Comparative Performance Evaluation of Near 3D Sound Field Reproduction System with Directional Loudspeakers and Wave Field Synthesis

Toshiyuki Kimura*, Yoko Yamakata†, Michiaki Katsumoto*,
Takuma Okamoto‡, Satoshi Yairi§, Yukio Iwaya‡ and Yôiti Suzuki‡
*Universal Media Research Center, National Institute of Information and Communications Technology,
Koganei, Tokyo 184–8795 Japan, Email: {t-kimura,katumoto}@nict.go.jp
†Academic Center for Computing and Media Studies, Kyoto University,
Sakyo-ku, Kyoto, 606–8501 Japan, Email: yamakata@media.kyoto-u.ac.jp
‡Research Institute of Electrical Communication / Graduate School of Information Sciences, Tohoku University,
Aoba-ku, Sendai, 980–8577 Japan, Email: okamoto@ais.riec.tohoku.ac.jp, {iwaya,yoh}@riec.tohoku.ac.jp
§Department of Electrical Engineering, Sendai National College of Technology,
Natori, Miyagi, 981–1239 Japan, Email: yairi@sendai-nct.ac.jp

***Abstract*—Near 3D sound field display systems are important toward realizing ultra-realistic communications systems such as 3D television. We have proposed the near 3D sound field reproduction systems using directional loudspeakers and wave field synthesis and developed the real system by constructing the surrounding microphone array and the radiated loudspeaker array. In this paper, to evaluate the performance of sound image localization in the developed system, sound image positions were estimated via computer simulation. The results indicated that listeners can accurately localize a sound image in any listening position around the radiated loudspeaker array if the radiation directivity of the loudspeaker units is sharpened, and that listeners can accurately localize a sound image close to the listening position if they listen to a sound in a particular listening position, even if the radiation directivity of the loudspeaker units is not sharpened. Results of estimation of a sound image position by acoustical measurement similarly showed that listeners can accurately localize a sound image close to the listening position if they listen to a sound in a particular listening position.**

I. Introduction

We have been investigating ultra-realistic communication techniques [1]. If, by applying these techniques, realistically reproduced 3D video and audio can appear in a 3D space, as shown in Fig. 1, and several people can view an object anywhere in its vicinity without having to wear equipment such as glasses, this will enable more realistic forms of communication (e.g., 3D television) than those currently offered by conventional video and audio techniques (HD video and 5.1-channel audio).

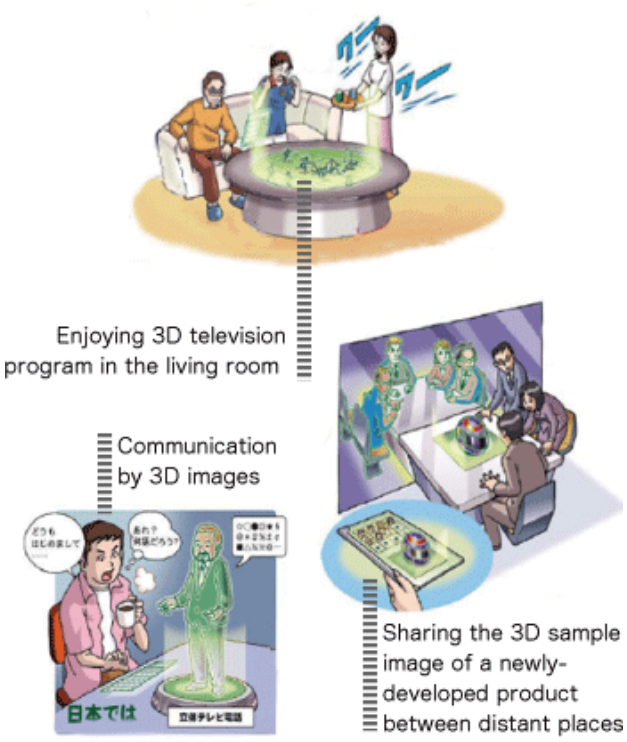

Fig. 1. Future image of ultra-realistic communications [1].

To realize 3D audio systems that are compatible with 3D video systems for 3D television, 3D sound field reproduction systems must be developed that enable multiple listeners to listen to an object's sound anywhere in its vicinity without the need for headphones. Wave field synthesis [2], [3], [4] is a 3D sound field reproduction technique for reproducing wave fronts from a control area in a different area (the listening area), according to the Kirchhoff-Helmholtz integral equation [5]. Microphones placed along the boundary of the control area record the original sound, which is replayed by loudspeakers placed along the boundary of the listening area. The loudspeakers and microphones are in the same respective positions. Multiple listeners can listen to the replayed sounds anywhere in the listening area without wearing devices such as headphones because the technique reproduces the sound field

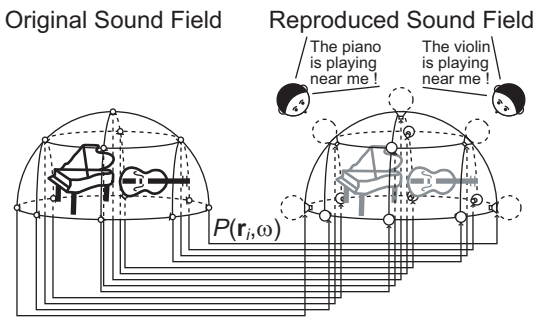

Fig. 2. Diagram of proposed near 3D sound field reproduction systems using directional loudspeakers and wave field synthesis.

of a 3D space rather than that of binaural positions.

In order to realize the 3D sound field reproduction system that multiple listeners can listen to a sound anywhere in its vicinity without having to wear equipment such as headphones, we have proposed the near 3D sound field reproduction systems using directional loudspeakers and wave field synthesis [6] and developed the real system by constructing the surrounding microphone array and the radiated loudspeaker array [7]. In this paper, in order to evaluate the performance of the sound image localization in the developed system, the positions of sound images are estimated by two methods (computer simulation and acoustical measurement). First, the positions of sound images are estimated via a computer simulation in the same condition as the developed system. Second, the positions of sound images are estimated by acoustical measurement in the developed system and the estimated results are compared with those from the computer simulation.

II. Diagram of Proposed System [6]

A diagram of the proposed system is shown in Fig. 2. First, the surrounding microphone array, which consists of M omnidirectional microphones, is placed around the sound sources in the original sound field and M audio signals are recorded. The positions of the microphones are on the boundary surface. Second, the radiated loudspeaker array, which consists of M directional loudspeaker units, is placed in the reproduced sound field and the M recorded audio signals are played. The positions of the loudspeaker units are the same as those of the microphones. The directivity of the loudspeaker units is toward the outside of the boundary surface. Since the 3D sound field on the outside of the radiated loudspeaker array is accurately controlled, listeners on the outside of array can resultantly feel that the sound sources are being played on the inside of the array. In the case of Fig. 2, listeners close to the piano can feel that they are listening to the sound near the piano, and the same for violin.

III. Developed System [7]

A. Surrounding Microphone Array

For the surrounding microphone array, we used a surrounding microphone array room at the Research Institute

Fig. 3. Image of surrounding microphone array placed in the room [8].

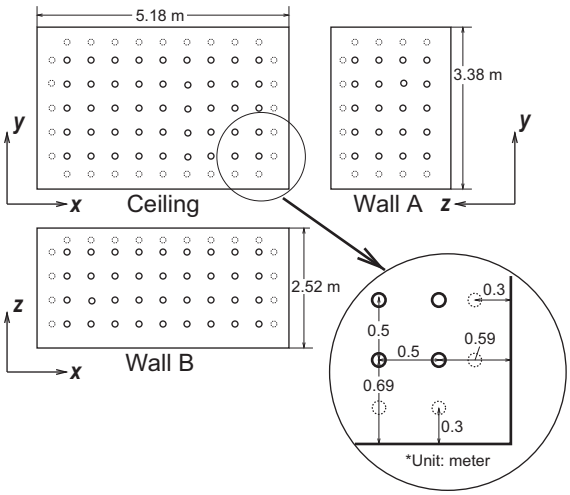

Fig. 4. Arrangement of microphones in the surrounding microphone array [8].

of Electrical Communication, Tohoku University [8]. The reverberation time of the room was approximately 150 ms, and 157 omnidirectional microphones (Brüel & Kjær: Type 4951) were attached to the five planes of the room, as shown in Fig. 3. These microphones mainly record the direct sound from sound sources placed in the room because sound absorption panels are also attached to all of the room's planes.

The arrangement of microphones is shown in Fig. 4. To the two narrow sidewall planes (Wall A), 20 $(= 5\times4)$ microphones are attached. To the two wide sidewall planes (Wall B), 36 $(= 9\times4)$ microphones are attached. And to the ceiling plane, 45 $(= 9\times5)$ microphones are attached. The interval between the microphones is 0.5 m. The microphones are connected to 10 microphone preamplifiers for 16 channels (Brüel & Kjær: Type 2694).

B. Radiated Loudspeaker Array

For the radiated loudspeaker array, we newly manufactured that shown in Fig. 5. Its size is one-fourth that of the

Fig. 5. Image of radiated loudspeaker array.

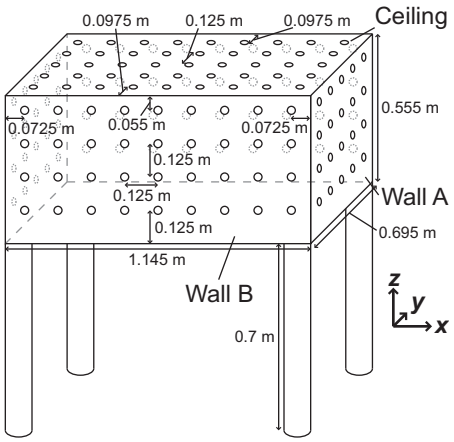

Fig. 6. Arrangement of loudspeaker units in the radiated loudspeaker array.

surrounding microphone array. It has a rectangular enclosure, 157 loudspeaker units, and a stand. The size of the rectangular enclosure is 1.145 m (width) $\times$ 0.695 m (depth) $\times$ 0.555 m (height), made of plywood and aluminum panels. The 157 loudspeaker units, with a size of 1 inch (AURASOUND: NSW1-205-8A suitable), are directly attached to the five planes of the rectangular enclosure. The directivity of each loudspeaker unit is toward the outside of the rectangular enclosure because the sound radiated from each unit to the inside of the enclosure does not leak to its outside.

The arrangement of the loudspeaker units is shown in Fig. 6. To the two narrow sidewall planes (Wall A), 20 $(= 5\times4)$ loudspeaker units are attached. To the two wide sidewall planes (Wall B), 36 $(= 9\times4)$ loudspeaker units are attached. And to the ceiling plane, 45 $(= 9\times5)$ loudspeaker units are attached. The interval between the loudspeaker units is 0.125 m, which is less than half the wavelength of 1000-Hz sound waves ($= \frac{340\,\mathrm{m}}{1000\,\mathrm{Hz}}$=34 cm). Thus, the spatial sampling theorem that pertains to the reproduction of wave fronts of sound waves with frequencies below 1000 Hz is satisfied. These loudspeaker units are connected to a preamplifier with 157 channels (custom-made). The radiated loudspeaker array is elevated 0.7 m by the stand.

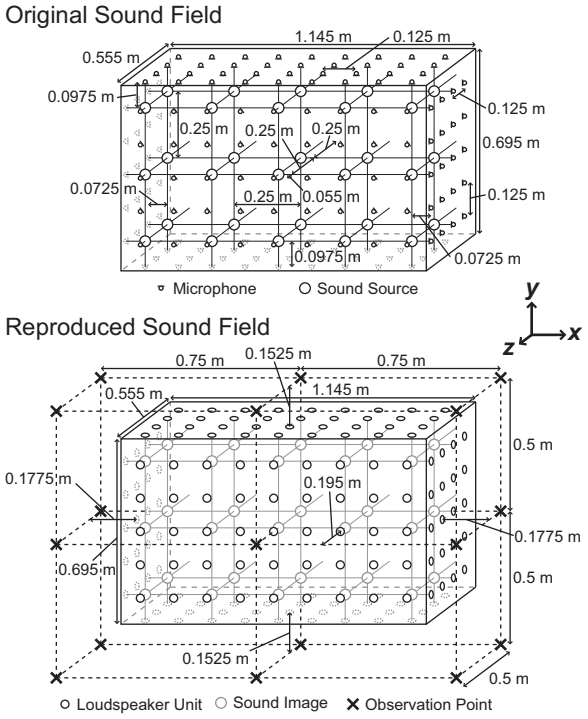

Fig. 7. Positions of the sound sources, sound images, microphones, loudspeaker units, and observation points used in the computer simulations.

IV. Performance Evaluation by Sound Image Position Estimation

A. Computer Simulation

1) Simulation Environment: As shown on the upper part of Fig. 7, 30 sound sources were placed in the surrounding microphone array in the original sound field. The original sound field was a free field with no reflection sounds. On the other hand, as shown in the lower part of Fig. 7, the radiated loudspeaker array was constructed in the reproduced sound field. The reproduced sound field was a free field in order to simplify the computer simulation. Seventeen observation points were placed outside the radiated loudspeaker array.

The sound source signal $s(t)$ was an octave-band noise with central frequency f_{cent}. The $x_i(t)$ (microphone signal of the ith microphone) was denoted as follows from the sound source signal $s(t)$ and the impulse response between the sound source and the microphone $g_i(t)$:

$$x_i(t) = g_i(t) * s(t) = \frac{1}{|\mathbf{r}_i - \mathbf{r}_0|} s\left(t - \frac{|\mathbf{r}_i - \mathbf{r}_0|}{c}\right), \quad (1)$$

where $*$ denotes the convolution operation, $\mathbf{r}_i$ and $\mathbf{r}_0$ are the position vectors of the ith microphone and the sound source, respectively, and c is the sound velocity. The $p(\mathbf{R}_j, f_{\mathrm{cent}}, t)$ (instantaneous sound pressure of the jth observation point $\mathbf{R}_j$)

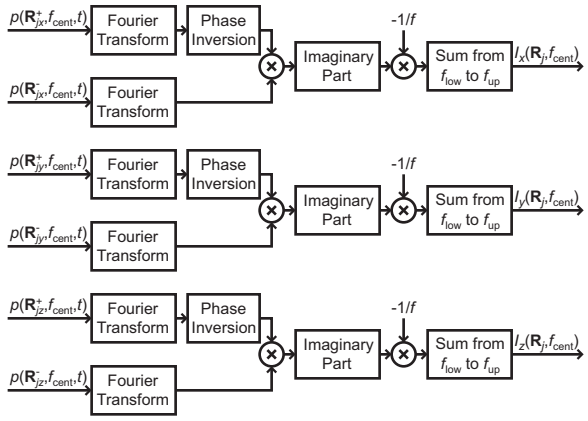

Fig. 8. Block diagram showing the calculation of sound intensities in the computer simulation.

was calculated from $x_i(t)$ as follows:

$$p(\mathbf{R}_j, f_{\text{cent}}, t) = \sum_{i=1}^{M} \frac{D_{si}}{|\mathbf{R}_j - \mathbf{r}_i|} x_i\left(t - \frac{|\mathbf{R}_j - \mathbf{r}_i|}{c}\right) = \sum_{i=1}^{M} \frac{D_{si}}{|\mathbf{R}_j - \mathbf{r}_i||\mathbf{r}_i - \mathbf{r}_0|} s\left(t - \frac{|\mathbf{R}_j - \mathbf{r}_i| + |\mathbf{r}_i - \mathbf{r}_0|}{c}\right), \quad (2)$$

where M is the total number of loudspeaker units, and D_{si} is the radiation directivity of the ith loudspeaker unit.

The arrival direction of sound sources corresponding to the performance of the sound image localization at the jth observation point $\mathbf{R}_j$ is not evaluated by the $p(\mathbf{R}_j, f_{\text{cent}}, t)$ (instantaneous sound pressures of the jth observation point $\mathbf{R}_j$). Since the direction of sound intensity vectors corresponds to the arrival direction of the sound sources [9], the sound intensity vectors at the jth observation point $\mathbf{R}_j$ were also calculated in order to evaluate the arrival direction of sound sources at that point.

The sound intensity vectors were calculated using the cross-spectral method, as shown in Fig. 8. Note that $I_x(\mathbf{R}_j, f_{\text{cent}})$, $I_y(\mathbf{R}_j, f_{\text{cent}})$ and $I_z(\mathbf{R}_j, f_{\text{cent}})$ in Fig. 8 are the x, y and z components of the sound intensity vectors $\mathbf{I}(\mathbf{R}_j, f_{\text{cent}})$, and $p(\mathbf{R}^+_{jx}, f_{\text{cent}}, t)$, $p(\mathbf{R}^-_{jx}, f_{\text{cent}}, t)$, $p(\mathbf{R}^+_{jy}, f_{\text{cent}}, t)$, $p(\mathbf{R}^-_{jy}, f_{\text{cent}}, t)$, $p(\mathbf{R}^+_{jz}, f_{\text{cent}}, t)$, and $p(\mathbf{R}^-_{jz}, f_{\text{cent}}, t)$ in Fig. 8 are the instantaneous sound pressure at six points ($\mathbf{R}^+_{jx}$, $\mathbf{R}^-_{jx}$, $\mathbf{R}^+_{jy}$, $\mathbf{R}^-_{jy}$, $\mathbf{R}^+_{jz}$ and $\mathbf{R}^-_{jz}$). The position vectors of the six points were set as follows:

$$\mathbf{R}^+_{jx} = \mathbf{R}_j + (\Delta, 0, 0)^T, \quad (3)$$
$$\mathbf{R}^-_{jx} = \mathbf{R}_j - (\Delta, 0, 0)^T, \quad (4)$$
$$\mathbf{R}^+_{jy} = \mathbf{R}_j + (0, \Delta, 0)^T, \quad (5)$$
$$\mathbf{R}^-_{jy} = \mathbf{R}_j - (0, \Delta, 0)^T, \quad (6)$$
$$\mathbf{R}^+_{jz} = \mathbf{R}_j + (0, 0, \Delta)^T, \quad (7)$$
$$\mathbf{R}^-_{jz} = \mathbf{R}_j - (0, 0, \Delta)^T, \quad (8)$$

where Δ is 0.001 m.

Parametric conditions are shown in TABLE I. The $\mathbf{r}_0 = (r_{0x}, r_{0y}, r_{0z})^T$ and $\mathbf{R}_j = (R_{jx}, R_{jy}, R_{jz})^T$ (position vector of the sound sources and the observation points) were set in a three-dimensional coordinate as shown in TABLEs II–III. The $\mathbf{r}_i$ (position vector of microphones and loudspeaker units)

TABLE I
PARAMETRIC CONDITIONS IN THE COMPUTER SIMULATION.

Central frequency of sound sources (f_{cent})	250, 500, 1000
Lower frequency of sound sources (f_{low})	$f_{\text{cent}} \div \sqrt{2}$
Upper frequency of sound sources (f_{up})	$f_{\text{cent}} \times \sqrt{2}$
Sound velocity (c)	340 m/s
Number of microphones and loudspeaker units (M)	157
Radiation directivity of loudspeaker units (D_{si})	Omnidirectional, Decay 20 dB, Unidirectional, Shotgun

TABLE II
POSITION COORDINATES OF SOUND SOURCES IN THE COMPUTER SIMULATION.

Index	r_{0x}	r_{0y}	r_{0z}	Index	r_{0x}	r_{0y}	r_{0z}
1	0.5	0.25	0.25	16	0.5	0.25	0
2	0.5	0	0.25	17	0.5	0	0
3	0.5	-0.25	0.25	18	0.5	-0.25	0
4	0.25	0.25	0.25	19	0.25	0.25	0
5	0.25	0	0.25	20	0.25	0	0
6	0.25	-0.25	0.25	21	0.25	-0.25	0
7	0	0.25	0.25	22	0	0.25	0
8	0	0	0.25	23	0	0	0
9	0	-0.25	0.25	24	0	-0.25	0
10	-0.25	0.25	0.25	25	-0.25	0.25	0
11	-0.25	0	0.25	26	-0.25	0	0
12	-0.25	-0.25	0.25	27	-0.25	-0.25	0
13	-0.5	0.25	0.25	28	-0.5	0.25	0
14	-0.5	0	0.25	29	-0.5	0	0
15	-0.5	-0.25	0.25	30	-0.5	-0.25	0

TABLE III
POSITION COORDINATES OF OBSERVATION POINTS IN THE COMPUTER SIMULATION.

j	R_{jx}	R_{jy}	R_{jz}	j	R_{jx}	R_{jy}	R_{jz}
1	0.75	0.5	0.5	10	0.75	0.5	0
2	0.75	0	0.5	11	0.75	0	0
3	0.75	-0.5	0.5	12	0.75	-0.5	0
4	0	0.5	0.5	13	0	0.5	0
5	0	0	0.5	14	0	-0.5	0
6	0	-0.5	0.5	15	-0.75	0.5	0
7	-0.75	0.5	0.5	16	-0.75	0	0
8	-0.75	0	0.5	17	-0.75	-0.5	0
9	-0.75	-0.5	0.5				

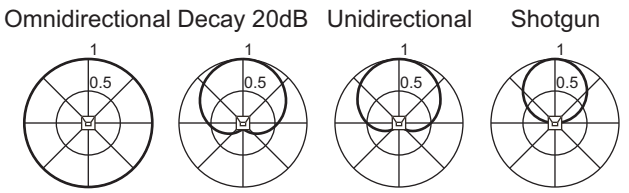

Fig. 9. Radiation directivity patterns of the loudspeaker units used in the computer simulation.

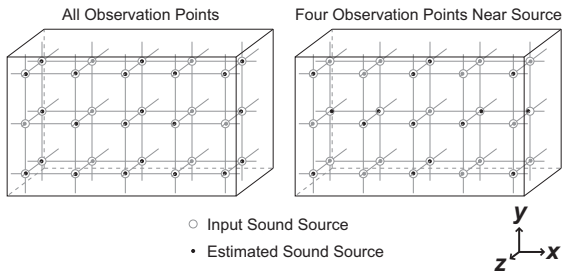

Fig. 10. Results of estimated sound image positions in the original sound field in the computer simulation.

was set in a three-dimensional coordinate as follows:

$$\mathbf{r}_i = \begin{cases} \begin{pmatrix} 0.5725\times(-1)^{i-1} \\ 0.125\times Q(\mathrm{fix}(\frac{i-1}{2}),4) - 0.25 \\ 0.125\times R(\mathrm{fix}(\frac{i-1}{2}),4) - 0.125 \end{pmatrix} \\ \qquad (i = 1\sim 40) \\ \begin{pmatrix} 0.125\times Q(\mathrm{fix}(\frac{i-41}{2}),4) - 0.5 \\ 0.3475\times(-1)^{i-41} \\ 0.125\times R(\mathrm{fix}(\frac{i-41}{2}),4) - 0.125 \end{pmatrix} \\ \qquad (i = 41\sim 112) \\ \begin{pmatrix} 0.125\times Q(i-113,5) - 0.5 \\ 0.125\times R(i-113,5) - 0.25 \\ 0.305 \end{pmatrix} \\ \qquad (i = 113\sim 157) \end{cases}, \quad (9)$$

where $Q(u,v)$ and $R(u,v)$ denote the quotient and remainder when u is divided by v.

The D_{si} (radiation directivity of the ith loudspeaker unit) is defined as shown in Fig. 9. Under the "Decay 20dB" condition, the radiation directivity of the loudspeaker units is simulated on the basis of the best results of acoustical measurements performed in the developed system [7]. Definitional equations are denoted as follows:

$$\text{(Omnidirectional)} \quad D_{si} = 1, \quad (10)$$

$$\text{(Decay 20dB)} \quad D_{si} = 0.55 + 0.45\cos\theta_{si}, \quad (11)$$

$$\text{(Unidirectional)} \quad D_{si} = \frac{1+\cos\theta_{si}}{2}, \quad (12)$$

$$\text{(Shotgun)} \quad D_{si} = \begin{cases} \cos\theta_{si} & (|\theta_{si}| \le 90^\circ) \\ 0 & (|\theta_{si}| > 90^\circ) \end{cases}, \quad (13)$$

where $\cos\theta_{si} = \frac{\mathbf{n}_{si}\cdot(\mathbf{R}_j - \mathbf{r}_i)}{|\mathbf{n}_{si}||\mathbf{R}_j - \mathbf{r}_i|}$ and $\mathbf{n}_{si}$ (the directional vector of the ith loudspeaker unit) is defined as follows:

$$\mathbf{n}_{si} = \begin{cases} \begin{pmatrix} (-1)^{i-1} & 0 & 0 \end{pmatrix}^T & (i = 1\sim 40) \\ \begin{pmatrix} 0 & (-1)^{i-41} & 0 \end{pmatrix}^T & (i = 41\sim 112) \\ \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}^T & (i = 113\sim 157) \end{cases} \quad (14)$$

The sound image positions $\mathbf{r}_E$ were estimated from the calculated instantaneous sound pressures $p(\mathbf{R}_j, f_{\text{cent}}, t)$ and sound intensity vectors $\mathbf{I}(\mathbf{R}_j, f_{\text{cent}})$ ($|\mathbf{I}(\mathbf{R}_j, f_{\text{cent}})| = 1$) according to following equation:

$$\mathbf{r}_E = \frac{1}{FN} \sum_{f_{\text{cent}}}^{250,500,1000} \sum_{j=1}^{N} \left\{ \mathbf{R}_j - \frac{\mathbf{I}(\mathbf{R}_j, f_{\text{cent}})}{p(\mathbf{R}_j, f_{\text{cent}})} \right\}, \quad (15)$$

where $p(\mathbf{R}_j, f_{\text{cent}})$ is the RMS of the sound pressures calculated from the instantaneous sound pressure $p(\mathbf{R}_j, f_{\text{cent}}, t)$ in $\mathbf{R}_j$ according to following equation:

$$p(\mathbf{R}_j, f_{\text{cent}}) = \sqrt{\frac{1}{T}\int_0^T \{p(\mathbf{R}_j, f_{\text{cent}}, t)\}^2 dt}, \quad (16)$$

where T(=1 s) is the time length. $\mathbf{r}_E$ was calculated after $p(\mathbf{R}_j, f_{\text{cent}})$ was normalized so that the variance of the estimated sound image positions was minimized. $F(=3)$ and N denote the number of octave-band noises and observation points used in the estimation of the sound image positions, respectively. In this paper, two cases ("all observation points" and "four observation points near source") were set in the number of observation points N. In the case of "all observation points," the sound image positions were estimated by using the $p(\mathbf{R}_j, f_{\text{cent}})$ and $\mathbf{I}(\mathbf{R}_j, f_{\text{cent}})$ in all observation points. In the case of "four observation points near source," the sound image positions were estimated by using the $p(\mathbf{R}_j, f_{\text{cent}})$ and $\mathbf{I}(\mathbf{R}_j, f_{\text{cent}})$ in four observation points near input sound sources.

In order to evaluate the performance of the sound image position estimation method used in this paper, the sound source positions were directly estimated in the original sound field. The instantaneous sound pressure $p_0(\mathbf{R}_j, f_{\text{cent}}, t)$ at the jth observation point $\mathbf{R}_j$ is denoted as follows:

$$p_0(\mathbf{R}_j, f_{\text{cent}}, t) = \frac{1}{|\mathbf{R}_j - \mathbf{r}_0|} s\left(t - \frac{|\mathbf{R}_j - \mathbf{r}_0|}{c}\right). \quad (17)$$

The estimated results are shown in Fig. 10. Gray open circles and black filled circles in Fig. 10 respectively denote the position of input sound images listed in Table II and estimated sound images. Note that gray filled circles are plotted instead black filled circles if black filled circles are behind gray open circles. The sound image is shown to be accurately estimated if the position of filled circles is close to gray open circles.

The positions of estimated sound images are almost the same as those of input sound sources in both cases. Thus, the

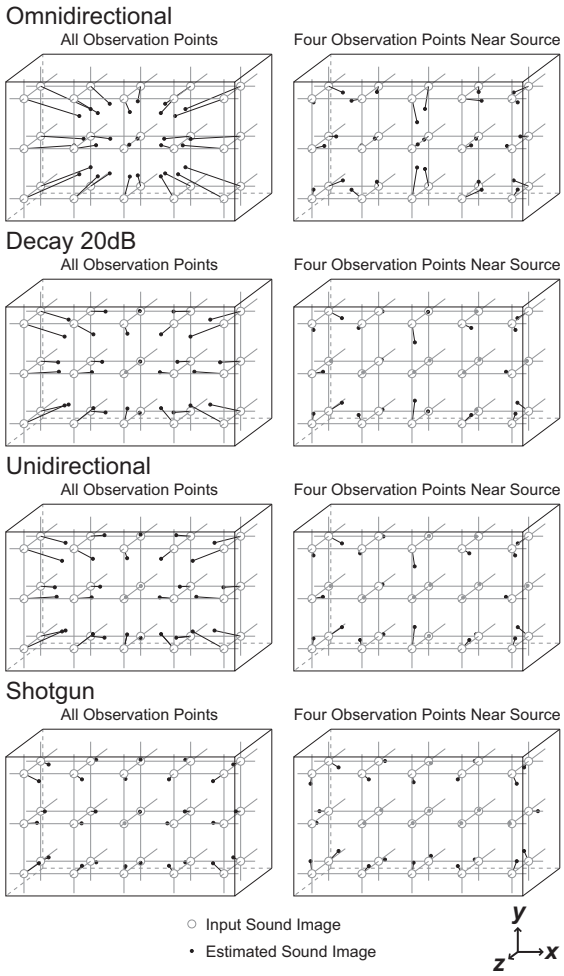

Fig. 11. Results of estimated sound image positions in the reproduced sound field in the computer simulation.

sound image position estimation method used in this paper can accurately estimate the position of sound images.

2) Simulation Results: The results of the estimated sound image positions in the radiation directivity of loudspeaker units and the number of observation points are shown in Fig. 11. Lines from gray open circles to black filled circles denote estimation errors.

When the sound image positions are estimated in all observation points, the sound images are accurately estimated if the radiation directivity of loudspeaker units is sharpened. This is because the 3D sound field is accurately controlled outside of the radiated loudspeaker array by sharpening the radiation directivity of the loudspeaker units. Therefore, if the radiation directivity of the loudspeaker units is sharpened, since sound images are accurately estimated at any observation point, listeners can accurately localize them at any listening position around the radiated loudspeaker array.

On the other hand, when the sound image positions are estimated in four observation points near the input sound source, the sound images are estimated approximately even

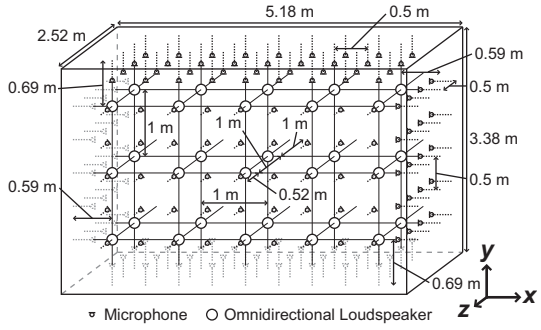

Fig. 12. Arrangement of microphones and omnidirectional loudspeakers for the room impulse response measurement in the acoustical measurement.

if the radiation directivity of the loudspeaker units is not sharpened. This is because the 3D sound field is not accurately controlled in the observation points away from the sound sources, though it is accurately controlled in the observation points near the sound sources when the radiation directivity of loudspeaker units is not sharpened. Therefore, even if the radiation directivity of loudspeaker units is not sharpened, since the sound image close to a particular observation point is accurately estimated, listeners can accurately localize the sound image close to the listening position if they listen to the sound in a particular listening position.

B. Acoustical Measurement

1) Measurement Environment: First, room impulse responses were measured in the surrounding microphone array room. Omnidirectional loudspeakers (Solid Acoustics: SA-355) were placed at the 30 positions shown by the black open circles in Fig. 12. The reverberation time, temperature, and background noise level of the room were respectively about 150 ms, 20°C, and 18.4 dB(A). TSP signals [10] (sampling frequency: 48 kHz, length: 65536 samples) were emitted from the playing equipment (M-Audio: FireWire 410) and omnidirectional loudspeakers, and captured by 157 omnidirectional microphones and the recording equipment (Mark Of The Unicorn: HD192×14). The recording software (Steinberg: Nuendo 3) was installed on four PCs (Apple: Power Mac G5). The sound pressure level was set to 85.6 dB(A) at a distance of one meter from the loudspeakers. Measured room impulse responses (sampling frequency: 48 kHz, length: 14400 samples) were obtained by processing the recorded audio signals on a computer. The number of synchronous repetitions was 16. In order to match the size of the surrounding microphone array to that of the radiated loudspeaker array, measured room impulse responses were regarded as FIR filters, of which the sampling frequency was 192 kHz. As a result, the reverberation time was about 37.5 ms since it became one-fourth.

Second, 157-channel audio signals were synthesized by convolving measured room impulse responses to a sound source signal on a computer. The sound source signal was the three octave-band noises listed in Table I. The length, sampling frequency and quantization bit of the sound source signal were

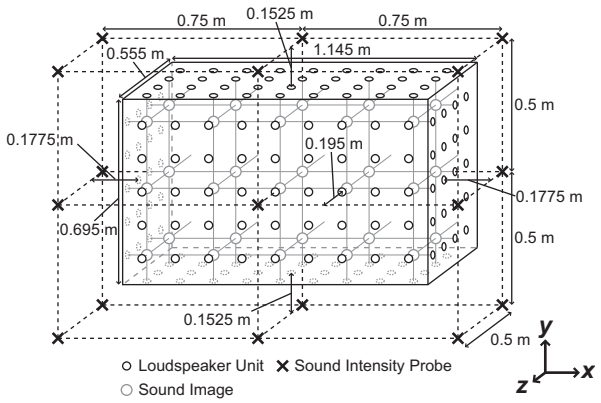

Fig. 13. Arrangement of loudspeaker units, sound images and intensity probes for the sound intensity measurement in the acoustical measurement.

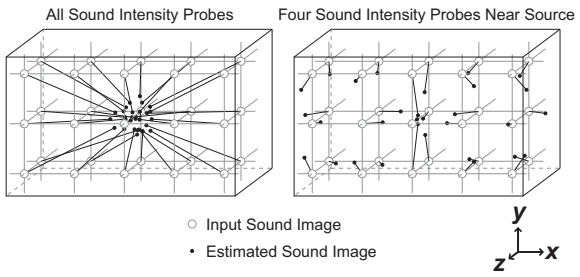

Fig. 14. Results of estimated sound image positions in the acoustical measurement.

10 seconds, 48 kHz, and 16 bits. The top and bottom of the sound source signal was faded in and out at a length of 1 ms. The sampling frequency of measured room impulse responses was transformed from 192 kHz to 48 kHz before the synthesis because the measured room impulse responses were regarded as the FIR filters of which the sampling frequency is 192 kHz.

Third, 157-channel audio signals (sampling frequency: 48 kHz, quantization bit: 16 bits) were played by the playing equipment (Digidesign: ProTools HD) and the radiated loudspeaker array. The channels were assigned so that the position of microphones was the same as that of the loudspeaker units. The playing software (Digidesign: ProTools HD) was installed in a PC (Apple: Mac Pro). The reverberation time, temperature, and background noise level of the playing room were respectively about 180 ms, 22°C, and 22 dB(A). As a result, the sound images were generated at the 30 positions shown by gray open circles in Fig. 13. The sound pressure level was set to 71 dB(A) at a distance of one meter from the sound image generated at the center of the radiated loudspeaker array. On the outside of the radiated loudspeaker array, sound intensity probes (Ono Sokki: MI-6420) were placed at the 17 positions shown in Fig. 13. Sound pressure levels and sound intensity vectors were measured by using the equipment (Ono Sokki: DS-2100 & DS-0285) and software (Ono Sokki: DS-0225) for sound intensity analysis.

The sound image positions $\mathbf{r}_E$ were estimated from the measured sound pressure levels and sound intensity vectors according to following equation:

$$\mathbf{r}_E = \frac{1}{FN} \sum_{f_{\text{cent}}}^{250,500,1000} \sum_{j=1}^{N} \left\{ \mathbf{R}_j - \frac{\mathbf{I}(\mathbf{R}_j, f_{\text{cent}})}{p(\mathbf{R}_j, f_{\text{cent}})} \right\}, \qquad (18)$$

where $\mathbf{R}_j$ denotes the position vector of the jth sound intensity probe, $\mathbf{I}(\mathbf{R}_j, f_{\text{cent}})$ denotes the sound intensity vector measured in $\mathbf{R}_j$ ($|\mathbf{I}(\mathbf{R}_j, f_{\text{cent}})| = 1$), and $p(\mathbf{R}_j, f_{\text{cent}})$ denotes the sound pressure level measured in $\mathbf{R}_j$. $\mathbf{r}_E$ was calculated after $p(\mathbf{R}_j, f_{\text{cent}})$ was normalized so that the variance of the estimated sound image positions was minimized. $F(= 3)$ and N denote the respective number of octave-band noises and sound intensity probes used in the estimation of the sound image positions. In this paper, two cases were set in the number of sound intensity probes N. The conditions of the two cases were the same as those in the computer simulation.

2) Measurement Results: The results of the estimated sound image positions in each number of the sound intensity probes are shown in Fig. 14. When the sound image positions are estimated in four sound intensity probes near the input sound source, the sound images are estimated approximately. This result is almost the same as that when the radiation directivity of loudspeaker units is omnidirectional in the computer simulation. Thus, the radiation directivity of loudspeaker units in the developed system is indicated as generally omnidirectional, though the result of the radiation directivity of loudspeaker units in the developed system is "Decay 20dB," as shown in Fig. 9. However, listeners can accurately localize the sound image close to the listening position if they listen to the sound in a particular listening position, as described in the computer simulation results.

When the sound image positions are estimated in all observation points, the sound images are not accurately estimated and the estimated positions are biased to the center of the radiated loudspeaker array compared with those when the radiation directivity of loudspeaker units is omnidirectional in the computer simulation. This is because the radiation directivity of loudspeaker units is generally omnidirectional and the effect of reflection sounds of the measured rooms is added as a measurement error. However, listeners can accurately localize the sound image in any listening position around the radiated loudspeaker array if the radiation directivity of loudspeaker units is sharpened, as described in the computer simulation results.

V. Conclusion

In this paper, in order to evaluate the performance of the sound image localization in the developed near 3D sound field display systems using directional loudspeakers and wave field synthesis, the positions of sound images were estimated via computer simulation. The results indicated that listeners can accurately localize the sound image in any listening position around the radiated loudspeaker array if the radiation directivity of loudspeaker units is sharpened, and listeners can accurately localize the sound image close to the listening

position if they listen to the sound in a particular listening position, even if the radiation directivity of the loudspeaker units is not sharpened. Results of estimation of a sound image position by acoustical measurement similarly showed that listeners can accurately localize the sound image close to the listening position if they listen to a sound in a particular listening position.

A subject for future work is development of a system with better performance in a real environment by sharpening the radiation directivity of the loudspeaker units. On the other hand, a cubic 3D visual display system, in which several people can view an object anywhere in its vicinity without having to wear equipment such as glasses, has also been developed [11]. Therefore development, by combining the proposed system and the system described above, is needed for a 3D visual and audio system in which several people can view an object and listen to a sound anywhere in its vicinity without having to wear equipment such as glasses and headphones.

REFERENCES

[1] *3D Spatial Image and Sound Group, Universal Media Research Center, National Institute of Information and Communications Technology*, http://www2.nict.go.jp/x/x171/index_e.html.

[2] M. Camras, "Approach to recreating a sound field," *J. Acoust. Soc. Am.*, vol. 43, no. 6, pp. 1425–1431, June 1968.

[3] 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, May 1993.

[4] T. Kimura and K. Kakehi, "Effects of directivity of microphones and loudspeakers in sound field reproduction based on wave field synthesis," in *Proc. ICA*, no. RBA-15-011, Madrid, Spain, September 2007, pp. 1–6.

[5] B. B. Baker and E. T. Copson, *The Mathematical Theory of Huygens' Principle*, 2nd ed. London, UK: Oxford University Press, 1950, pp. 23–26.

[6] T. Kimura, Y. Yamakata, and M. Katsumoto, "Theoretical study of near 3D sound field reproduction based on wave field synthesis," in *Proc. Acoustics'08*, no. 1763, Paris, France, June/July 2008, pp. 4585–4590.

[7] T. Kimura, Y. Yamakata, M. Katsumoto, T. Okamoto, S. Yairi, Y. Iwaya, and Y. Suzuki, "Development of real system in near 3D sound field reproduction system using directional loudspeakers and wave field synthesis," in *Proc. WESPAC*, no. 0164, Beijing, China, September 2009, pp. 1–6.

[8] T. Okamoto, R. Nishimura, and Y. Iwaya, "Estimation of sound source positions using a surrounding microphone array," *Acoust. Sci. & Tech.*, vol. 28, no. 3, pp. 181–189, May 2007.

[9] F. J. Fahy, *Sound Intensity*. UK: Spon Press, 1995.

[10] 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, February 1995.

[11] S. Yoshida, R. Lopez-Gulliver, S. Yano, H. Ando, and N. Inoue, "gCubik: Implementation of graspable cubic auto-stereoscopic display for multi-user environments," *J.Inst.Image Inform.TV.Engnr*, vol. 64, no. 4, pp. 570–576, July 2010.