2013 Ninth International Conference on Intelligent Information Hiding and Multimedia Signal Processing

Personal Auditory Tele-existence System Using a TeleHead

Tatsuya Hirahara
Faculty of Engineering
Toyama Prefectural University
Imizu, Japan
hirahara@pu-toyama.ac.jp

Daisuke Morikawa
School of Information Science
JAIST
Nomi, Japan
morikawa@jaist.ac.jp

Yukio Iwaya
Faculty of Engineering
Tohoku Gakuin University
Tagajo, Japan
iwaya.yukio@tjcc.tohoku-gakuin.ac.jp

Abstract— **A TeleHead is a steerable dummy head that tracks a listener's head movement quickly and quietly. We made a personal auditory tele-existence system connecting Toyama Prefectural University and Research Institute of Electrical Communication Tohoku University by using a TeleHead over the Internet. The remote TeleHead can provide dynamic binaural signals associated with the head-movement data input, and the local listener can hear the sounds represented by them. The listener perceives the three-dimensional sound image as if he were at the remote place even when the system has a total latency of 223 ms.**

Tele-existence; TeleHead; Dynamic binaural signal; 3D sound; Internet (key words)

I. Introduction

Telephones enable us to communicate verbally over long distances. They virtually put the speaker's mouth close to the listener's ear by bringing to a remote receiver the electrical signals transduced by a local microphone. Although a traditional telephone uses monaural sound, recent technologies enable us to handle three-dimensional (3D) sound. One that is promising for realizing a personal 3D sound communication system is binaural technology.

In the digital era, binaural signals are usually synthesized with a head-related transfer function (HRTF)-based digital binaural system. Most of the digital binaural systems, however, take something away from the synthesized binaural signal and add something to them. Some are static systems, which do not handle dynamic binaural signals associated with a listener's head movement. And even the dynamic binaural systems usually remove early reflections as well as reverberations because head-related impulse responses (HRIRs) reduced in length are used to synthesize binaural signals. Doppler effects caused by the listener's head movement are also usually ignored. In a dynamic binaural system, the spatial resolution of HRIRs should be less than 1 degree in order to avoid impulsive noises due to waveform discontinuities [1]. HRIRs of such fine spatial resolution are not measured, but are obtained by interpolating the measured HRIRs of coarse spatial resolution (e.g. 5 to 10 degrees). The use of interpolated HRIRs of course makes the synthesized binaural signals different from the actual binaural signals. In addition, any digital systems synthesizing binaural signals should know the locations of the sound sources. In the virtual world the locations of sound sources are given, but in the real world they are not. And determining the locations of sound sources in the real world is not an easy task.

If we had a dummy head that had the same shape as the listener's head and tracked the listener's head movement, it could provide perfect dynamic binaural signals for the listener. A TeleHead is a steerable personal dummy head that tracks a listener's head movement in real time quietly [2]. Two small microphones at its right and left ears provide dynamic binaural signals associated with the listener's head movement. Nothing is taken away from or added to the electrical signals they produce. The signals represent the actual binaural sounds obtained physically without using information about the locations of the sound sources. When the binaural signals are reproduce by earphones at the listener's ear, the listener perceives 3D sound as if he/she were at the remote place. A TeleHead can be an ideal auditory tele-existence device.

In this paper, we describe a personal auditory tele-existence system that uses a TeleHead over the Internet. We also demonstrate its performance by showing the results of sound localization experiments with the system.

II. Architecture of TeleHead

Figure 1 shows the architecture of TeleHead IV [3]. A listener’s head movement is tracked by a motion sensor (Ascension Technology's Flock of Birds) fixed at the top of the head. The head rotation angle data sampled at 120 Hz are sent to a servo amplifier via a control PC. TeleHead V has the same architecture, but its motor and servo amplifier are different from those of TeleHead IV, and its motion sensor is Polhemus FASTRAK.

The dummy head is a personalized one accurately reproducing the shape of an actual listener's head. In the work reported here, each listener was an adult male listener and the dummy heads were made by using a rapid prototyping system. The 3D shape of each listener's head was measured by a 1.5 T magnetic resonance imaging (MRI) system.

A small electret condenser microphone (SONY ECM77B) embedded in an earplug made of silicone impression material was placed in the vicinity of each of the outer-ear canal entrances (left and right) of the dummy head. The output signals of the microphones were amplified 35 dB by microphone amplifiers (Earthworks 1021) and headphone

978-0-7695-5120-3/13 $26.00 © 2013 IEEE
DOI 10.1109/IIH-MSP.2013.88

amplifiers (audio-technica AT-HA20), and then reproduced with closed-type circumaural dynamic headphones (Sennheiser HDA200) [4].

III. TeleHead over the Internet

To use the TeleHead over the Internet, one needs a two-channel wideband sound codec for binaural signals and a serial data communication device for head movement data (Figure 2).

A. *Binaural-signal transmission device 1*

One of the binaural signal transmission devices we used was a video teleconference system (SONY, PCS-XG55) with a built-in an MPEG-4 advanced audio encoding (AAC) stereo codec (192 kbps). When the two video conference systems were connected to the local area network (LAN) of Toyama Prefectural University (TPU), the latency of the binaural signals transmission was 370 ms. The round-trip packet delay between the two systems measured by the ping command was less than 1 ms. Thus, most of the 370 ms was the encoding and decoding time in the AAC codec.

When the two video conference systems were connected over the Internet between TPU and the Research Institute of Electronic Communication (RIEC) Tohoku University, the round-trip latency of the binaural signals transmission including encoding and decoding time at both side was 1268 ms. Thus, the one-way latency was 264 ms. On the other hand, the one-way packet delay between the two systems at that time was about 23 ms.

The interaural time difference (ITD) of the binaural signals diffed by as much as 20 µs between input and output signals, and the corresponding interaural level difference (ILD) was as much as 0.5 dB. The frequency response of the system was ±2 dB between 200 Hz and 5 kHz but dropped to −6 dB at 100 Hz and 12 kHz. This difference in frequency response must be due to the codec used in the system.

B. *Binaural-signal transmission device 2*

Another binaural signal transmission device we used was a low-latency stereo-sound codec system (APT, WorldCast Astral), which offered high-quality audio transport over IP networks using RTP/UDP. From several types of sound codec available with the codec system, we chose the 20-kHz bandwidth 16-bit Linear PCM codec.

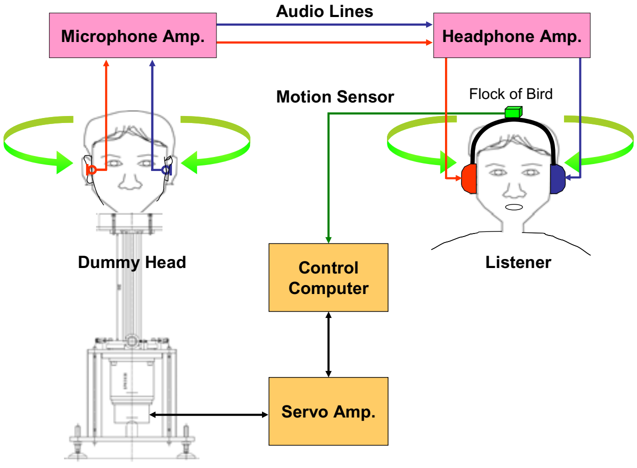

Figure 1. Architecture of TeleHead

The one-way latency for transferring the binaural signals from the RIEC to TPU was 80 ms with the low-latency codec system. The one-way packet delay measured by the ping command was also about 23 ms at that time. There was no ITD difference between input and output binaural signals, but the ILD difference between them was 1.9 dB. The origin of this ILD difference must be the difference of gain between the right- and left-channel analog amplifiers in the codec system. The frequency response of the system was flat form 100 Hz to 4 kHz but dropped to -6 dB at 20 kHz.

C. *Head-movement-data transmission device*

Head-rotation-angle data of a listener measured every 1/120 seconds by the head tracker were sent to the TeleHead controller via an RS-232C serial interface. The local and the remote RS-232C lines were connected over the Internet via UDS2100s device servers (Lantronix).

When the listener's head-movement data were loop-backed RIEC-TPU-RIEC over the Internet with TeleHead V, it took about 145 ms to reproduce the listener's head movement. As the mechanical motion delay of TeleHead V is about 100 ms, the round-trip latency of the data transfer was estimated to be 45 ms, which was also the round- trip latency for a single packet.

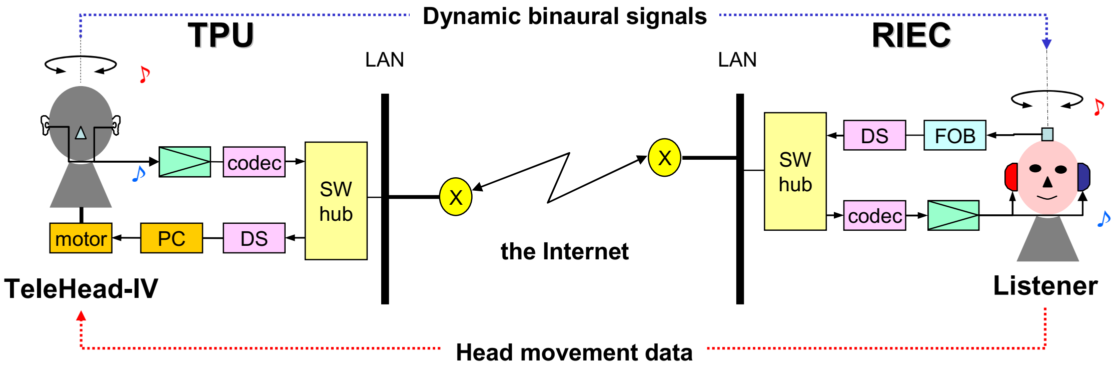

Figure 2. Tele-existence system connecting a TeleHead IV in Imizu with a listener in Sendai.

D. *Total transmission latency over the Internet*

When the video conference systems were used, the time required for the binaural signals received by two microphones in a remote TeleHead being reproduced at a local listener's ears was 634 ms, which was the sum of 370 ms for coding and decoding and 264 ms for the binaural-signal-data transmission over the Internet. The time required for the local listener's head movement being reproduced at the remote TeleHead was 143 ms, which was the sum of 23 ms for the data transmission over the Internet and 120 ms for the mechanical motion delay of the TeleHead-IV. It therefore took 777 ms for the listener to hear the dynamic binaural sounds provided by the remote TeleHead, in which his head movement was reflected.

When the low-latency sound codec systems were used, the latency for the binaural signals was only 80 ms, which included the time for coding, the data transmission over the Internet, and the time for decoding. Therefore the total latency of the system was 223 ms.

IV. PERFORMANCE EVALUATION

The performance of the tele-existence system using TeleHead IV was evaluated by conducting horizontal sound localization experiments.

A. *Procedure*

The TeleHead IV was placed in an experimental room at TPU and 12 loudspeakers were placed around it in the horizontal plane at intervals of 30°. The distance from the loudspeakers to the center of the dummy head was 1 m. The A-weighted noise level of the room was 22 dB and the reverberation time was about 50 ms.

White noise was used as the sound stimulus. It was generated on a PC with a sampling frequency at 48 kHz, D/A converted by USB audio interface units (Roland UA-101), amplified by an amplifier (Bose 1705 II), and fed to a loudspeaker. The binaural signals captured by the dummy head at TPU were transferred over the Internet to headphones worn by a listener sitting in a sound-proof room at the RIEC. The listener at the RIEC thus listened to the sound stimuli through the remote dummy head's ears, as if he sat at the center of the speaker array in the experiment room at TPU.

The duration of each stimulus was 3 s and the inter-stimulus interval was 5 s. The sound pressure level of the stimulus presented from the front speaker was set at 70 dB, and the stimulus reproduced by headphones at the RIEC was presented to the listener at 70 dB.

Sound localization experiments for the head-still and head-movement conditions were conducted separately. Each experiment consisted of four sessions, and every session consisted of 60 trials in which white noise was randomly presented five times from each of the 12 azimuthal angles. Twenty responses were thus obtained to sound from each of the 12 angles. Listeners were asked to judge from where the sound was presented and to indicate, on an answer sheet, which of the 12 angles it came from as well as whether or not the sound image was in-head.

In the first experiment, conducted on July 2011, two video conference systems were used to connect the TeleHead IV at TPU and the listener at the RIEC. Three male listeners participated in the experiment. The dummy head put on the TeleHead was not their own personalized dummy head but someone else's. In the second experiment, conducted on March 2012, two low-latency sound codec systems were used to connect the TeleHead IV at TPU and the listener at the RIEC. Three other male listeners participated in the experiment. The dummy head used was each listener's own.

In the third experiment, conducted on March 2012, the TeleHead IV and the listener were both at TPU and were connected directly by wires. The total latency of the system was 120 ms. The listeners in this experiment were the same three listeners who participated in the second experiment. The dummy head used was each listener's own.

B. *Results*

Figure 3 shows the pooled results for the three listeners in the head-still and head movement conditions in the first experiment. In each panel the area of the blue-filled circles is proportional to the in-head localization rate, and that of the red-filled circles is proportional to the out-of-head localization rate.

In the head-still condition most stimuli were localized out-of-head but some presented at a front position were localized in-head. There were also some front-back confusions and near-miss orientation-judgment errors. As a result, the mean correct out-of-head localization rate P_c was 45%. In the head-movement condition, there were fewer in-head localizations and front-back confusions than there were in the head-still condition. Near-miss localization errors, however, remained, resulting in a P_c of 51%. All of the listeners reported that it took so long for the head movement to be reflected in the sound-image movement that sound localization was not easy in the head-movement condition.

Figure 4 shows the pooled result for the three listeners in the head-movement conditions in the second experiment in which the head-still condition was skipped. Since it was the head-movement condition, there were fewer in-head localizations, front-back confusions, and near-miss orientation-judgment errors, resulting in a P_c of 82%. All of the listeners reported that they could localize each stimulus easily even though the delay between the head movement and sound-image movement was clearly perceptible.

Figure 5 shows the pooled results for the three listeners in the head-still and head-movement conditions in the third experiment. In the head-still condition, most stimuli were localized out-of-head but some presented at the front position were localized in-head. Some front-back confusions and near-miss orientation-judgment errors occurred. As a result, the mean correct out-of-head localization rate P_c in the head-still condition was 84%. In the head-movement condition, the P_c was 88%. Although in the head-movement condition there were no in-head localizations and there were fewer front-back confusions than there were in the head-still condition, near-miss localization errors remained.

C. *Discussion*

Even when the total latency of the system was 777 ms, and the listener used someone else's dummy head, front-back confusions were diminished by the use of dynamic binaural signals provided by the remote TeleHead. The correct out-of-head sound localization rate P_c in the head-movement condition, however, was only 51%. This relatively low value is probably due to the long system latency. The results of the first experiment suggest that the impact of the head movement in sound localization [5, 6] is reduced when the system latency is too long. This might be because then the listener's brain cannot associate the listener's head movement with movement of the sound image [3].

When the total latency of the system was 223 ms and the listener used his personalized dummy head, the P_c in the head-movement condition was 85%. This rate was almost the same as that obtained using a local TeleHead IV. The use of the dynamic binaural signals overcame the long system latency. The results of the second experiment demonstrate that a TeleHead used with a low-latency codec system works over the Internet as a personal auditory tele-existence system with satisfactory performance.

V. Conclusion

A personal auditory tele-existence system between TPU and the RIEC was implemented over the Internet by using a low-latency codec to connect the remote TeleHead IV and the local listener. The total delay time between the listener's head movement and the corresponding sound image movement was 223 ms: 23 ms for sending the head movement data through an RS-232C port, 80 ms for sending the binaural signals, and 120 ms for the mechanical motion delay of TeleHead IV. The mean correct-out-of-head sound localization rate for the horizontal sound sources in the head-movement condition was 85%, which was comparable with that obtained using a local TeleHead IV. The total system latency with the video conference system was about 780 ms, yielding poorer performance. When the system latency is too long, the listener's voluntary head movement is less effective in sound localization.

Acknowledgments

Part of this work was carried out under the Cooperative Research Project Program of the RIEC, Tohoku University. Authors would like to thank Professor Yôiti Suzuki for supporting the project. Part of this work was also supported by JSPS KAKENHI Grant Number 22300061.

References

[1] M. Otani, T. Hirahara, "Auditory Artifacts due to Switching Head-Related Transfer Functions of a Dynamic Virtual Auditory Display," IEICE Trnas. on Fundamentals of Electronics, Communications and Computer Sciences E91-A No.6, pp.1320-1328 (2008).

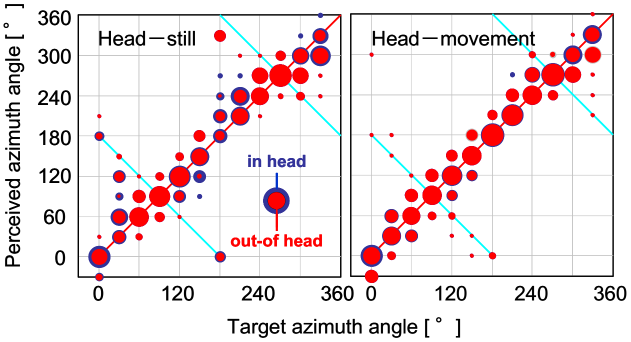

Figure 3. Results of the first experiment with a remote TeleHead and local listener connected by a video conference system.

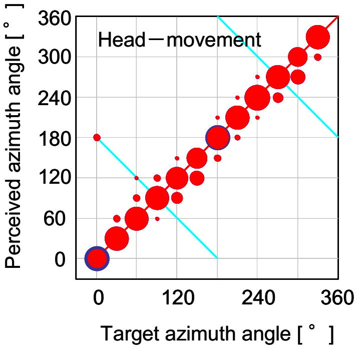

Figure 4. Results of the second experiment with a remote TeleHead and local listener connected by a low-latency audio codec system.

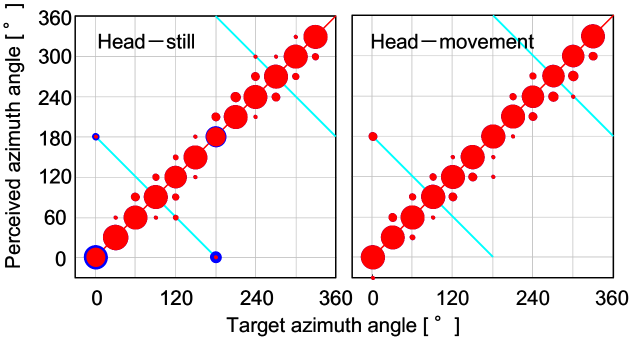

Figure 5. Results of the third experiment with a TeleHead at TPU and listener at TPU connected directly by wire.

[2] I. Toshima, S. Aoki, T. Hirahara, "Sound localization using an auditory telepresence robot: TeleHead II," Presence, Vol. 17, No. 4, pp. 392-404, (2008).

[3] T. Hirahara, D. Yoshisaki, D. Morikawa, "Impact of dynamic binaural signal associated with listener's voluntary movement in auditory spatial perception," Proceedings of Meetings on Acoustics, Vol. 19, 050130, pp. 1-8 (2013).

[4] T Hirahara, "Physical characteristics of headphones used in psychophysical experiments," Acoustical Science and Technology Vol. 25, No.4, pp. 276-285 (2004)

[5] D. Morikawa, T. Hirahara, "Effect of head rotation on horizontal and median sound localization of band-limited noise," Acoustical Science and Technology Vol.34, No.1, pp.56-58, (2013)

[6] Y. Iwaya, Y. Suzuki, D. Kimura, "Effects of head movement on front-back error in sound localization," Acoustical Science and Technology Vol. 24, No. 5, pp. 322–324 (2003).