The 2009 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 11-15, 2009 St. Louis, USA

Possibility of Simplifying Head Shape with the Effect of Head Movement for an Acoustical Telepresence Robot: *TeleHead*

Iwaki Toshima and Shigeaki Aoki, *Member, IEEE*

Abstract – **We built an acoustical telepresence robot named TeleHead, which has a user-like dummy head and whose movement is synchronized with the user's head movement in real time. An accurate-shape user-like dummy head improves sound localization accuracy, but making an accurate-shape user-like dummy head for all users is not realistic. There have been many efforts to simplify dummy heads without head movement in order to make a dummy head suitable for all users. Head movement also improves sound localization accuracy. Therefore, we are trying to simplify TeleHead's head shape by taking the effect of head movement into consideration. In this work, we made two types of simplified dummy heads, a ball-like dummy head and a ball-like dummy head with a user-like pinna, and used them in sound localization experiments. The experimental results show that the pinna is very important for sound localization in the median plane. Head movement can improve sound localization and subjects can localize sound with another person's pinna. However, it is hard for subjects to localize a sound without a pinna even with head movement. In addition, the acoustical characteristics of each dummy head are significantly different. The results indicate the possibility of using a ball-like dummy head with a generic pinna for acoustical telepresence robots.**

I. Introduction

One of the ultimate goals of telecommunications research is the development of technology that allows users to feel as if they are at a remote place. This is called telepresence technology [1]. A telepresence robot, which is an important technology for telepresence, works at a remote place instead of a human. For users to be able to feel as if they are indeed at a remote place, a telepresence robot should be able to work as if the user is at the remote place. It should also be able to transmit information about the environment, such as visual information and auditory information, correctly. Having a physical body at the remote place makes it possible for the user to have physical interactions. In general, telecommunications technology based on signal processing cannot provide physical interactions, at least not without some new equipment. Therefore, telerobotics technology can play an important role in realizing telepresence. To explore the possibilities of telepresence technology and solve the problems in achieving acoustical telepresence, we have built an acoustical telepresence robot named TeleHead [2, 3].

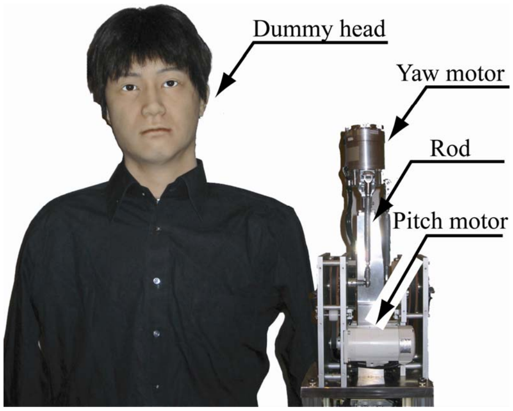

Fig. 1. Acoustical telepresence robot: TeleHead. It has a user-like dummy head and synchronizes with user's head movement in three degrees of freedom in yaw, roll, and pitch.

To realize acoustical telepresence, we need an understanding of human auditory characteristics. The sound localization function is one of the most fundamental functions. Previous studies have clarified that the acoustical characteristics of head shape, head-related transfer functions (HRTF) [4, 5], and head movement are important for sound localization [6, 7], and many three-dimensional acoustical display technologies using signal processing methods or dummy heads have been proposed [8 - 11]. However, because of the large individual differences in HRTFs, individual HRTFs or dummy heads have to be prepared. This is a critical demerit for promoting a telepresence robot like TeleHead for practical use. In addition, measuring correct HRTFs is difficult [12]. Therefore, we face many challenges in avoiding or solving the problems, such as exploring new HRTF measuring methods [13], generating a general HRTF [14, 15], selecting a suitable HRTF, calculating HRTFs from head shape [16], customizing from another person's HRTF using head shape [17, 18, 19], comparing and reviewing many dummy heads [20], and making simplified HRTFs or dummy heads [21]. In spite of these huge efforts, we still cannot avoid the problem of individual HRTF differences. On the other hand, head movement improves sound localization accuracy. In a head-movement situation, the HRTF is sometimes not so important [7]. Therefore, we are now trying to simplify TeleHead's dummy head, considering the effect of head movement. In this paper, we examine the relationships between simplified head shapes and sound localization accuracy in the head movement condition using our robot TeleHead.

I. Toshima is with NTT Communication Science Laboratories and Tokyo Institute of Technology, 3-1 Morinosato-Wakamiya, Atsugi-shi Kanagawa, Japan, (corresponding author to provide phone: +81-46-240-3575; fax: +81-46-240-4716; e-mail: toshima@brl.ntt.co.jp)
S. Aoki is with NTT Communication Science Laboratories, now: Kanazawa Institute of Technology. (e-mail: aoki_s@neptune.kanazawa-it.ac.jp)

978-1-4244-3804-4/09/$25.00 ©2009 IEEE

II. Outline of TeleHead

As shown in Fig. 1, the dummy head is driven with three motors, one each for yaw, roll, and pitch. Figure 2 outlines TeleHead. Head posture data of the user is measured with a six-dimensional position and posture sensor (Fastrak, Polhmus), and TeleHead is driven depending on the posture data. TeleHead has degrees of freedom in the yaw, roll, and pitch directions. The ranges of movement are sufficient for yaw, but smaller than human ranges of movement for roll and pitch (See [3] for details). There is an omni-directional microphone in each ear of TeleHead. Sounds are collected by these microphones and transmitted to the user through amplifiers and headphones (HDA200, Sennheiser). The dummy head is made as an accurate replica of the user to avoid the problems of HRTF individuality. Construction methods, a quantitative evaluation of the dummy head, the selection and effect of headphones, and the effects of the head shape and head movement are reported in another paper [3]. In that paper, we also confirmed that the accuracy of sound localization in the horizontal plane is almost the same when using TeleHead and when listening to the sound stimuli directly.

III. Simplified dummy heads

Clarifying the effect of reducing the physical accuracy of the dummy head in the head movement condition for sound localization is important for realizing an acoustical telepresence robot. For this purpose, we made the following simplified dummy heads: ball-like dummy heads with a user-like pinna (Fig. 3) and a ball-like dummy head without a pinna (Fig. 4). The dummy head in Fig. 3 has user-like pinna because the pinna has a large influence on the HRTF. The height of the dummy heads is 231 mm and the width is 155 mm. These dimensions are the same as those of the head of one of the subjects (subject 1). The pinna is set at the center point on each side of the ball-like dummy head. The size of the base for the pinna is shown in Fig. 3. The size of the base for the microphones is shown in Fig. 4. The pinna is made of

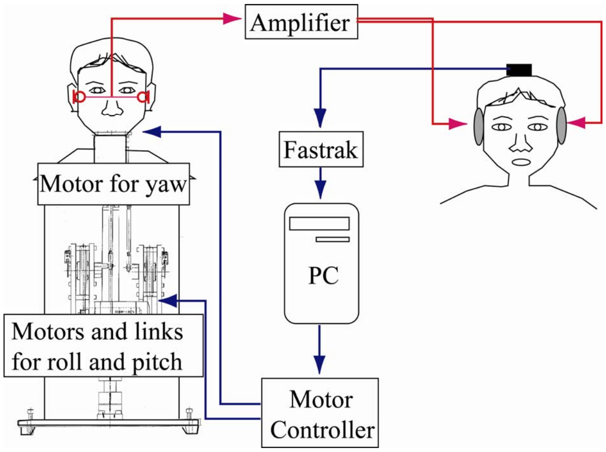

Fig. 2. Outline of TeleHead. TeleHead is synchronized with the user's head movement and the sound collected with microphones in the dummy head is transmitted to the user by headphones. Blue lines are the flows of head posture data. Red lines are the flows of acoustical signal.

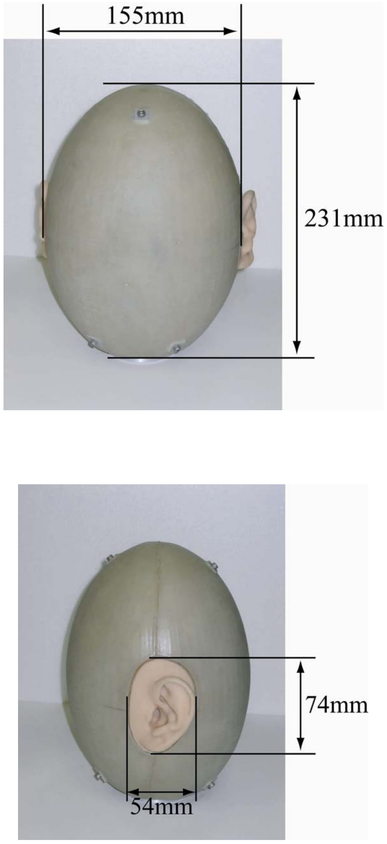

Fig. 3. Simplified dummy head with accurate pinna, in a front view (top panel) and side view (bottom panel).

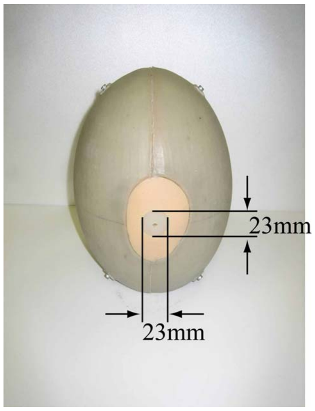

Fig. 4 Ball-like dummy head. There is no pinna on the head.

silicon. The ball-like body is made of FRP covered with silicon. To discuss the possibility of simplifying head shape for an acoustical telepresence robot, we performed sound localization experiments using these dummy heads with and without head movement.

IV. Sound localization Experiments

A. Method

Nine loudspeakers were arranged, -45 to 75, -50 to 70, or -40 to 80 degrees, at intervals of 15 degrees in the median plane. The distance between TeleHead and the loudspeakers was 1.2 m. A photograph of TeleHead and the loudspeakers is shown in Fig. 5. Subjects knew that there were loud speakers in the frontal median plane but could not know their exact positions. TeleHead was placed in an anechoic room and the subject was in a soundproof room. Sound stimuli were presented to subjects with headphones. The subjects could see neither TeleHead nor the loudspeakers. The duration of each stimulus was 8 s. The subjects reported the sound direction after the stimulus ended. The interval between stimulus presentations was 5 s. The stimulus was white Gaussian noise. Sound pressure was changed in every trial so that the subjects would not be able to use it in judging the direction of a sound. The range of the sound pressure level of the stimulus was roughly 55 to 65 dB SPL. Correct answers were not disclosed during the experiment or after it. Therefore, subjects could not check their replies and correct answers. One session consisted of five trials in each direction. Therefore, the median plane experiment consisted of 45 trials. One session was done for each loudspeaker arrangement. Therefore, there were three sessions in one condition. We did not do experiments in the horizontal plane because the effect of head shape in the median plane is stronger than that in the horizontal plane [3, 4].

Three subjects (subject 1, subject 2 and subject 3) took part in the experiments. These subjects have participated in a huge number of psychophysical experiments, including sound localization experiments. Since they are not naïve for such experiments, we should take learning effects into consideration. Making dummy heads is not so easy, which makes doing such experiments with naïve subjects impractical. Audiometer readings confirmed all subjects have normal hearing. We used four kinds of simplified dummy heads: a ball-like dummy head with subject 1's pinna (DH1a) (shown in Fig. 3), a ball-like dummy head with subject 2's pinna (DH2b), a ball-like dummy head with subject 3's pinna (DH3b), and a ball-like dummy head without a pinna (DH0b) (shown in Fig. 4). In addition, since head movement is also an important factor for sound localization, we did sound localization experiments with the dummy heads kept stationary and with them synchronized with the user's head movement.

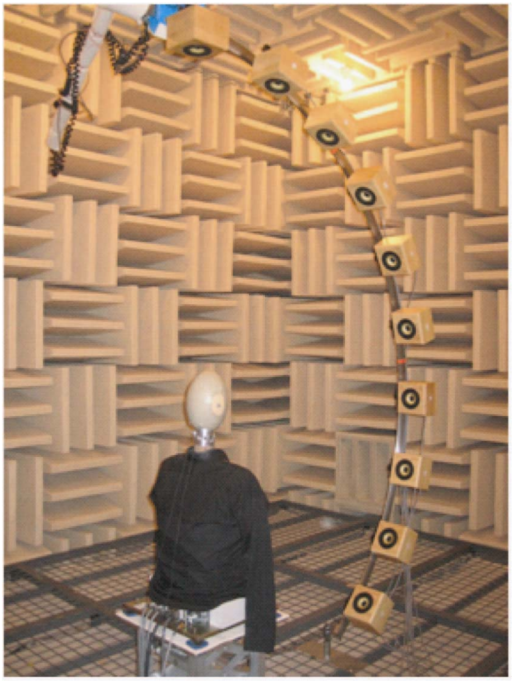

Fig. 5. Photograph of the setup for sound localization experiments. TeleHead with simplified dummy head is set in the anechoic room. Loudspeakers are set in the median plane from -45 deg to 75 deg at intervals of 15 degrees 1.2 m in front of TeleHead.

B. Result

Figure 6 shows examples of results. In Fig. 6, the abscissa represents the stimulus direction in degree and the ordinate represents the directions reported by the subjects. The stimulus direction was set to zero degrees at the front of a subject. The degree of the angle increases upwardly in the median plane. Therefore, correct answers are plotted on a diagonal line (red solid line). Figure 6 shows the results in the synchronized condition (with head movement). Left columns show the results for subject 1, center columns show those for subject 2, and right columns show those for subject 3. Top panels show the results for the ball-like dummy head with the user-like pinna. Bottom panels show those for the ball-like dummy head without a pinna. The others show those for the ball-like dummy head with a non-user-like pinna.

The ball-like dummy head with the user-like pinna (top panels) gave the best results. This means that a user-like pinna and head movement improve sound localization accuracy. The results with DH0b are a little better for subjects 1 and 3. This means that for two of the three subjects, head movement improved sound localization accuracy even when the ball-like dummy head was used. The result for subject 1 with DH1b is worse than with DH2b, but there is no significant difference. The results with the user-like pinna tend to be better than or, at least, comparable to those with the non-user-like pinna.

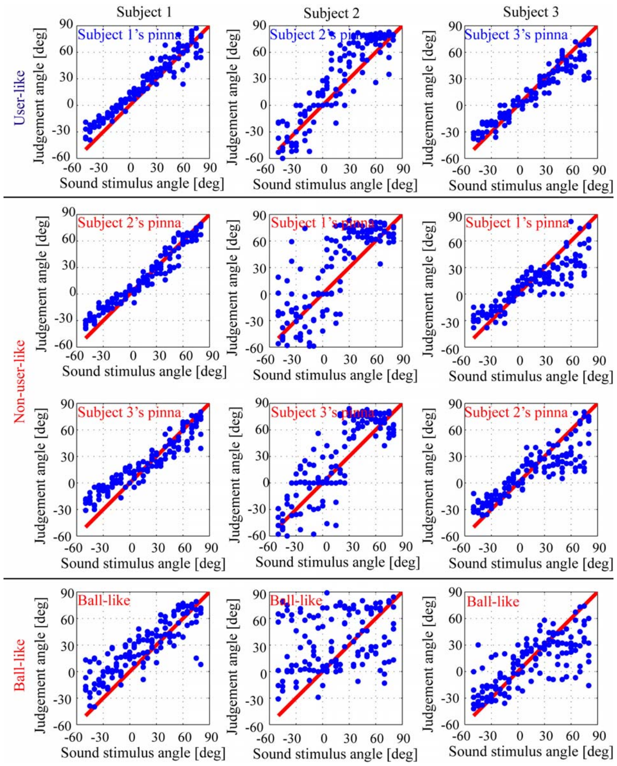

Fig. 6. Results of sound localization experiments in synchronized (head-movement) condition. Top panels show the results using dummy heads with a user-like pinna. Bottom panels show the results using the ball-like dummy head. The others show the results using the dummy heads with a nonuser-like pinna. Left column shows the results for subject 1, center column shows those for subject 2, and right column shows those for subject 3.

We have already done sound localization experiments with more accurate dummy heads and reported the effects of head shape and head movement. The results showed almost the same tendency as above. The physical accuracy of the dummy head, which is the user-like dummy head for subject 1 (Fig. 1), is roughly less than 2 mm. This accuracy improves sound localization considerably (See [3] for details).

To evaluate the sound localization accuracy, we calculated the correlation coefficient of each result. The calculation results are shown in Fig. 7. Results that were clearly up-down confusion [4] were omitted from the calculation. In Fig. 7, the upper panel shows the results without head movement and the lower panel shows those with head movement. Blue bars show the results for the ball-like dummy head with the user-like pinna. Green bars show the averaged results for two kinds of dummy head with the non-user-like pinna, and brown bars show those for the ball-like dummy head. The results for subject 1 tend to be better than for the other subjects in the synchronized condition. The heights and widths of all dummy heads are the same as those of subject 1. This may cause this tendency.

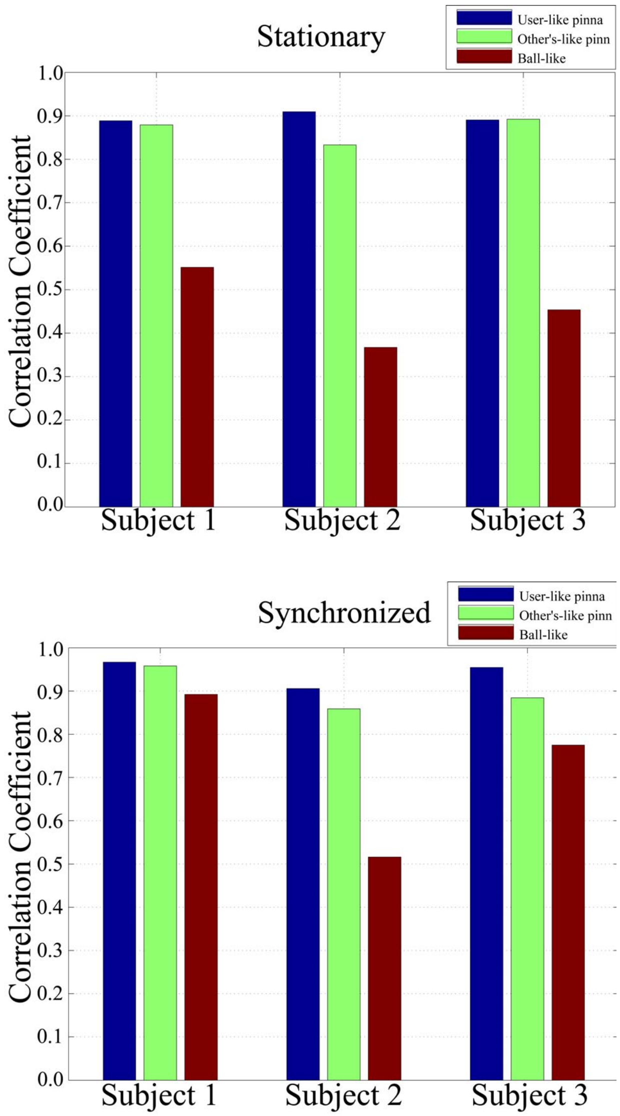

Fig. 7. Correlation coefficient of each result. Upper panel shows the results in the stationary condition and lower panel shows those in the synchronized condition. Blue bars show the results for the ball-like dummy head with user-like pinna. Green bars show the averaged results for two kinds of dummy head with the nonuser-like pinna, and brown bars show those for the ball-like dummy head.

Figure 8 shows the averaged correlation coefficient for all subjects. Error bars shows standard deviations of each result. Data for the stationary condition and synchronized condition are shown in the left and right graphs, respectively. There are significant differences between the results for the dummy head with the user-like pinna and those for the ball-like dummy head ($p<0.05$ in T-test), and between those for the non-user-like pinna and those for the ball-like dummy head ($p<0.05$ in T-test). The results show that only the sound localization results for the ball-like dummy head are worse than those for the other dummy heads. This means that omitting the pinna is not acceptable, even with head

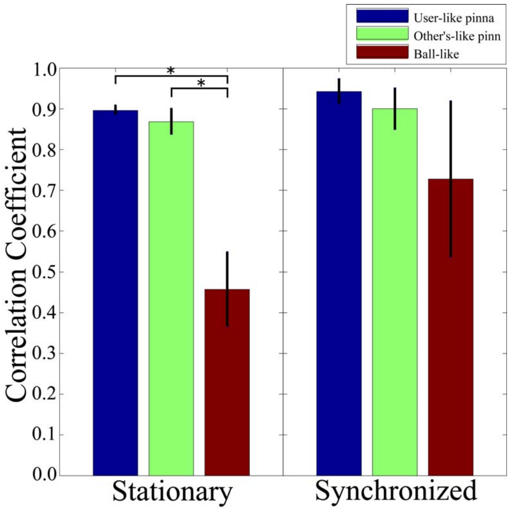

Fig. 8 Averaged correlation coefficient for all subjects. The left graph shows in the stationary condition, and the right graph shows in the synchronized condition. Error bars shows standard deviations of each result. The results for the ball-like dummy head are the worst. The results for the dummy head with user-like pinna and with non-user-like pinna are almost the same. There are significant differences between the result using the ball-like dummy head and the other results in the stationary condition.

movement. On the other hand, the results for the user-like pinna and non-user-like pinna are good. This suggests the possibility that the dummy head can be simplified under the head movement condition.

V. Acoustical Characteristics of Head Shape

The acoustical characteristics of head shape are important for discussing the effect of head shape. They are expressed as the HRTF [2-5, 13-21], which is defined as

$$HRTF(\omega,\theta,\phi,r)=\frac{H_{sp-l\ or\ -r}(\omega,\theta,\phi,r)}{H_{sp-center}(\omega,\theta,\phi,r)} \quad (1)$$

where $H_{sp-center}$ is the transfer function from a far-field sound source point to the head center point in a free field, and H_{sp-r} or H_{sp-l} is the transfer function from the sound source point to the ear-canal entrance of a specific listener's right or left ear. The ω is frequency, and θ, φ, and, *r* are the relative positions of the sound source. By convolving a pair of HRTFs of a certain azimuth and elevation with a sound source signal, the listener can localize a virtual sound source [5]. HRTFs of the real heads and dummy heads were measured in an anechoic room. The distance from the center of the head to the sound source was 1.2 m. The range of the measurement was zero to 360 degrees in azimuth and -40 to 90 degrees in elevation. Measurements were performed at 143 points. Each measurement point in the median plane and the horizontal plane was set at intervals of 10 degrees. Other measurement points were set so as their HRTFs could be interpolated from neighboring measurement points located less than 20 degrees in the vertical or horizontal directions. Here, we used 143 measurement points to keep the measuring time below 90 minutes and thereby reduce the burden on the subjects. Details of the method of measurement are described in [3].

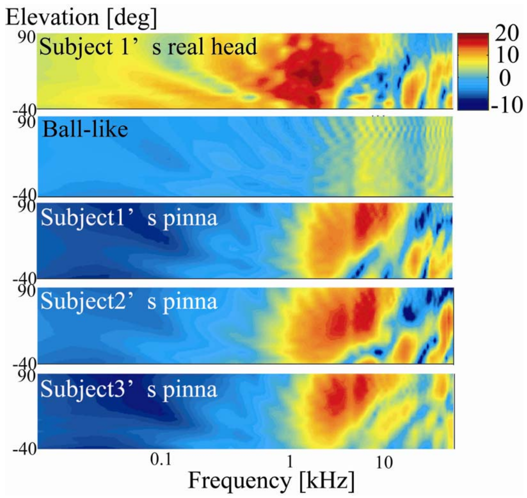

Fig. 9 Results of HRTF measurement for subject 1's real head, the ball-like dummy head, that with subject 1's pinna, that with subject 2's pinna, and that with subject 3's pinna.

Figure 9 shows the results of HRTF measurement. Red areas are high gain and blue areas are low gain. In the low-frequency area, the gain of the HRTFs of real heads is higher than the others'. This may be mainly due to the influence of the shoulders and body. The shapes of blue areas, which are higher than 8 kHz, are important for sound localization in the median plane [17, 18]. The shape of subject 1's HRTF and that of the HRTF of the ball-like dummy head with subject 1's pinna are similar. In contrast, the shape of the blue area for the ball-like dummy head is very different from that for the other dummy heads. Quantitative difference between two HRTFs of dummy or real heads i and j is defined as the spectral difference D_{HRTF}^{FFT} for all directions as

$$D_{HRTF}^{FFT}=\sum_d(\sqrt{\sum_\omega(|H_i-H_j|)^2/N_\omega})/N_d \quad (2)$$

where *d* is measurement direction, ω is angular frequency, N_ω the number of points of FFT, and N_d the total number of directions. The HRTF magnitude in direction *d*, is denoted as $H_i(\omega,d)$ and $H_j(\omega,d)$, abbreviated H_i and H_j. The frequency range for the calculation was 8 to 20 kHz, which is the important range for sound localization in the median plane. The average gains of HRTFs were equalized before we calculated the difference between the subjects and the dummy heads. Each measurement was done three times and Ds were calculated for all combinations of all measurements and averaged.

Table 1 Spectral differences between HRTFs [dB]

	RH2	RH3	DH1	DH2	DH3
RH1	7.38	7.90	7.50	7.87	7.66
RH2	−	8.67	8.85	8.48	8.30
RH3		−	8.23	8.30	6.42
DH1			−	6.79	6.63
DH2				−	6.59

Table 1 shows the spectral differences (SDs) [calculated by (2)] between HRTFs. RH stands for real head; DH stands for dummy head. Numbers just after RH or DH are the numbers of subjects. Blue blocks are SDs between the real head and the ball-like dummy head with the user-like pinna. The SDs between subject’s real-head and the ball-like dummy head with the user-like pinna are smaller than those between the subject’s real head and the ball-like dummy head with another’s pinna for subject 1 and subject 3, all of these differences are statically significant ($p<0.001$ in T-test). For subject 2, there is no significant difference. This means that the HRTFs of two of the three ball-like dummy heads with user-like pinna are similar to subject’s HRTF. Almost all HRTFs of subjects’ real heads and those of the simplified dummy heads are significantly different. Of course, SDs between RH1, 2, and 3 are all significantly different [3].

Considering the results of sound localization (Fig. 8), the results with head movement are not significantly different, even though the difference in the SDs of the HRTFs between user and dummy head is significant. This suggests that there is a possibility of simplifying the dummy head with the effect of head movement.

VI. CONCLUSIONS

We made two kinds of simplified dummy heads: a ball-like dummy head with a user-like pinna and a ball-like dummy head without a pinna. We performed sound localization experiments using them and measured their acoustical characteristics, HRTFs. The results suggest the following:

● With or without head movement, sound localization accuracy for the ball-like dummy head without a pinna is not good. This means that omitting the pinna is not acceptable even with head movement.

● In the head movement condition, sound localization accuracy is not significantly different between the ball-like dummy head with the user-like pinna and with another’s pinna in spite of the fact that these dummy heads’ HRTFs are significantly different. This indicates the possibility of using a ball-like dummy head with a generic pinna for acoustical telepresence robots.

ACKNOWLEDGMENT

We thank Dr. Masafumi Okada, associate professor at the Tokyo Institute of Technology, for his valuable advice.

REFERENCE

[1] R. M. Held, and N. I. Durlach, “Telepresence”, Presence: Teleoperators and Virtual Environments Vol. 1, pp. 109-112, 1992.

[2] I. Toshima, H. Uematsu, and T. Hirahara, “A steerable dummy head that tracks three-dimensional head movement: *TeleHead*”, Acoustical Science and Technology, Vol. 24, No. 5, pp. 327-329, 2003. (Online free site, http://www.jstage.jst.go.jp/article/ast/24/5/327/_pdf)

[3] I. Toshima, S. Aoki, and T. Hirahara, “Sound Localization Using an Acoustical Telepresence Robot: *TeleHead II*”, Presence, Vol. 17, No. 4, pp. 392-404, 2008.

[4] J. Blauert, "Spacial hearing: The psychophysics of human sound localization", MIT Press, Cambridge, Mass. , 1997.

[5] Henrik Moller and Michael Friis Sorensen and Dorte Hammershoi and Clemen Boje Jensen, "Head-Related Transfer Functions of Human Subjects", J. Audio Eng. Soc., Vol. 43, pp. 300-321, 1995.

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

[7] F. L. Wightman, and D. J. Kistler, "Resolution of front-back ambiguity in spatial hearing by listener and source movement", J. Acoust. Soc. Am. , vol. 105, no. 5, pp. 2841-2853, 1999.

[8] E. M. Wenzel, “Localization in virtual acoustical display”, Presence: Teleoperators and Virtual Environments, vol. 1, pp. 80-107, 1992.

[9] W. E. Kock, “Binaural Localization and Masking”, Journal of the Acoustical Society of America, vol. 22, no. 6, pp. 801-804, 1950.

[10] D. N. Zotkin, R. Duraiswami, and L. S. Davis, “Rendering localized spatial audio in a virtual auditory space”, IEEE trans. on Multimedea, vol. 6, no. 4, pp. 553-564, 2004.

[11] V. R. Algazi, R. O. Duda, and D. M. Thompson, “Motion-Tracked Binaural Sound”, Journal of the Audio Engineering Society, vol. 52, no. 11, pp. 1142-1156, 2004.

[12] K. A. J. Riederer, "Repeatability analysis of head-related transfer function measurements", 105th Audio Eng. Soc., Audio Engineering Society, No. 4846, 1998.

[13] D. N. Zotokin, R. Duraiswami, E. Grassi, and N. A. Gumerov, "Fast head-related transfer function measurement via reciprocity", J. Acoust. Soc. Am., Vol, 120, pp. 2202-2215, 2006",

[14] T. Nishino, M. Ikeda, K. Takeda, F. Itakura, “Interpolating Head Related Transfer Functions”, The seventh western pacific regional acoustics conference, pp. 293-296, 2000.

[15] D. W. Grantham, J. A. Willhite, K. D. Frampton, and D. H. Ashmead, "Reduced order modeling of head related impulse response for virtural acoustic displays", J. Acoust. Soc. Am., vol. 117, pp. 3116-3125, 2005.

[16] B. F. G. Katz, “Boundary element method calculation of individual head-related transfer function.”, J. Acoust. Soc. Am., vol. 110, no. 5, pp. 2440-2448, 2001.

[17] J. C. Middlebrooks, “Individual differences in external-ear transfer functions reduced by scaling in frequency”, Journal of the Acoustical Society of America, vol. 106, no. 3, pp. 1480-1492, 1999.

[18] J. C. Middlebrooks, “Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency”, Journal of the Acoustical Society of America, vol. 106, no. 3, pp. 1493-1510, 1999.

[19] J. B. Melick, V. R. Algazi, R. O. Duda,and D. M. Thompson, “Customization for personalized rendering of motion-tracked binaural sound”, 117th Convention of the Audio Engineering Society, Paper 6255, 2004.

[20] M. Pauli, F. M. Sorensen O. Krarup, C. Flemming, and H. Moller, "Localization with Binaural Recordings from Artificial and Human Heads", J. Audio Eng. Soc. vol. 49, pp. 323-336, 2001.

[21] D. W. Grantham, J. A. Willhite, K. D. Frampton, and D. H. Ashmead, "Reduced order modeling of head related impulse response for virtural acoustic displays", J. Acoust. Soc. Am., vol. 117, pp. 3116-3125, 2005.