ACOUSTICAL LETTER

Development of dynamic crosstalk cancellation system for multiple-listener binaural reproduction

Hiroaki Kurabayashi[1], Makoto Otani[2,*], Masami Hashimoto[3] and Mizue Kayama[3]

[1]*Graduate School of Science and Technology, Shinshu University,*
4–17–1 Wakasato, Nagano, 380–8553 Japan
[2]*Graduate School of Engineering, Kyoto University,*
Kyoto daigaku-Katsura, Nishikyo-ku, Kyoto, 615–8540 Japan
[3]*Faculty of Engineering, Shinshu University,*
4–17–1 Wakasato, Nagano, 380–8553 Japan

1. Introduction

Transaural[®][†] reproduction [1,2] is one approach to realize binaural presentation of auditory scenes to a listener using loudspeakers and crosstalk cancellation (CTC), instead of a set of headphones. Wireless and non-contact binaural reproduction achieved by transaural reproduction is the approach's main advantage over headphone presentation.

One shortcoming of conventional transaural reproduction is that CTC filters vary depending on the listener's head orientation and position. Therefore, if static CTC filters are used, then the listener is forced to remain motionless during listening. However, recent research has indicated that dynamic transaural reproduction using a head tracking device and adaptive CTC filter processing provides accurate binaural presentation even when a listener rotates his/her head [3–5]. Another concern is that most conventional transaural reproduction systems were developed to present binaural signals to the ears of a single listener, thereby disabling its use by multiple listeners. In theory, simultaneous binaural reproduction to multiple listeners by transaural reproduction is realizable using crosstalk cancellation assuming multiple listeners, but development of such multi-listener transaural reproduction is insufficient [6,7]. No attempt has been reported to apply multi-listener crosstalk cancellation to dynamic transaural reproduction.

This paper presents the development of a dynamic transaural reproduction system for multiple listeners, based on a single-listener dynamic transaural reproduction system that was developed previously [5]. This report describes results of a subjective evaluation of localization performance using the developed system.

2. Multi-listener dynamic transaural reproduction system

A dynamic transaural reproduction system for two listeners was developed as an extension of a single-listener

system developed previously by the authors [5]. Figure 1 presents a schematic of the multi-listener dynamic transaural reproduction system. Pure Data (Pd) on Apple OSX (Apple Computer Inc.) was used to operate real-time generation of CTC filters and convolution. Based on the head position and orientation of each listener detected by Microsoft Kinect operating on Windows OS (Microsoft Corp.), Pd synthesizes binaural signals that are to be reproduced at each listener's ears, generates CTC filters, and convolves source signals to head-related transfer functions (HRTFs) and CTC filters. OpenSound Control (OSC) protocol [8] on UDP (User Datagram Protocl) is used for data communication between Windows and OSX PCs. The CTC filters are obtained in real time by least-norm-solution in a frequency domain [9,10] to reproduce the synthesized binaural signals at four control points: four ears of two listeners. In CTC filter design, the filter length is 1,024 pt. The regularization parameter is 0.000,1 for all frequencies. Acoustic influences between two listeners are neglected. An HRTF database used for both binaural synthesis and CTC filter generation was obtained from the boundary element simulation [11] and a computer head and pinnae model that was captured by magnetic resonance imaging. The HRTFs were prepared for point sources located between 0.15 m and 2 m from the center of the head with 0.05-cm intervals and 5-deg intervals of azimuth and elevation. The system includes a loudspeaker array consisting of 24 loudspeaker units with vibration surfaces facing upward to simulate omni-directivity in the horizontal plane.

A measurement showed that signal processing delay is approximately 9 ms for both CTC filter generation and convolution of CTC filter and binaural signals, respectively. Assuming approximately 100 ms delay of Kinect and other delays, the total system latency would exceed 100 ms, which is larger than the acceptable system latency of a virtual auditory display [12]. Such large latency might degrade the presentation of binaural signals and perceived auditory space when listeners move quickly. This issue shall be addressed in future works.

---

*e-mail: otani@archi.kyoto-u.ac.jp
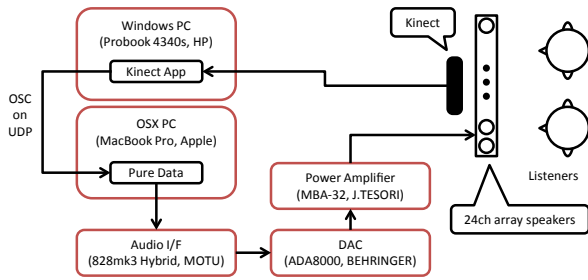[†]Transaural is a trademark of Cooper Bauck Corporation.

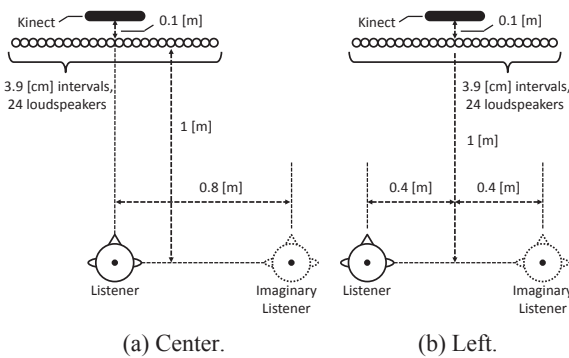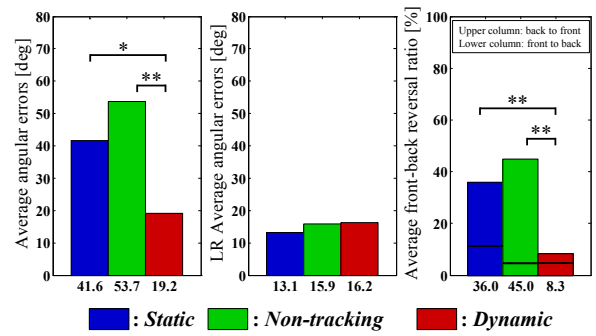**Fig. 1** Schematic of the multi-listener dynamic trans-aural reproduction system.

(a) Center.   (b) Left.

Fig. 2 Experimental conditions: (a) center and (b) left listening positions.

(a) Center.

(b) Left.

Fig. 3 Experimental results: Averaged angular errors, left-right (LR) angular errors, and front-back reversal.

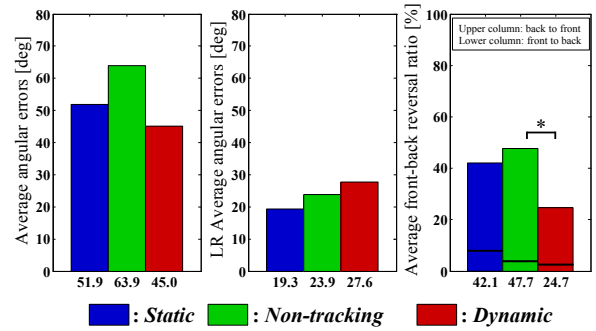## 3. Sound localization experiment

3.1. Method

Six subjects participated in a sound localization experiment performed in an anechoic chamber at the Faculty of Engineering, Shinshu University. Figure 2 presents configurations of a line loudspeaker array, Kinect, and listeners. The line loudspeaker array consisted of 24 loudspeaker units (ϕ 25 mm, NSW1-205-8A; Aurasound) installed to a wooden frame (1,143 mm × 87 mm × 38 mm) at 39-mm intervals. To eliminate visual cues, the loudspeaker array was veiled by a black cloth that was confirmed to be acoustically transparent before the experiment. Other equipment is presented in Fig. 1. One non-individualized set of HRTFs was used for all subjects. The stimulus was pink noise of 3-s duration. A-weighted sound pressure level was approximately 61 dB when a sound image was presented from the frontal direction and 1-m distance.

Sound images were presented at 12 horizontal directions with 30-deg. intervals. Trials were performed by single subjects, but the transaural system was operated assuming two listeners so that same sound images were presented to the second "imaginary" listener. Each stimulus for one sound image position was presented five times. Therefore one session consisted of 60 (= 5 × 12) trials in all, which were presented in a randomized order. The session was performed for three conditions: ***Static***, ***Non-tracking***, and ***Dynamic***. In the ***Static*** condition, subjects were instructed not to move their head during listening to each stimulus. In ***Non-tracking*** and ***Dynamic*** conditions, subjects were instructed to rotate their head from the front to left (−30 deg), right (30 deg), and then

to the front while listening to each stimulus. In the ***Dynamic*** condition, the system responded to the subject's head rotation, although it did not in the ***Non-tracking*** condition. In all the conditions, the imaginary listener was assumed to be still. The experiment was performed for two listening positions, as presented in Fig. 2: participants were located on the center or left side of the loudspeaker array to assess the effect of listening position relative to the loudspeaker array.

3.2. Results and discussion

Figures 3(a) and 3(b), respectively portray experimental results for the center and left listening positions. The left, center, and right panels in each figure respectively represent an angular error [deg], left–right (LR) angular error [deg], and front–back reversal ratio [%], which are averaged among all subjects. The angular error denotes an absolute error between the presented and answered angles. The LR angular error represents an absolute error by which the answered angle is reversed with respect to the subject's transversal plane when front–back reversal occurs. For the front–back reversal ratio, upper and lower columns respectively show front-to-back and back-to-front misjudgements. Asterisks (∗) denote the result of multiple comparison (Tukey method) among all conditions (∗: $p < 0.05$, ∗∗: $p < 0.01$).

For the center listening position depicted in Fig. 3(a), the angular error and front–back reversal ratio are significantly smaller in the ***Dynamic*** condition than in other conditions, although no significant difference was found in LR angular error among the conditions. Especially, the front–back reversal ratio is greatly reduced to less than 10% in the ***Dynamic*** condition, which indicates that the dynamic trans-

aural reproduction functions successfully even when applied to a multi-listener use.

However, for the left listening position depicted in Fig. 3(b), localization errors are greater than those in the center listening position, although the front–back reversal ratio is significantly smaller in the **Dynamic** condition than in **Non-tracking** condition, which indicates that, for the experimental setup of the current study, the performance of the multi-listener dynamic transaural reproduction depends on the listening position. Such degradation would be attributable to an increased error of Kinect tracking that might be prominent when a listener is located close to an edge of Kinect's field of view or to an increased error of binaural signals reproduced at both a listener's ears caused by inadequate positional relationship between loudspeakers and both a listener's ears leading to ill-conditioned CTC filter generation.

4. Summary

This paper presents a development of dynamic transaural reproduction system for multiple listener. A sound localization experiment was conducted to evaluate the performance of the developed system subjectively. The results reveal that the multi-listener dynamic transaural reproduction functions properly when a listener is located immediately in front of the loudspeaker array, whereas showing a degraded performance when a listener is located off-center. Such degradation would be eliminated by improvements in the tracking sensor or loudspeaker arrangement, which should be addressed in future works.

Acknowledgment

References

[1] M. R. Schroeder and B. S. Atal, "Computer simulation of sound transmission in rooms," *IEEE Conv. Rec.*, **7**, 150–155 (1963).

[2] D. H. Cooper and J. L. Bauck, "Prospects for transaural recording," *J. Audio Eng. Soc.*, **37**, 3–19 (1989).

[3] W. G. Gardner, *3-D Audio Using Loudspeaker*, Massachusetts Institute of Technology, Ph.D. Thesis (1997).

[4] T. Lentz, "Dynamic crosstalk cancellation for binaural synthesis in virtual reality environment," *J. Audio. Eng. Soc.*, **54**, 283–294 (2006).

[5] H. Kurabayashi, M. Otani, K. Itoh, M. Hashimoto and M. Kayama, "Sound image localization using dynamic transaural reproduction with non-contact head tracking," *IEICE Trans.*, **E97-A**, 1849–1858 (2014).

[6] Y. Kahana, P. A. Nelson, O. Kirkeby and H. Hamada, "Objective and subjective assessment of systems for the production of virtual acoustic images for multiple listeners," *Audio Eng. Soc. 103rd Conv.*, pre-print 4573 (1997).

[7] Y. Kim, O. Deille and P. A. Nelson, "Crosstalk cancellation in virtual acoustic imaging systems for multiple listeners," *J. Sound Vib.*, **297**, 251–266 (2006).

[8] OpenSound Control: http://opensoundcontrol.org/ (accessed 2015-09-25).

[9] A. Kaminuma, *Wide Area Sound Field Reproduction System Design Using Inverse Filter*, Nara Institute of Science and Technology, Ph.D. Thesis (2001).

[10] A. Tanaka, N. Otani and M. Miyakoshi, "Regularized inverse filter design for sound reproduction system," *IEICE Trans.*, **J87-A**, 1466–1467 (2004).

[11] M. Otani and S. Ise, "Fast claculation system specialized for head-related transfer function based on boundary element method," *J. Acoust. Soc. Am.*, **119**, 2589–2598 (2006).

[12] S. Yairi, Y. Iwaya and Y. Suzuki, "Estimation of detection threshold of system latency of virtual auditory display," *Appl. Acoust.*, **68**, 851–863 (2007).