Chapter 2
Development of a Sound Field Sharing System for Creating and Exchanging Music

Shiro Ise

Abstract Many interactions in modern society have become virtual. To form deep bonds between people in our highly connected society, a mechanism for sharing one's true feelings is required. Sound is one of the strongest tools that can be used to emulate an in-person experience, making 3D auditory interface technology a necessary area of exploration. There are several reasons for such a simple solution, realistic sound, not yet being realized. First, development of visual technology has been prioritized owing to its clarity and ease of understanding. Second, text-based technology has rapidly grown because of commercial requirements. Finally, the conventional theoretical basis of sound field reproduction is not sufficiently realizable. However, the principle of boundary surface control (BoSC) was proposed during the 1990s as a new theoretical basis of 3D sound field reproduction. After several trials of developing the sound field reproduction system based on the BoSC principle, this project started to clarify the applicability of the BoSC principle. As a result of this project over five years, we succeeded in being the first in the world to develop an immersive auditory display, which is named "sound cask" from its cask-shaped appearance. Since producing sound and listening to sound has been important to human interaction since early humanity, the sound cask will be an invaluable tool.

Keywords Immersive auditory display · Stereophonics · Virtual reality · 3D sound technology · Sound sharing system

2.1 Prologue

While bodily cues, especially emotions, are lost as soon as speech is transformed into text, text simultaneously ensures longevity and reproducibility. Sound emanating from the vocal chords vibrating reverberates through the air and dissipates in an instant. The listener reacts accordingly, interpreting each little sound wave one after another. On the other hand, written words will remain largely static regard-

S. Ise (✉)
Tokyo Denki University, Chiba, Japan
e-mail: iseshiro@mail.dendai.ac.jp

less of passing time and, furthermore, allows for duplication and proliferation of the contents to other places. Furthermore, if the written word reveals new and useful knowledge, it tends to be proliferated even more. The development of informational technology in recent years has dramatically increased the possibility of distributing written words compared to prior printing technology. Text put on the internet forms a digital medium that can be reproduced and viewable from any connected device in the world. Although unwritten knowledge is everywhere, written-down partial truths, due to being easily proliferated on a large scale, have a significant effect on our physical environment as momentary social phenomena.

Modern humans have lost a significant number of opportunities to be physically active and create sound, by passing time interacting with others across virtualized worlds using mobile devices. Even in the office, instead of talking to one's neighbor, email, SNS, or the like are quite often used. Since communication that does not involve the exchange of emotions does not enhance feelings of unity, anxieties about social life cannot be expunged very easily. No matter how much superficial information is exchanged, the deep interpersonal communication that is necessary cannot be formed. Therein lies the root of what ails modern society.

As a start to solving this problem, we now focus on the interface between individuals and this information environment. Specifically, I suggest a sound interface that is able to sufficiently handle the means by which people create and broadcast sound in information space. A classic example of using sound in a physical space is the skill of performing with a musical instrument in a concert hall. The audience in the concert hall can then appreciate the musical arts, as a result of the musical group acting in coordination on stage. Modern-day rock music, theater, and other such various performances endure under a single cultural and economic value. Therefore, the fundamental position of this research project is that, while humanity's natural inclination involves the gathering of people into a shared space for such activities, the ideal for a society with harmony of information is for the same to take place for such activities in virtual space.

2.2 Theoretical Basis

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. More academic or exact methods include various binaural reproduction methods, wave field synthesis (WFS) [3–7], and, more recently, the 6-channel system [8, 9] and higher-order ambisonics [10–12]. The boundary surface control (BoSC) technique proposed in 1993 by Ise is an academically important reproduction method [13–15]. Because the actual construction of a reproduction system using 62-channel loudspeakers is at an early stage, development has not only been a function of sound field reproduction, but the system has also been investigated as a function of sound field sharing at a remote location [16–21]. The experimental results indicate the
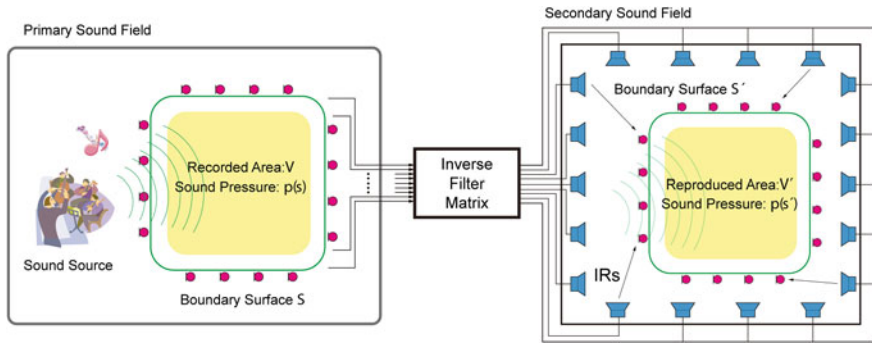
Fig. 2.1 Concept of the boundary surface control principle with an inverse filter matrix. The sound pressures at surface S are reproduced at surface S' in the secondary sound field. The inverse filter matrix, which is calculated from impulse responses from all possible combinations of loudspeakers and microphones, is introduced to reproduce the sound pressures at S'

validity of the system. The boundary surface control principle is described using the Kirchhoff-Helmholtz integral equation and the inverse system [15]. Figure 2.1 shows its basic concept.

We are considering the reproduction of a sound field within a recorded area V in the primary field into a reproduction area V' in the secondary field. Given that V is congruent with V', the following equation holds:

$$|\mathbf{r}' - \mathbf{s}'| = |\mathbf{r} - \mathbf{s}| \qquad (\mathbf{s} \in V, \mathbf{r} \in S, \mathbf{s}' \in V', \mathbf{r}' \in S') \tag{2.1}$$

where S and S' denote the boundary of the recorded area and the boundary of the reproduction area respectively. If we denote the sound pressure in V and V' as $p(\mathbf{s})$ and $p(\mathbf{s}')$ respectively, $p(\mathbf{s})$ and $p(\mathbf{s}')$ are denoted by the following equations

$$p(\mathbf{s}) = \iint_S G(\mathbf{r}|\mathbf{s}) \frac{\partial p(\mathbf{r})}{\partial n} - p(\mathbf{r}) \frac{\partial G(\mathbf{r}|\mathbf{s})}{\partial n} \delta S, \qquad (\mathbf{s} \in V) \tag{2.2}$$

$$p(\mathbf{s}') = \iint_{S'} G(\mathbf{r}'|\mathbf{s}') \frac{\partial p(\mathbf{r}')}{\partial n}' - p(\mathbf{r}') \frac{\partial G(\mathbf{r}'|\mathbf{s}')}{\partial n}' \delta S', \qquad (\mathbf{s}' \in V') \tag{2.3}$$

where n and n' denote normal vectors on S and S' respectively. By applying Eq. 2.1, we obtain the following relationships of Green's function and its gradient:

$$G(\mathbf{r}|\mathbf{s}) = G(\mathbf{r}'|\mathbf{s}') \tag{2.4}$$

$$\frac{\partial G(\mathbf{r}|\mathbf{s})}{\partial n} = \frac{\partial G(\mathbf{r}'|\mathbf{s}')}{\partial n'} \tag{2.5}$$

Hence, it follows that if the sound pressure and its gradient on each boundary are equal to each other, then the sound pressures in each area are also equal to each other from Eqs. 2.2 and 2.3. This is expressed as

Fig. 2.2 3D sound field recording in a concert hall using a C80 fullerene microphone

$$\forall \mathbf{r} \in S \quad \forall \mathbf{r}' \in S'$$
$$p(\mathbf{r}) = p(\mathbf{r}') \quad \frac{\partial p(\mathbf{r})}{\partial n} = \frac{\partial p(\mathbf{r}')}{\partial n'}$$
$$\implies \quad \forall \mathbf{s} \in V \quad \forall \mathbf{s}' \in V' \quad p(\mathbf{s}) = p(\mathbf{s}'). \tag{2.6}$$

Considering this as a boundary value problem, the uniqueness of the solution follows in that either the sound pressure value or its gradient value are sufficient to determine the value for both [22]. To construct the recorded and reproduction areas, a microphone array is generally used. As shown in Fig. 2.2 the C80-shaped fullerene microphone array was adopted in our project. One additional 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 Fig. 2.1) are measured in advance, and the inverse filter matrix is calculated [15]. These filters are applied to the signals for reproducing the sound pressures of surface S at the target surface of S'. Another well-known sound reproduction method using the Kirchhoff-Helmholtz integral equation is wave field synthesis (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 exist, e.g., 6-channel system [8, 9] and ambisonics system [10–12]. However, the sound cask has practical advantages compared with other systems in the following points:

- Sound image along the depth direction can be controlled even in the vicinity of the head of the listener;

- The whole system can be easily moved into any place.
- A theoretically assured combination with a recording system, an 80-channel fullerene-shaped microphone array in our case, can be constructed.

2.3 Sound Cask

The main characteristic of the BoSC system is its ability to reproduce a sound field, not by points but in three dimensions. A listener can freely move his head, and the system can provide high performance of spatial information reproduction such as conveying sound localization and sound distance [17]. Based on 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 process of the sound field reproduction system based on the BoSC principle, space design, which is suitable for inverse filter calculation, will be important, since the quality of these filters directly affect the total performance of the system. The previous system [17] consisted of 62 loudspeakers mounted on a dome-shaped wooden frame arranged inside the music practice chamber whose floor space was around $2\,m \times 2\,m$. Several experiments indicated that the following items inevitably and adversely affected the performance of the inverse filter:

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

Measures against these items are:

- no parallel planes inside the enclosure except for ceiling and floor 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 to have many people 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 is aimed for in order to realize an immersive environment.
- The inner space must be large enough to play musical instruments.
- The basic performance of sound acquiring and reproducing devices, such as the microphone and loudspeaker, should be higher, to the extent possible, to achieve so-called Hi-Fi reproduction.

Fig. 2.3 Sound cask

- The space density of reproducing loudspeakers should also be higher, to the extent possible, to achieve higher resolution of sound localization. At the same time, a larger dimension of the loudspeaker unit would be preferred for better response in the low frequency range.
- The numbers of channels should be practically controllable from a commonly available computer and digital audio workstation (DAW).

As a practical and reasonable compromise of the conditions above, 96 loudspeakers are allocated inside the sound cask. Figure 2.3 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 a diameter of 1950 mm in the central horizontal plane and a height of 2150 mm, the sound cask has a sufficiently large internal space to play wind and string instruments. Ninety-six full-range loudspeakers are installed on the walls and ceiling but not the floor plane. Six loudspeakers are installed on the ceiling plane. The speaker is 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 adjoining heights is around 350 mm. The average interval between adjoining loudspeakers in the horizontal direction is around 540 mm for the top and bottom heights, and around 330 mm for all other heights. In our previous BoSC system, 62 loudspeakers were installed around the upper body of the listener. However, in the sound cask, loudspeakers were installed to cover the whole body of the 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. Each wall of this sound cask can be disassembled within around 30 to 40 min, and can be assembled within around two hours by several workers. For a sound-absorbing material, poly-wool, recycled material from plastic bottles (thickness 120 mm, density 32 kg/m³), 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 the newly designed digital amplifiers (Fig. 2.4). Ninety-six channels of audio signals are transmitted from the PC through only two MADI (multichannel audio digital interface) optical lines as shown in Fig. 2.5. The 128 channels of data, which consist of 124 channels of audio signals and 4 channels of control signals, are distributed to the 12 serially connected digital amplifiers using an optical MADI cable. Each 8ch D-class amplifier extracts appropriate 8ch audio signals from the MADI optical lines, and creates amplified PWM signals. After PWM signals pass through LPF, loudspeakers are driven in at 76 W at 10% THD+N 8Ω. In total, 12 D-class amplifiers, 12 third-order passive
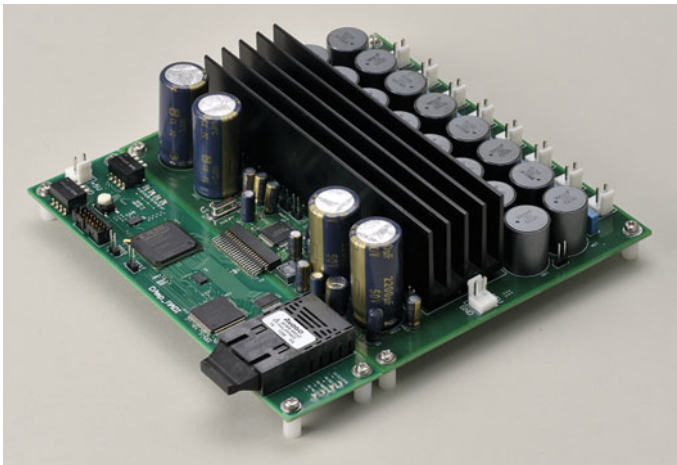
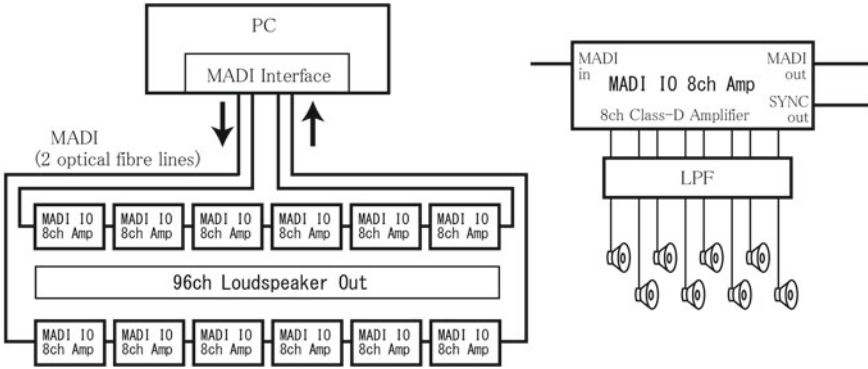**Fig. 2.4** Newly designed 8ch digital amplifiers with a MADI interface

Fig. 2.5 Structure of the 96ch digital amplifier network

Fig. 2.6 Amplifiers and power supplies at the bottom of the cask

LPFs, 4 stabilized power supplies of 36 V (6.7 A), a 12 V (13 A) power supply, and a 5 V (15 A) power supply are installed at the bottom of the sound cask as shown in Fig. 2.6.

2.4 Performance Evaluation of the System

2.4.1 Physical Performance

In order to visualize the sound field reproduction, a microphone traverse system is located inside the sound cask as shown in Fig. 2.7. By moving a microphone, it can measure the sound pressure of the cylindrical region having a diameter of 1 m and a height of about 2 m. As the primary sound field, a loudspeaker and fullerene microphone were located in a soundproof room at a distance of 1.5 m. The loudspeaker was driven from five directions: front direction; right direction horizontal +30 and +120°; and left direction horizontal −30 and −120° by pulse signals with limited frequencies lower than 1 kHz. In the reproduced sound field in the sound cask, the sound pressure signals were measured iteratively at the moving point at each 4 cm distance and 4°. As depicted in Fig. 2.8, the wave front in the circle of 1 m in diameter can be seen, in order from the top, 30° left direction (L), 30° right direction (R), front direction (C), 120° left direction (Ls), and 120° right direction (Rs). In Fig. 2.8, the left-to-right change indicates the passage of time. It was found that the wave front was reproduced in a wider region than the fullerene microphone diameter (45 cm) [23].
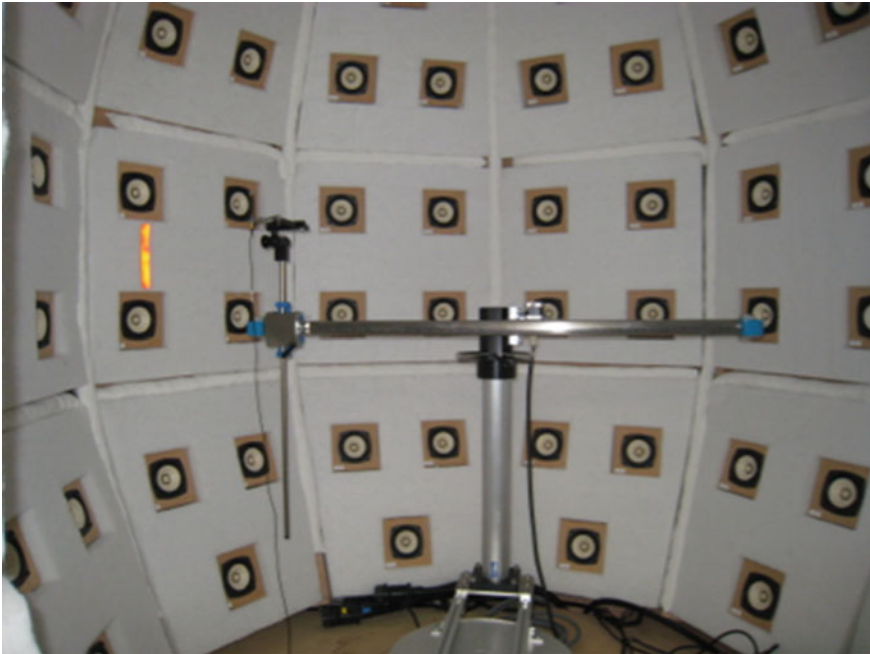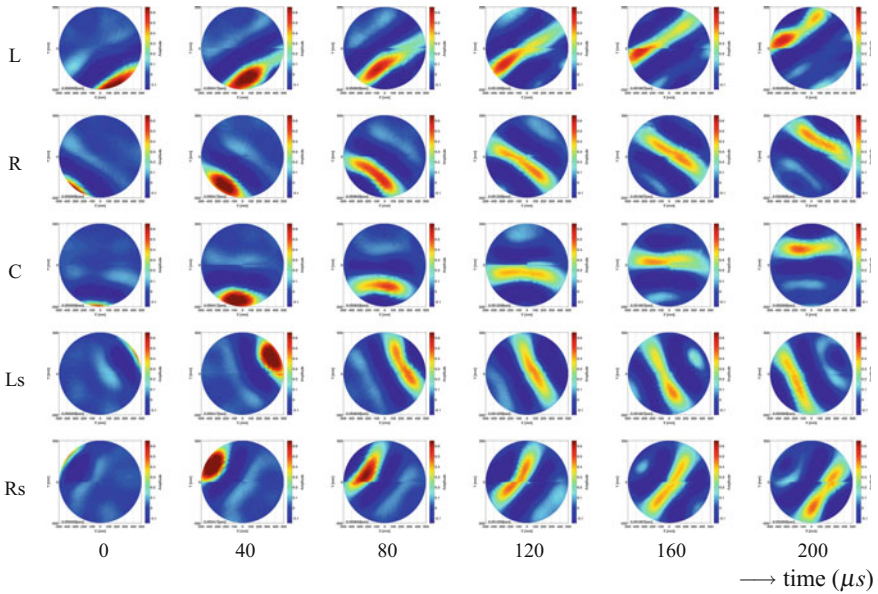

Fig. 2.7 Microphone traverse system located in the sound cask

Fig. 2.8 Measured reproduced sound field

2.4.2 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 20–22, four 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 each impulse response, in order to simulate the signals at each control point at the primary boundary surface in free space. The impulse responses were calculated assuming a point source was located at a distance of 2 m from the center of the microphone array and at the angle of each direction. The A-weighted sound pressure level of stimulus was adjusted to 60 dB at the center of a head to eliminate the level differences between directions or distances.

2.4.2.1 Procedures

Testing was conducted in the sound cask. The participants sat down in the 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 $0-345°$ with $15°$ intervals. In the vertical session, the stimuli were presented from angles of $0-90°$ each with $15°$ intervals. In both sessions, the

distance was 2 m. The participants were asked to illustrate on the answer sheets their perceived direction after listening to the stimuli. In the distance session, we used the magnitude estimation method [24]. The standard stimulus was at a 100 cm distance and was given a numerical value of 100. For subsequent stimuli, participants were asked to report numerically their perceived distance relative to the standard so as to preserve the ratio between the sensations 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°. As the perceived distance for each participant, we calculated the geometric average across repeated trials for each distance condition. The session order was the horizontal, the vertical, and the distance session for all participants. The trial was repeated 10 times for each condition, 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. Intervals between the trials were 5 s.

2.4.2.2 Results and Discussion

Horizontal session: Fig. 2.9a 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 error angle was 14°. Previous studies showed that the minimum error angle was 2° and the maximum error angle was almost 15° in the horizontal localization with real sources [26]. These results indicate that the listeners can perceive the sound image at the presented locations in the horizontal plane.

Vertical session: Fig. 2.9b shows the mean perceived location versus actual locations in the vertical session. The mean error angle was ±3° across participants, the maximum error angle was 30° in the 0° condition, and the minimum error angle was 15° in the 45° condition. Previous studies showed that the minimum error angle was 3° and the maximum error angle was 15° in vertical localization with real sources [26]. It is difficult for the listeners to localize the sound image into the presented locations in the vertical plane at this time.

Distance session: Fig. 2.9c shows the mean perceived distance versus actual distance across all participants. It is clear that the listeners tended to overestimate the distance to the sound image within 100 cm in the sound cask. At more than 100 cm, the participants could not discriminate the difference in distances. In this study, we used auditory stimuli that simulate the impulse responses of the free field, and set the same sound pressure level across all conditions. Under these conditions, it is considered that the participants used the cue of HRTFs to estimate the distance from the stimuli. However, the HRTFs lost validity at a distance of more than 100 cm [27]. Thus, participants could not estimate the distance at more than 100 cm. In addition, previous studies also showed that listeners tend to overestimate the distance of the sound image within 100 cm in the front direction [28]. From these results, we
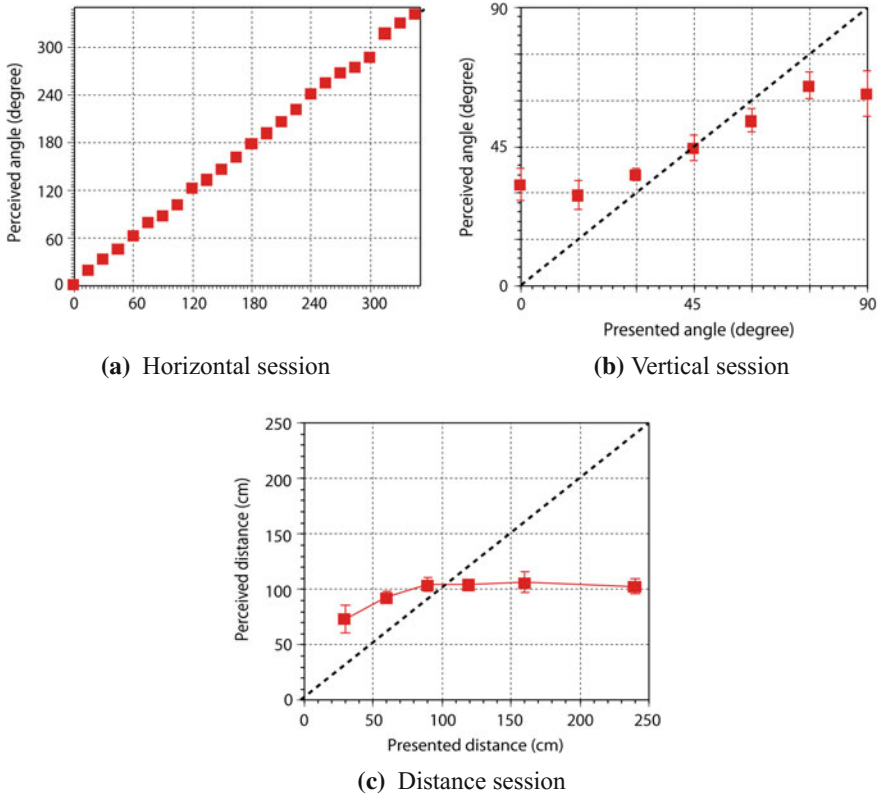
(a) Horizontal session

(b) Vertical session

(c) Distance session

Fig. 2.9 Results of the localization tests [25]

consider that the sound cask can provide reasonable reproduction performance to perceive the distance from the apparent source.

2.4.3 Psychological and Physiological Evaluation of Feeling of Reality in a 3D Sound Field

In order to be confident that others are present in the same space, sound is an essential element. Rather, it has been found that humans have the potential ability to feel the presence from a trace of someone's existence using auditory sensation. This means that it is possible that some physiological phenomenon occurs when you understand the atmosphere of the space. Therefore, three psychological and physiological experiments were carried out focusing on the "presence of the speaker (pronounced body)" changing the experimental conditions of the 3D sound field. From these results, reproducing a 3D sound field with high accuracy can create a feeling of extremely

high reality of the speaker (pronounced body). At the same time, the psychological and physiological method is effective for evaluating the performance of recreating the atmosphere of the space using a 3D sound field system.

2.4.3.1 Autonomic Responses Caused by Acoustic Information of the Speaker's Tiny Movements

Many engineers have created communication systems that enhance the sense of presence or reality to realize natural communication with those in remote locations. However, there are very few methods by which a sense of presence or reality can be quantitatively evaluated. Here, we describe a quantitative evaluation of the acoustic sensation of presence by measuring autonomic responses. We examined the effects of acoustic information about speakers' movements on the sense of presence in personal communication by using a three-dimensional sound field reproduction system based on the boundary surface control principle, by which listeners can experience a highly realistic sensation of speakers. We prepared two types of speech stimuli, "dynamic" and "static." In the dynamic condition, the speakers' speech was recorded along with their subtle unconscious movements. In the static condition, the speech stimulus in the dynamic condition presented from a mouth simulator was recorded to remove any information about the speakers' movements. The sense of the speakers' presence and friendliness was assessed subjectively by the participants. In physiological experiments, we evaluated the autonomic responses by measuring blood volume pulse amplitude and the skin conductance response during the speakers' voice presentation. We found that a higher sense of presence was observed in the dynamic condition than in the static condition, and that the participants expressed greater friendliness towards speakers in the dynamic condition. Moreover, there were differences in the autonomic nervous system activities between the dynamic and the static conditions. These findings suggest that a sense of presence is influenced by acoustic information about speakers' unconscious subtle movements and that the existence or non-existence of speakers' movements can be detected from the autonomic responses [29, 30].

2.4.3.2 Physiological Response Due to the Approach of Moving Sound Sources

By observing time variation of activity of the autonomic nervous system and subjective assessment due to the approach of moving sound sources, we confirmed that the accuracy of 3D sound field reproduction affects the feeling of reality of human presence. In the experimental result, by listening to the approach of moving sound sources with high accuracy of 3D sound field reproduction, the activity of the sympathetic nervous system of the listener increases. Because this phenomenon occurs

in the case of invasion of personal space, it is indicated that personal space can also be invaded by a sound in a virtual space [31].

2.4.3.3   Activity of the Mirror Neuron System Caused by Action-Related Sound

The motor cortical area is often activated in the presence of auditory stimuli in the human brain. In this section, we examine whether the motor area shows differential activation for action-related and non-action-related sounds and whether it is susceptible to the quality of the sounds. A three-dimensional sound field recording and reproduction system based on the boundary surface control principle (BoSC system) was used for this purpose. We measured brain activity while hearing action-related or non-action-related sounds with electroencephalography using μ rhythm suppression (μ-suppression) as an index of motor cortical activation. The results showed that μ-suppression was observed when the participant heard action-related sounds, but it was not evident when hearing non-action-related sounds. Moreover, this suppression was significantly larger in the 3D sound field (BoSC reproduction), which generates a more realistic sound field, than in the single-loudspeaker condition. These results indicate that the motor area is indeed activated for action-related sounds and that its activation is enhanced with a 3D realistic sound field. It is indicated that the mirror neuron system is related to the subjective sense of reality, not only in real space but also in virtual space (Fig. 2.10) [32–34].
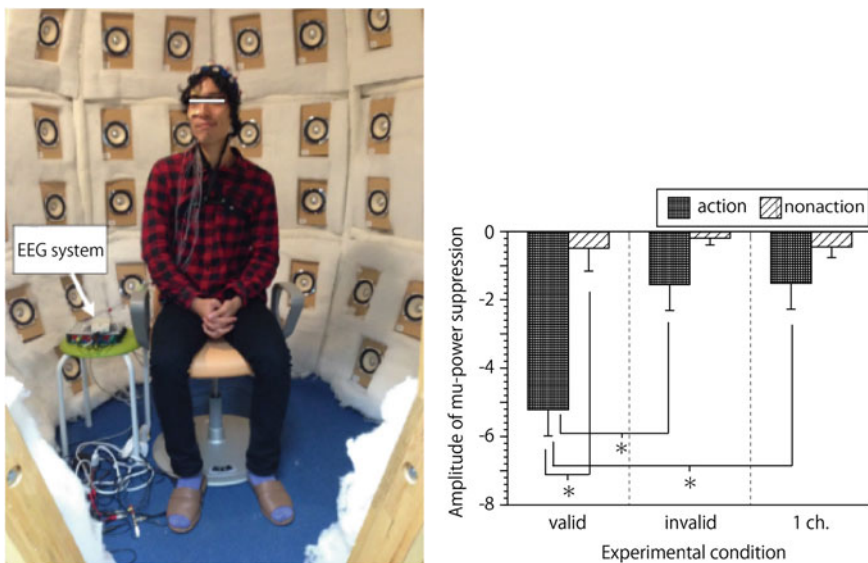

Fig. 2.10 Experimental setup of brain activity in the sound cask and the result of μ-suppression [35]

2.5 Application of the Sound Cask

2.5.1 Sound Field Simulator

As shown in Fig. 2.11, the source signal U generated from a loudspeaker, assumed an instrument of a player on a concert hall stage in the primary field, passes through the transfer functions $[F_j]$. The BoSC microphone is located at the position of the player. The transfer function is given by $[F_j] = [D_j + R_j](\in \mathbf{C}^{1 \times M})$, where $[D_j]$ is the direct sound and $[R_j]$ is the reverberant sound. In the simulated field, as in the primary field, after the source signal U is picked up by a microphone for the musical instrument, the output signal $\widehat{X}$ of the microphone is convolved using the FIR filter $[Q_i](\in \mathbf{C}^{1 \times N})$ in real time. Driving the loudspeakers in the sound cask with the filter output $[S_i](\in \mathbf{C}^{1 \times N})$ reproduces the same sound field as the primary field in the region surrounding the head of the listener (Fig. 2.12). From these conditions, FIR filter $[Q_i]$ is obtained as follows,

$$[Q_i] = \frac{[R_j][G_{ij}]^{-1}}{\widehat{D}_0 + [R_j][G_{ij}]^{-1}[G_{i0}]} e^{j\omega\tau_1}. \tag{2.7}$$

Fig. 2.11 Impulse response measurement in the primary sound field
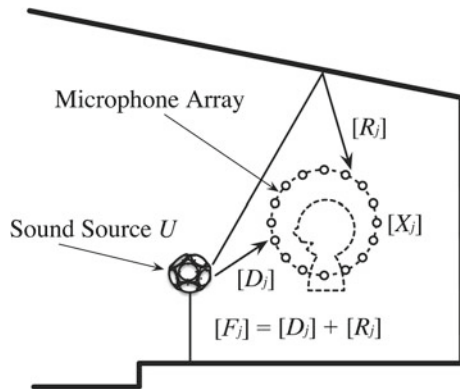
Fig. 2.12 Structure of the sound field simulator system using the sound cask
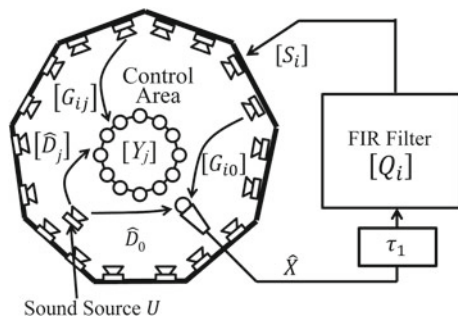
where τ_1 is the delay time of real-time computing of the FIR filter $[Q_i]$. Assuming that the microphone for the musical instrument is located near the sound source, we obtain $\widehat{D}_0 >> [R_j][G_{ij}]^{-1}[G_{i0}]$. Furthermore, the transfer function from the source to the microphone for the musical instrument is just the delay τ_2, i.e., $\widehat{D}_0 = e^{-j\tau_2\omega}$. Then, the FIR filter $[Q_i]$ is expressed as

$$[Q_i] \simeq [R_j][G_{ij}]^{-1} e^{j\omega(\tau_1+\tau_2)} \tag{2.8}$$

where $[G_{ij}](\in \mathbf{C}^{N\times M})$ is the transfer function matrix from i-th loudspeaker in the reproduced sound field to j-th microphone on the surface S'. By considering the causality of the inverse system, we need to assume the delay time τ_h caused by the inverse system. Instead, the reflective sound $[R_j]$ can be shifted τ_r earlier, i.e., $[R'_j] = [R_j]e^{j\omega\tau_r}$. Therefore, actual FIR system $[Q'_i]$ is given as

$$[Q'_i] = [R'_j][H_{ji}] = [R_j][G_{ij}]^{-1} e^{j\omega(\tau_r-\tau_h)} \tag{2.9}$$

where $[H_{ji}](\in \mathbf{C}^{M\times N})$ is the inverse system of the transfer function $[G_{ij}]$ considering the causality mentioned above. $[Q_i] = [Q'_i]$ in Eqs. 2.8 and 2.9 holds when $\tau_1 + \tau_2 = \tau_r - \tau_h$.

2.5.1.1 Experimental Condition

As the primary sound field, the impulse responses in a multi-purpose hall with a reverberation time of 1.5 s are measured, and the reverberant filter is calculated by assuming the starting time of the reflective component of the impulse response to be 25 ms. $[H_{ji}]$ in Eq. 2.9 is designed using the regularization parameter method with an FIR tap length of 4096 and a latency of 2048 points (about 42.7 ms) after truncating the impulse response with 2048 points and converting it into a frequency domain signal by discrete Fourier transformation with 8192 points. Figure 2.13 shows the experimental setup of the sound field simulator. A small omnidirectional microphone (DPA-4060) as the instrument microphone is located on the wall inside the sound cask at a height of 145 cm. A BoSC microphone located at the center in the sound cask at a height of 120 cm is used to obtain the indicator of room acoustics. As a feedback suppression method we adopt the inverse design method so that the sound at the instrumental microphone in the sound cask is canceled [36]. In order to confirm the effectiveness of this method, the performance of the sound field simulator with the feedback canceler is evaluated using parameters of room acoustics.

2.5.2 Experimental Result

From the impulse response measured in the sound field simulator (Fig. 2.13), the reverberation time in the frequencies of each octave band, early reflection energy L_{er}, and late reflection energy L_{rev} were calculated. In these calculations, the direct sound
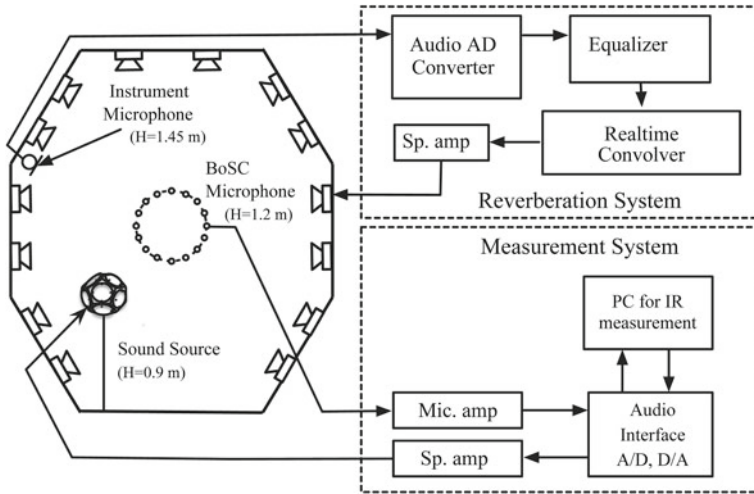
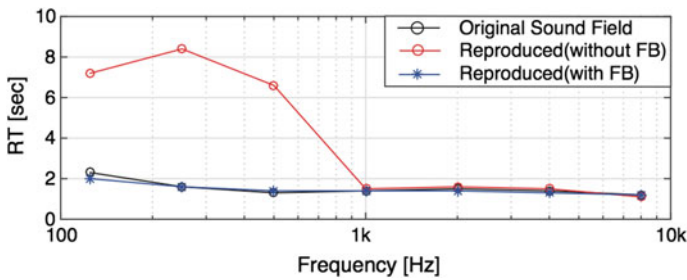**Fig. 2.13** Experimental setup of the sound field simulator

Fig. 2.14 Reverberation time in the frequencies of each octave band

energy of the impulse response is calculated from the energy within the time span of the direct sound arrival time to 10 ms after. The early reflection energy and the late reflection energy are calculated from the proportion of energy of 25 ms to 100 ms and after 100 ms, respectively, to the direct sound energy. The reverberation time is shown in Fig. 2.14. As shown in Fig. 2.14, without the feedback cancellation system, large error can be seen in the frequency range below 500 Hz between the primary sound field and the simulated sound field. However, with the feedback cancellation system, these errors become small. The early reflection energy and the late reflection energy are depicted in Figs. 2.15 and 2.16, respectively. Unlike the case of the reverberation time, large error can be seen in the frequency range below 2 kHz between the primary sound field and the simulated sound field regardless of whether or not feedback cancellation is employed. This is thought to be caused by the assumption that the frequency response of $\hat{D}_0$ is flat in the formulation previously described. Therefore, we correct the frequency characteristic of the instrumental microphone using a one-
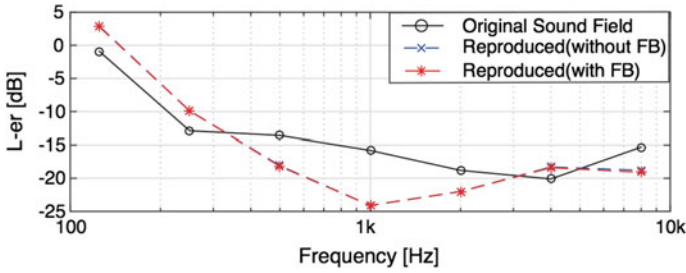
Fig. 2.15 Comparison of the reflection energy (before correction): early reflection energy
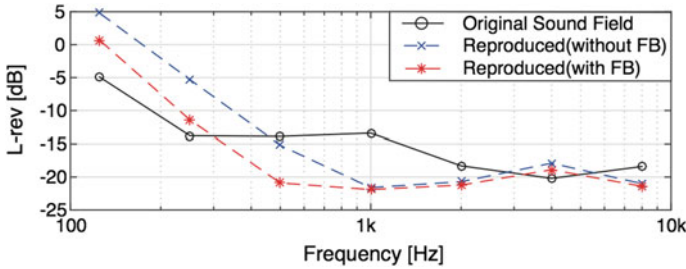

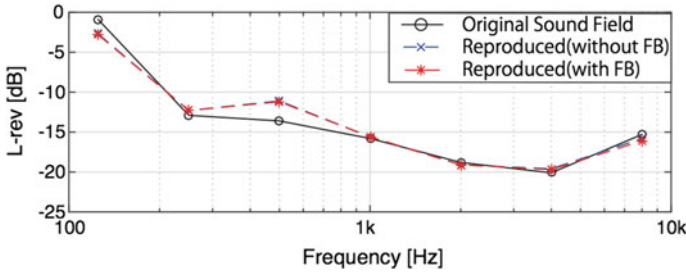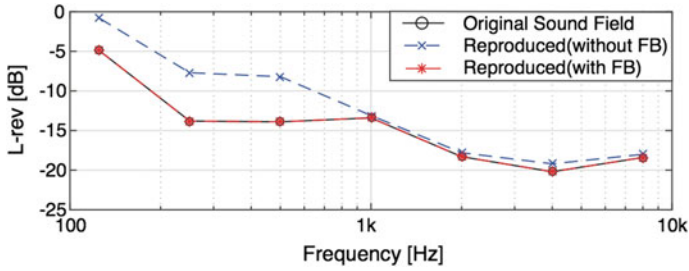**Fig. 2.16** Comparison of the reflection energy (before correction): late reflection energy

Fig. 2.17 Comparison of the reflection energy (after correction): early reflection energy

third-octave band equalizer so that the sum of the early reflection energy and the late reflection energy in the simulated sound field become equal to that in the primary field. The early reflection energy and the late reflection energy corrected in the manner mentioned above are depicted in Figs. 2.17 and 2.18, respectively. We can see that both the early reflection energy and the late reflection energy in the simulated sound field basically correspond to those in the primary field.

As described above, the sound field simulator can be realized by using the immersive auditory display, the sound cask. In the formulation to design the sound field simulator, the frequency characteristic of the instrumental microphone is not flat because of the frequency characteristic inside the sound cask. To improve the per-

Fig. 2.18 Comparison of the reflection energy (after correction): late reflection energy

formance of the sound field simulator, it was found that the equalizer is required to correct the frequency response of the instrument microphone input.

2.5.2.1 Evaluation of Instrument and Musical Performance

In order to reveal the applicability of the sound cask to music players and instrument makers, we investigated the auditory impression of a music performance with flutes in a concert hall using the sound cask regarding whether there is a feeling of near-field sonority and far-field sonority (Fig. 2.19). In the experiment, sounds produced by eight flutes, recorded on the stage and the audience seats of a concert hall, were presented to the participants; they were asked to evaluate both the near-field sonority and the far-field sonority of each flute. In addition, we interviewed them about the definition of near-/far-field sonority and its evaluation points. The results showed that the definition conforms to the general definition in that the clarity of sounds in the audience seats is an important factor while determining the near-/far-field sonority of the instrument. Moreover, the sonority was not similarly evaluated by all participants, whereas several other instruments have been similarly, and highly, evaluated. Acoustic analyses of the sounds produced by flutes showed the possibility that their evaluation was related to the physical characteristics of the sound level and overtone spectrum [37].

2.5.3 Sound Table Tennis

A virtual sound table-tennis system was developed using the sound cask. Sound table tennis is a modified table tennis for visually impaired people, in which players need to roll a ball from one end of the table to the other, instead of hitting the ball over a net. And by using a sound-ball and a special racket, the player may hit the ball by listening for the direction in which the ball is rolling [38] (Fig. 2.20).

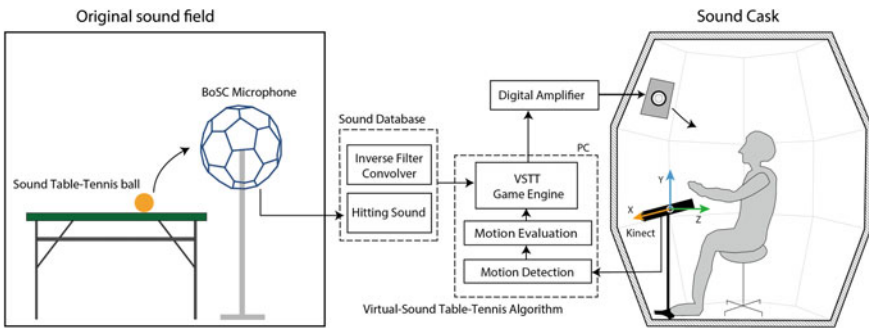**Fig. 2.19** Recording of flute performance and listening test [37]

Fig. 2.20 Sound table tennis system

2.5.4 Sound Field Sharing

As an application of the BoSC system, the concept of a sound field sharing system using more than one system has been introduced. This system is a distant communication system that allows us to communicate with each other as if we were in the same room. A similar sharing system has already been designed and tried as shown in [39]. In this reference, the concept and the practical scheme, and examples of distant field sharing, are shown. However, the methods of sound field capture and re-construction are quite basic; e.g., a 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 are unclear). The advantageous of our proposed system can be summarized as follows:

- capability of capturing directional information with high accuracy by the microphone array;
- theoretically proved reproduction method;
- capability of moving sources;
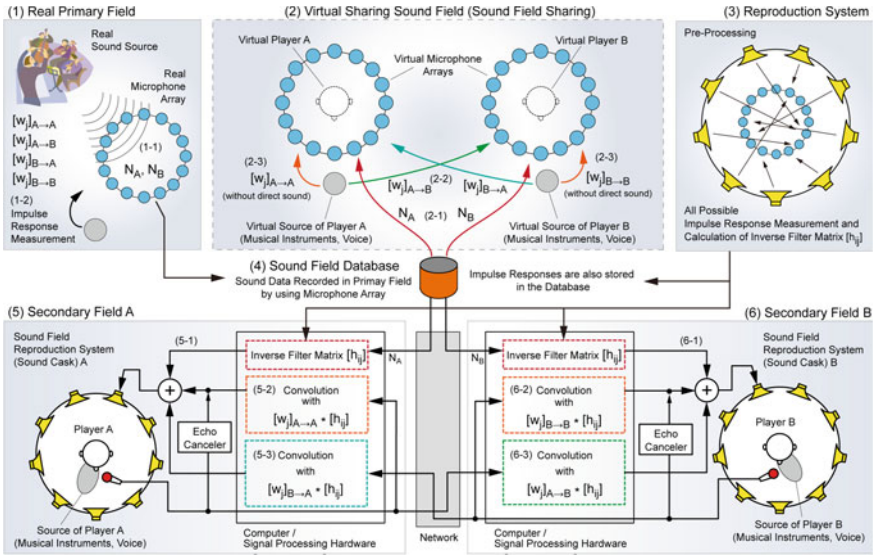- real immersive environment by enclosed space, the cask.

Fig. 2.21 Sound field sharing system

Figure 2.21 shows a conceptual diagram of the sound field sharing system, in which Player A and B share a sense of being in the same primary field. Space (1) in the figure is the existing real primary field (Fig. 2.22) and is the sound field that a subject aims to share. In this primary field, for example, music performances are recorded using a microphone array as (1-1). The array is installed in places in which it is assumed that Player A and B stand and recorded signal N_A or N_B can be played back by a reproducing system. In addition, impulse responses between sound sources and the adjacent microphone array in this real space are measured as indicated (1-2). These are necessary for yielding musical sounds or voices, which can be recognized as if they are being 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 at Player B is indicated as $\left[w_j\right]_{A \to B}$ in the figure. Space (2) in the figure is a virtual sound sharing 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 by using a reproduction system based on the BoSC principle. 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 $\left[w_j\right]_{A \to B}$ in the figure. The reverse process is the same, too, and expressed as $\left[w_j\right]_{B \to A}$ as indicated at (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) $\left[w_j\right]_{A \to A}$

Fig. 2.22 Concert hall measurement for the ensemble experiment

Fig. 2.23 Appearance of the ensemble test with two casks located side by side

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-type microphone array. The impulse responses between all possible combinations of loudspeakers and microphones are measured at 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 dedicated line, on demand. The sound fields, where Players A and B exist, are indicated as Space (5), Secondary Field A; and Space (6), Secondary Field B, respectively. More precisely, these are the inner space of the sound field reproduction system, the sound cask, in this case. 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 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 $[w_j]_{B \to B}$, 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 using the BoSC system with 62-channel loudspeakers was developed [18, 19]. The BoSC system enables perception of the direction of the reproduced voice. The system transmits voice direction in a three-person conversation by changing the transfer functions in accordance with the angle the speaker is facing [20]. Only 24 loudspeakers were used to reproduce voices between the systems to avoid a large amount of calculation [21]. Basic experiments on communication between separate casks with necessary convolutions of signals in real time have been started recently as shown in Fig. 2.23. At the current stage, two casks are located side by side and players inside the casks have tried an ensemble while hearing the other play, which is reproduced by the necessary convolutions ((5-3) and (6-3) in Fig. 2.21). The possibility of a remote ensemble with high presence has been confirmed. For further improvement of performance, continuous experimental examinations are conducted.

2.5.4.1   Latency Reduction

In the case of playing in an orchestra, players on the stage of a concert hall listen to other players' sounds with delay caused by sound propagation. For example, if players are positioned at a distance of 10 m, sound from other players is delayed by 34 ms because of the sound propagation speed of 340 m/s. Therefore, the conductor is required to synchronize and control the orchestra's performance. When there is no conductor, such as in an ensemble performance, a delay of 20 ms is the maximum limit to play music naturally [40]. Thus, this sound field sharing system aims to suppress delay to less than 20 ms. In the sound field sharing system, delay is caused in the telecommunication system, the inverse system, and the audio input/output system. To reduce the latency of the audio input/output system, we developed an FPGA board, which can convolve 96ch data from two MADI input signals with 1071 FIR

coefficients in real time. The impulse response after 1071 points is convolved using a PC and is added to the FPGA results. Then, 96ch impulse responses longer than 4 s can be convolved in real time with almost no latency [41].

2.5.4.2 Feedback Cancellation

In the sound field sharing system, the voice or musical performance is first recorded in one of the sound casks, and it is transmitted and reproduced in others. At the same time, the same recording and reproduction procedure is implemented for one of the other sound casks, providing the listener with a feeling of a shared-sound field. In this case, two types of acoustic feedback occur owing to the installation of microphones inside the sound cask. This feedback causes an echo and leads to instability of the system, thereby degrading the accuracy of the reproduced sound field. In this section, we introduce an acoustic feedback suppression method by manipulating the inverse system design algorithm, in which we introduce an additional control point, called a "null space," where summation of all signals fed from the speakers is equal to zero. Figure 2.24 shows the results of an octave band analysis at the reference microphone using the musical signal. The signal can be suppressed over all ranges of frequency bands. Specifically, the suppression level of the center frequency of 500 Hz is about 30 dB in both measurement signals. However, the suppression level is low in the higher frequency band [36, 42].

2.5.4.3 Ensemble Experiment

In order to confirm the performance of the system, an ensemble experiment of the sound field sharing system using two sound casks was performed. Generous cooperation was given by five ensemble groups (10 players in total). As a result, we found that the time delay of the system caused by transferring and calculating data

Fig. 2.24 Octave analysis of the observed signal at the microphone when using the orchestra as the primary signal. [43]
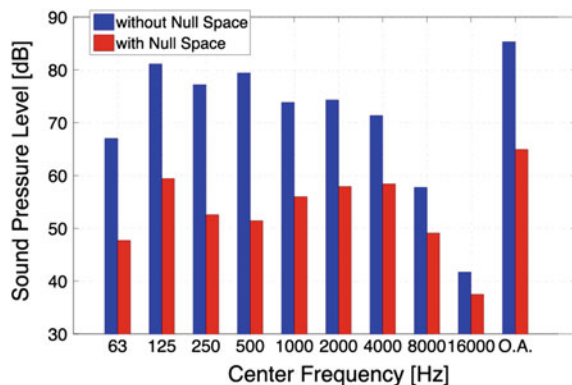
can be entirely ignored, and we received the favorable comment that the sound field sharing system using two sound casks can realize a stage in a virtual concert hall. Furthermore, we received constructive comments concerning another application, for example, "This system can be used for remote music education." On the other hand, it is pointed out that the directivity of musical instruments is nearly imperceptible and the visual monitor should be larger if the system is to be used for music education.

2.6 Further Improvement

We outline two topics for future work. First, we propose a method to include information on the body movement of the source in the primary sound field of the sound sharing system, which can increase the presence of the source (e.g. music players in an ensemble performance or speakers in a conversation) in the receiver's sound field. Second, two other types of sound field reproduction rooms are introduced: an open system and a small system. Long-term perspectives of the immersive auditory display based on BoSC principle are also described focusing on cost reduction, increase in the internal space and enhancement of the sound quality, smaller and lighter equipment and practical application.

2.6.1 Reproduction of Sound Source Directivity

One of the features of the BoSC system is that it allows a listener to move his/her head. In addition, the players and speakers can move their bodies during the communication process as well. It is possible that the minor changes in the sound caused by the speaker's body movements stimulate the sense of the presence of the other party [30]. A change in the sound directivity is specified as one of the physical changes caused by the players' or speakers' movements. In this section, we propose a sound directivity reproduction method that estimates the radiation from a sound source by solving an inverse problem between the secondary sound sources enclosing the sound source and a microphone array outside the secondary sound sources. We demonstrate the effectiveness of this method by simulations and measurements made in an anechoic chamber.

Figure 2.25 shows the concept of the proposed method of sound directivity reproduction. First, we consider the radiation from sound sources inside a three-dimensional volume, V_1, that is bounded by surface S_1. On the basis of the external Helmholtz integral equation (HIE) [44], the sound pressures of volume V_O outside of V_1 are
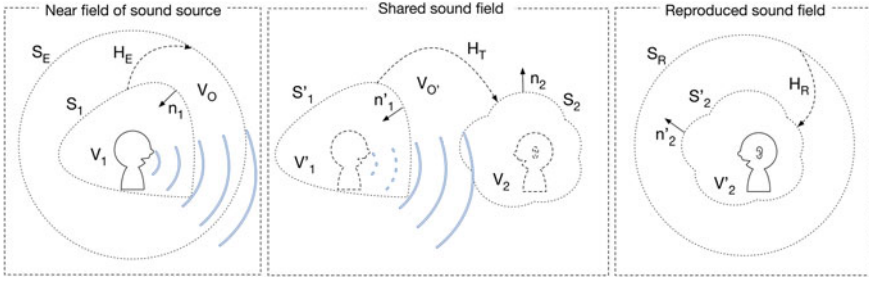
Fig. 2.25 Concept of sound directivity reproduction

$$p\left(r'_1\right) = \iint_{S_1} \left(G\frac{\partial p\left(r_1\right)}{\partial n_1} - p\left(r_1\right)\frac{\partial G}{\partial n_1}\right) dS_1 \tag{2.10}$$

$$\left(r_1 \in S_1, r'_1 \in V_O\right),$$

where G is a Green's function, and n_1 is a normal vector to surface S_1. This equation implies that the radiation of sound sources is expressed by the sound pressures and particle velocities on closed surface S_1.

Now we consider observation surface S_E that is outside of surface S_1. Because $S_E \subset V_O$, the sound pressures on S_E are also given by Eq. 2.10. We discretize surfaces S_1 and S_E into N_1 and M_1 small elements of areas $\Delta S_{1,k}(k = 1, \ldots, N_1)$ and $\Delta S_{E,j}(j = 1, \ldots, M_1)$, respectively. From Eq. 2.10, the sound pressure in area $\Delta S_{E,j}$ is

$$p_E\left(j\right) = \sum_{k=1}^{N_1} \left(G_{j,k}\frac{\partial p_1\left(k\right)}{\partial n_1} - p_1\left(k\right)\frac{\partial G_{j,k}}{\partial n_1}\right) \Delta S_{1,k}, \tag{2.11}$$

where $p_1(k)$ is the sound pressure in $\Delta S_{1,k}$, and $G_{j,k}$ is the Green's function between areas $\Delta S_{E,j}$ and $\Delta S_{1,k}$.

Let $\Delta S_{IN,1,k}$ and $\Delta S_{OUT,1,k}$ be small elements of the areas that are inside and outside of $\Delta S_{1,k}$, respectively, in the direction normal to S_1 and at distance h from its surface. When distance h is short enough, the sound pressures and particle velocities in small area $\Delta S_{1,k}$ are

$$p_1\left(k\right) \cong \frac{p_{IN,1}\left(k\right) + p_{OUT,1}\left(k\right)}{2}, \tag{2.12}$$

$$\frac{\partial p_1\left(k\right)}{\partial n_1} \cong \frac{p_{IN,1}\left(k\right) - p_{OUT,1}\left(k\right)}{2h}, \tag{2.13}$$

where $p_{IN,1}(k)$ and $p_{OUT,1}(k)$ are the sound pressures in $\Delta S_{IN,1,k}$ and $\Delta S_{OUT,1,k}$ respectively.

Inserting Eqs. 2.12 and 2.13 into Eq. 2.11 yields

$$p_{\text{E}}(j) = \frac{1}{2} \sum_{k=1}^{N_1} \left(\left(\frac{G_{j,k}}{h} - \frac{\partial G_{j,k}}{\partial n_1} \right) p_{\text{IN},1}(k) \right.$$
$$\left. - \left(\frac{G_{j,k}}{h} + \frac{\partial G_{j,k}}{\partial n_1} \right) p_{\text{OUT},1}(k) \right) \Delta S_{1,k}. \tag{2.14}$$

Therefore, we obtain a matrix form of Eq. 2.14:

$$\mathbf{p}_{\text{E}} = \mathbf{H}_{\text{E}} \mathbf{p}_1, \tag{2.15}$$

where

$$\mathbf{p}_1 = \left[p_{\text{IN},1}(1), \ldots, p_{\text{IN},1}(N_1), \right.$$
$$\left. p_{\text{OUT},1}(1), \ldots, p_{\text{OUT},1}(N_1) \right]^T,$$
$$\mathbf{H}_{\text{E}} = \frac{1}{2} \mathbf{GS},$$
$$\mathbf{G} = [\mathbf{G}_1 \quad \mathbf{G}_2], \quad \mathbf{S} = \begin{pmatrix} \mathbf{S}_d & 0 \\ 0 & \mathbf{S}_d \end{pmatrix},$$
$$\mathbf{G}_1(j,k) = \frac{G_{j,k}}{h} - \frac{\partial G_{j,k}}{\partial n_1}, \mathbf{G}_2(j,k) = \frac{G_{j,k}}{h} + \frac{\partial G_{j,k}}{\partial n_1},$$
$$(j = 1, \ldots, M_1, \quad k = 1, \ldots, N_1).$$

Here, $\mathbf{p}_{\text{E}}$ is the column vector of the sound pressures in all small areas $\Delta S_{\text{E},j}$, $\mathbf{H}_{\text{E}}$ is an $M_1 \times 2N_1$ matrix, $\mathbf{S}_d$ is a diagonal matrix $diag(\Delta S_{1,1}, \ldots, \Delta S_{1,N})$, and $[\cdot]^T$ denotes the transpose.

According to Eq. 2.15, the sound pressure vector of surface S_1 is represented by the following Eq. using the inverse matrix of $\mathbf{H}_{\text{E}}$:

$$\mathbf{p}_1 = \mathbf{H}_{\text{E}}^{-1} \mathbf{p}_{\text{E}}. \tag{2.16}$$

Equation 2.16 implies that we can obtain the sound pressures and particle velocities on surface S_1 from the sound pressures on surface S_{E} by solving the inverse problem. From Eq. 2.11, we also find that the radiation from the sound source is obtained through Eq. 2.15.

Next, we consider the reproduction of the sound source radiation in a shared sound field. Let V_1' and S_1' be the volume and surface in the shared sound field that are congruent with V_1 and S_1, respectively. On the basis of the external HIE, the sound pressures in a volume $V_{O'}$ which is outside of V_1' are given by an equation similar to Eq. 2.10 using the sound pressures and particle velocities on surface S_1'. Considering the congruency, we find that the radiation from the sound source in V_1 is reproduced in $V_{O'}$ when the sound pressures and particle velocities on S_1' correspond

to those on S_1: $\mathbf{p}_1 = \mathbf{p}'_1$, where $\mathbf{p}'_1$ is the column vector of the sound pressures on S'_1. That is, when this Eq. is satisfied, there is a virtual sound source in V'_1 in the shared sound field.

In the shared sound field, we consider volume V_2 where a virtual listener is located and which is bounded by surface S_2. In the reproduced sound field, we also consider a volume V'_2 and surface S'_2 that are congruent with V_2 and S_2, respectively. On the basis of BoSC, the sound pressures in V'_2 correspond to those in V_2 when the sound pressures on S'_2 are matched with those on S_2 using secondary sources located on a surface S_R outside of S'_2: $\mathbf{p}_2 = \mathbf{p}'_2$ where $\mathbf{p}_2$ and $\mathbf{p}'_2$ are the vectors of the sound pressure in the small areas obtained by discretizing surfaces S_2 and S'_2 into N_2, respectively.

The relationship between the sound pressures at the secondary sources and those on surface S'_2 is

$$\mathbf{p}'_2 = \mathbf{H}_R \mathbf{p}_R, \tag{2.17}$$

where $\mathbf{p}_R$ is the column vector of the sound pressures in the small areas obtained by discretizing surface S_R into M_2, and $\mathbf{H}_R$ is an $N_2 \times M_2$ matrix corresponding to the transfer matrix between the two surfaces.

Finally, we consider the relationship between the sound pressures on surfaces S'_1 and S_2 to reproduce the virtual sound source for the virtual listener. This relationship can be derived in the same way as Eq. 2.15, and the column vector of the sound pressures on surface S_2 is

$$\mathbf{p}_2 = \mathbf{H}_T \mathbf{p}'_1, \tag{2.18}$$

where $\mathbf{H}_T$ is an $N_1 \times N_2$ matrix.

Therefore, from Eqs. 2.16–2.18, we can obtain

$$\mathbf{p}_R = \mathbf{H}_R^{-1} \mathbf{H}_T \mathbf{H}_E^{-1} \mathbf{p}_E. \tag{2.19}$$

That is, when we control the sound pressures at the secondary sources on surface S_R to satisfy Eq. 2.19, the radiation of the sound source in volume V_1 is reproduced in volume V'_1 and then reproduced in volume V'_2 after propagating in the shared sound field.

Well-known methods that trace the sound radiation back by solving the inverse problem are acoustical holography [45] and near-field acoustical holography [44]. In this paper, an inverse problem that traces back to the sound source is applied to a telecommunication system using an immersive sound reproduction system based on BoSC. To apply it to a telecommunication system, we install secondary sources between the measurement surface and the sound source and solve the inverse problem using the measured impulse response matrix. Note that solving the inverse problem makes it easier to remove the effect of the characteristic features of the loudspeakers, microphones, and room acoustics.

We consider a telecommunication system based on the concept described in the preceding section. Figure 2.26 shows the system, which reproduces a speaker's or player's original sound directivity into the other party's system.
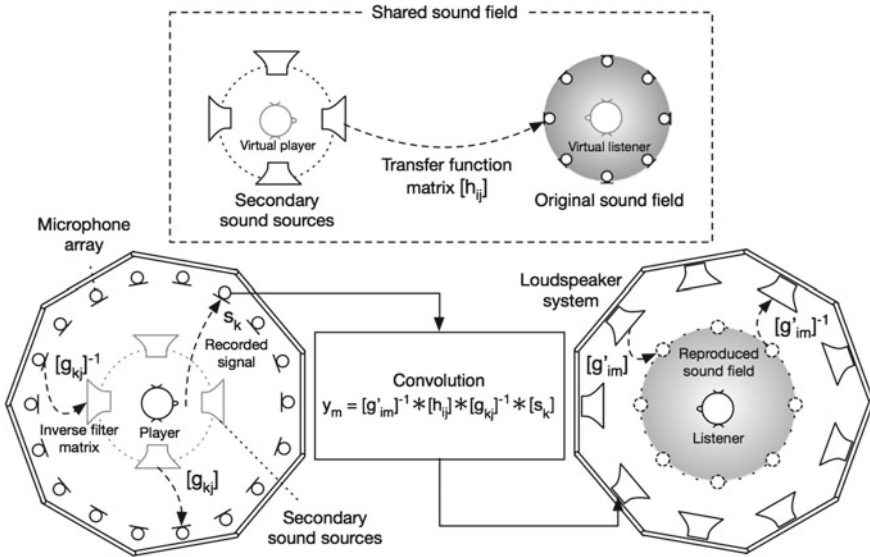
Fig. 2.26 Sound directivity reproduction system for use in telecommunication

The sounds produced by the player are recorded with a microphone array that is installed so as to enclose the player. Let the signal recorded by the k-th microphone of the array be $s_k (k = 1, \ldots, N_1)$ in the time domain. First, signals $[s_k]$ are convolved with inverse matrix $[g_{kj}]^{-1}$ to estimate the radiation from the sound source. This inverse matrix is derived from an impulse response matrix $[g_{kj}](j = 1, \ldots, M_1)$, which is measured using the microphone array, and M_1 secondary sound sources, which are placed so as to enclose the position of the original sound source. By this convolution, we obtain the signals of secondary sources to reproduce the radiation of the sound source.

Next, in the shared sound field, a loudspeaker array and a microphone array for BoSC sound reproduction are installed so as to enclose the position of a virtual player and a virtual listener, respectively. We measure impulse response matrix $[h_{ij}](i = 1, \ldots, N_2)$ from the loudspeaker array to the microphone array. On the basis of BoSC, we reproduce the sound field at the virtual listener into the area where a real listener is. Therefore, the reproduction of the sound field using the inverse system $[g'_{im}]^{-1}$ requires loudspeaker system signals $y_m (m = 1, \ldots, M_2)$ in the reproduction area to form the following equation:

$$y_m = [g'_{im}]^{-1} * [h_{ij}] * [g_{kj}]^{-1} * [s_k]. \tag{2.20}$$

Considering that the system is time invariant, and that secondary sources for the reproduction of the sound source's radiation are used only to measure impulse responses, the loudspeaker array can be replaced by a single loudspeaker that is moved to achieve

the same result. The configuration of the secondary sources must be determined so as to control the sound pressures and particle velocities on the specific closed surface.

2.6.2 Other Systems for Sound Field Reproduction

2.6.2.1 Open System

To reproduce the spatial impression including feeling of distance with a high degree of accuracy, an enclosed environment for single user is required. However, an open environment for multiple users will be convenient in a sound field requiring lower accuracy, such as in a sound-surrounding environment. So, an open system for sound field reproduction based on the BoSC principle was developed. Figure 2.27 shows an octagonal-shaped room with a height of 1.8 m, eight walls (opposite-side distance is 3 m), with each wall consisting of six cuboid enclosures (1.5 m × 0.15 m × 0.15 m) stacked vertically. Loudspeaker unit FOSTEX FX120 is mounted on an enclosure filled with poly-wool. Though listeners can be in the area wider than the sound cask, the actual sweet spot of the listening area is the same size as the sound cask because, theoretically, the sound reproduction area is limited to inside of the BoSC microphone based on the boundary surface control principle. Compared with the sound cask, the open system has high portability and users can listen to the sound without an oppressive feeling.

Fig. 2.27 Open-type BoSC system

Fig. 2.28 Small-type BoSC
system
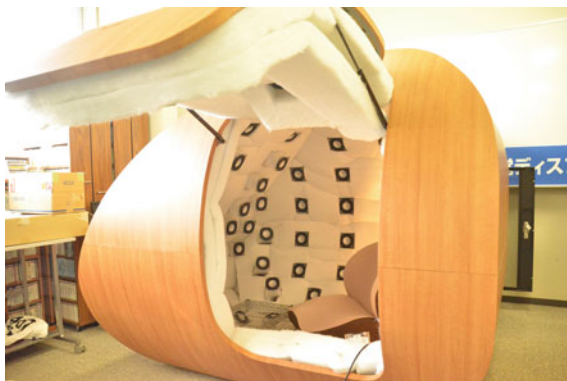
2.6.2.2 Small System

The sound cask is designed so that musicians can play instruments inside it. If the
purpose is just to listen to the sound, a more compact BoSC reproduction room can
be built. Therefore, we designed the small-size BoSC reproduction room shown in
Fig. 2.28. Some people say that the small-size BoSC reproduction room can create a
more composed space than the "Sound Cask".

2.6.3 Long-Term Perspectives

The immersive auditory display developed in this research constitutes a new medium
that expands the interface between humans and information technology. That is, it
represents a new means of using sound, which is a classical human communication
method, and also of creating content with new value in the information society.
Developing the technology to record and reproduce 3D sound has been a dream in
the field of acoustics since the invention of the loudspeaker about a hundred years
ago, far before the advent of "Virtual Reality". Recently, multichannel digital signal
processing technology finally broke through. As a next step, we can consider the
following improvements:

1. Cost reduction,
2. Increase in the internal space and enhancement of the sound quality,
3. Smaller and lighter equipment,
4. Practical application of the sound cask.

2.6.3.1 Cost Reduction

In this project, we used a C80-shaped fullerene microphone array, which consists of 80 omnidirectional microphones, for the BoSC recording system. Because each microphone costs tens of thousands of yen, the total cost is very high. On the other hand, the acoustic performance of the MEMS microphones used in smartphones is increasing year by year. If MEMS microphones could be used in the BoSC recording system, a significant cost reduction could be achieved. Regarding the sound field reproduction system, costs could be cut by lowering the performance of the loudspeaker unit and also by using a rectangular parallelepiped outer shape (like commercially available soundproof chambers) rather than a cask shape. However, in both cases, it is necessary to take into account that designing a stable inverse system would become difficult.

2.6.3.2 Increase in the Internal Space and Enhancement of the Sound Quality

By increasing the size of the microphone array while keeping the density of microphones constant, it is possible to enlarge the region where the sound field is reproduced, for example, to surround multiple listeners. Conversely, by increasing the density of microphones while keeping the size of the microphone array constant, the frequency range of the reproduced sound field can be extended, which will enhance the sound quality. Since the number of microphones increases in both cases, it would be necessary to increase the number of loudspeakers in the reproduction system in order to design a stable inverse system.

2.6.3.3 Smaller and Lighter Equipment

Regarding the size of the recording system, connecting 80 microphones to an 80-channel audio recorder requires an enormous amount of wiring, so there is much room for improvement. Its dimensions could be reduced if, instead of transmitting the microphone output as a weak electric analog signal using a huge cable, we convert it to a digital signal and transmit it through an optical cable. As for the sound field reproduction system, its size and weight could be reduced by using a smaller and lighter loudspeaker unit. In this case, however, it is necessary to consider the associated shortage of low frequencies—although this may be compensated for by drastically increasing the number of loudspeakers. Since the performance of electromagnetic loudspeakers built in smartphones has significantly improved, it is worth to try to develop a more reasonably sized immersive auditory display using small and light loudspeaker units.

2.6.3.4   Practical Application of the Sound Cask

Using the sound cask, it is possible to practically realize an acoustic virtual space, such as a high-end audio room precisely tuned by an audio expert, a music practice room where players can confirm their performance at the listener's position in a concert hall or a church, or a recording studio for 22.2ch audio where recording engineers can confirm the 22.2ch sound created by their mix-down signal.

2.7   Conclusion

The newly developed immersive sound field reproduction system, named "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 the basic test indicate that the sound cask has excellent performance for localization of reproduced sound sources in the horizontal plane. However, there is room for improvement in the performance of the vertical direction and recognition at distance. Further examinations for improvement, e.g., with various types of inverse filter, are currently being conducted. At the present stage of the project, four casks have been constructed at Kyushu, Kyoto, and Tokyo in Japan. Several hundred people have experienced the performance of sound field reproduction in the sound cask. A series of psychological experiments is shown. Including recordings in concert halls, many types of content such as environmental sound outdoors are being continuously accumulated and stored in a database for reproduction [46]. In order to make the sound cask commercially viable, the principal members of this project started a company called Cask Acoustics Co. Ltd. We are currently focusing on sound engineers for 3D sound creation, musicians for education and audiophiles for listening to high-end audio as target users of the sound cask.

References

1. ITU-R BS.775.1, Multichannel stereophonic sound system with and without accompanying picture, 1992–1994
2. SMPTE 2036-2-2008, Ultra high definition television–Audio characteristics and audio channel mapping for program production, 2008
3. A.J. Berkhout, D. de Vries, P. Vogel, Acoustic control by wave field synthesis. J. Acoust. Soc. Am. **93**(5), 2764–2778 (1993)
4. P.A. Gauthier, A. Berry, Adaptive wave field synthesis with independent radiation mode control for active sound field reproduction: Theory. J. Acoust. Soc. Am. **119**(5), 2721–2737 (2006)
5. P.A. Gauthier, A. Berry, Adaptive wave field synthesis for sound field reproduction: theory, experiments and future perspectives. J. Audio Eng. Soc. **55**(12), 1107–1124 (2007)
6. P.A. Gauthier, A. Berry, Adaptive wave field synthesis for active sound field reproduction: experimental results. J. Acoust. Soc. Am. **123**(4), 1991–2002 (2008)

7. G. Theile, H. Wittek, Wave field synthesis: a promising spatial audio rendering concept. Acoust. Sci. Tech. **25**(6), 393–399 (2004)
8. K. Ueno, K. Yasuda, H. Tachibana, T. Ono, Sound field simulation for stage acoustics using 6-channel system. Acoust. Sci. Tech. **22**(4), 307–309 (2001)
9. S. Yokoyama, K. Ueno, S. Sakamoto, H. Tachibana, 6-channel recording/reproduction system for 3-dimensional auralization of sound fields. Acoust. Sci. Tech. **23**(2), 93–103 (2002)
10. D.H. Cooper, T. Shiga, Discrete-matrix multichannel stereo. J. Audio Eng. Soc. **20**(5), 346–360 (1972)
11. M.A. Gerzon, Hierarchical system of surround sound transmission for HDTV, in *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–(in Japanese), in *Proc. of Mtg. of Acoust. Soc. Japan* (1993), pp. 479–480
14. S. Ise, A principle of active control of sound based on the Kirchhoff-Helmholtz integral equation and the inverse system theory (in Japanese). J. Acoust. Soc. Jpn. **53**(9), 706–713 (1997)
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. Toyodaet, S. Enomoto, S. Nakamura, An attempt of sound field sharing system for profound communication–concept and basic stance of the project–(in Japanese), in *Proc. Mtg. Acoust. Soc. Japan* (2007), pp. 585–586
17. S. Enomoto, Y. Ikeda, S. Ise, S. Nakamura, Three dimensional sound field reproduction and recording system based on boundary surface control principle, in *ICAD 2008—14th International Conference on Auditory Displays* (June 2008)
18. S. Enomoto, 3D sound field recording/reproduction system for telecommunication (in Japanese). Arch. Acoust. Noise Cont. **38**(4), 37–42 (2009)
19. S. Enomoto, Y. Ikeda, S. Ise, S. Nakamura, 3D sound field reproduction system for the sound field shared communication based on the boundary surface control principle (in Japanese), in *Proc. Mtg. Acoust. Soc. Japan* (2009), pp. 1411–1414
20. Y. Ikeda, S. Enomoto, S. Ise, S. Nakamura, Three-party sound field sharing system based on the boundary surface control principle, in *ICA 2010—20th International Congress on Acoustics* (2010)
21. S. Enomoto, Y. Ikeda, S. Ise, S. Nakamura, Optimization of loudspeaker and microphone configurations for sound reproduction system based on boundary surface control principle, in *ICA 2010—20th International Congress on Acoustics* (August 2010)
22. R. Kleinman, G. Roach, Boundary integral equations for the three dimensional Helmholtz equation. SIAM Rev. **16**, 214–236 (1974)
23. S. Enomoto, Measurements of synthesized sound eld in vertical plane for sound reproduction system based on boundary surface control (in Japanese), in *Proc. of Mtg. of Acoust. Soc. Japan* (2014), pp. 841–842
24. S.S. Stevens, Problems and method of psychophysics. Psychol. Bull. **55**, 177–196 (1958)
25. M. Yamashita, H. Nakajima, M. Kobayashi, Y. Ikeda, S. Enomoto, K. Ueno, S. Ise, Sound localization experiment of "sound cask"–Comparison between 62-channel and 96-channel sound field reproduction system based on the BoSC principle (in Japanese), in *Proc. of Mtg. of Acoust. Soc. Japan* (March 2013)
26. J.C. Makous, J.C. Middlebrooks, Two-dimensional sound localization by human listeners. J. Acoust. Soc. Am. **87**(5), 2188–200 (1990)
27. P. Zahorik, Assessing auditory distance perception using virtual acoustics. J. Acoust. Soc. Am. **111**(4), 1832–1846 (1990)
28. D.S. Brungart, N.I. Durlach, W.M. Rabinowitz, Auditory localization of nearby source. II. Localization of a broadband source. J. Acoust. Soc. Am. **106**(4), 1956–1968 (1999)
29. M. Kobayashi, K. Ueno, M. Yamashita, S. Ise, S. Enomoto, Subjective evaluation of a virtual acoustic system: trials with three-dimensional sound field reproduced by the 'Sound Cask', in *ICA 2013—21st International Congress on Acoustics* (June 2013)

30. M. Kobayashi, Y. Ooishi, S. Enomoto, N. Kitagawa, K. Ueno, S. Ise, M. Kashino, Effects of the movement of speakers on the sense of presence Study by using the three-dimensional sound field reproduction system. in *Proceedings of the Auditory Research Meeting sponsored by the Technical Committee of Psychological and Physiological Acoustics*, vol. 42, (Feburary 2012), pp. 41–46

31. M. Kobayashi, K. Ueno, S. Ise, The effects of spatialized sounds on the sense of presence in auditory virtual environments: a psychological and physiological study. Presence **24**(2), 163–174 (2015)

32. K. Tsuchida, K. Ueno, S. Shimada, Motor area activity for action-related and nonaction-related sounds in a three-dimensional sound field reproduction system. NeuroReport **26**(5), 291–295 (2015)

33. K. Tsuchida, K. Ueno, S. Shimada, Brain activity to action- and non-action-related sounds in a three-dimensional sound reproduction system. in *JSST 2013 International Conference on Simulation Technology (JSST2013)* (2013), pp. 45

34. K. Tsuchida, K. Ueno, S. Shimada, Modulation of EEG mu and beta rhythm in the mirror neuron system to action-related sounds in a three-dimensional sound reproduction system, in *44th Annual Meeting of the Society for Neuroscience (Neuroscience2014)* (2014)

35. M. Kobayashi, K. Tsuichida, K. Ueno, S. Shimada, Measurement of mirror neuron system activation for the 3-dimentional reproduction sound fields (in Japanese). Trans. Virtual Real. Soc. Jpn. **21**(1), 73–79 (2016)

36. S. Kohno, Y. Watanabe, H. Yoshida, Y. Ikeda, S. Ise. Acoustic feedback cancellation for a sound-field simulation system based on boundary-surface control principle (in Japanese), in *Proc. of Mtg. of Acoust. Soc. Japan* (March 2015), pp. 653–654

37. M. Kobayashi, S. Tamura, K. Ueno, Auditory impression of flutes at a concert hall (in Japanese), in *Proc. of Mtg. of Acoust. Soc. Japan* (September 2013), pp. 889–890

38. A. Nagai, S. Numakami, Y. Ikeda, Y. Watanabe, S. Ise, K. Ueno, Development of virtual table tennis system between the two persons for sound cask (in Japanese), in *Proc. of Mtg. of Acoust. Soc. Japan* (March 2016)

39. W. Woszczyk, J. Cooperstock, J. Roston, W.L. Martens, Shake, rattle, and roll: getting immersed in multisensory, interactive music via broadband networks. J. Audio Eng. Soc. (AES) **53**(4), 336–344 (2005)

40. T. Nagao, T. Watanabe, Y. Ikeda, K. Ueno, S. Ise, Study on the effect of sound delay conditions on ensemble performance (in Japanese), In *Proc. of Mtg. of Acoust. Soc. Japan* (March 2012), pp. 997–998

41. H. Yoshida, Y. Kitagawa, Y. Watanabe, S. Ise, Implementation of low-latency convolver using FPGA and its application for sound-eld sharing system (in Japanese), in *Proc. of Mtg. of Acoust. Soc. Japan* (September 2015), pp. 515–518

42. Y. Watanabe, H. Yoshida, S. Kohno, Y. Ikeda, S. Ise, Development of a sound-field simulation system and its evaluation on room acoustics—Examination of acoustic feedback cancellation (in Japanese), in *Proc. of Mtg. of Acoust. Soc. Japan* (March 2015), pp. 1061–1062

43. H. Aso, K. Ueno, M. Takahashi, M. Kobayashi, Examination on performance of the sound filed sharing system based on boundary-surface control principle,–Subjective evaluation by performing musicians–(in Japanese), in *Proc. of Mtg. of Acoust. Soc* (September 2016). (Japan (in Press))

44. E.G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic Press, London, 1999)

45. B.P. Hildebrand, B.B. Brenden, *An Introduction to Acoustical Holography* (Plenum Press, New York, 1974)

46. A. Omoto, S. Ise, Y. Ikeda, K. Ueno, S. Enomoto, M. Kobayashi, Sound field reproduction and sharing system based on the boundary surface control principle. Acoust. Sci. Technol. **36**(1), 1–11 (2015)