Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Architectural Acoustics
Session 3aAAa: Virtual Concert Hall Acoustics I

3aAAa7. Subjective evaluation of a virtual acoustic system: Trials with three-dimensional sound field reproduced by the 'Sound Cask'

Maori Kobayashi*, Kanako Ueno, Mai Yamashita, Shiro Ise and Seigo Enomoto

***Corresponding author's address: Meiji University, Setagaya, 155-0031, Tokyo, Japan, te11001@meiji.ac.jp**

It has been necessary to establish subjective measures for the performance of the virtual acoustic systems. In this paper, we report our trials to evaluate the performance of a three-dimensional sound field reproduction system based on the boundary surface control principle, the 'Sound Cask'. First, we introduce our investigations for the experts of audio engineering in order to clarify the difference of auditory impression between the Sound Cask and conventional audio systems. Second, we report psychological and physiological experiments focusing on the advantageous points of the Sound Cask, localization performance and a clear sense of reality, that were pointed out in the investigations for the experts. Finally, we discuss the issues to be considered for subjective evaluation of virtual acoustic systems for future studies.

Published by the Acoustical Society of America through the American Institute of Physics

Received 29 Jan 2013; published 2 Jun 2013

INTRODUCTION

Many acoustic media technologies have been proposed for reproducing acoustic environments realistically. At present, media technology is expected to transmit not only voice and music but also “presence,” which refers to the subjective experience of being there in the acoustic environments. It is necessary to establish objective and quantitative measures of presence and/or realism, in order to evaluate listeners’ experience in virtual acoustic environments.

In this study, we report our attempt at evaluating the performance of a three-dimensional sound field reproduction system called the “Sound Cask,” based on the boundary surface control principle. First, we introduced our investigations to experts in audio engineering in order to clarify the differences of auditory impressions between the Sound Cask and conventional audio systems. Second, we report psychological and physiological experiments focusing on the advantages of the Sound Cask, localization performance, and a clear sense of presence, which were factors pointed out in the investigations by the experts.

THE SOUND CASK AND THE BOSC SYSTEM

In this study, we used the “Sound Cask” (Fig. 1 bottom) to present acoustic virtual environments that are close to the real environments. The Sound Cask is a new version of the boundary surface control (BoSC) sound reproduction system that enables us to record and reproduce a sound field area.

Ise proposed the boundary surface control principle (BoSC principle) [1]. By integrating the Kirchhoff-Helmholtz integral and inverse system, the BoSC system can accurately reproduce a three-dimensional (3-D) sound field surrounded by a closed boundary surface. The BoSC system comprises the BoSC microphone system (Fig. 1 top) and the reproduction room, namely the Sound Cask (Fig. 1 bottom). First, the BoSC microphone system records the sound pressure on the surface of a volume defined by the BoSC microphone array. Next, the recorded signals are convoluted with a set of inverse filters. Finally, the loudspeaker array in the Sound Cask accurately recreates the sound field in other locations by reproducing the convoluted signals. In the BoSC system, the inverse filters are determined by an inverse system of a transfer function matrix measured between each loudspeaker and microphone pair.

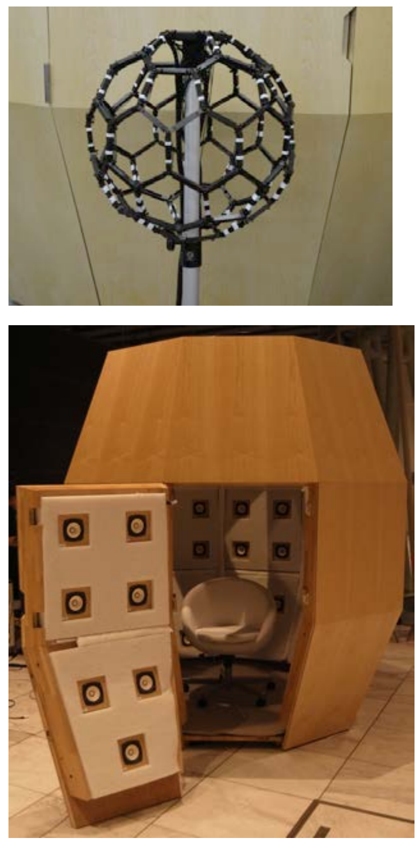

FIGURE 1. The BoSC system. Top panel is theBoSC microphone system. The bottom panel is the Sound Cask.

The BoSC microphone array has the same configuration as a *C80* fullerene. Eighty omnidirectional microphones (DPA 4060BM) are installed at the nodes of the fullerene. The diameter of the microphone array is about 46 cm. The reproduction area in the BoSC system is the inner region enclosed by the microphone array. Therefore, the system can reproduce a 3-D sound field surrounding the listener’s head. Consequently, listeners can perceive sound images created within sound fields without being adversely affected by their own head movements.

The Sound Cask has a nonagonal horizontal cross section, and the perimeter surfaces are set at three different angles in the vertical direction to avoid parallel surfaces except for the floor and ceiling planes. It has a 96-channel sound field reproduction system that uses a total of 96 full-range loudspeakers (FOSTEX FE103EN) installed in all wall surfaces except for the floor plane. The loudspeakers are installed in the wall surfaces, which are divided into six layers based on height, with 9 loudspeakers allotted to each of the top and the bottom layers, and 18 loudspeakers allotted to each of the remaining layers. For the details of the BoSC system and the Sound Cask, see the study by Ikeda and Ise [2] and previous studies [3].

EXPERIMENT 1: CHARACTERISTICS REPORTED BY ACOUSTIC EXPERTS

We conducted research among audio and acoustics experts in order to examine the characteristics of the sound fields reproduced by the Sound Cask. In this research, experts were presented some sound fields reproduced by the Sound Cask and asked to describe the characteristics of the Sound Cask compared to other systems in their own words.

Method

A total of 54 adults participated in the experiment. They were expert listeners, either studying or working in the field of acoustics.

We used seven different sound fields that were recorded by the BoSC recording system (see Table 1). Table 1 lists details of the sound fields that were used in this experiment. The test stimuli were 30 to 65 s long, presented at 60 to 76 dB (A).

Testing was conducted in the Sound Cask. Participants sat in the chair and listened to the auditory stimuli. After presentation of all stimuli, they were asked to answer questions about the types of audio systems they had previously used and about the characteristics of the Sound Cask compared to other audio systems. Participants answered by choosing options for the first question and by describing the characteristics in their own words for the second question. The presentation order was fixed: guitar playing, cicada buzzing, conversation in a forest, sound of clapping, symphony 1, symphony 2, and sounds during a Noh (classic Japanese musical drama). Participants were permitted to move their heads while listening to the stimuli. Intervals between the stimuli were 5 s, and the total duration of experiment was about 15 min.

TABLE 1. Details of the seven sound stimuli used in Experiment 1.

Name	Description	Recording position
Guitar	A man walking around the microphone while playing a guitar	Center of the player's pathway in a multipurpose hall
Cicada	Buzz of cicadas in a forest	In the forest
Conversation	People talking while walking in trees	In the trees
Clapping[*2]	5 men approaching the microphone while clapping	Center of the pathway of the men in a multipurpose hall
Symphony	String orchestra	In front of the conductor
Symphony	String orchestra	Audience seat in a hall
Noh[*1]	Noh (position close to performer)	Below the stage

[*1]classic Japanese musical drama, [*2] used in experiment 3

Results and discussion

Figure 2 shows the types of the auditory systems that participants have used prior to this experiment. These results suggested that headphones, 2-channel stereo speaker systems, or 5.1-channel surround sound systems were the target systems for comparison with the Sound Cask.

A semantic analysis was conducted on the spontaneous description of reproduced acoustic environments. A total of 72 phrasings were classified in semantic categories emerging from the free verbalizations. For the first classifications, we asked three independent blind raters to classify the 72 phrasings into 10 different category groups. Next, we summarized these 10 category groups into 6 groups. Then, we asked other two raters to classify the 72 phrasings into these 6 groups, and judge whether each phrasing was positive, negative or neither. Finally, we counted the phrasings that the two raters classified into the same category group. Mean kappa coefficients between the raters was 0.68, suggesting that the categorization was reliable.

Table 2 shows the results of this categorization of the verbal data. The results show that experts used terms such as “presence” (sense of being there), “localization,” “envelopment,” “reverb,” and “sound quality” to evaluate the sound fields. It is also suggested that the Sound Cask is able to present virtual acoustic fields with a high presence. However, the evaluation of localization was found to be varied across participants; most participants gave horizontal localization a positive evaluation, some participants gave vertical and distance localizations a negative evaluation.

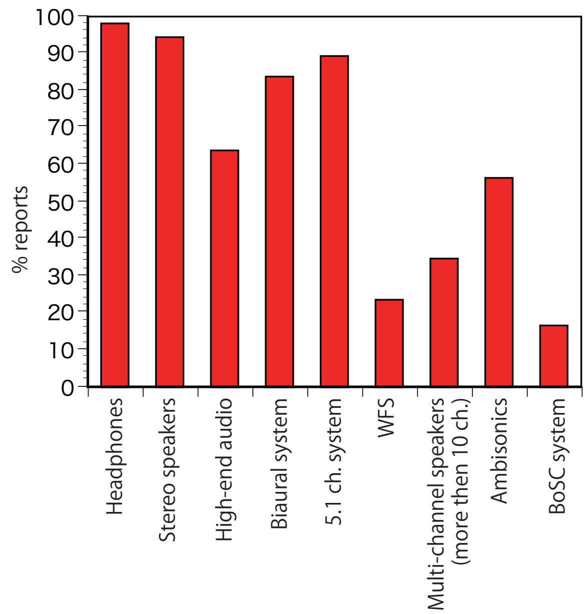

FIGURE 2. Audio systems that participants have used prior to the experiment. The vertical axis shows the percentage of participants who reported having used the systems indicated on the horizontal axis.

TABLE 2. The results of the categorization the verbal data in the experiment 1.

Category group	Description	Positive	Negative	Neither
Sense of presence or realism	Being there, Natural	14	3	0
Localization	Perceive sounds position	10	7	1
Spatial reproducibility except for localization	Envelopment by sounds, Surround, Reverb	11	6	0
Sound quality	Frequency range, Volume	1	2	0
Other	Size of the system, Impression	1	11	0

EXPERIMENT 2: LOCALIZATION TEST

From the results of Experiment 1 with experts, it was suggested that the Sound Cask is able to present acoustical virtual environments with high reality. However, these results showed both positive and negative evaluation of localization. In Experiment 1, since all auditory stimuli that were recorded by the BoSC microphone system in the real world had many sound sources, it might have been difficult for participants to judge the sound localization. Further, the localization test is a conventional measure for evaluating virtual 3-D acoustic environments. Therefore, we conducted Experiment 2 to examine auditory localization in the Sound Cask.

Method

A total of eight adults with normal hearing participated in this experiment. Auditory stimuli reproduced using the BoSC system were generated from the convolution of pink noise (1 s on-time and 0.4 s off-time, three burst) and the impulse response. The impulse responses were simulated from the free-space Green's function. The source location was assumed as 2 m away from the center of the microphone array in the simulation. The sound pressure level of stimulus was adjusted to 60 dB (A) at the center of the microphone array to eliminate the level differences between directions.

Testing was conducted in the Sound Cask. Participants sat in the chair and listened to the auditory stimuli. They were asked to illustrate their perceived direction after listening to the stimuli. The experiment was divided into two sessions: horizontal and vertical session. In the horizontal session, stimuli were presented from angles of 0–345° with 15° intervals. In the vertical session, the stimuli were presented from angles of 0–90° with 15° intervals. All participants underwent the horizontal session before the vertical session. The trial was repeated 10 times for each angle condition, the presentation order was randomized in each session. Participants were permitted to move their heads and bodies while listening to the stimuli. The participants took part in a practice session, which was followed by the experimental session. Intervals between the trials were 5 s, and the total duration of the experiment was about 120 min, including rest time.

Results and Discussion

The A panel of Fig. 3 shows the mean perceived angle and the standard error across all participants in the horizontal session. The mean localization error was almost 6°. The sound localization performance was extremely good for all angles of the horizontal plane. Front-back error was not found for any participants. The B panel of Fig.3 shows the mean perceived angle and the standard error across all participants in the vertical session. The means the localization error was almost 15°. These results show that localization performances were different between horizontal and vertical directions.

We consider that the variation of localization evaluation in experiment 1 was explained by these results: although the participants accurately assessed horizontal localization, their accuracy in assessing vertical localization was low. It is unclear why the performance was lower for vertical localization compared to the horizontal localization performance. This can be investigated in future research.

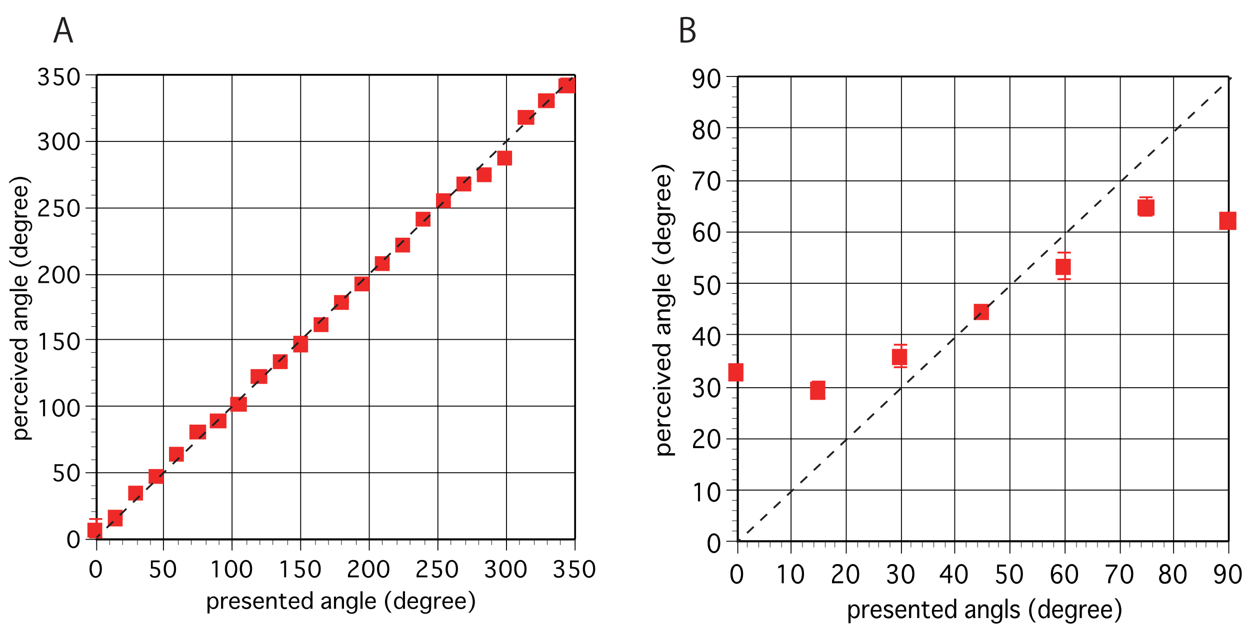

FIGURE 3. The results of Experiment 2. (A). The panel shows the results for horizontal localization. (B). The right panel shows the results for elevation. The horizontal and vertical axes show the presented and perceived directions, respectively.

EXPERIMENT 3: PHYCHOLOGICAL AND PHYSIOLOGICAL TEST FOR SENSE OF PRESENCE

In Experiment 1, experts evaluated the high presence of the sound fields reproduced by the Sound Cask. Presence is considered to be important for evaluating a virtual environment media system. Presence is generally defined as a user's subjective sensation of "being there" in a scene depicted by a medium [4]. Since presence is a subjective experience, it has most commonly been measured by self-reporting, either during a virtual environment experience or immediately afterwards, by using questionnaires. However, self-reporting or questionnaires have limitations owing to verbal constraints such as lack of knowledge of terms that evaluate presence. Previous studies have proposed methods for evaluating presence (questionnaire: [5-7], behavior test: [8], physiological test: [4, 9]) in visual virtual environments. However, few studies have examined using various methods in acoustic virtual environments without visual information. More reliable and objective measures are needed for presence in acoustic virtual environments.

We conducted Experiment 3 to evaluate presence by multiple methods, two physiological indexes and four psychological indexes, by using the auditory stimulus that men were approaching the microphone. This stimulus was the same as the one in Experiment 1, which was highly evaluated for presence in Experiment 1. It was seen that many participants retracted their necks when they listened to this stimulus. In order to measure the subjective inner response objectively, we used the blood volume pulse amplitude (BVPA) and the skin conductance level (SCL) as physiological indexes. Both indexes have been used to indicate parasympathetic nervous system activities, and they are often used to measure the arousal level or tension level of participants [10]. In addition, we used four simple questions to measure subjective experiences in the sound field of the Sound Cask.

Method

A total of 12 healthy adults (5 male, 7 female) between 20 and 27 years of age with normal hearing participated in the experiment. Auditory stimulus recorded by the BoSC recording system was the sound of 4 men preceded by a guitar player approaching and moving away from the microphone while clapping. The auditory stimuli duration was 42 s, and it was presented at 70 ±2 dB (A) at the center of listening position. We set two conditions: the valid and the invalid condition. In the valid condition, the input signals of each channel of the BoSC microphone were convoluted with valid inverse filters. In the invalid condition, these input signals were convoluted with randomized inverse filters. The valid condition and the invalid condition were created from the same original signals, but the spatial relation among control points on the boundary surface in the reproduction room were sufficiently different between these conditions such that participants could not perceive correct sound positions in the invalid condition.

Finger photoplethysmographs and the skin conductance changes (SCCs) were used to measure the blood volume pulse (BVP) in the finger and the skin conductance levels (SCLs), respectively. Analog data were amplified and digitized with a BIOPAC MP150 data acquisition and analysis system (BIOPAC Systems, USA). The photoplethysmograph transducer was attached to the left index finger, the Ag-AgCl electrodes were attached the left middle and ring fingers of the participant. The sampling rate was 1,250 Hz for both the finger photoplethysmographs and the electrodes. In the finger photoplethysmograph measurements, amplified analog data were processed with a high-pass (0.5 Hz) filter. A reduction in the BVP amplitude indicated the vasoconstriction of the arterioles at the finger mediated by a-adrenergic receptors [11]. In the SCCs measurement, amplified analog data were processed with a high-pass (0.05 Hz) filter..

Testing was conducted in the Sound Cask. Participants sat in the chair and listened to the auditory stimulus. While the stimulus was being presented, participants were asked not to move their left fingers in order to measure BVP and SCL accurately.

After each auditory stimulus ended, participants were asked to answer four questions: Question 1, Did you feel the sound stimulus was unnatural? Question 2, Did you feel as if you were there? Question 3, Did you feel as if someone passed by there? Question 4, What was the distance between you and the clapping man when he was closest to you? A 5-point rating scale was used: 1 was labeled "not at all"; 5 was labeled "very much"; and 3 was labeled "neither." For Question 4, 1 was labeled "within 50 cm"; 2 was labeled "50–100 cm"; 3 was labeled "100–150 cm"; 4 was labeled "150–200 cm"; 5 was labeled "more than 200 cm." Each condition was repeated 4 times, and the order of presentation was randomized across the participants. The interval between the stimuli was 40 s. The total experimental duration was approximately 12 min including relaxation time before the experiment.

Results

Psychological test

Figure 4 shows mean score and standard error of each questionnaire for all the participants. A repeated-measure ANOVA (2 stimulus conditions × 4 questions) was performed on the mean score of all participants. A main effect of the question and the interaction between the stimuli condition and the question was found (main effect: $F(3,33) = 11.71$, $p < 0.01$; interaction: $F(3,33) = 12.49$, $p < 0.01$). The *post-hoc* test revealed the significant differences between the valid and the invalid condition in Question 2 ($p < 0.01$), Question 3 ($p < 0.05$), and Question 4 ($p < 0.01$). These results show that participants perceived higher presence of spaces and objects for the valid condition than the invalid condition. This is consistent with previous studies stating that the accuracy of the spatialized sounds is important for presence in acoustic virtual environments [12].

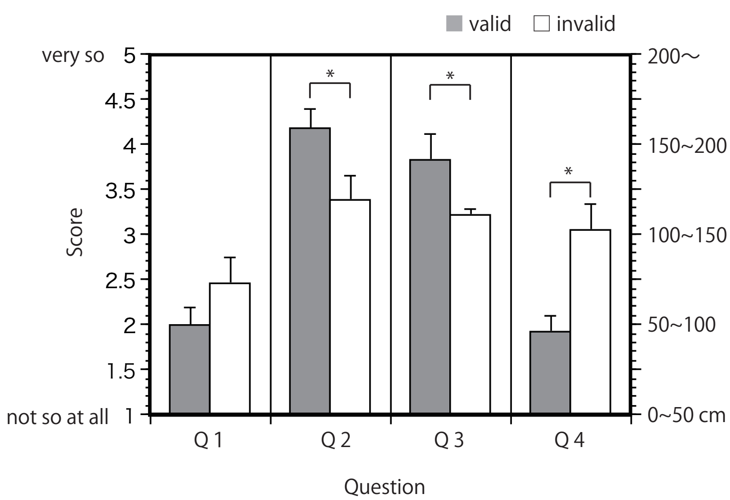

FIGURE 4. The results of Experiment 3. Mean score and standard error across participants for each question. The horizontal axis shows each question. The left vertical axis shows the score for Questions 1–3. The right vertical axis shows the score for Question 4.

Physiological test

BVP: To measure the BVP amplitude, the time series of the positive and negative peak values for each photoplethysmograph measurement were resampled at 100 Hz. The time-series data of the BVP amplitude were expressed as a percentage of the mean value during baseline recording. The baseline was defined as the mean amplitude for 5 s before the presentation of the sound stimulus in each trial by refering [13]. We determined the task-related BVP by calculating the mean BVP amplitude during the stimulus presentation for each condition. Some participants showed a difference between the conditions for the time-series data of BVP amplitude. Figure 5 show that a typical case of a time-series data of BVP amplitude during the sound presentation. The top panel of Figure 5 shows the temporal waveform of the auditory stimulus, the bottom shows the time-series data of BVP amplitude of a participant. Although the levels of the auditory stimuli were the same between conditions, the time-series of BVP amplitude was different between conditions. Generally, reductions of BVP amplitudes were induced by sounds presented at higher sound pressure levels and it showed rapid habituation [14]. The present case may suggest that the BPN amplitude was affected by the accuracy of the spatial reproduction.

SCL: In first 40 s of stimulus duration, 50,000 data points had been recorded. To reduce this to a reasonable number of points for analysis, the time series of the SCCs were resampled at 100 Hz. The mean SCL from the stimulus presenting duration (42 ms) was the activated arousal level. The base line SCL consisted of the average skin conductance during the 5 seconds prior to each stimuli onset. The difference between these two estimated arousal level (activated − baseline) was taken as the task-related SCL. The task-related SCLs were calculated for each condition.

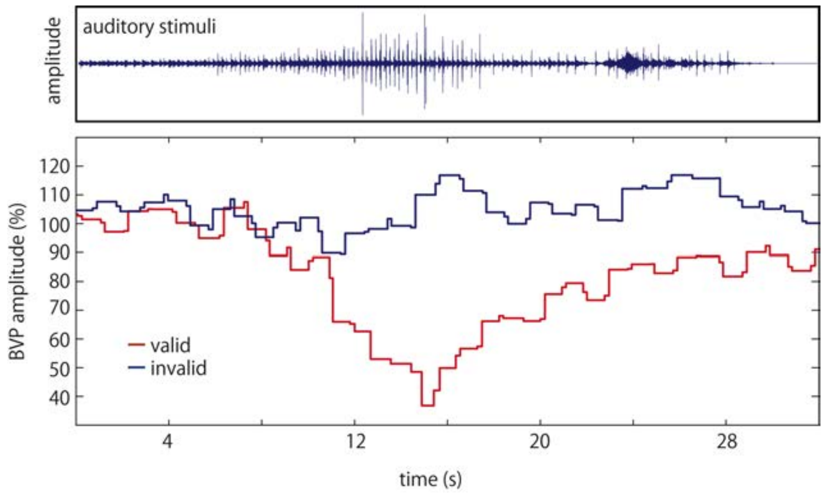

FIGURE 5. Results of Experiment 3. The top panel shows the temporal waveform of the auditory stimuli; the bottom panel shows the time-series data of BPN amplitude.

TABLE 3. The results of experiment 3. Mean BPN and SCL across participants.

	BPN (%)	SCL (μS)
Valid	86.9	1.51
Invalid	97.5	3.08
P value	0.02	0.03

Table 3 shows mean amplitudes of both BPN and SCL across participants. The BPN decreased in the valid condition more than in the invalid condition (BPNs: t (11) = 6.02, $p < 0.05$). The SCLs increased in the valid condition more than in the invalid condition (SCLs: t (11) = 6.02, $p < 0.05$). These results suggest that the BPN and SCL was influenced by the spatial reproduced condition in the auditory environments.

GENERAL DISCUSSION

In this study, we report three trials to evaluate listeners' experience in an acoustic virtual environment in a quantitative and objective way by using the Sound Cask.

In Experiment 1, we investigated characteristics of the Sound Cask using verbal reports from experts in audio engineering in order to clarify the differences of auditory impression between the Sound Cask and conventional audio systems. As a result, the experts pointed out localization performance and a clear sense of presence as advantageous features of the Sound Cask in the investigations. However, evaluation of the localization performance varied across participants. Then, we conducted the localization test to judge localization performance in the Sound Cask. The results show that the performance varied depending on the dimension: the performance of the horizontal dimension was almost perfect, while the elevation was not so good. It is considered that these results were consistent with the evaluation by the experts in Experiment 1. In Experiment 3, we examined the methods to quantify the sense of presence by psychological and physiological measures. As a result, both indexes showed different responses between conditions of spatial reproduction accuracy, suggesting the possibility to be the objective measures to evaluate presence or subjective experience in virtual acoustic environments.

In experiment 3, we used an approaching sound as an auditory stimulus. The reason of the usage of this stimulus was that listeners reported the stimulus provided a higher sense of presence in pre-experiments. The basic verbal reports also suggested that listeners perceived a higher presence for human voices in the Sound Cask. We also need to examine the methods to quantify and externalize listeners' experience for these stimuli as a future task. We consider that this examination contributes to the evaluation of the quality of virtual acoustic system.

ACKNOWLEDGMENTS

This work was supported by CREST from Japan Science Technology Agency. We thanks to Dr. Yusuke Ikeda for his technical assistance.

REFERENCES

1. Ise, S. “A principle of sound field control based on the kirchhoff-helmholtz integral equation and the theory of inverse systems”. *Acoustica*, **85**, 78–87 (1999).
2. Ikeda, Y., Ise, S. “Sound Cask: Music and voice communications system with three-dimensional sound reproduction based on boundary surface control principle” Proc. ICA 2013, in press
3. Enomoto, S., Ikeda, Y., Ise, S., Nakamura, S. “Three-dimensional sound field reproduction and recording systems based on boundary surface control principle”, Proc. 14^{th} ICAD, 24-27 (2008)
4. Freeman, J., Avons, S.E., Pearson, D.E. and IJsselsteijn,W.A. “Effects of sensory information and prior experience on direct subjective ratings of presence” Presence: Teleoperators and Virtual Environments, **8**, 1-13 (1999).
5. Lessiter, J., Freeman, J., Keogh, E., Davidoff, J. “A cross-media presence questionnaire: The ITC Sense of Presence Inventory” Presence: Teleoperators and Virtual Environments, **10**, 282-298 (2001).
6. Slater, M., Usoh, M., and Steed, A. “Depth of presence in virtual environments” Presence: Teleoperators and Virtual Environments, **3**, 130-144 (1994).
7. Witmer, B. G., and Singer, M. J. “Measuring presence in virtual environments: A presence questionnaire” Presence: Teleoperators and Virtual Environments, **7**, 225-240 (1998).
8. Emoto, M., Masaoka, K., Sugawara, M., Nojiri, Y. “The viewing angle dependency in the presence of wide field still image viewing and its relationship to the evaluation indices.” J. Image info. Television Eng. (in Japanese) **60**, 1288-1295 (2006).
9. Baumgartner, T., Volko, L., Esslen, M., Jancke, L. “Neural correlate of spatial presence in an arousing and noninteractive virtual reality: An EEG and Psychophysiology study” Cyberpsychol. Behav., **9**, 30-45 (2006)
10. Barry, R.J., Clark A.R., McCarthy, R., Seikowitz, M., Rushby, J. A. “Arousal and activation in continuous performance task: An exploration of state effects in normal children” J. Psychohysiol., **19**, 91-99, (2005)
11. Grote, L., Zou, D., Kraiczi, H., Hedner, J. “Finger plethysmography—a method for monitoring finger blood flow during sleep disordered breathing”. Respir Physiol Neurobiol, **136**, 141–152 (2003).
12. Lombard, M., Ditton, T. “At the hear of it all: the concept of presence” J Computer-Mediated Com 3.
13. Ohishi, Y., Kashino, M. “Habituation of rapid sympathetic response to aversive timbre eliminated by change in basal sympathovagal balance” Psychophysiology, **49**, 1059-1071, (2012).
14. Barry, R. J., & James, A. L. “Fractionation of phasic responses in a dishabituation paradigm”. Physiology & Behavior, **26**, 69–75 (1981).