PAPER

Real Environment Performance of Recording System for Personal 3D Sound Field Reproduction Using Hyperdirectional Microphones and Wave Front Synthesis

Toshiyuki KIMURA[†], ***Member***

SUMMARY This study presents a recording system utilizing an array of eight hyperdirectional microphones designed for personal three-dimensional (3D) sound field reproduction via wave front synthesis. The recording positions of the hyperdirectional microphones were identified through impulse response measurement, enabling microphone array construction. To evaluate the localization performance of the constructed microphone array, the impulse responses were measured, replay sounds were synthesized, and the localization experiment was performed. Results demonstrated that the developed recording system outperformed ambisonic microphone, a standard conventional 3D sound field recording , in localization accuracy.

key words: *personal 3D sound field reproduction, hyperdirectional microphone, wave front synthesis, localization test, ambisonic microphone*

1. Introduction

In recent years, three-dimensional (3D) sound field reproduction technologies have significantly progressed. When applied to remote operation systems, such technologies can enhance effciency and situational awareness by creating a "realistic sensation" that enables operators to experience remote environments with enhanced perceptual accuracy.

To realize the remote operation systems, 3D sound field reproduction technologies satisfy the following technical requirements:

1. Accurate 3D sound field representation;
2. Minimum number of transmission channels;
3. Limited acoustic signal processing delays to enable interaction through operation, like in telexistence (the remote operation of a robot) [1], where delays should not exceed 74 ms.

Two widely studied 3D sound field reproduction technologies are binaural [2] and transaural [3] technologies. These technologies involve the use of headphones positioned at the ears of a dummy head (representing a operator in a remote location) to transmit the recorded acoustic signals. While requiring only two transmission channels, which is advantageous, the use of headphones may lead to intra-head localization or front-back errors if the dummy's head shape does not match the operator's. Transaural technology avoids intra-head localization by replaying sounds through loudspeakers. However, it needs to convolute the recorded acoustical signal to the inverse filter that cancels the acoustical path between the loudspeaker and the left and right ear positions of operator's head. This process introduces a time delay equivalent to one-quarter of the reverberation time in the room where the operator is located to completely cancel the acoustical path [4]. Thus, The inverse filter may introduce excessive delay in cases where the reverberation time is about 300 ms (equivalent to the reverberation time in a normal conference room). Moreover, both technologies require sensors to detect the orientation of the operator's head and dummy head rotation equipments for synchronizing the dummy and operator head orientations [5].

Another 3D sound field reproduction approach is ambisonics [6], which uses an array (called an "ambisonic microphone") of four unidirectional microphones to record a 3D sound field. The sound field is replayed using a loudspeaker array after processing with matrix calculation. Although this technology allows free head movement of the listeners within the loudspeaker array, the localization accuracy of the 3D sound field is poor when only audio information is presented.

It is possible to apply surround recording techniques [7]. However, the currently proposed microphone arrangement in surround recording technology assumes a two-dimensional (2D) plane for the 5.1ch system [8]. Thus, it is not suitable for conveying acoustic information captured during remote operation (sounds appear to originate from a specific direction in 3D space). Although surround recording technologies that address a 3D space (22.2ch system [9], higher order ambisonics [10], and boundary surface control [11]) have been proposed, they require large number of microphones to accurately express acoustic information, thereby increasing the number of transmission channels.

On the other hand, wave front synthesis technology using directional microphones [12]–[15] has been proposed. This technology reproduces a 3D sound field such that the wave front of the control area is accurately reproduced in the listening area, leveraging the Kirchhoff-Helmholtz integral equation [16].

Camras [12] placed twelve unidirectional microphones on the boundary surfaces of a control area in a highly reverberant environment (e.g., concert hall). The sound field was then reproduced using loudspeakers in relatively less reverberant spaces (e.g., conference room and outdoors). This gave

Manuscript received September 25, 2024.
Manuscript revised January 20, 2025.

[†]Faculty of Informatics, Tohoku Gakuin University, Sendai-shi, 984-8588 Japan.

DOI: 10.1587/transfun.E98.A.1

listeners an impression of being in a concert hall, even though they were in a relatively less reverberant space. However, the theoretical justification for using unidirectional microphones was not addressed.

Berkhout et al. [13] synthesized a wavefront by placing an omnidirectional microphone on the boundary surface, using the second-kind Rayleigh integral an approximation of the Kirchhoff-Helmholtz integral equation. However, because the boundary surface is limited to an infinite plane, the acoustic information for the remote operation cannot be presented.

Kimura and Kaheki [14] showed that a wavefront can be synthesized in a control area with no shape restrictions by placing unidirectional or hyperdirectional microphones on the boundary surface, using the Fresnel-Kirchhoff diffraction formula another approximation of the Kirchhoff-Helmholtz integral equation. However, as this study was based on a computer simulation, 630 microphones were required for a circular control area with a radius of 2 m, and 800 for a square-control area with 4-m sides.

Kimura et al. [15] showed that even when the wave front is reproduced only up to about 300-400 Hz based on the spatial sampling theorem in the technique described by Kimura and Kaheki [14], there is no perceptual difference even if the number of microphones is increased beyond that. For a circular control area with a radius of 2 m, the number of microphones was reduced to 24.

To further reduce the number of transmission channels, a personal 3D sound field reproduction technique in which the listening area is limited to the vicinity of the listener's head has been developed [17]. The configuration of the technique is shown in Fig. 1. First, in the original sound field, eight hyperdirectional microphones are placed at the vertices of a cubic control area for recording sounds (shown on the left-hand side of Fig. 1). The hyperdirectional microphones are then directed outward of the control area. Second, in the reproduced sound field, the recorded sound is replayed using eight loudspeakers (shown on the right-hand side of Fig. 1). Each loudspeaker is placed at the same position as each hyperdirectional microphone. The cubic loudspeaker array synthesizes the wave fronts within the control area, allowing listeners to perceive accurate 3D sound movements, as shown on the right-hand side of Fig. 1. For instance, when a sound moves above the microphone array, the listener perceives the same movement above their head within the loudspeaker array. The listener's field of vision in the horizontal direction is not blocked by loudspeakers. In future, it will be possible to build the remote control systems combined with video because the operator's line of sight is often horizontal in a remote operation. The localization experiments were performed in order to evaluate the performance of this technique. The results showed that the localized performance was suffcient for building a system if a cubic array of hyperdirectional microphones are used with sides 0.4 m.

An analysis is conducted to determine whether the personal 3D sound field reproduction technique satisfies the three technical requirements for remote operations. In this technique, the wave front is synthesized in the frequency domain up to 425 Hz (= $\frac{340}{0.4\times2}$) based on the spatial sampling theorem, as the microphone interval is 0.4 m. Only eight transmission channels are required. Although this number is higher compared to binaural and transaural systems (two), and ambisonics (four), it is small enough to be practical. This technique allows free head movement within the listening area. In addition, the technique does require any acoustic signal processing like inverse filters. Therefore, this technique satisfies the three technical requirements of a remote operating system.

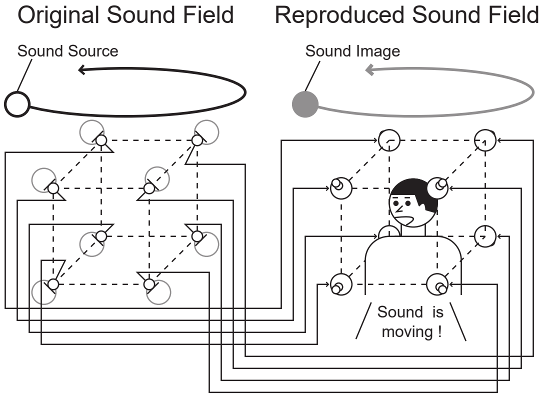

Fig. 1 Basic configuration of personal 3D sound field reproduction [17].

This study develops a recording system for personal 3D sound field reproduction using an array of eight hyperdirectional microphones. Section 2 describes the construction of the microphone array after identification of the recording position of the hyperdirectional microphones.

Section 3 evaluates the localization performance of the constructed microphone array. Impulse responses are measured, replay sounds are synthesized, and a localization experiment is performed. The result of the localization experiment are compared with those of ambisonic microphones - a conventional 3D sound field recording technique.

2. Recording system

In Kimura's study [17], sounds replayed from the eight loudspeakers during the localization experiment were synthesized on a computer rather than recorded in real environment. Therefore, to evaluate the performance of the personal 3D sound field reproduction technique in a real environment setting, a microphone array using eight hyperdirectional microphones is constructed.

2.1 Construction of microphone array

The constructed microphone array is shown in Fig. 2. To build a remote operation system that includes video, the array frame for holding the hyperdirectional microphones is supported by four plates positioned on the sides such that the pillars are not placed in front of and behind the horizontal plane. This design also leaves the center of the array open

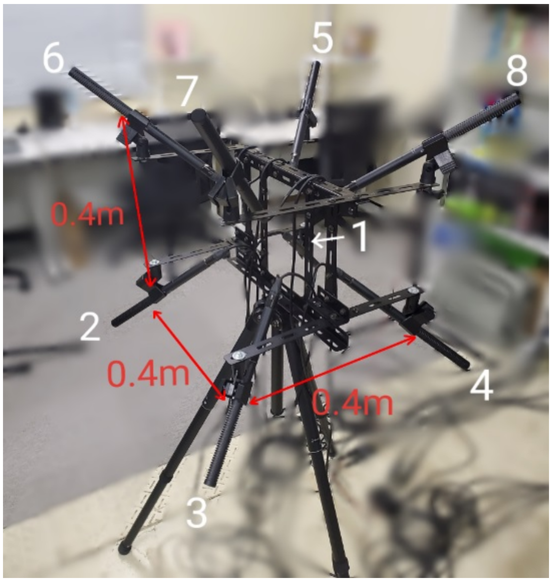

Fig. 2 Microphone array of the personal 3D sound field reproduction.

to house the video camera.

Eight hyperdirectional microphones (AZDEN: SGM-1000) were placed at the vertices of a cube with sides measuring 0.4 m. The directional characteristics of the hyperdirectional microphones were oriented outward of the control area by visually aligning pairs of the microphones in a straight line (1 and 7, 2 and 8, 3 and 5, 4 and 6 in Fig. 2).

Accurate placement of the hyperdirectional microphones at the vertices of the cube is crucial for the personal 3D sound field reproduction technique. However, it is not possible to determine the recording positions of the hyperdirectional microphones just by visual observations due to the cylindrical design of the commercially available hyperdirectional microphones. Therefore, it is necessary to identify the recording position of the hyperdirectional microphones in order to construct the microphone array. In Section 2.2, using the recording position identification method was employed via the impulse response measurement.

2.2 Identification of the recording position

2.2.1 Principle of the method

The configuration of the conventional recording position identification method is shown in Fig. 3. A hyperdirectional microphone is placed in front of the loudspeaker and the impulse response is measured. The recording position (d_m, the distance from the tip of the hyperdirectional microphone in Fig. 3) is calculated using the initial delay time t of the impulse response and the distance as measured between the loudspeaker and the hyperdirectional microphone. However, the environment for the impulse response measurement is generally not constant; in particular, the sound velocity c changes considerably when the temperature in the space changes. Thus, in order to accurately identify the recording positions with this method, disturbances due to the environmental factors (especially temperature) must be considered using some other method.

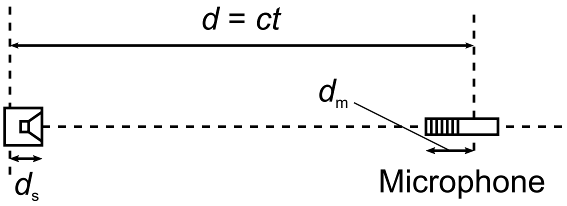

Fig. 3 Basic configuration of the conventional method for identifying the recording position.

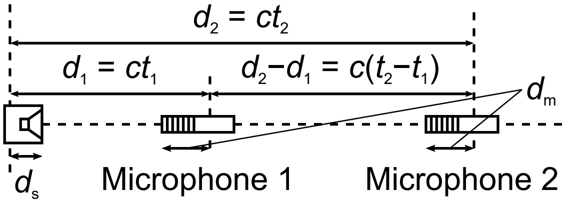

Fig. 4 Basic configuration of the modified method for identifying recording positions.

The configuration of the identification method is modified (Fig. 4) to calculate the sound velocity from the impulse response measurements. Two hyperdirectional microphones are placed in front of the loudspeaker and the impulse responses are measured. The direction of the two hyperdirectional microphones were assumed to be identical. Using the difference between the initial delay times in the impulse responses of the two hyperdirectional microphones ($t_1 - t_2$) and the measured distance between the two microphones ($d_1 - d_2$), the sound velocity $c(= \frac{d_2-d_1}{t_2-t_1})$ was calculated. The recording position can be identified by using the initial delay time t_1 of the impulse response and the distance as measured between the loudspeaker and the hyperdirectional microphone.

However, the playing position of the loudspeaker used for the measurement of the impulse response (d_s, the distance from the loudspeaker surface in Fig. 4) was located behind the loudspeaker, as it usually corresponds with the diaphragm's vibration point inside the loudspeaker. Moreover, it is not possible to identify the playing position of the loudspeaker from visual observations alone, as surface of the loudspeaker is obscured by a protective mesh, depending on the type of loudspeaker. Therefore, to address this, the impulse responses are measured again using the configuration shown in Fig. 5. Two small microphones were placed in front of the loudspeaker and the impulse responses are measured. Using the difference between the initial delay times of the impulse responses of the two small microphones ($t_1 - t_2$) and the measured distance between the two small microphones ($d_1 - d_2$), the sound velocity $c(= \frac{d_2-d_1}{t_2-t_1})$ was calculated. The playing position of the loudspeaker can be identified by using the initial delay time t_1 of the impulse re-

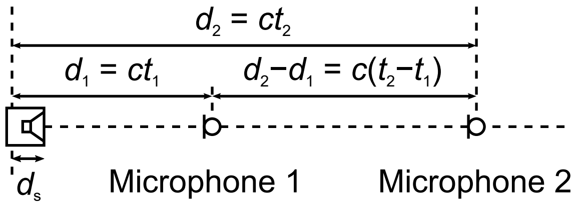

Fig. 5 Basic configuration of the method for identifying the playing position of the loudspeakers.

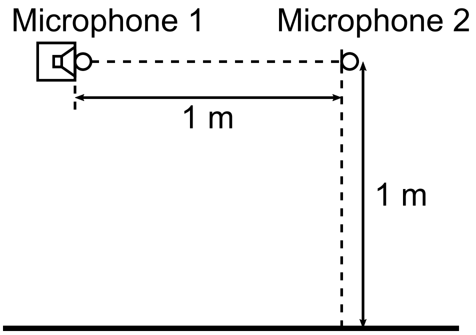

Fig. 6 Position of small microphones and the loudspeaker during the identification of the playing position.

sponse and the measured distance between the loudspeaker and the small microphone.

Considering the above, the procedure of the recording position identification method used in this paper is as follows. First, the impulse response is measured using the configuration shown in Fig. 5. The playing position of the loudspeaker (d_s, the distance from the loudspeaker surface) is calculated using the sound velocity ($c = \frac{d_2 - d_1}{t_2 - t_1}$) and the measured distance. Next, the impulse response is measured again using the configuration shown in Fig. 4, and the sum of the distance from the tip of the hyperdirectional microphone and the distance from the loudspeaker surface ($d_\mathrm{s} + d_\mathrm{m}$) is calculated from the sound velocity ($c = \frac{d_2 - d_1}{t_2 - t_1}$) and the measured distance. Finally, the recording position (d_m, the distance from the tip of the hyperdirectional microphone) is identified by calculating the difference between the distances calculated from two measurements.

2.2.2 Identification of playing position of the loudspeaker

To identify the playing position of the loudspeaker, the impulse responses from the loudspeaker to the two small microphones were first measured in the laboratory. As shown in Fig. 6, two small microphones were placed at distances 0 and 1 m from the loudspeaker. A closed loudspeaker (Ohm Denki: ASP-204N-K), which is shown in Fig. 7, with a size of 7 cm×7 cm×7 cm was used as the loudspeaker. An ultra-compact microphone (Audio-Technica: AT9903) was used as the small microphone.

The measurement conditions are shown in Table 1. The value of the sound pressure level was measured at 1 m distance from the loudspeaker. Since the purpose of this measurement is to determine the initial delay time of the impulse response, the gain of the microphone amplifier was adjusted so that no clipping occured for the microphone placed 0 m from the loudspeaker.

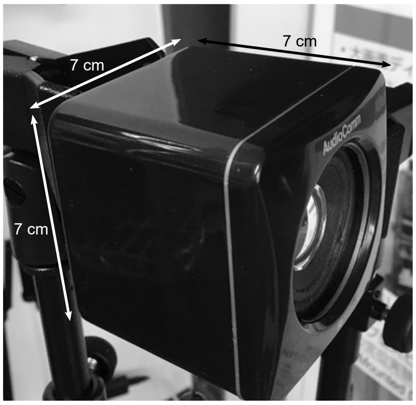

Fig. 7 Loudspeaker used in the measurement.

Table 1 Measurement conditions of impulse responses.

Background noise level	40.2 dBA
Sound pressure level	70.0 dBA
Sampling frequency	48 kHz
Time stretched pulse (TSP) length	65536 samples
Repetition number	9

After the analysis of the measured impulse responses, the arrival time t_1 from the loudspeaker to the small microphone 0 m distance was found to be 0.104 ms $\left(= \frac{5}{48000} \times 1000\right)$. The arrival time t_2 from the loudspeaker to the small microphone 1 m distance was 2.979 ms $\left(= \frac{143}{48000} \times 1000\right)$. Because the distance between the two small microphones is 1 m, the sound velocity in this measurement is 347.83 m/s $\left(= \frac{1 \times 48000}{143 - 5}\right)$. I considered that the calculated sound velocity value was valid because the calculated value was close to the theoretical value.

The playing position of the loudspeaker was identified from the calculated sound velocity. As the arrival time t_1 from the loudspeaker to the small microphone 0 m distance is 0.104 ms, the playing position of the loudspeaker is 3.623 cm $\left(= \frac{5}{143 - 5} \times 100\right)$ behind the loudspeaker surface. I considered that the identified playing position of the loudspeaker was appropriate because the identified value was smaller than the depth of the loudspeaker (7 cm).

2.2.3 Identification of the recording position of the hyperdirectional microphone

Next, in order to identify the recording position of the hyperdirectional microphone, the impulse responses from the loudspeaker to the two hyperdirectional microphones were

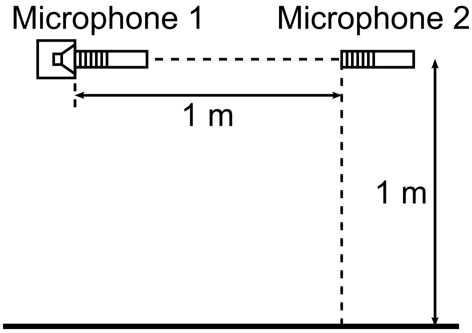

Fig. 8 Position of the hyperdirectional microphones and the loudspeaker used for identifying the recording position.

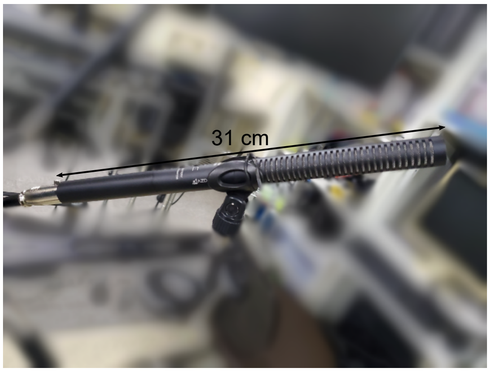

Fig. 9 Hyperdirectional microphone used for the measurement.

measured in the laboratory. As shown in Fig. 8, the two hyperdirectional microphones were positioned so that the tip of the hyperdirectional microphones were 0 m and 1 m, respectively, from the loudspeaker. The direction of the two hyperdirectional microphones was toward the loudspeaker. The same loudspeaker (Ohm Denki: ASP-204N-K) used in the measurement described in Section 2.2.2 was used as the loudspeaker. Hyperdirectional microphones (AZDEN: SGM-1000) having a total length of 31 cm, as shown in Fig.9, were used as the hyperdirectional microphones. The measurement conditions are the same as those in Section 2.2.2.

After the analysis of the measured impulse responses, the arrival time t_1 from the loudspeaker to the hyperdirectional microphone positioned 0 m away from the loudspeaker was found to be 0.521 ms $\left(= \frac{25}{48000} \times 1000\right)$. The arrival time t_2 from the loudspeaker to the hyperdirectional microphone 1 m away was 3.521 ms $\left(= \frac{169}{48000} \times 1000\right)$. Because the distance between the hyperdirectional microphones was 1 m, the sound velocity in this measurement was 333.33 m/s $\left(= \frac{1 \times 48000}{169-25}\right)$. I considered that the calculated sound velocity value was valid because the calculated value was close to the theoretical value.

The recording position of the hyperdirectional microphone is identified based on the calculated sound velocity. Since the arrival time t_1 from the loudspeaker to the hyperdirectional microphone 0 m distance is 0.521 ms, the distance from the playing position of the loudspeaker to the recording position of the hyperdirectional microphone is 17.361 cm $\left(= \frac{25}{169-25} \times 100\right)$. The distance from the tip of the hyperdirectional microphone to the recording position of the hyperdirectional microphone is 13.7 cm (= 17.361 − 3.623 = 13.738 ≈ 13.7) because the distance from the playing position of the loudspeaker to the loudspeaker surface is 3.623 cm. It is considered that the identified recording position of the hyperdirectional microphone is appropriate because the identified value is smaller than the total length of the hyperdirectional microphone (31 cm).

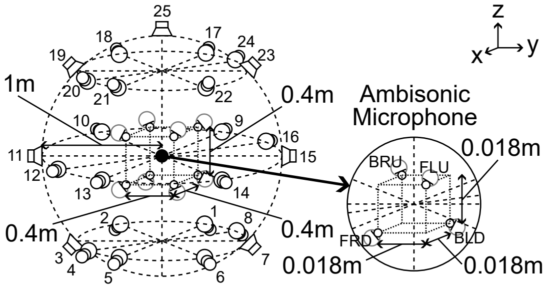

Fig. 10 Placement of loudspeakers and microphones for the measument of impulse responses.

3. Evaluation of localization performance

3.1 Measurement of the impulse response

First, in order to create the replaying sound that participants heard in the evaluation experiment, impulse responses were measured using the constructed microphone array and the ambisonic microphone (ZOOM: H3-VR), which is a conventional 3D sound field recording technique.

Measurements were performed in the laboratory. The reverberation time of the laboratory was 500 ms, and the background noise level was 32.6 dBA during the measurement. The sound pressure level was set to 75.1 dBA at 1 m from the sound source. As shown in Fig. 10, an ambisonic microphone was placed at the center of the microphone array, and 25 sound source positions were set. The azimuth and elevation angles of the sound source positions are shown in Table 2.

An image of the measurement is shown in Fig. 11. A 65536-point TSP signal [18] with a sampling frequency of 48 kHz was played from a loudspeaker (Ohm Electric: ASP-204N-K). The number of repetitions was nine. The sound source loudspeaker was moved after each measurement at one sound source position, and the impulse response measurements were repeated 25 times.

3.2 Synthesis of replayed sound

Next, the eight-channel replay sounds were synthesized by

Table 2 Azimuth and elevation angles of sound source position for the measument of impulse responses.

Index	θ	ϕ	Index	θ	ϕ
1	-180°	-45°	14	45°	0°
2	-135°	-45°	15	90°	0°
3	-90°	-45°	16	135°	0°
4	-45°	-45°	17	-180°	45°
5	0°	-45°	18	-135°	45°
6	45°	-45°	19	-90°	45°
7	90°	-45°	20	-45°	45°
8	135°	-45°	21	0°	45°
9	-180°	0°	22	45°	45°
10	-135°	0°	23	90°	45°
11	-90°	0°	24	135°	45°
12	-45°	0°	25	—	90°
13	0°	0°			

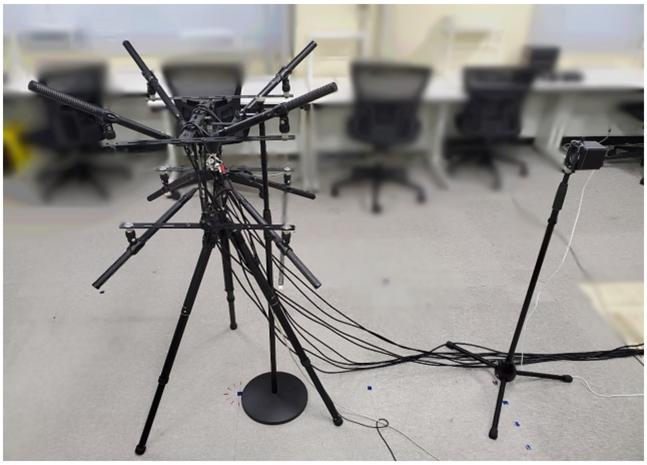

Fig. 11 Image showing the measument of impulse responses. (Sound position index: 9).

convolving the measured impulse responses with the sound sources used in the previous study [19]. The sound source was the white noise and speech, of which the duration was 4 s. To eliminate the effect of reverberation in the measurement environment on localization performance, the reverberant sound in the measurement environment was removed by setting the amplitude of the component that was later than the direct wave in the impulse responses to 0.

The replay sounds of the constructed microphone array were synthesized by convolving the sound source signal with the impulse response measured at each microphone in the array. On the other hand, based on previous research [20], the replay sounds of the ambisonic microphone were expanded to eight channels according to the following equations after convolving the sound source signal with the measured impulse responses:

$$h_{\mathrm{BRD}}(n) = w(n) + 0.707\{-x(n) - y(n) - z(n)\}, \quad (1)$$

$$h_{\mathrm{FRD}}(n) = w(n) + 0.707\{\ \ x(n) - y(n) - z(n)\}, \quad (2)$$

$$h_{\mathrm{FLD}}(n) = w(n) + 0.707\{\ \ x(n) + y(n) - z(n)\}, \quad (3)$$

$$h_{\mathrm{BLD}}(n) = w(n) + 0.707\{-x(n) + y(n) - z(n)\}, \quad (4)$$

$$h_{\mathrm{BRU}}(n) = w(n) + 0.707\{-x(n) - y(n) + z(n)\}, \quad (5)$$

$$h_{\mathrm{FRU}}(n) = w(n) + 0.707\{\ \ x(n) - y(n) + z(n)\}, \quad (6)$$

$$h_{\mathrm{FLU}}(n) = w(n) + 0.707\{\ \ x(n) + y(n) + z(n)\}, \quad (7)$$

$$h_{\mathrm{BLU}}(n) = w(n) + 0.707\{-x(n) + y(n) + z(n)\}, \quad (8)$$

where $w(n)$, $x(n)$, $y(n)$, and $z(n)$ are composed as follows [21]:

$$w(n) = g_{\mathrm{W}}(n) * s(n), \quad (9)$$

$$x(n) = g_{\mathrm{X}}(n) * s(n), \quad (10)$$

$$y(n) = g_{\mathrm{Y}}(n) * s(n), \quad (11)$$

$$z(n) = g_{\mathrm{Z}}(n) * s(n), \quad (12)$$

$$\begin{pmatrix} g_{\mathrm{W}}(n) \\ g_{\mathrm{X}}(n) \\ g_{\mathrm{Y}}(n) \\ g_{\mathrm{Z}}(n) \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} g_{\mathrm{FLU}}(n) \\ g_{\mathrm{FRD}}(n) \\ g_{\mathrm{BLD}}(n) \\ g_{\mathrm{BRU}}(n) \end{pmatrix}, \quad (13)$$

where $s(n)$ is the sound source signal, and $*$ is the convolution operation; $g_{\mathrm{FLU}}(n)$, $g_{\mathrm{FRD}}(n)$, $g_{\mathrm{BLD}}(n)$, and $g_{\mathrm{BRU}}(n)$ denote the impulse responses measured with the ambisonic microphone in Fig. 10. In both setups for the constructed microphone array and the ambisonic microphone, the delay time caused by the convolution of the impulse responses was the same.

3.3 Localization test

Localization tests were performed using the synthesized sound. The tests were conducted in a corner of the laboratory that was used to measure the impulse responses described in Section 3.1. The participants sat on a chair within the array of loudspeakers shown in Fig. 12, wearing a head-mounted display (Microsoft: HoloLens2). The loudspeakers in the loudspeaker array were placed at the vertices of a cube whose side length was 40 cm. A loudspeaker unit (ELECOM: diverted from MS-P06A) attached to a sealed enclosure (Ohm Electric: diverted from ASP-204N-K) was used as the loudspeaker. The height of the center of the participant's head was the same as that of the center of the array (1.4 m from the floor), and the sound pressure level was set to approximately 70 dBA at the center of the array.

A video of 3D virtual space, which is shown in Fig. 13, was created using Unreal Engine 4.25 and presented on the head-mounted display during the test. In the 3D virtual space, the camera was fixed at the position of the participant's head, and 25 spherical objects with numbers were placed in the direction shown in Table 2. The distance between the camera and the objects was 10 m.

The participants in the test were 10 people with no abnormalities of vision or hearing in their daily life. The gender and age of the participants were not collected because the experiment director judged that this should not be collected for privacy reasons. Participants received an explanation about the experimental ethics before the experiment. In addition to the purpose and contents of the experiment, the following items were explained during the explanation:

- Measurements are performed while taking a break

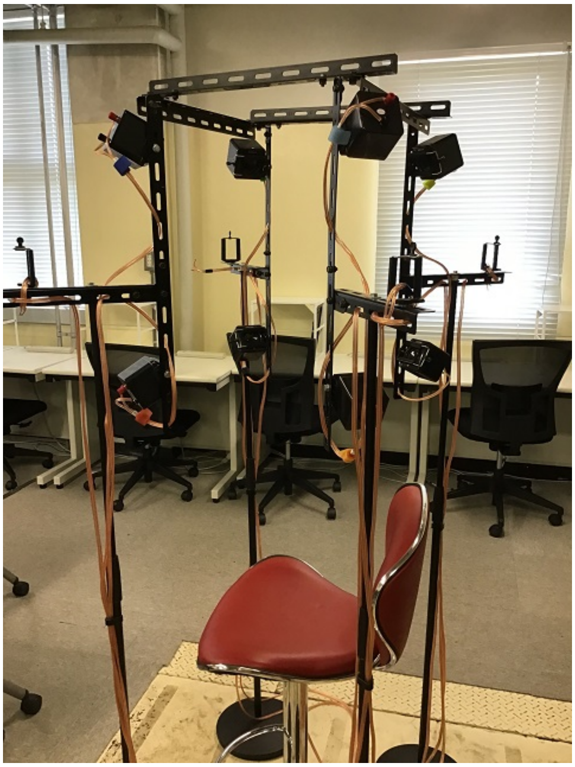

Fig. 12 Loudspeaker array in the localization test.

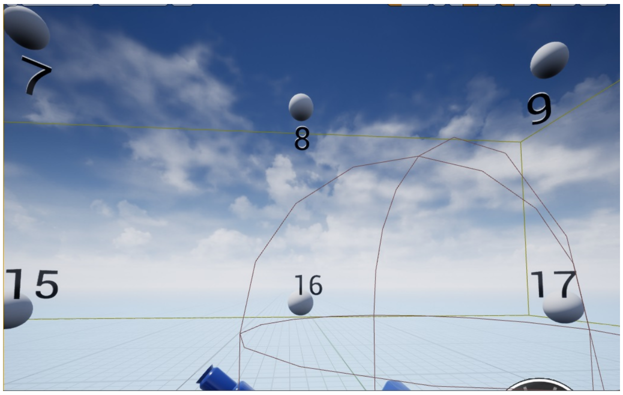

Fig. 13 3D virtual space in the localization test.

- There is no pain or invasion in the measurements
- The measurement procedure can be terminated at participants' request
- The obtained data is processed in such a way that individuals cannot be identified
- Processed data are statistically used

After the explanation, if participants agreed, they signed the prepared agreement form.

The flowchart for the test is shown in Fig. 14. In the test, sessions were set up for each sound source (white noise and speech). In each session, the participants underwent four practice trials and then performed 100 main trials (=25 sound image directions × 2 recording systems × 2 repetitions). The order of presentation of the sessions, the sound image directions, and the recording systems were randomized for each participant. In each trial, the participants listened to a sound for 4 seconds and then gave the number for the direction from which the sound was coming. The experimenter recorded the number given by the participant.

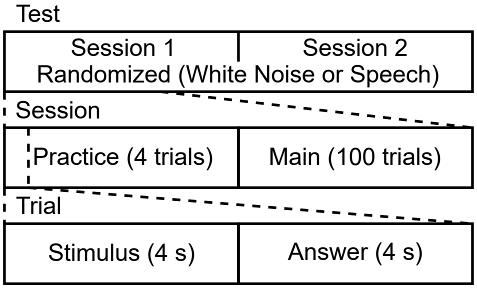

Fig. 14 Flowchart of the localization test.

3.4 Result of the localization test

First, the results of the localization test were analyzed based on the direction of the perceived sound image. In this case, the perceived direction was generally evaluated using the azimuth and elevation angles. However, when the presented direction was directly above (elevation angle 90°) and the perceived direction was not directly above, the displacement of the sound image from directly above could not be analyzed using the azimuth and elevation angles. Therefore, when the perceived direction of the localization was analyzed to include the sound signals presented directly above, it was necessary to convert the experimental results to a coordinate system to assume that all the presented directions were analyzed under the same conditions.

In this paper, the polar coordinate system was rotated according to the following formulae, as was done in a previous study [17]. As a result, the azimuth and elevation angles (θ, ϕ) of the answer direction were converted to the horizontal and vertical angles (θ', ϕ'):

$$\theta' = \tan^{-1}\frac{y'}{x'}, \tag{14}$$

$$\phi' = \sin^{-1}\frac{z'}{\sqrt{x'^2 + y'^2 + z'^2}}, \tag{15}$$

where the three-dimensional coordinates (x', y', z') of the converted answer direction are defined as follows:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \mathbf{R}_y(-\phi_0)\mathbf{R}_z(-\theta_0)\begin{pmatrix} \cos\theta\cos\phi \\ \sin\theta\cos\phi \\ \sin\phi \end{pmatrix}, \tag{16}$$

$$\mathbf{R}_y(-\phi_0) = \begin{pmatrix} \cos\phi_0 & 0 & \sin\phi_0 \\ 0 & 1 & 0 \\ -\sin\phi_0 & 0 & \cos\phi_0 \end{pmatrix}, \tag{17}$$

$$\mathbf{R}_z(-\theta_0) = \begin{pmatrix} \cos\theta_0 & \sin\theta_0 & 0 \\ -\sin\theta_0 & \cos\theta_0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{18}$$

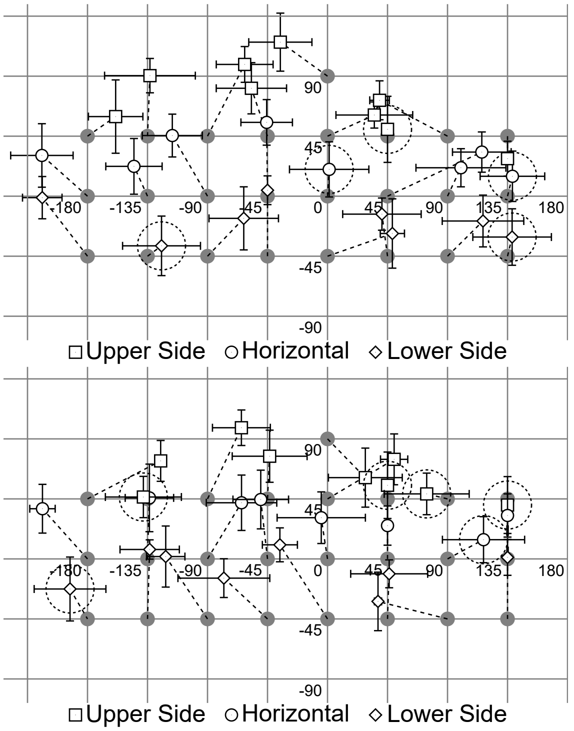

Fig. 15 Results for the perceived directions of the ambisonic microphone during the localization test (Upper: White Noise, Lower: Speech).

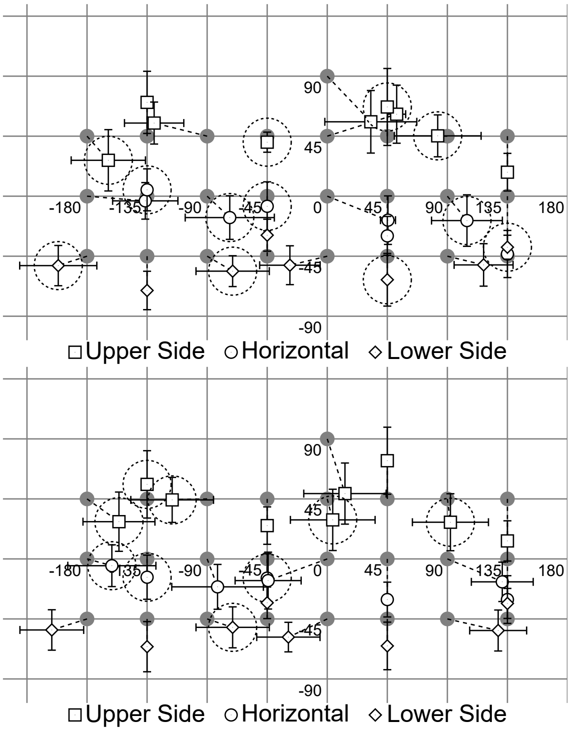

Fig. 16 Results of the perceived directions of the developed recording system during the localization test (Upper: White Noise, Lower: Speech).

where $\mathbf{R}_z(-\theta_0)$ corresponds to the horizontal rotation operation, and $\mathbf{R}_y(-\phi_0)$ corresponds to the vertical rotation. θ_0 and ϕ_0 denote the azimuth and elevation angles of the presented direction. The converted horizontal angle corresponds to the horizontal displacement of the sound image when the listener turns his or her head in the direction of the presented sound stimulus. However, the converted vertical angle corresponds to the vertical displacement of the sound image when the listener turns his or her head in the presented direction. Note that the data for the converted horizontal angle θ' are not used in following analysis if the converted vertical angle ϕ' becomes ±90°.

The converted horizontal and vertical angles of the ambisonic microphone are shown in Fig. 15. The error bars of the horizontal and vertical directions denote the 95% confidence intervals for horizontal and vertical angles. To make it easier to understand the vertical and horizontal relationship of the presented directions in the figure, the averages are shifted horizontally and vertically by the azimuth and elevation angles of the direction of the presented signal. The presented descriptions are also connected to the answer directions by black dotted lines. In addition, the presented directions, in which gray circles that indicate the presented direction are inside the vertical and horizontal error bars, are surrounded by black dashed lines. In other words, in the presented direction surrounded by the black dashed line, the *t*-test of the population mean for each converted horizontal and vertical angle showed that there were no significant difference between the presented and the perceived directions.

The localization performance of the ambisonic microphone cannot be judged as good because the number of presentation directions surrounded by the black dashed line is few for both sound sources (white noise: 5; speech: 6). This is because ambisonics is a sound field reproduction method that uses only the 0th and 1st order terms in the spherical harmonic analysis of the sound field, and also because the ambisonic microphone cannot be placed at the center, although each microphone should ideally be placed at the center of the array.

The converted horizontal and vertical angles of the developed recording system are shown in Fig. 16. It can be observed that the localization performance of the developed recording system is better than that of the ambisonic microphones because the number of presented directions that are encircled by the black dashed line is greater than that of the ambisonic microphones for both sound sources (white noise: 12; speech: 9). This is thought to be due to the fact that the sound field was reproduced in the loudspeaker array by placing the hyperdirectional microphones at the vertices of the cubes having a side length of 0.4 m, according to a previous study [17].

To conduct further quantitative analysis, the correct rates for each experimental condition were calculated. The number of answers was 500 (= 25 directions × 2 repetitions × 10 participants). In the correct answers, the participants answered correctly in response to the presented direction.

The correct rates in each experimental condition are shown in Fig. 17. The Fisher's exact test for each sound source using js-STAR XR+ [22] shows that there are significant

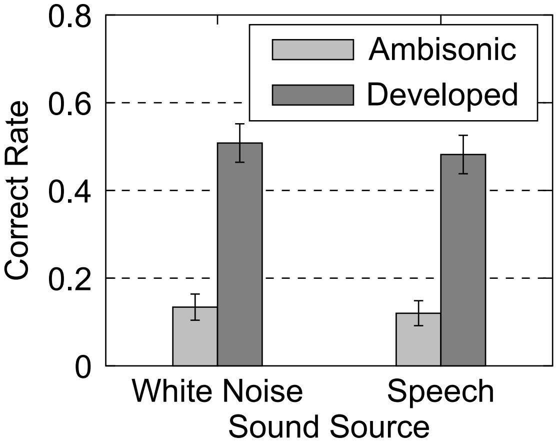

Fig. 17 Results showing the correct rates in the localization test.

differences at the 0.1% level. Therefore, it can also be said that, in terms of the correct rate, the localization performance of the developed recording system is better than that of the ambisonic microphone.

4. Conclusion

In this paper, a recording system for personal 3D sound field reproduction using a microphone array with eight hyperdirectional microphones was developed. The recording position of the hyperdirectional microphone was identified by using a recording position identification method via impulse response measurement, and the microphone array was constructed. To evaluate the localization performance of the constructed microphone array, the impulse responses were measured, the sound was reproduced, and the localization experiment was performed. During the localization experiment, the localization performance was shown to be better than that of a conventional ambisonic microphone.

In the future, it will be necessary to build a remote operating system that combines the developed recording system with video and to evaluate its performance. However, based on the audio-visual localization performance, the conditions necessary for three-dimensional sound field reproduction may be relaxed. Thus, if an audio-visual presentation system that combines a head-mounted display and a loudspeaker array is used, the conditions that are necessary for three-dimensional sound field reproduction can be also evaluted.

Acknowledgment

This paper is based on three IEICE technical reports [23]–[25]. The author would like to thank Mr. K. Hagino, Ms. F. Hanyu and Mr. K. Hirata for their continued support and encouragement during the research.

References

[1] K. Takeshita, K. Watanabe, K. Sato, K. Minamizawa, and S. Tachi, “Study on telexistence LXIII -study of tolerance of network delay for TELESAR3-,” Proceedings of Annual Conference of Virtual Reality Society of Japan, no.1C1-4, pp.146–149, September 2010.
[2] J. Blauert, Spatial Hearing, revised ed., pp.372–392, MIT Press, Cambridge, Mass, 1997.
[3] M.R. Schroeder, D. Gottlob, and K.F. Siebrasse, “Comparative study of european concert halls: Correlation of subjective preference with geometric and acoustic parameters,” J. Acoust. Soc. Am., vol.56, no.4, pp.1195–1201, 1974.
[4] T. Kimura, K. Kakehi, K. Takeda, and F. Itakura, “Spatial coding of multi-channel audio signals in sound field reproduction,” Trans. VR Soc. Jpn., vol.8, no.4, pp.433–442, 2003.
[5] I. Toshima, S. Aoki, and T.Hirahara, “Sound localization using an auditory telepresence robot: TeleHead II,” Presence, MIT Press, vol.17, no.4, pp.392–404, 2008.
[6] R.K. Furness, “Ambisonics - an overview,” Proc. Audio Eng. Soc. 8th International Conference, pp.181–190, May 1990.
[7] A. Fukada, “The microphone technique for music recording,” The Journal of The Institute of Image Information and Television Engineers, vol.64, no.9, pp.1344–1348, 2010.
[8] ITU-R Recommendation BS.775-1, Multichannel Stereophoneic Sound System with and without Accompanying Picture, 1992-1994.
[9] K. Ono, T. Nishiguchi, K.Matsui, and K. Hamasaki, “Portable spherical microphone for Super Hi-Vision 22.2 multichannel audio,” Proc. Audio Eng. Soc. 135th Convention, no.8922, October 2013.
[10] T. Okamoto, Y. Iwaya, S. Sakamoto, and Y. Suzuki, “Implementation of higher order ambisonics recording array with 121 microphones,” Proc. 3rd Student Organizing Int. Mini-Conf. on Inf. Electr. Syst., pp.71–72, October 2010.
[11] H. Kashiwazaki and A. Omoto, “Sound field reproduction system using narrow directivity microphones and boundary surface control principle,” Acoust. Sci. & Tech., vol.39, no.4, pp.295–304, 2018.
[12] M. Camras, “Approach to recreating a sound field,” J. Acoust. Soc. Am., vol.43, no.6, pp.1425–1431, 1968.
[13] A. J. Berkhout, D. de Vries and P. Vogel, “Acoustic control by wave field synthesis,” J. Acoust. Soc. Am., vol.93, no.5, pp. 2764—2778, 1993.
[14] T. Kimura and K. Kakehi, “Effects of microphone and loudspeaker directivity on accuracy of synthesized wave fronts in sound field reproduction with wave field synthesis,” Trans. VR Soc. Jpn., vol.12, no.2, pp.191–198, 2007.
[15] T. Kimura, K. Kakehi, K. Takeda and F. Itakura, “Subjective Effect of the Number of Channel Signals in Wave Field Synthesis - In the Case of Sound Sources of Frontal Direction -,” Trans. VR Soc. Jpn., vol.10, no.2, pp. 257–266, 2005.
[16] B.B. Baker and E.T. Copson, The Mathematical Theory of Huygens’ Principle, second ed., pp.23–26, Oxford University Press, London, UK, 1950.
[17] T. Kimura, “Personal compact 3D sound field reproduction system based on concept of wave front synthesis technique using eight directional microphones,” IEICE Trans. Fundamentals, vol.J97-A, no.4, pp.284–294, 2014.
[18] Y. Suzuki, F. Asano, H.Y. Kim, and T. Sone, “An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses,” J. Acoust. Soc. Am, vol.97, no.2, pp.1119–1123, 1995.
[19] T. Kimura and H.Ando, “3D audio system using multiple vertical panning for large-screen multiview 3d video display,” ITE Transactions on Media Technology and Applications, vol.2, no.1, pp.33–45, 2014.
[20] D.G. Malham and A. Myatt, “3-D sound spatialization using ambisonic techniques,” Computer Music Journal, vol.19, no.4, pp.58–70, 1995.
[21] K. Farrar, “Soundfield microphone,” Wireless World, vol.85, pp.48–51, 1979.
[22] js-STAR XR+ Website. https://www.kisnet.or.jp/nappa/software/star/.
[23] T. Kimura and K. Hagino, “Identification method of the recording

position of hyper-directional microphones," IEICE Technical Report, no.EA2017-27, pp.1–4, August 2017.

[24] F. Hanyu and T. Kimura, "Subjective evaluation of recording system for personal 3D sound field reproduction," IEICE Technical Report, no.EA2020-81, pp.128–133, March 2021.

[25] T. Kimura, "3D localization evaluation of recording system for personal 3D sound field reproduction," IEICE Technical Report, no.EA2022-30, pp.13–18, August 2022.

Toshiyuki Kimura received the B.E., M.A. and Ph.D. degrees from Nagoya University, Japan in 1998, 2000 and 2005, respectively. He was a research fellow of Japan Society for the Promotion of Science, a research fellow of Nagoya University, Japan, a research associate of Tokyo University of Agriculture and Technology, Japan and a limited term researcher of National Institute of Information and Communications Technology, Japan from 2003 to 2015. He is currently an associate professor of Tohoku Gakuin University, Japan from 2015. His research interests include ultra-realistic communication, spatial perception and array signal processing. He is a member of the IEICE, IPSJ, ASJ, VRSJ, and AES.