2007 IEEE International Conference on
Robotics and Automation
Roma, Italy, 10-14 April 2007

Quantitative evaluation of delay time of head movement for an acoustical telepresence robot: *TeleHead*

Iwaki Toshima, and Shigeaki Aoki, *Member, IEEE*

***Abstract* – We built an acoustical telepresence robot, *TeleHead*, which has a user-like dummy head and is synchronized with the user's head movement in real time. We are trying to clarify the effects of reproducing head movement. In this paper, we evaluated the sense of incongruity induced by the delay time in reproducing head movement. The results indicate that head movement control should have a dead time shorter than 27 ms. In addition, this dead time does not depend on a head shape of an acoustical telepresence robot in terms of guidelines for building an acoustical telepresence robot. The results also suggest that the cue for the discrimination of delay is not the delay time itself. They suggest that subjects might discriminate the difference between the perception of auditory sound localization and somatosensory perception of their head posture.**

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 have two functions: The robot should be able to work as if the user is at the remote place, should be able to transmit the 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, no other telecommunications technology using signal processing can provide physical interactions, at least not without some new equipment. Therefore, telerobotics technology can play an important role for realizing telepresence.

We face many challenges in achieving acoustical telepresence [2], [3]. We are trying to build an acoustical telepresence robot. Auditory functions are important in helping humans understand an environment. Auditory functions, such as caution, work for all directions and play an important role for understanding environment. In addition, they are very important for communication. Therefore, transmission of the sound environment is necessary for humans to understand the environment using telepresence.

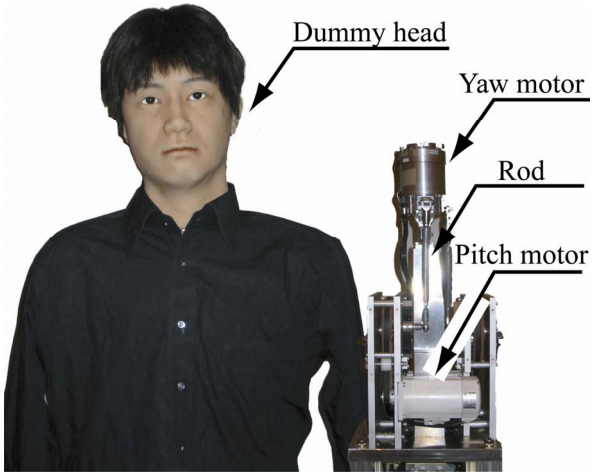

Fig. 1. Acoustical telepresence robot: *TeleHead*. It has a user-like dummy head and synchronizes with user's head movement in three degrees of freedom.

Another merit of using robots, besides the possibility of interaction at the remote place, is that the body can be used for understanding a sound environment. Human beings understand an environment through their senses. In the case of audition, the acoustical characteristics of the body are important for understanding a sound environment. The acoustical characteristics of head shape, head-related transfer functions (HRTFs), are especially important for understanding the direction of a sound, which is called sound localization function [4]. In addition, for sound localization, not only stationary cues but also dynamic cues are important [5][6]. Taken together, the above-mentioned facts indicate that a robot for realizing telepresence technology should have a body and should act at a remote place.

We built an acoustical telepresence robot that has a user-like dummy head and is synchronized with the user's head movement in real time. Therefore, users can use the information obtained from head shape and head movement. We named the acoustical telepresence robot *TeleHead* (left panel of Fig. 1). We have evaluated the effects of head shape and head movement quantitatively in sound localization experiments. The results showed that both the user-like

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@avg.brl.ntt.co.jp)
S. Aoki is with NTT Communication Science Laboratories (e-mail: aoki@avg.brl.ntt.co.jp)

dummy head and synchronization with user's head movement improve the accuracy of sound localization [7]. Moreover, *TeleHead* does not require information about the sound source beforehand. This is also one of the merits compared with methods using HRTFs or the binaural recording method [8].

While there are many merits in using a robot for acoustical telepresence, as pointed out above, there are some demerits as well, such as noise and delay. Although, with robots, noise and delay problems are largely unavoidable, we have been able to essentially solve the former for *TeleHead*, The delay problem still remains though, and we think there are two ways for robotics researchers to solve it. One is to try to reduce the delay to zero. The other is to try to make it so the user does not feel the delay. The former method has a limitation, and if we increase feedback gain, the noise will increase too. Moreover, auditory is one of the most severe sensors of time. Therefore, we consider it would be better to minimize the effect of delay from the perceptual point of view for controlling the robot on the basis of the characteristics of human auditory and motor perception. For this purpose, we have been trying to clarify the characteristics of human auditory perception and have proposed two type of quantitative evaluation for acoustical telepresence robots. One is to measure the accuracy of sound localization and the other is to measure differential thresholds. The accuracy of sound localization means how accurately users can judge the directions of sound. This is one of the basic functions of the human auditory system. It can be used to evaluate the working efficiency using an acoustical telepresence robot. On the other hand, we measure the difference threshold by discriminating between two different time delays of head movement. If the subject can perceive the difference in the two delays, it means that the subject can feel a sense of incongruity.

The delay of head movement generates the sense of incongruity and also decreases the accuracy of sound localization. The sense of incongruity deteriorates the system from the perceptual point of view. Accuracy of sound localization affects the efficiency of work using the robot. Therefore, on the basis of the results of sound localization experiments and discrimination experiments, the delay can be classified into four regions: regions where subjects can (or cannot) localize sound accurately, and those where they can (or cannot) discriminate the delay of head movement. Logically, there are four regions, but it is never the case that subjects cannot localize sound accurately but cannot feel the delay of head movement. Therefore, the delay is classified into three regions: the region where subjects can localize sound accurately and cannot feel the sense of incongruity, that where they can localize sound accurately but feel the sense of incongruity, and that where they cannot localize sound accurately and can also feel the sense of incongruity. We know that subjects can localize sound accurately using

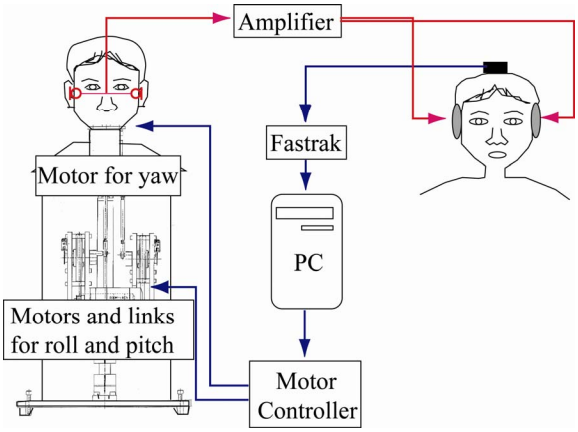

Fig. 2. Outline of *TeleHead*. *TeleHead* is synchronized with the user's head movement and the sound collected with microphones in the dummy head is transmitted to the user by headphones. Blue lines are the flows of head posture data. Red lines are the flows of acoustical signal.

TeleHead. Now, we want to determine how much time delay is required for the subjects to feel incongruity. This can be one of the guidelines for making an acoustical telepresence robot. As a first step to clarifying this, we should clarify how humans feel the sense of incongruity. In this paper, we will discuss what causes the sense of incongruity. Then, we will propose a guideline for delay to build an acoustical telepresence robot.

II. Outline of *TeleHead*

As shown in Fig. 1, the dummy head is driven with three motors for yawing, rolling, and pitching. 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 three degrees of freedom, which are yaw, roll and pitch. The ranges of movement of *TeleHead* are sufficient for yaw, but smaller than human ranges of movement for roll and pitch (See [7] for details). There is a omni-directional microphone in each ear of *TeleHead*. Sounds are collected by these microphones and transmitted to the user through amplifiers and headphones. 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, and the effect of the head shape and head movement are reported in another paper [7]. 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. Discrimination Task

A. Method

TeleHead was set in an anechoic room (Fig. 3). Sound stimuli were generated from a loudspeaker (Vifa, MG10SD09-08) set 1.2 m in front of *TeleHead*. The duration of stimuli was 8 s, and the interval between stimuli was 6 s. The sound level was adjusted to a comfortable one for discrimination, roughly about 65 dBA (A weighting filtered sound pressure level) at the dummy head and also roughly about 65 dBA at the subjects. Subjects sat in a soundproof room and listened to the sound stimuli through headphones (Sennheiser, HDA200). *TeleHead* followed the subjects' head movement while subjects listened to the sound. We used a constant method. There were three standard stimuli, 80-, 100-, and 120-ms delay. There were seven comparison stimuli, standard stimuli plus 0, 10, 20 …, and 60 ms. The sound stimuli were ordered randomly, and subjects did not know the order. Subjects listened to a pair of stimuli (standard stimulus and comparison stimulus). Then, they were required to judge which stimulus was natural, with the expectation that if they could feel a delay of the dummy head movement, they would judge the stimulus as unnatural. The method was forced choice; therefore, even if they could not feel any delay, they could possibly answer correctly 50% of the time. We decided that a 75 % correct-answer rate would mean subjects could feel the difference between the standard stimuli and comparison stimuli. Each session consisted of five pairs of stimuli in each condition. Therefore, thirty-five pairs of stimuli (70 stimuli) were generated. Ten sessions were done for each condition. We used white noise as an acoustical stimulus. In case of using *TeleHead* in the real world, of course, white noise is not preferable for transmitting sound environment. Speech would be better for evaluating *TeleHead* or clarifying guidelines for building an acoustical telepresence robot. However, speech contains some silent periods. Because a task using speech is more difficult than using white noise, it may make the thresholds larger. Smaller thresholds lead to severer guidelines for making an acoustical telepresence robot. Therefore, we chose white noise, which is the easiest stimulus and may provide the smallest thresholds as acoustical stimuli. First, we will discuss the results for white noise. After that, we will also mention experiments using speech. Then, we will discuss the difference between using white noise and using speech.

Head-shape conditions are also important for discussing the auditory perception. A user-like dummy head should be used for *TeleHead*. However, a perfect user-like dummy head is difficult to make. Therefore, clarifying the effects of head shape is important for discussing guidelines for building an acoustical telepresence robot. In the present experiments, we used two subjects (subject 1 and subject 2)

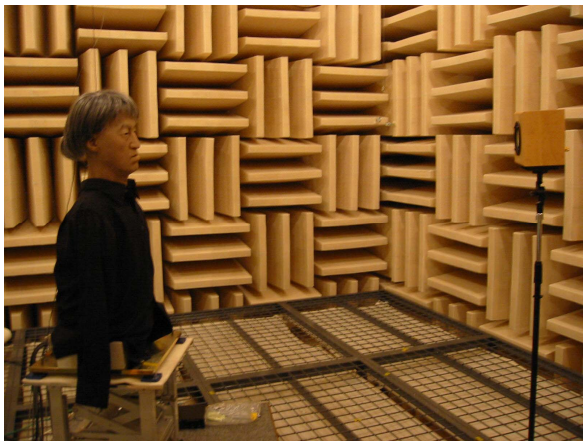

Fig. 3. Photograph of the experiment. Dummy head 3 (DH3) was set on *TeleHead* in an anechoic room. A loudspeaker was set 1.2 m in front of *TeleHead*.

and two dummy heads [dummy head 1 (DH1) and dummy head 3 (DH3), which are respectively shown in Figs. 1 and 3]. DH1 is a user-like dummy head of subject 1. DH3 is not user-like dummy head of either subject. The two dummy head are different from each other in physical shape and acoustical characteristics [7]. Therefore, there are four conditions: subject 1 with DH1/DH3 and subject 2 with DH1/3. By comparing the results obtained in these conditions, we can discuss the effects of head shape and individual difference.

B. Results

Figure 4 shows the raw results of using white noise. The upper panel shows the results for subject 1/DHI (DH1 is very alike of subject 1). The second panel shows the results for subject 1/ DH3. The third panel shows the results for subject 2/ DH1. The bottom panel shows the results of subject 2/ DH3. Of course, subject 2 differs from DH1/DH3 in the physical and acoustical characteristics [7]. Dots show the raw results of the experiments, and lines show the psychometric functions presumed from the dots. We assumed that the psychometric functions are logistic curves such that

$$P(X \le x) = \frac{1}{1+e^{-(a+bx)}} \qquad (1)$$

and we calculated a and b to minimize the squared error. The psychometric functions were evaluated using R as

$$R = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2} \qquad (2)$$

where y is the result of measurement, $\hat{y}_i$ is presumed data, and $\bar{y}_i$ is averaged data. They were fitted with more than 95 % probability for subject 2 and with more than 86% probability for subject 1. The results in the four panels are almost the same. However, there are small differences between the results for the user-like dummy head and non-user-like dummy head. This suggests that difference thresholds of head movement delay and the accuracy of reproduction of head shape are independent in terms of guidelines for building an acoustical telepresence robot.

The Weber fraction (ratio of standard stimuli to threshold) is shown in Fig. 5. If subjects judge the delay time itself, the Weber fraction should be constant. However, in this case, it was not constant. It is clear that the Weber fraction is smaller for the longer standard stimuli, suggesting that subjects did not judge the delay time itself. Difference thresholds for each standard stimulus are shown in Fig. 6. The results for each condition do not have the same tendency. The results for each condition are almost the same value. The average of the difference thresholds is about 27 ms.

C. *Characteristics of head movement*

To put discussion of the results in context, we should first look at the characteristics of head movement. A histogram of head movement speed is shown in Fig. 7. This is a typical example of head movement speeds during an experiment classified by head speed every 50 deg/s. We also measured average of the total time of head movement speed during a trial (8 s). The histogram is based on the total time. Maximum head speed was about 400 deg/s. Subjects can not move their head constantly at a speed of over 360 degrees/s. Average head speed was about 200 deg/s. This shows that subjects moved their heads quickly to feel sense of incongruity of head movement. We measured the fastest head movement without any additional task for each subject. Even in that case, the head movement speed of each subject was almost the same as it was in the experiments in this paper.

D. *Experiments using speech*

Results of the experiments using speech are shown in Fig. 8. All conditions were the same as those using the white noise. The standard stimulus was 80 ms. We did not use the other standard stimuli. The psychometric function was also calculated in the same manner. Difference thresholds using the white noise and speech are shown in Fig. 9. For subject 1, the results using white noise and speech are almost the same. In addition, the results using DH1 and DH3 are almost the same. This indicates that both head shape and acoustical stimuli do not contribute the discrimination task for subject 1 in terms of guidelines for building an acoustical telepresence robot. In contrast, for subject 2, using the white noise was easier than using the speech. We think that there are many

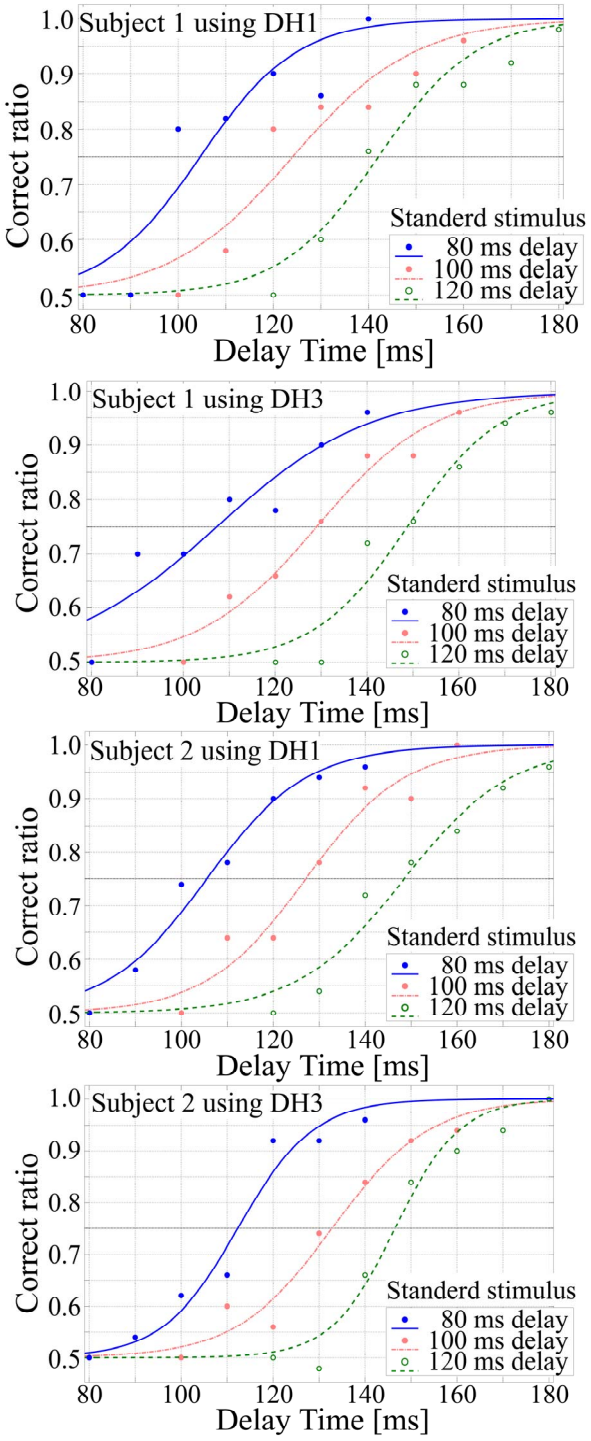

Fig. 4. Psychometric functions for delay of head movement: subject 1/DH1 (top panel), subject 1/DH3 (second panel), subject 2/DH1 (third panel), and subject 2/DH3 (bottom panel). Difference thresholds are crossing points of the psychometric functions and 0.75 correct ratio.

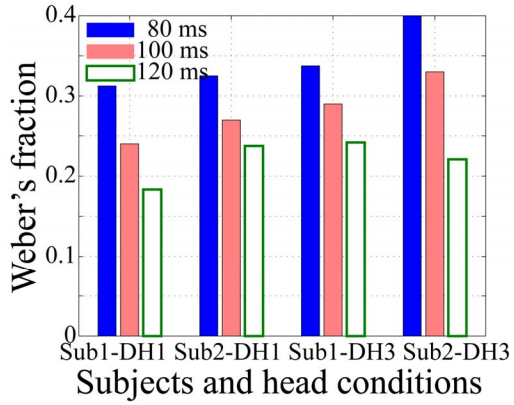

Fig. 5. Weber's fraction. Weber's fractions were different with each standard stimulus, indicating that subjects did not feel the time delay itself.

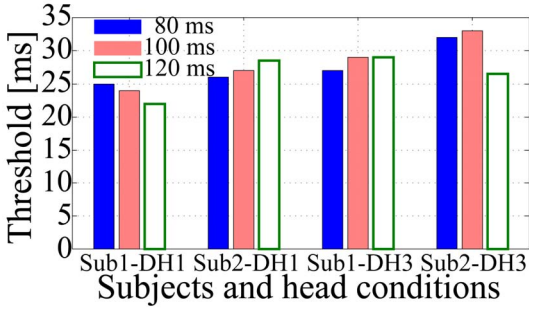

Fig. 6. Difference thresholds of delay. The results were almost constant between conditions.

disadvantages to discriminate using speech. For example, speech has some silent periods, and humans cannot avoid thinking of the meaning of the phrase [9].

E. Discussion

The results indicate that subjects did not feel the time delay between head movement and *TeleHead*'s movement itself but did feel the difference from indirect information of sound localization. The difference between the head posture of the subjects and *TeleHead* was calculated from the head movement speed and delay time. It was about 6 degrees on average and about 11 degrees maximum. These ranges of angles are larger than the minimum audible angle, which is about 1 degree (There are many researches and results about that, for example [10]). On the other hand, the ranges are smaller than minimum audible movement angle (In this case, this does not mean that the minimum angle at which the subject can feel movement of sound source [11], but the minimum angle at which the subject can feel a change of the position of a moving sound source [12].) depends on the speed of head movement. In this experiment, subjects' head speed was about 200 deg/s on average, and about 360 degr/s

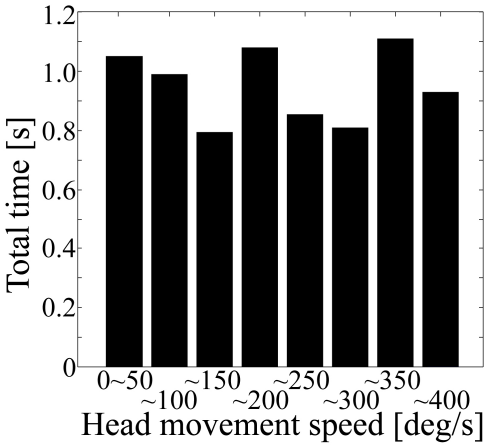

Fig. 7. Histogram of head movement speed. It is a typical example of head movement speed during an experiment classified by head speed every 50 deg/s. The histogram shows that subject moved their head almost as quickly as possible.

maximum. The minimum audible movement angle around this speed ranges from 15 to 20 degrees. Therefore, the result of the experiment is smaller than the minimum audible movement angle. There is no research for perception of movement sound source with listener's head movement. We think this result suggests that the minimum audible angle with a moving sound source and subjects' moving their head by themselves would be different. We hypothesize that subjects judge the delay not by the delay time but by the spatial information of the sound stimulus and head movement, which would explain all the present results. Moreover, the difference between the shape of the subject's head and the dummy head decrease accuracy of sound localization for each subject. This would also explain the small differences between the results for subject 1/ DH1 and subject 1/ DH3 (Difference thresholds using DH1 are a little smaller than using DH3).

In this paper, we discussed the delay time of robot's head movement. However, there were two types of delays in these experiments. One is *TeleHead*'s original delay of control (80 ms). The other comprises additional dead times for the experiments (0, 20, and 40 ms). Therefore, in the other words, 27-ms dead time is the threshold of dead time that does not depend on the dead time of the standard stimulus (0, 20, and 40 ms in this case). Dead time is important for building an acoustical telepresence robot, because it can be considered as a transmission delay. In addition, with respect to head movement prediction for control, this dead time can be considered the maximum time that can be used to take data and perform calculation for the prediction.

The 80-ms delay derived from *TeleHead* is too long compared with that for other robots. However, this long delay is closely related to realizing with the low driving noise. High gain and stiffness make vibration stronger and strong

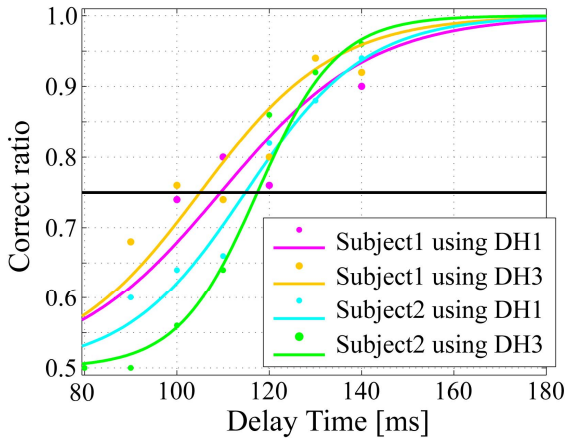

Fig. 8. Difference thresholds using speech. Standard stimulus in all experiments was 80-ms delay. There were two subjects and two types of dummy head, for a total of four conditions.

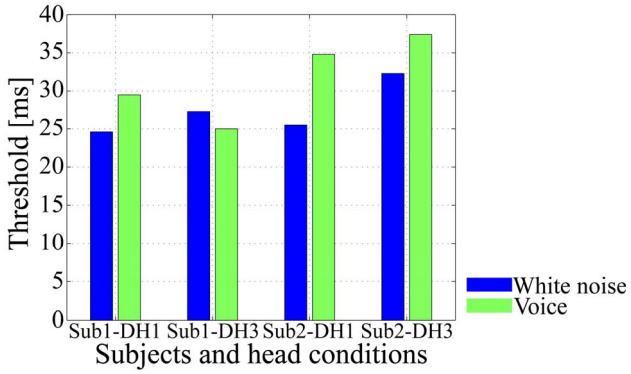

Fig. 9. Difference thresholds using white noise and speech. The standard stimulus was 80 ms.

vibration results in large driving noise. The microphones in the dummy head are directly set on the robot. Therefore, if the robot vibrates, the microphones pick up the vibration as loud noise. Noise problems should be solved by modifying the robot mechanisms and control methods. The results indicate that 27-ms delay is acceptable for humans to use an acoustical telepresence robot. They also indicate that we can use the 27-ms dead time for solving the noise problems and of course any other problems in building an acoustical telepresence robot. Moreover, if the sense of incongruity from using an acoustical telepresence robot is mainly derived from the delay, 27 ms has very important meaning in acoustical telepresence robot construction. The results depend on head shape a little. They suggest that an acoustical telepresence robot should be moved with dead time lower than 27 ms. In addition, this condition for our experiment is the most severe case. Subjects moved their head as fast as possible to detect the delay. In the case of slower and smaller head movements, subjects may not readily feel the sense of incongruity because this detection depends on the differences in the head posture of the subject and *TeleHead*. Therefore, dead time of more than 27 ms may be acceptable in general using situation.

VI. Conclusions

We conducted an experiment in which the task was for subjects to discriminate the delay of head movement of an acoustical telepresence robot with *TeleHead,* which has a user-like dummy head and is synchronized with user's head movement in real time. The results suggest three things as follows

- An acoustical telepresence robot should be built with less than 27-ms dead time.
- Difference threshold of the delay-discrimination task does not change largely with the conditions of head shape in terms of guidelines for building an acoustical telepresence robot. However, with a user-like dummy head, the difference thresholds are a little smaller than with a non-user-like dummy head.
- The cue of the discrimination of delay is not the delay time itself. Results suggest that subjects might discriminate the difference between the perception of auditory sound localization and somatosensory perception of their head posture.

These results can be used for building and controlling an acoustical telepresence robot, especially the control considering transmission delay and head movement prediction.

Reference

[1] R. M. Held, and N. I. Durlach, "Telepresence", Presence: Teleoperators and Virtual Environments Vol. 1, pp. 109-112, 1992.
[2] E. M. Wenzel, "Localization in virtual acoustical display", Presence: Teleoperators and Virtual Environments, vol. 1, pp. 80-107, 1992.
[3] 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.
[4] J. Blauert, "Spacial hearing: The psychophysics of human sound localization", MIT Press, Cambridge, Mass. , 1997.
[5] H. Wallach, "On sound localization" , J. Acoust. Soc. Am., vol. 10, pp. 270-274, 1939.
[6] 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.
[7] I. Toshima, S. Aoki, T. Hirahara, "An acoustical tele-presence robot: *TeleHead* II", Proc. of International conference on intelligent robots and systems (IROS) 2004, pp. 2105-2110, 2004.
[8] H. Moller, "Fundamentals of binaural technology", Applied Acoustics, vol. 36, pp. 171-218, 1992.
[9] I. Kinoshita, and S. Aoki, "Continuous apparent motion by successive presentation of sound for miscellaneous sources", J. Acoust. Soc. Jan. vol. 18, no. 3, pp. 139-141, 1997.
[10]A. W. Mills, "On the Minimum Audible Angle", J. Acoust. Soc. Am. , vol. 30, no. 3, pp. 237-246, 1958.
[11]K. Saberi, and D. R. Perrott, "Minimum audible movement angles as a function of sound source trajectory", J. Acoust. Soc. Am. , vol. 88, pp. 2639-2644, 1990.
[12]D. R. Perrott, and A. D. Musicant, "Minimum auditory movement angle: Binaural localization of movement sound sources", J. Acoust. Soc. Am. , vol 62, pp. 1436-1466, 1977.