put into a microphone amplifier (MA-2016C; Thinknet) and were sent to a set of headphones (HDA200; Sennheiser Electronic GmbH) in a local place through an amplifier (HD35R; CEC). A 6-DOF magnetic position sensor (Fastrak; Polhemus) with a 120 Hz sampling rate was attached to the top of the headphones to sense a listener's head movements. The sensed data were analyzed using a personal computer and a command message was sent to a servomotor to control the TeleHead head angle. The personal computer collected listeners' responses of sound localization and controlled sound stimuli radiated from one of the 35 loudspeakers. The TeleHead head was controlled so that the horizontal angle of the TeleHead was identical to the horizontal angle of the listener in the local place. The maximum latency between the horizontal rotation in both remote and local places was about 50 ms when the frequency of the movement was about 0.5 Hz. The latency was shorter than the detection threshold [21], indicating that the TeleHead would have sufficient performance in synchronization with the listener's head rotation.

Method

In the remote place, 16 of 35 loudspeakers, from front bottom -60 degrees to rear -60 degrees with an interval of 20 degrees, were used. From one of 16 loudspeakers, A stimulus was radiated from one of 16 loudspeakers. A pink noise of 10-s duration, to which inverse filter characteristics of the headphones' response was convolved, was used as the stimulus. The weighted sound pressure level of stimulus was 60 dB. A stimulus for each angle was played five times, resulting in 90 trials in all, which were played in a randomized order. A participant was asked to select an elevation angle at which she localizes the stimulus from 16 angles using a graphical user interface drawn on a computer screen. Two conditions were tested in the experiment: One was a restricted condition (Restricted), in which the TeleHead did not move its head during the experiment, even when the participant moved his head. Another was a free condition (Free), in which the TeleHead moved its head with synchronized to the rotation of the participant. Each participant was asked to move his head freely during the session.

Results and Discussion

Results of the experiment of Listener 1 are shown in Fig. 5 as examples. The horizontal axis elevation direction of the loudspeaker, which radiated the stimulus and the vertical axis, represents the perceived elevation direction. The upper and bottom figures respectively show Restricted and Free conditions. Characters in both axes 'F' and 'B' respectively denote 'front' and 'bottom'. The area of each black circle is proportional to the frequency of responses. Frequent front–back confusion errors are seen in a Restricted condition. In contrast, perceived elevation angles in free condition are nearly identical to correct ones. All the participants showed the same tendency. These results suggest that the accuracy of sound localization can be improved with synchronization of an avatar robot in a remote place even when synchronization of a head movement to a local listener is limited to horizontal rotation.

EXPERIMENT II: SOUND LOCALIZATION WITH THE TELEHEAD IN VARIOUS SYNCHRONIZATION RATIOS

In experiment I, horizontal rotation of the TeleHead was synchronized; other movements were not. Nevertheless, the localization performance of source elevation in the median plane was improved, which might indicate a robustness of sound localization with head movement. The authors investigated further robustness of sound localization in experiment II by modifying the ratio between the actual angle of a listener and the controlled angle of the TeleHead.

Participants

Three males participated. All had participated in experiment I as well.

Experimental setup

The apparatus used in the experiment II was the same as that in experiment I. However, the control commands from the personal computer to the servomotor were modified so that the ratio between local and remote horizontal angle of heads was set to be one of three constant values: 0.5, 1.0, or 1.5. When the value was 0.5, the head of TeleHead moved its head so that

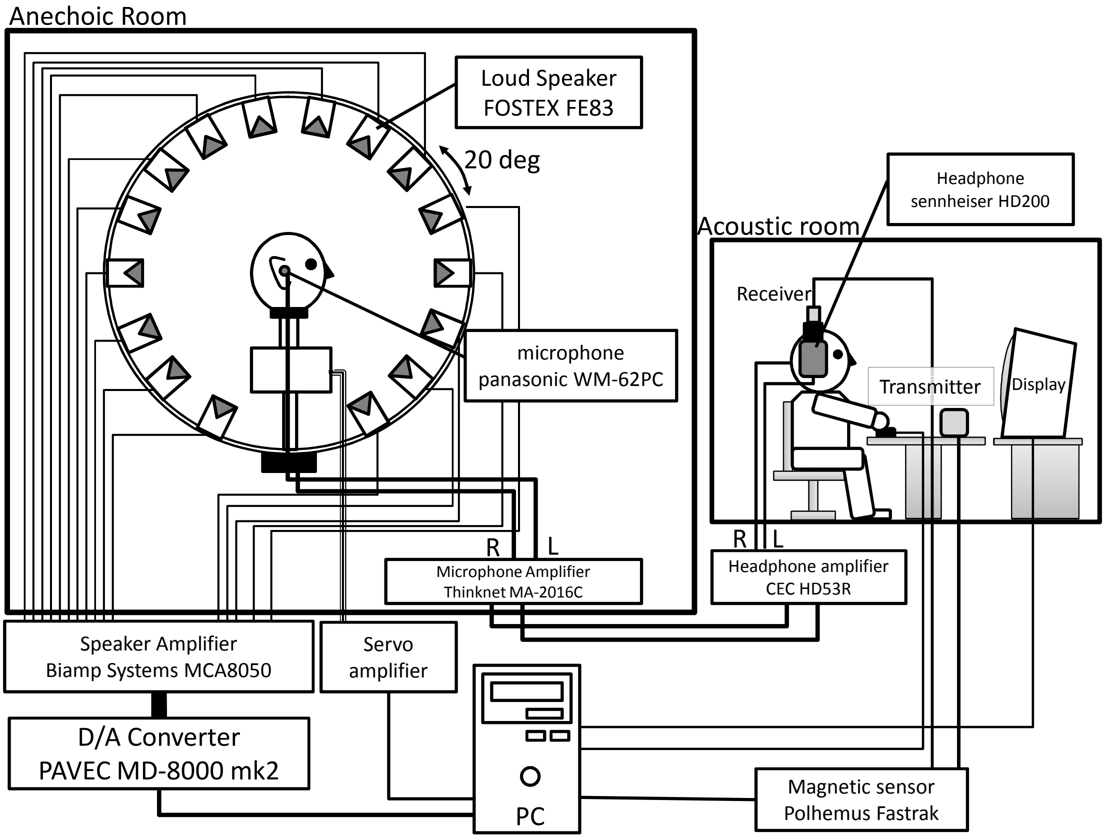


Figure 3: Experimental setup

Figure 4: TeleHead and a speaker array

the angle of rotation was a half of the actual head's rotation angle. When the value was 1.0, the condition was the same in the experiment I. Finally, when the value was 1.5, the angle of rotation of TeleHead was 1.5 times the actual head's rotation angle.

Method

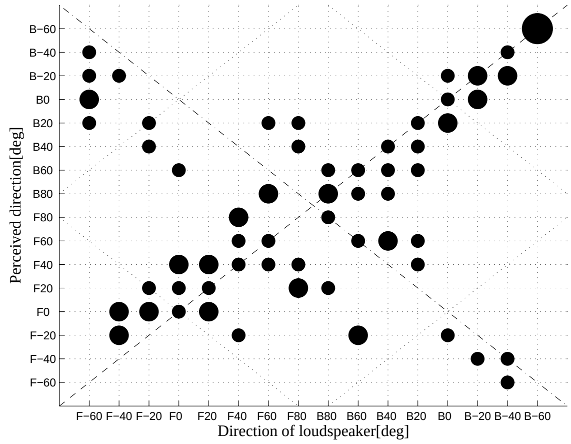
Participants were asked to answer the elevation direction that was perceived with the GUI used in experiment I. The ratio used in each session was fixed during the session. Participants did not know that the ratio was modified at all. Each participants was asked to move his head during each trial so that the head was moved at least one swing: center–right–center–left–center or center–left–center–right–center.

Results and Discussion

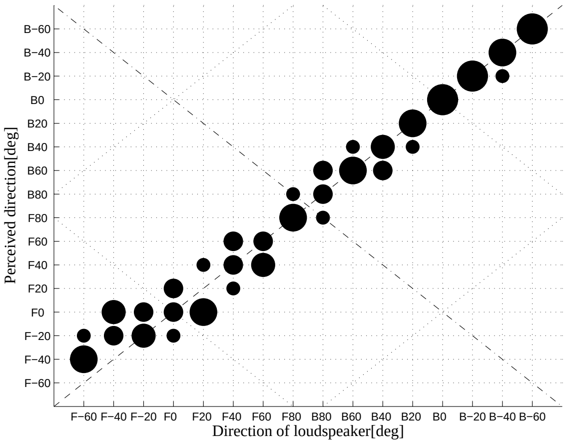
Results of experiment II of Listener 1 are shown in Fig. 6. Although some front-back confusions are in all conditions, almost all perceived elevation directions are roughly equal to presented directions. All participants showed the same tendency, which indicates that the directional vector of head motion is more important than complete synchronization of head movements between local and remote places in elevation localization with the head movement.

CONCLUSION

As described in this paper, sound localization accuracy in a median plane using an avatar robot 'TeleHead' with synchronization of listener's horizontal head rotation was investigated. The simplified TeleHead, which was responsive only to the listener's horizontal rotation, was developed and six dummy heads were constructed using stereolithography for all participants in experiments. Sound localization tests were performed



Restricted Condition



Free Condition

Figure 5: Results of experiment I (Listener 1)

in a median plane with control of synchronization of the avatar robot to the actual head rotation. The accuracy of sound localization in the median plane was improved even when the synchronization was limited to horizontal rotation. Furthermore, the ratios between the listener's actual movement and that of the TeleHead do not affect the accuracy of sound localization, suggesting remarkable robustness in using cues provided by head rotation. Consequently, the sound localization accuracy can be improved when the avatar head in a remote place rotates synchronously with a unity correlation coefficient to the listener's head rotation.

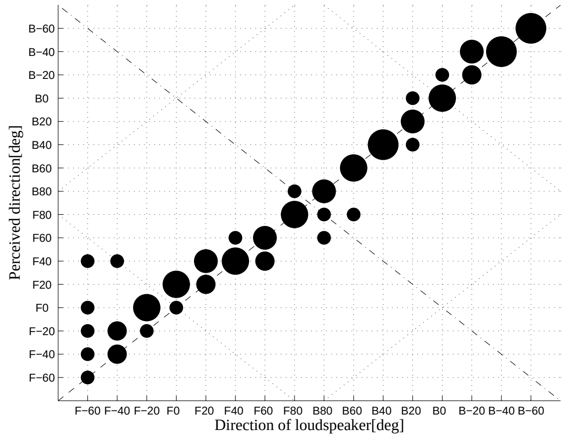
ACKNOWLEDGMENT

This study was partially supported by a Kakenhi Grant-in-Aid for Specially Promoted Research (to Y. Suzuki, 19001004) and a Kakenhi Grant-in-Aid for Scientific Research (C)(General) (to Y. Iwaya, 20500110). The authors thank Prof. T. Hirahara (Toyama Pref. University) for his cooperation in developing the TeleHead.

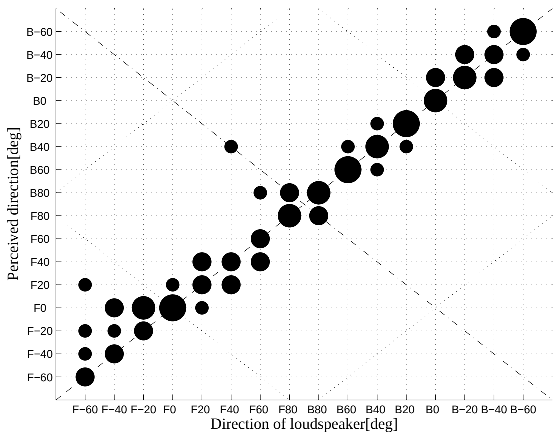
REFERENCES

[1] W. R. Thurlow, J. W. Mangels and P. S. Runge. "Head movement during sound localization," *J. Acoust. Soc. Am.*, Vol. 42, No. 2, pp. 489-493, 1967.

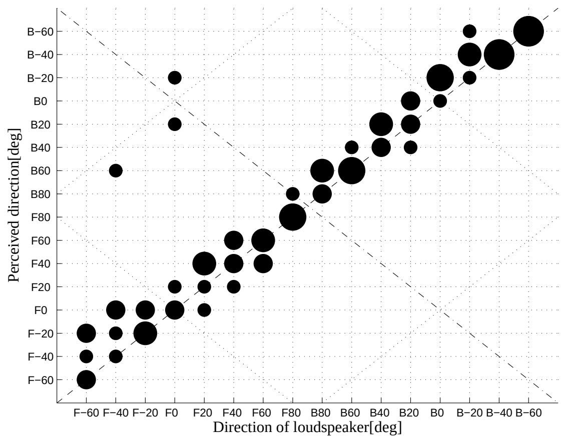
[2] W. R. Thurlow and P. S. Runge. "Effect of induced head movement on localization of direction of sounds," *J. Acoust. Soc. Am.*, Vol. 42, No. 2, pp. 480-488, 1967.



(a) ratio value = 0.5



(b) ratio value = 1.0



(c) ratio value = 1.5

Figure 6: Results of experiment II (Listener 1)

[3] S. Perrett and W. Noble, "The effect of head rotations on vertical plane sound localization," *J. Acoust. Soc. Am.*, Vol. 102, No. 4, pp. 2325-2332, 1997.

[4] Y. Iwaya, Y. Suzuki, and D. Kimura, "Effects of head movement on front-back error in sound localization," *Acoust. Sci. & Tech.*, Vol. 24, No. 5, pp. 322-324, 2003.

[5] Y. Iwaya, "Effects of head movement on sound space perception," *Proc. of Mini RIEC Workshop on Multimodal Perception*, 2009.

[6] Y. Iwaya, "Sound space perception in virtual environments with head movements," *Proc. International Workshop on*

the Principles and Applications of Spatial Hearing (IW-PASH), (pdf paper of 4 pages), 2009.

[7] S. Yairi, Y. Iwaya, and Y. Suzuki, "Influence of large system latency of virtual auditory display on sound localization task," *Acta Acustica united with Acustica*, Vol. 94, No. 6, pp. 1016–1023, 2008.

[8] D.-G. Kang, Y. Iwaya, R. Miyauchi, and Y. Suzuki, "Comparison of sound localization performance between visual and real three-dimensional immersive sound field," *Acoust. Sci. & Tech.*, Vol. 30, No. 3, pp. 216-219, 2009.

[9] A. J. Berkout, D. de Vries, and P. Vogel, "Acoustics control by wave field synthesis," *J. Acoust. Soc. Am.*, Vol. 93, No. 5, pp. 2764–2778.

[10] S. Ise, "A principle of sound field control based on the Kirchhoff–Helmholtz integral equation and the theory of inverse systems," *Acustica – Acta Acustica*, Vol. 85, No. 1, pp. 78–87, 1999.

[11] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics,“ *J. Audio Eng. Soc.*, Vol. 53, No. 11, pp. 1004–1025, 2005.

[12] S. Sakamoto, S. Hongo, R. Kadoi, and Y. Suzuki, "SENZI and ASURA: New high-precision sound-space sensing systems based on symmetrically arranged numerous microphones," *Proc. of Second International Symposium on Universal Communication (ISUC2008)* pp. 429–434, 2008.

[13] J. Blauert, "Spatial Hearing," The MIT Press, Cambridge, Massachusetts, 1997.

[14] Y. Iwaya, K. Saito, and Y. Suzuki, "Individualization method of HRTFs using a tournament-style listening test of virtual sound images," *Proc. of WESPAC IX* (in 8 pages in Conference CD-ROM), 2006.

[15] M. Otani and S. Ise, "Fast calculation system specialized for head-related transfer function based on boundary element method," *J. Acoust. Soc. Am.*, Vol. 119, No. 5, pp. 2589–2598, 2006.

[16] B.F.G. Katz, "Boundary element method calculation of individual head-related transfer function – I. Rigid model calculation," *J. Acoust. Soc. Am.*, Vol. 110, No. 5, pp. 2440–2448, 2001.

[17] Y. Iwaya and Y. Suzuki, "Numerical analysis of the effects of pinna shape and position on the characteristics of head-related transfer functions," *Proc. of Forum Acusticum 2008*, 2008.

[18] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Comparison of simulated and measured HRTFs: FDTD simulation using MRI head data," *Audio Engineering Society 123rd Convention* Paper 7240, New York, NY.

[19] T. Iwaki, S. Aoki, and T. Hirahara, "An acoustical telepresence robot: TeleHead 2," *Proc. of 2004 IEEE International Conference on Intelligent Robots and Systems*, 2004.

[20] T. Iwaki and S. Aoki, "Sound localization during head movement using an acoustical telepresence robot: TeleHead," *Advanced Robotics*, Vol. 23, pp. 289–304, 2009.

[21] S. Yairi, Y. Iwaya, and Y. Suzuki, "Estimation of detection threshold of system latency of virtual auditory display," *Applied Acoustics*, Vol. 68, No. 8, pp. 851–863, 2007.