Acoust. Sci. & Tech. **36**, 1 (2015)

PAPER

Sound field reproduction and sharing system based on the boundary surface control principle

Akira Omoto[1,5,*], Shiro Ise[2,5], Yusuke Ikeda[2,5], Kanako Ueno[3,5], Seigo Enomoto[4,5] and Maori Kobayashi[3,5]

[1]*Kyushu University, 4–9–1 Shiobaru, Minami-ku, Fukuoka, 815–8540 Japan*
[2]*Tokyo Denki University, 2–1200 Muzai Gakuendai, Inzai, 270–1382 Japan*
[3]*Meiji University, 1–1–1 Higashi-Mita, Tama-ku, Kawasaki, 214–8571 Japan*
[4]*National Institute of Information and Communication Technology, 3–5 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619–0289 Japan*
[5]*Japan Science and Technology Agency, CREST, 5–3, Yonbancho, Chiyoda-ku, Tokyo, 102–8666 Japan*

(*Received 23 March 2014, Accepted for publication 29 July 2014*)

Abstract: In this paper, we introduce a newly developed sound-field-reproducing and -sharing system. The system consists of an 80-channel fullerene-shaped microphone array and a 96-channel loudspeaker array mounted in an enclosure called a sound cask, so named because of its shape. The cask has two functions. First, it functions as a precise sound field reproduction system. The sound signals acquired from a microphone array in any sound field can be reproduced in the sound cask after passing through filters that modify the amplitude and phase on the basis of the boundary surface control principle. The large number of loudspeakers result in the precise orientation and depth of sound images. Second, it functions as a platform for a sound-field-sharing system. Several casks located remotely can appear to exist in the same sound field for subjects inside a cask. In addition, the cask is large enough for one to be able to play a musical instrument inside it. The musical sound or voices produced by subjects can be shared by subjects in a distant cask after convoluting the impulse responses of the original sound field. The concept of the system is explained in detail.

Keywords: Sound field reproduction, Boundary surface control, Multichannel system

PACS number: 43.38.MD, 43.60.Pt [doi:10.1250/ast.36.1]

1. INTRODUCTION

Currently, several types of sound-field-reproducing systems are known to be in use. A 5.1 surround [1] is an example of a system in general use, and a 22.2-channel system [2] is a well-known example of an ideal reproduction environment. The more academic or exact methods include various binaural reproduction methods, wave field synthesis (WFS) [3–7], the more recently developed 6-channel system [8,9], and higher-order ambisonics [10–12]. The boundary surface control (BoSC) technique proposed in 1993 by Ise is academically one of the important reproduction methods [13–15]. Since the actual construction of a reproduction system using 62-channel loudspeakers at an early stage, development has not only been in terms of sound field reproduction, but the system has also been investigated in terms of sound field sharing at a remote location [16–21]. The experimental results indicate the validity of the system.

In this paper, the concept and concrete structure of a 96-channel immersive sound field reproduction system newly developed on the basis of the BoSC principle are introduced. Because the system resembles a cask in shape, it is named the "sound cask." A cross section in the horizontal plane of the cask is nonagonal and does not have parallel wall surfaces except for at the floor and ceiling. The signals recorded using a C-80 fullerene-shaped microphone array are reproduced after passing through an inverse filter matrix between the loudspeakers and microphones. Additionally, the concept and assumed procedures for the sound-field-sharing system at distant places with a sound cask are also described.

2. PRINCIPLE AND CONCEPT OF THE SYSTEM

In 1993, Ise proposed the BoSC principle, which is a

*e-mail: omoto@design.kyushu-u.ac.jp

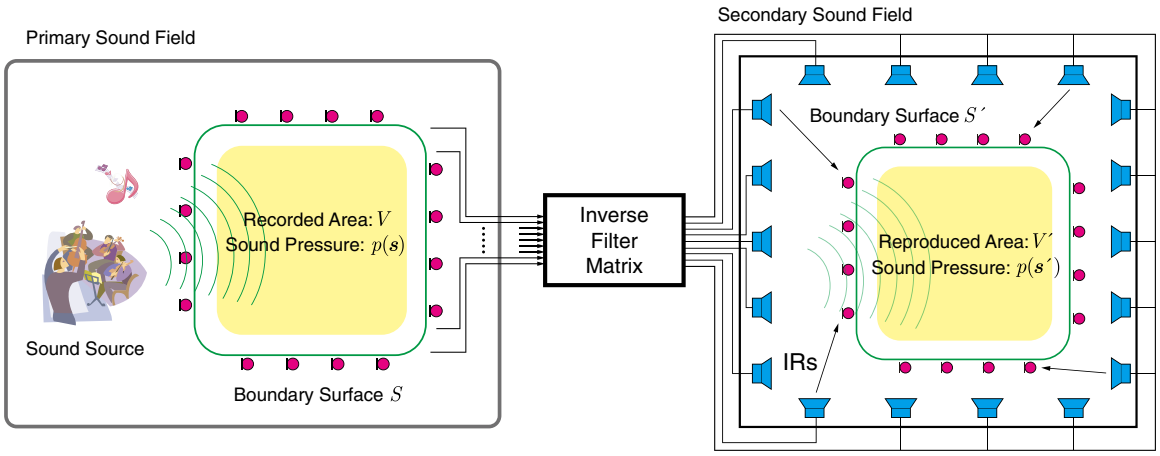

Fig. 1 Concept of the boundary surface control principle with an inverse filter matrix. The sound pressures at the surface S are reproduced at the surface S' in the secondary sound field. The inverse filter matrix, which is calculated from impulse responses from all possible combinations of loudspeakers and microphones, is introduced to reproduce the sound pressures at S'.

3-dimensional sound reproduction method based on the Kirchhoff-Helmholtz integral equation and inverse system [13–15]. Figure 1 shows the basic concept of the BoSC principle.

The recorded area V in the primary sound field is the target for reproduction. This reproduced field is expressed as V' in the secondary sound field. Assuming V is congruent with V', the following equation holds:

$$|\boldsymbol{r}' - \boldsymbol{s}'| = |\boldsymbol{r} - \boldsymbol{s}| \quad (\boldsymbol{s} \in V, \boldsymbol{r} \in S, \boldsymbol{s}' \in V', \boldsymbol{r} \in S'), \quad (1)$$

where $\boldsymbol{s}$ and $\boldsymbol{r}$ are arbitrary position vectors in the volume V and on the surface S surrounding V, respectively. $\boldsymbol{s}'$ and $\boldsymbol{r}'$ are position vectors in the volume V' and on the surface S' surrounding V', respectively. For the sound pressures p in V and V', the following equations can be established on the basis of the Kirchhoff-Helmholtz integral equation:

$$p(\boldsymbol{s}) = \iint_S \left[G(\boldsymbol{r}|\boldsymbol{s}) \frac{\partial p(\boldsymbol{r})}{\partial \boldsymbol{n}} - p(\boldsymbol{r}) \frac{\partial G(\boldsymbol{r}|\boldsymbol{s})}{\partial \boldsymbol{n}} \right] dS \quad (2)$$

$$: \boldsymbol{s} \in V$$

$$p(\boldsymbol{s}') = \iint_{S'} \left[G(\boldsymbol{r}'|\boldsymbol{s}') \frac{\partial p(\boldsymbol{r}')}{\partial \boldsymbol{n}'} - p(\boldsymbol{r}') \frac{\partial G(\boldsymbol{r}'|\boldsymbol{s}')}{\partial \boldsymbol{n}'} \right] dS' \quad (3)$$

$$: \boldsymbol{s}' \in V',$$

where $\boldsymbol{n}$ and $\boldsymbol{n}'$ are normal vectors on S and S', respectively, and $G(\boldsymbol{r}|\boldsymbol{s})$ is Green's function between $\boldsymbol{s}$ and $\boldsymbol{r}$. By applying Eq. (1), we obtain the following relationship for G and its gradient in Eqs. (2) and (3):

$$G(\boldsymbol{r}|\boldsymbol{s}) = G(\boldsymbol{r}'|\boldsymbol{s}'), \quad \frac{\partial G(\boldsymbol{r}|\boldsymbol{s})}{\partial \boldsymbol{n}} = \frac{\partial G(\boldsymbol{r}'|\boldsymbol{s}')}{\partial \boldsymbol{n}'}. \quad (4)$$

If the sound pressure and its gradient on each boundary are equal, then the sound pressures in all areas are also identical, as determined from Eqs. (2) and (3). This is expressed as follows:

$$\forall \boldsymbol{r} \in S, \ \forall \boldsymbol{r}' \in S',$$

$$p(\boldsymbol{r}) = p(\boldsymbol{r}'), \ \frac{\partial p(\boldsymbol{r})}{\partial \boldsymbol{n}} = \frac{\partial p(\boldsymbol{r}')}{\partial \boldsymbol{n}'}$$

$$\Longrightarrow \forall \boldsymbol{s} \in V, \ \forall \boldsymbol{s}' \in V', \ p(\boldsymbol{s}) = p(\boldsymbol{s}'). \quad (5)$$

Considering this as a boundary value problem, the uniqueness of the solution follows from the fact that either the sound pressure or its gradient is sufficient to determine the values of both [22].

To construct the recorded and reproduction areas, a microphone array is generally used. The C80 fullerene-shaped microphone array is adopted in our project. Another feature of the BoSC principle is the introduction of the inverse filter matrix. Impulse responses between all possible combinations of loudspeakers and microphones in the secondary sound field (IRs in the figure) are measured in advance, and the inverse filter matrix is calculated [15]. These filters are applied to the signals for reproducing the sound pressure at surface S at the target surface of S'.

Another well-known sound reproduction method using the Kirchhoff-Helmholtz integral equation is WFS [3–7]. However, a characteristic of the boundary surface control principle is that the configuration of the closed surface is not restricted because of the introduction of the inverse system.

In addition to WFS, several stereophonic systems, that can reproduce all surrounding primary sound fields exist, e.g., the 6-channel system [8,9] and ambisonics system [10–12]. However, the sound cask has practical advantages compared with other systems from the following viewpoints:

– The sound image along the depth direction can be controlled even in the vicinity of the head of the listener.
– The entire system can easily be moved into any location.
– A theoretically assured combination with a recording system, a 80-channel fullerene-shaped microphone array in our case, can be constructed.

3. SOUND CASK

The main characteristic of the BoSC system is its ability to reproduce a sound field, not at small points but in a finite region. A listener can freely move his head, and the system can provide a high performance of spatial information reproduction such as conveying sound localization and sound distance [17]. On the basis of these system features, as an example of a more effective application of the BoSC system, we propose the design of a sound cask. In the design of the sound field reproduction system based on the BoSC principle, space design, which is suitable for inverse filter calculation, is important since the quality of these filters directly affects the total performance of the system.

A previously developed system [17] consisted of 62 loudspeakers mounted on a dome-shaped wooden frame arranged inside a music practice chamber whose floor space was approximately $2 \times 2\,\mathrm{m}^2$. Several experiments indicated that the following items inevitably and adversely affect the performance of the inverse filter:

– reflection from the wooden frame,
– strong normal modes of the outer rectangular chamber,
– uneven distribution of loudspeakers; a dense distribution only at positions higher than the listener's head.

The measures employed to resolve these items are as follows:

– irregularly shaped enclosure to suppress dominant acoustic modes,
– no reflective material inside the enclosure, e.g., loudspeakers are mounted directly on the walls with surrounding absorbing material,
– evenly distributed loudspeakers covering the whole body of the listener.

Additional design guidelines are summarized as follows:

– To increase opportunities for many people to experience the sound field reproduction system, easy disassembly, transportation, and assembly should be ensured.
– Smaller-scale hardware is also preferred for easy transportation.
– A completely enclosed space should be provided to realize immersive environment.

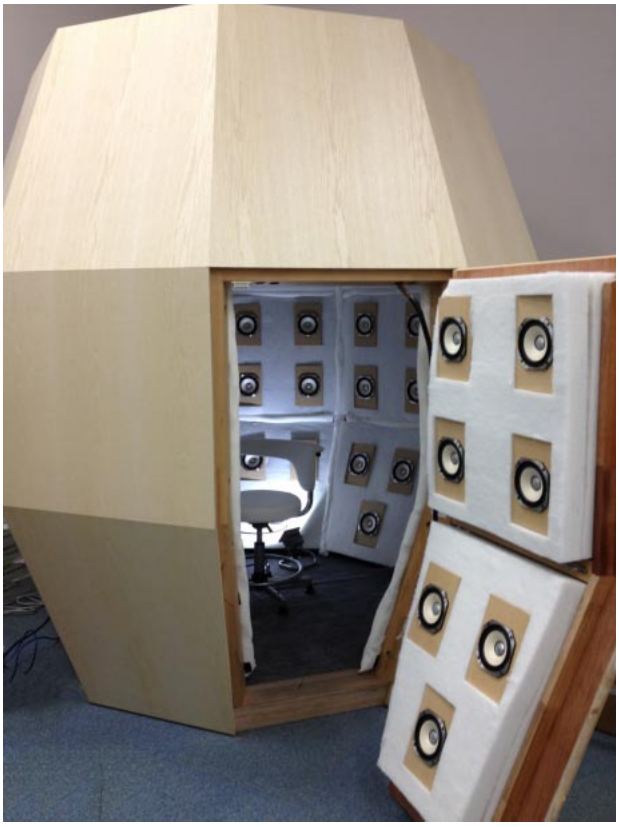

Fig. 2 Overview of the sound cask with 96 loudspeakers. The inside of the cask has enough area to play a musical instrument.

– The inner space must be large enough to play musical instruments.
– The basic performance of sound-acquiring and -reproducing devices, such as microphones and loudspeakers should be as high as possible to achieve so-called Hi-Fi reproduction.
– The space density of reproducing loudspeakers should also be as high as possible to achieve higher-resolution sound localization. At the same time, larger dimensions of the loudspeaker unit would be preferred for a better response in the low-frequency range.
– The number of channels should be practically controllable from commonly available computers and digital audio workstations (DAWs).

As a practical and reasonable compromise to achieve the conditions above, 96 loudspeakers are allocated inside the sound cask. Figure 2 shows a picture of the practically designed sound cask. In particular, a higher-grade loudspeaker unit (Fostex FX120) was adopted in the current version of the sound cask after several listening tests.

The horizontal cross section of the sound cask is the shape of a regular nonagonal cask. Hence, except for the floor and ceiling planes, the sound cask has no parallel sides. This shape has the effect of suppressing any dominant acoustic mode inside the sound cask.

With internal dimensions of 1,950 mm diameter in the central horizontal plane and 2,150 mm height, the sound cask has a large enough internal space to play a wind or string instrument. Ninety-six full-range loudspeakers are installed on the walls and ceiling but not the floor plane. Six of the loudspeakers are installed on the ceiling plane. The speakers are installed on the wall surface at six heights. Nine loudspeakers are allotted to the top and bottom heights, and 18 loudspeakers are allotted to each of the remaining heights. The average interval between the adjacent heights is approximately 350 mm. The average interval between adjacent loudspeakers in the horizontal direction is approximately 540 mm for the top and bottom heights and approximately 330 mm for all other heights. In the previous BoSC system, 62 loudspeakers were installed around the upper body of a listener. However, in the sound cask, loudspeakers were installed to cover the whole body of a listener. Therefore, this is expected to improve the sound reproduction performance in the vertical direction.

In addition, the wall parts of the sound cask are modularized and can be dismantled when transporting the system. The system is divisible into nine parts horizontally with each part forming a side of a regular nonagon. The walls of the sound cask are divisible into three; top, middle, and bottom in the vertical direction. This sound cask can be disassembled into separate walls within approximately 30 to 40 minute, and can be assembled within approximately 2 h by several workers.

For a sound-absorbing material, polywool, a recycled material from plastic bottles (thickness, 120 mm; density, 32 kg/m^3,) is used to achieve adequate absorption. The sound insulation performance of the wall of the cask is Dr-20. We also considered the ease of inverse system design by shortening the reverberation time.

The loudspeakers in the sound cask are driven by newly designed digital amplifiers installed at the bottom of the sound cask (Fig. 3). As shown in Fig. 4(a), 128 channels of audio signals are transmitted from the PC to the multi-channel audio digital interface (MADI) converter through only two MADI optical lines. The 144 channels of data, which consist of audio signals and 16 channels of control signals, are distributed to 12 serially connected network amplifiers Fig. 4(b) using a LAN cable. The LAN cable can transmit four low voltage differential signaling (LVDS) signals. As shown in Fig. 4(c), one of them is used for a bit-clock signal of 49.152 MHz, and the remaining three are used for audio and control signals. In each signal line, one sample has 40-bit header information, followed by 48 channels each with 20 bits of data, which were the 4B5B-transformed version of the transmitted 16-bit audio signals and 20-bit footer. Three LVDS signal lines can therefore transmit 144 (48×3) channels of data. Because only 96-channel signals are necessary for the sound cask, in practice two LVDS lines are used in the designed system and the remaining line is used for control signals such as volume information. At each network amplifier, the assigned eight-channel audio signals are picked up from the 96-channel signals, and a D-class amplifier is used for amplification. In the amplifier, word-clock and bit-clock signals are generated by a phase locked loop (PLL) synchronized to the header information.

Fig. 3 Amplifiers and power supplies at the bottom of the cask.

A two-channel 16-bit A/D converter is also included in the amplifier, as shown in Fig. 4(b). The A/D converted signals are 4B5B-transformed to 20 bits and can be transmitted through the lines by replacing the picked up signal information. One amplifier requires a five-bits buffer; therefore, a 100 ns delay occurs. Because the LAN cable is connected in a ring form, all signals are sent again to the MADI converter and the 128-channel data return to the PC. All the circuits and power supplies are hidden below the floor of the cask, as shown in Fig. 3.

4. SOUND RECORDING SYSTEM

4.1. Microphone Array

An 80-channel microphone array is adopted as the standard data acquisition device for the sound cask. This microphone array is constructed in the shape of a C80 fullerene structure and miniature microphones are located at each node.

At each nodal point of the C80 fullerene structure constructed by an aluminum frame, a miniature condenser microphone (DPA 4060) is embedded. Details of the microphone arrangement are shown in Fig. 5. The average distance between adjacent microphones is approximately 8 cm. The theoretical upper limit of precise reproduction (distance equals a half-wavelength) is therefore approximately 2 kHz in this system.

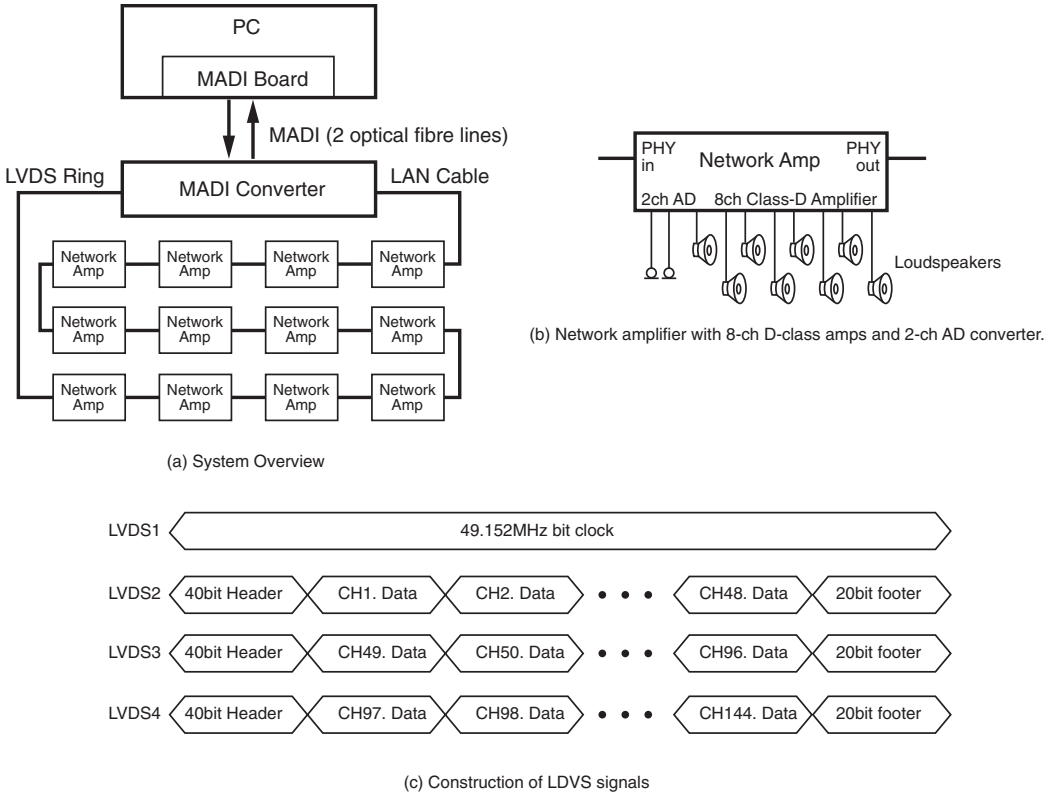

Fig. 4 Structure of the network amplifier and signals, (a) system overview, (b) details of the amplifier, and (c) structure of the LVDS signals.

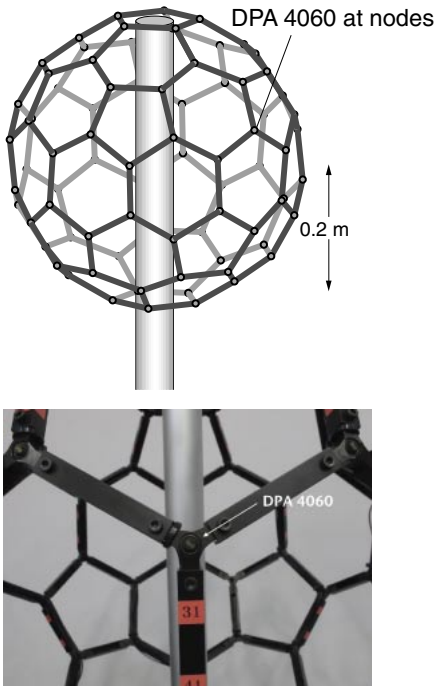

Fig. 5 Appearance of a C80 fullerene microphone array (top) and DPA 4060 miniature microphone mounted at the nodes (bottom). The microphone #31 is usually located in front.

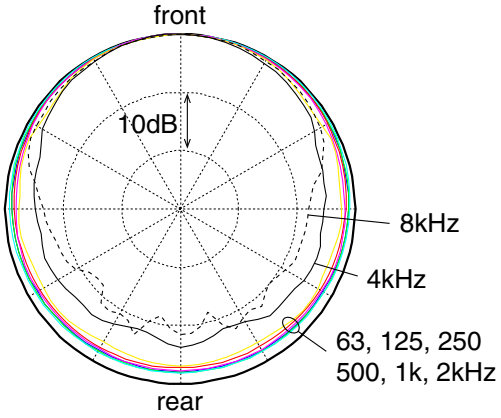

Fig. 6 Directivity of microphone #31 in the array. The results from 63 Hz to 8 kHz for every octave band are shown.

In an anechoic chamber, the sound source was located at a distance of 2 m and the impulse response to each microphone was measured using a turn table rotating in 5° steps. The directional characteristic of each microphone was then calculated in the frequency range from 63 Hz to 8 kHz. The results for microphone #31 is shown in Fig. 6. (This microphone is usually located in front in our recordings.)

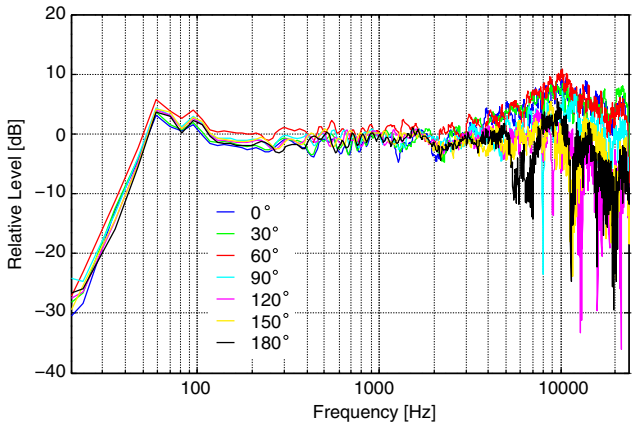

Fig. 7 Frequency response of microphone #31. The results for rotations of multiples of 30° are also plotted.

Since the center of rotation of the measurement coincided with the center of the array and not with the location of microphone #31, fluctuations of up to 1.5 dB could occur. Taking into account such fluctuations, the directivity of the microphone can be regarded as omni directional for low and mid-frequencies of up to 2 kHz. For higher frequencies, 4 and 8 kHz, a reduction of approximately 10 dB can be observed for the sound incident from the rear direction.

Figure 7 shows the frequency characteristics of the response. The results for the rotated cases for every 30° step are also plotted. An almost flat response from 100 Hz to 5 kHz and a slightly boosted response at approximately 10 kHz were observed. To ensure the uniform sensitivity of all microphones, a specific signal from a sound calibrator (B&K Type 4231 with a specially made jig for fitting the microphone) is recorded for every recording project.

4.2. Signal-Acquiring System

For convenience in outdoor recording, a battery-powered system was constructed. The system had 10 field recorders, which could simultaneously record eight channels. The Tascam HS-P82 was adopted. This recorder can supply eight channels of phantom power to the microphones. Signals were recorded in a Compact Flash card and because the file format was FAT32, the maximum size of one continuous recording file was limited to 2 GB. The recorded data could be transferred to a PC using a USB. The connection diagram for the SMPTE time code and word clock synchronization is shown in Fig. 8. The time code was distributed from the parent machine to all child machines directly and the word clock was distributed using a cascade connection. A sampling frequency of 48 kHz with 24-bit resolution was used in our recordings. Approximately 3 h of recording could be carried out using the batteries.

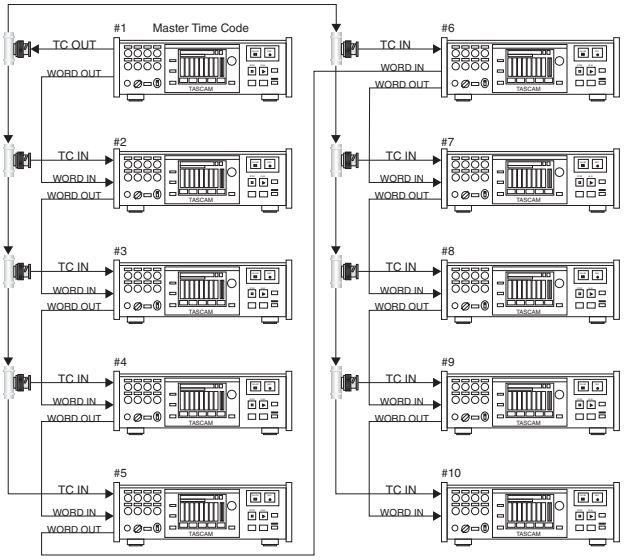

Fig. 8 Connection for the recorder for time code (TC) and word clock (WORD) synchronization. The time code was directly distributed from the first machine to all others and the word clock was distributed using a cascade connection.

Examples of recording are shown in Fig. 9. The upper photo shows recording in a concert hall. The lower photo was taken during the measurement of the sound of a cicada on a university campus.

As mentioned in the previous section, impulse responses are needed to calculate the inverse filter matrix. The total number of possible combinations of microphones and loudspeakers in the sound cask amounted to $80 \times 96 = 7{,}680$. The impulse responses were measured in advance to enable reproduction.

For data transfer and post-processing such as inverse filtering with MATLAB and playback with DAW software, an iMac with a clock frequency of 3.4 GHz (Intel Core i7) and 32 GB memory, or an equivalent Windows PC, is usually used. Nuendo 5.5 or higher is used in our project as the standard DAW. Note that the above specs are not the minimum requirements. A PC with a lower clock frequency or less memory would work without serious problems. However, as general software, one with a larger memory and a faster clock provides a more stable and reliable performance.

5. LOCALIZATION TEST

To verify the principal performance of the sound cask and the recording system, a simple localization test was carried out. Eight adults with normal hearing (age range 20–22; 4 males, 4 females) participated in this experiment. Informed consent was obtained after the nature and possible consequences of the studies were explained.

Auditory stimuli were pink noise (1 s on-time and 0.4 s off-time, three bursts). Each stimulus was convoluted with

Fig. 9 Examples of recording. Top: recording in a concert hall. Bottom: on a university campus.

each impulse response to simulate the signals at each control point at the primary boundary surface in free space. The impulse responses were calculated assuming that a point source was located at distance of 2 m from the center of the microphone array and in each direction. The A-weighted sound pressure level of stimulus was adjusted to 60 dB at the center of the head to eliminate the level differences between directions or distances [23].

5.1. Procedure

The test was conducted in the sound cask. Participants sat on a chair and listened to the auditory stimuli. The experiment was divided into three sessions: horizontal, vertical, and distance sessions. In the horizontal session, stimuli were presented from angles of 0–345° at 15° intervals.

In the vertical session, the stimuli were presented from angles of 0–90° at 15° intervals. In both sessions, the stimuli were presented at a distance of 2 m. The participants were asked to illustrate their perceived direction on answer sheets after listening to the stimuli.

In the distance session, we used magnitude estimation method [24]. The standard stimuli were presented at a distance of 100 cm and were given a numerical value of 100. For subsequent stimuli, the participants were asked to numerically report their perceived distance relative to the standard so as to preserve the ratio between the sensation and numerical estimates. We set seven conditions: 30, 60, 90, 120, 150, 180, and 240 cm. In this session, the horizontal angle was 0°, and the vertical angle was 0°. For the perceived distance of each participant, we calculated the geometric average across repeated trials under each distance condition.

The sessions were in order of horizontal, vertical, and distance for all participants. The trial was repeated 10 times at each distance and the presentation order was randomized in each session. The participants were permitted to move their heads and bodies during the stimulus presentation. The participants took part in a practice session, which was followed by the experimental session. The interval between the trials was 5 s.

5.2. Results and Discussion

Horizontal session: Figure 10(a) shows the mean perceived location versus actual locations in the horizontal session across all participants. The mean error angle was 7°, the minimum error angle was almost 3°, and the maximum was 14°. Previous studies showed that the minimum angle error was 2° and the maximum was almost 15° in horizontal localization with real sources [25]. These results indicate that the listeners were able to perceive the sound image at the presented locations in the horizontal plane.

Vertical session: Figure 10(b) shows the mean perceived location versus actual locations in the vertical session. The mean error angle was 23° across participants, the maximum error angle was 30° under the 0° condition, and the minimum was 15° under the 45° condition. Previous studies showed that the minimum error angle was 3° and the maximum was 15° in vertical localization with real sources [25]. It was thus difficult for the listeners to localize the sound image at the presented locations in the vertical plane.

Currently, the authors assume two reasons for the unsatisfactory results in the vertical direction compared with previous results such as those in [26]: there was theoretically limited reproduction performance at high frequencies due to the distance between adjacent microphones, and the performance of inverse filters was also limited owing to the calculation method.

For the first reason, previous studies of head related transfer functions (HRTFs) have suggested that notches with a frequency of more than 4 kHz are very important for vertical localization [27]. However, the present BoSC system, in which the distance between control points is 8 cm, cannot secure spatial reproduction performance at high frequencies such as 4 kHz. Therefore, the listeners

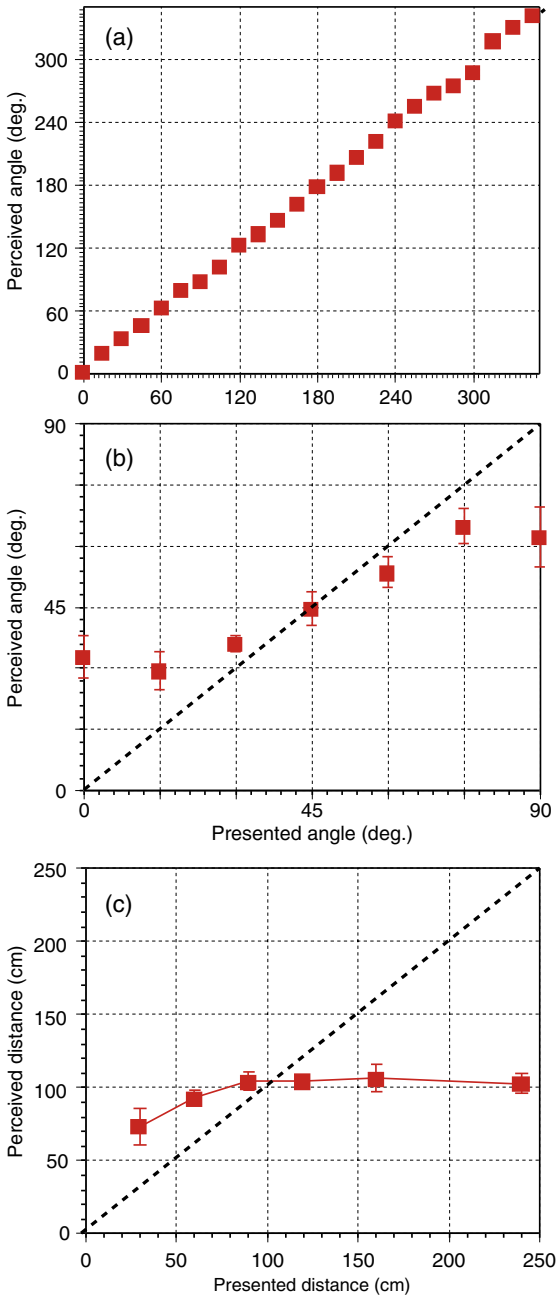

Fig. 10 The results of the localization tests. (a): the horizontal session, (b): the vertical session, (c): the distance session.

were unable to perceive the sound image at an assumed position in the vertical direction.

For the second reason, improvement by reasonable modification of the calculation, e.g., the method of calculation of the inverse filters, is needed. In the BoSC system, the characteristics of the inverse filter should be an important factor for the performance such as localization, since the filters directly affect the amplitude and phase of reproduced signals. Currently, the minimum-norm solution is used. However, there is still room for adding an operation. The relationship between the method of generating the inverse filter and the total performance is one of the important subjects of our ongoing research.

Distance session: Figure 10(c) shows the mean perceived distance versus actual distance across all participants. It is clear that the listeners tended to overestimate the distance from the sound image up to distances of 100 cm in the sound cask. Above 100 cm, participants were unable to discriminate the difference in distance. In this study, we used auditory stimuli that simulate the impulse responses of the free field, and used the same sound pressure level across all conditions. Under these conditions, it is considered that participants used the cue of HRTFs to estimate the distance of the stimuli. However, the HRTFs were less valid at distances of more than 100 cm [28]. Therefore, the participants were unable to estimate distances of more than 100 cm. In addition, previous studies also showed that listeners tended to overestimate the distance of a sound image up to distances of 100 cm in the front direction [29]. From these results, we consider that the sound cask can provide reasonable reproduction performance in terms of perception of the distance of the apparent source.

6. CONCEPT OF A SOUND FIELD SHARING SYSTEM

As one of the applications of the BoSC system, the concept of a sound-field-sharing system using more than one system has been introduced. This system was a distant communication system that allows people to communicate as if they were in the same room. A similar sharing system has already been designed and evaluated as shown in [30]. In [30], the concept, practical scheme, and examples of distant-field sharing were shown. However, the methods of sound field capture and reconstruction were quite basic; e.g., conventional recording technique with several microphones was used and reverberation sounds were reproduced from the loudspeakers located in the direction of incidence (details of the reproduction method were unclear). The advantages of our proposed system can be summarized as follows:

- directional information can be captured with high accuracy by the microphone array
- theoretically verified reproduction method
- applicability to moving sources
- a real immersive environment in an enclosed space, the cask

Figure 11 shows a conceptual diagram of the sound-field-sharing system, in which Players A and B share a sense of being in the same primary field. Space (1) in the figure is the existing real primary field and is the sound field that a subject aims to share. In this primary field, for example, the music performances are recorded using a microphone array as (1-1). The array is installed in the

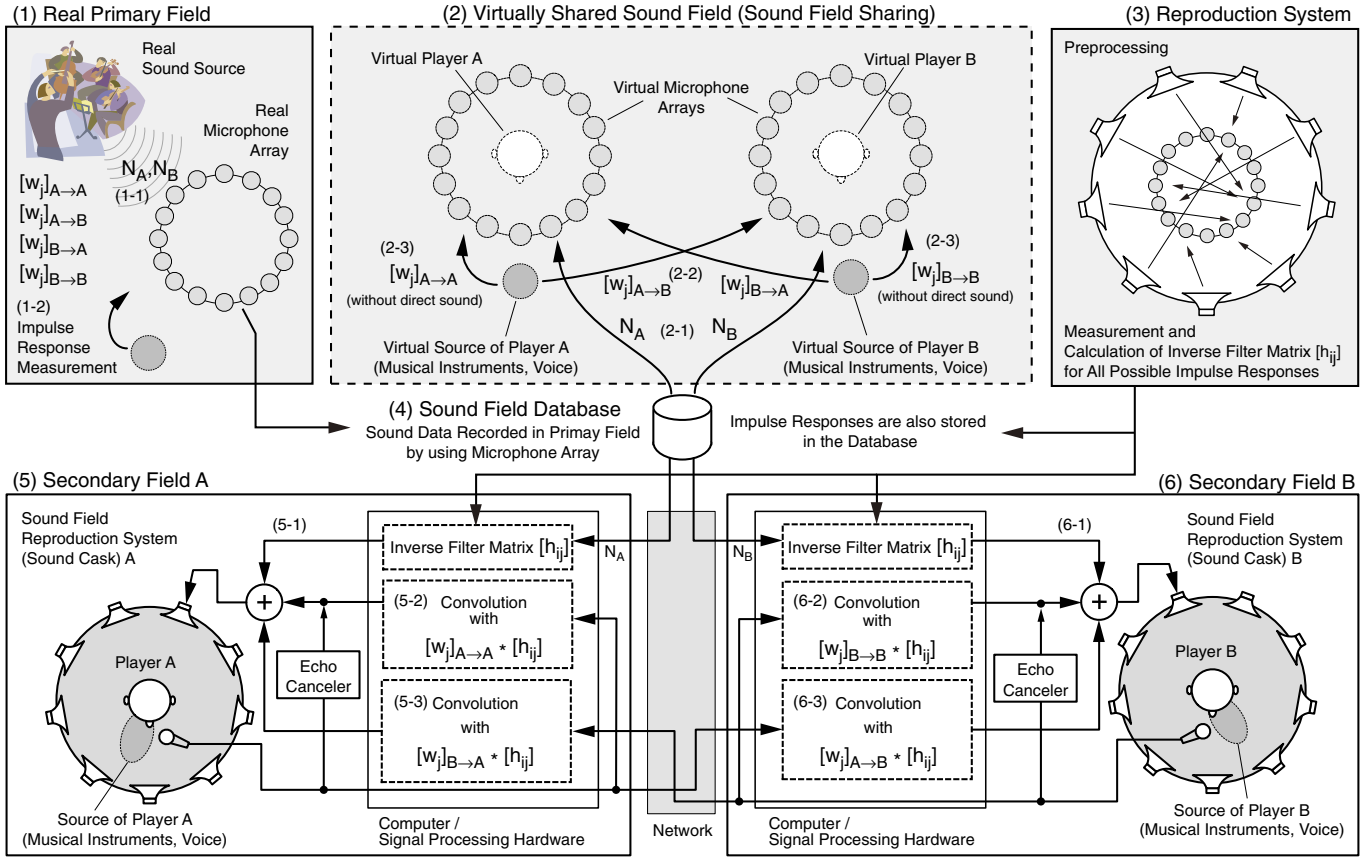

Fig. 11 Concept of a sound-field-sharing system, in which Players A and B inside the reproduction systems at distant locations share a sense of being in the same primary field.

places where Players A and B are assumed to stand and the recorded signal N_A or N_B can be played back by a reproduction system. In addition, the impulse responses between the sound sources and the adjacent microphone array in this real space are measured as indicated in (1-2). These are necessary to yield the musical sounds or voices, which can be recognized as if they are played in the primary sound field by the experiment participants, the players. For example, the impulse response from the instrument position of Player A to the j-th microphone of Player B is indicated as $[w_j]_{A\to B}$ in the figure.

Space (2) in the figure is a virtually shared sound field, and a sharing space for the players. Players A and B listen to music in the same primary sound field with a high degree of presence using a reproduction system based on the BoSC principle. The recorded signals N_A and N_B are played back as (2-1).

In addition, they feel as if they are playing as an ensemble in the same primary field. The sound played by Player A is transmitted to Player B after passing through the impulse responses between the source position of A and the listening position of B measured in the primary field. This is the signal flow shown as $[w_j]_{A\to B}$ in the figure. The reverse process is the same and expressed as $[w_j]_{B\to A}$ as indicated in (2-2).

Furthermore, the played signals are also transmitted to themselves after passing through the impulse responses of their own source position and listening position. These are shown as (2-3) and $[w_j]_{A\to A}$ and $[w_j]_{B\to B}$.

Space (3) in the figure is the sound-field-reproducing system. This consists of a sound cask and a C80 fullerene-shaped microphone array. The impulse responses between all possible combinations of loudspeakers and microphones are measured in the preliminary stage and used to calculate the inverse filter matrix necessary for the BoSC principle. The inverse filter matrix is indicated as $[h_{ij}]$ in the figure. The recorded signals and impulse responses are stored in the database. These are transmitted via a network or a dedicated line on demand.

The sound fields where Players A and B exist are indicated as spaces (5) (Secondary Field A), and (6) (Secondary Field B), respectively. More precisely, these are the inner spaces of the sound field reproduction system, i.e., the sound cask in this case. The real and detailed signal flows are provided in the lower part of the figure. The recorded signals in the primary field are played back after passing through the inverse filter matrix $[h_{ij}]$ and these are the signals of (5-1) and (6-1). In addition, the musical sound or voice of Player A is transmitted back after passing through the impulse responses $[w_j]_{A\to A}$ and inverse

Fig. 12 Appearance of ensemble test with two casks located side by side.

filter matrix $[h_{ij}]$, and to Player B with filters $[w_j]_{A\to B}$ and $[h_{ij}]$. These are indicated as (5-2) and (6-3), respectively. A similar process for Player B is indicated as (6-2) and (5-3). In the autoregression filters $[w_j]_{A\to A}$ and $[w_j]_{B\to B}$, the direct sound should be removed and an echo canceler introduced in the regression process. A sound from the real primary field (1) which is added to the sound, is heard by Players A and B and expected to evoke the environment in which they exist and play in the same sound field and playing ensemble.

In previous studies, a sound-field-sharing system was developed, which used the BoSC system with 62-channel loudspeakers [18,19]. The BoSC system enables the perception of the direction of the reproduced voice. The system transmits the voice direction in a three-person conversation by changing the transfer functions in accordance with the angle at which the speaker is facing [20]. Only 24 loudspeakers were used to reproduce voices between the systems to avoid a large number of calculations [21].

Basic experiments to evaluate the communication between separate casks with the necessary convolutions of signals in real time have been started recently as shown in Fig. 12. At the current stage, two casks are located side by side and players inside the casks try to play as an ensemble during which one can hear the other's play, which is reproduced by the necessary convolutions ((5-3) and (6-3) in Fig. 11). The possibility of a remote ensemble with a high sense of presence has been confirmed. For further improvement of performance, continuous experimental examinations are being conducted.

7. CONCLUSIONS

The newly developed immersive sound field reproduction system, named the "sound cask," was introduced. The system, which has 96-channel reproducing loudspeakers, realizes precise sound field reproduction by combination with an 80-channel microphone array and the principle of boundary surface control. The results of basic tests indicated that the cask had superior performance in terms of the localization of a reproduced sound source in the horizontal plane. However, there is room for improvement in the performance in the vertical direction and in the recognition of distance. Further examinations for improvement, e.g., using various types of inverse filter, are currently being conducted.

At the present stage of the project, four casks have been constructed at Fukuoka, Kyoto, and Tokyo in Japan. Several hundred people have experienced the performance of the sound field reproduction in the cask. A series of psychological experiments have been reported, for example, in [23]. In addition to recordings in concert halls, many types of content such as outdoor environmental sound have been continuously accumulated and stored in a database for reproduction.

ACKNOWLEDGMENT

The sound cask was designed and constructed as an important device in the research project entitled "Development of a sound field sharing system for creating and exchanging music" organized by Japan Science and Technology (JST) CREST, whose purpose was to build an information communication environment that supports the exchange of music, a universal language, with high fidelity.

REFERENCES

[1] ITU-R BS.775.1, "Multichannel Stereophonic Sound System With and Without Accompanying Picture," Geneva (1992–1994).
[2] SMPTE 2036-2-2008, "Ultra High Definition Television — Audio Characteristics and Audio Channel Mapping for Program Production" (2008).
[3] A. J. Bourkhout, D. de Vries and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Am.*, **93**, 2764–2778 (1993).
[4] P. A. Gauthier and A. Berry, "Adaptive wave field synthesis with independent radiation mode control for active sound field reproduction: Theory," *J. Acoust. Soc. Am.*, **119**, 2721–2737 (2006).
[5] P. A. Gauthier and A. Berry, "Adaptive wave field synthesis for sound field reproduction: Theory, experiments and future perspectives," *J. Audio. Eng. Soc.*, **55**, 1107–1124 (2007).
[6] P. A. Gauthier and A. Berry, "Adaptive wave field synthesis for active sound field reproduction: Experimental results," *J. Acoust. Soc. Am.*, **123**, 1991–2002 (2008).
[7] G. Theile and H. Wittek, "Wave field synthesis: A promising spatial audio rendering concept," *Acoust. Sci. & Tech.*, **25**, 393–399 (2004).
[8] K. Ueno, K. Yasuda, H. Tachibana and T. Ono, "Sound field simulation for stage acoustics using 6-channel system," *Acoust. Sci. & Tech.*, **22**, 307–309 (2001).
[9] S. Yokoyama, K. Ueno, S. Sakamoto and H. Tachibana, "6-channel recording/reproduction system for 3-dimensional auralization of sound fields," *Acoust. Sci. & Tech.*, **23**, 93–103 (2002).

[10] D. H. Cooper and T. Shiga, "Discrete-Matrix Multichannel Stereo," *J. Audio Eng. Soc.*, **20**, 346–360 (1972).
[11] M. A. Gerzon, "Hierarchical system of surround sound transmission for HDTV," *Proc. AES 92nd Convention*, Preprint 3339 (1992).
[12] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *J. Audio Eng. Soc.*, **53**, 1004–1025 (2005).
[13] S. Ise, "A study on the sound field reproduction in a wide area (1) — based on kirchhoff-helmholtz integral equation—," *Proc. Autumn Meet. Acoust. Soc. Jpn.*, pp. 479–480 (1993) (in Japanese).
[14] S. Ise, "A principle of active control of sound based on the Kirchhoff-Helmholtz integral equation and the inverse system theory," *J. Acoust. Soc. Jpn.* (*J*), **53**, 706–713 (1997) (in Japanese).
[15] S. Ise, "A principle of sound field control based on the Kirchhoff-Helmholtz integral equation and the theory of inverse systems," *Acustica*, **85**, 78–87 (1999).
[16] S. Ise, M. Toyoda, S. Enomoto and S. Nakamura, "An attempt of sound field sharing system for profound communication. Concept and basic stance of the project–," *Proc. Spring Meet. Acoust. Soc. Jpn.*, pp. 585–586 (2007) (in Japanese).
[17] S. Enomoto, Y. Ikeda, S. Ise and S. Nakamura, "Three-dimensional sound field reproduction and recording system based on boundary surface control principle," *Proc. Int. Conf. Aud. Disp. 2008*, 8 pages (2008).
[18] S. Enomoto, "3D sound field recording/reproduction system for telecommunication (in Japanese)," *Arch. Acoust. Noise Control*, **38**, 37–42 (2009).
[19] S. Enomoto, Y. Ikeda, S. Ise and S. Nakamura, "3-D sound field reproduction system for the sound field shared communication based on the boundary surface control principle," *Proc. Autumn Meet. Acoust. Soc. Jpn.*, pp. 1411–1414 (2009) (in Japanese).
[20] Y. Ikeda, S. Enomoto, S. Ise and S. Nakamura, "Three-party sound field sharing system based on the boundary surface control principle," *Proc. Int. Congr. Acoust. 2010*, 6 pages (2010).
[21] S. Enomoto, Y. Ikeda, S. Ise and S. Nakamura, "Optimization of loudspeaker and microphone configurations for sound reproduction system based on boundary surface control principle," *Proc. Int. Congr. Acoust. 2010*, 7 pages (2010).
[22] R. Kleinman and G. Roach, "Boundary integral equations for the three dimensional Helmholtz equation," *SIAM Review*, **16**, 214–236 (1974).
[23] M. Kobayashi, K. Ueno, M. Yamashita, S. Ise and S. Enomoto, "Subjective evaluation of a virtual acoustic system: Trials with three-dimensional sound field reproduced by the 'Sound Cask'," *Proc. Int. Congr. Acoust. 2013*, 9 pages (2013).
[24] S. S. Stevens, "Problems and method of psychophysics," *Psychol. Bull.*, **55**, 177–196 (1958).
[25] J. C. Makous and J. C. Middlebrooks, "Two-dimentional sound localization by human listeners," *J. Acoust. Soc. Am.*, **87**, 2188 (1990).
[26] J. Blauert, "An experiment in directional hearing with simultaneous optical simulation," *Acustica*, **23**, 118–119 (1970).
[27] K. Iida, M. Itoh, A. Itagaki and M. Morimoto, "Sound transmission to and within the human ear canal," *Appl. Acoust.*, **68**, 835–850 (2007).
[28] P. Zahorik, "Assesing auditory distance perception using virtual acoustics," *J. Acoust. Soc. Am.*, **111**, 1832–1846 (2002).
[29] D. S. Brungart, N. I. Durlach and W. M. Rabinowitz, "Auditory localization of nearby source. II. Localization of a broadband source," *J. Acoust. Soc. Am.*, **106**, 1956–1968 (1999).
[30] W. Woszczyk, J. Cooperstock, J. Roston and W. Martens, "Shake, rattle, and roll: Getting immersed in multisensory, interactive music via broadband networks," *J. Audio. Eng. Soc.*, **53**, 336–344 (2005).

Akira Omoto graduated from the Department of Acoustic Design, Kyushu Institute of Design, in 1987 and received his Ph.D. degree from the University of Tokyo in 1995. From 1987 to 1991, he worked as a research and development engineer at Nittobo Acoustic Engineering Co., Ltd., Tokyo. In 1991, he was appointed as a Research Assistant at Kyushu Institute of Design and was made an Associate Professor in 1997. He is currently a Professor of Kyushu University following the integration of universities in 2003. He is in charge of database construction in this project.

Shiro Ise graduated from the Department of Electronics and Communications Engineering, Waseda University, in 1984 and received his Ph.D. degree from the University of Tokyo in 1991. From 1984 to 1986, he worked as a research and development engineer at Korg Inc., Tokyo. In 1994, he was appointed as a Research Assistant at Nara Institute of Science and Technology and was made an Associate Professor at Kyoto University in 1998. He has been a full Professor of the School of Information Environment, Tokyo Denki University since April 2013. He is in charge of system design and the team leader in this project.

Yusuke Ikeda graduated from the Department of Information Science, Waseda University, in 2001, and received his Ph.D. degree from Waseda University in 2007. From 2007 to 2009, he worked as a researcher at Advanced Telecommunications Research Institute International (ATR-SLC). From 2009 to 2011, he worked as a researcher at National Institute of Information and Communications Technology (NICT). From 2011 to 2013, he worked as a researcher at Kyoto University and CREST of Japan Science and Technology Agency (JST). Since April 2013, he has been an Assistant Professor of Tokyo Denki University and a researcher of JST/CREST.

Kanako Ueno graduated from the Faculty of Engineering, University of Tokyo, in 1996, and received Doctor of Engineering degree from University of Tokyo in 2003. From 1999 to 2008, she worked as a Research Associate at the Institute of Industrial Science, University of Tokyo. In 2008, she was appointed as a Lecturer at the School of Science and Technology, Meiji University, where she has been an Associate Professor since 2010. She is in charge of the psychological evaluation of the system in this project.

Seigo Enomoto received his B.E. degree from Kinki University in 1997, his M.E. degree from Nara Institute of Science and Technology (NAIST) in 1999, and his Ph.D. degree from Kyoto University in 2005. From 2005 to 2009, he worked as a researcher at Spoken Language Communication Lab., Advanced Telecommunications Research Institute International (ATR-SLC). He is currently a researcher of Multisensory Cognition and Computation Lab., National Institute of Information and Communications Technology (NICT). He is a member of AES, ASA, ASJ, and IEICE.

Maori Kobayashi graduated from the Department of Psychology, Rikkyo University, in 2001, and received her Ph.D. from Rikkyo University in 2006. From 2006 to 2011, she worked as a researcher at the Research Institute of Electrical Communication, Tohoku University. Since April 2011, she has worked as a researcher at Meiji University and CREST of Japan Science and Technology Agency (JST).