Theoretical Study and Numerical Analysis of 3D Sound Field Reproduction System Based on Directional Microphones and Boundary Surface Control

Toshiyuki Kimura[1]

[1]*National Institute of Information and Communications Technology, Koganei, Tokyo, 184-8795, Japan*

Correspondence should be addressed to Toshiyuki Kimura (t-kimura@nict.go.jp)

ABSTRACT

Three-dimensional sound field reproduction using directional microphones and wave field synthesis can be used to synthesize wave fronts in a listening area using directional microphones and loudspeakers placed at the boundary of the area; the position of the loudspeakers is the same as that of microphones in this technique. Thus, it is very difficult to construct an audio-visual virtual reality system using this technique because the screen or display of the visual system cannot be placed at the position of loudspeakers. In order to reproduce the 3D sound field of the listening area even when the loudspeakers are not placed at the boundary of the area, this paper proposes a 3D sound field reproduction system using directional microphones and boundary surface control. Results of a computer simulation show that the proposed system can reproduce the 3D sound field in the listening area more accurately than the conventional system.

1. INTRODUCTION

Three-dimensional sound field reproduction techniques that could be used in auditory display systems were recently investigated. If these techniques are practically applied, people in different places can experience teleconferencing as though they are in the same conference room (teleconferencing system) or they can play music as though they are in the same concert hall (tele-ensemble system). Since these systems facilitate more realistic experiences than conventional auditory display systems (telephones and 5.1 ch audio systems), in general, telecommunication will be more realistic and popular if these systems are used.

Wave field synthesis [1, 2, 3] is a 3D sound field reproduction technique that synthesizes wave fronts on the basis of Huygens' principle. In this technique, the original sound in a control area is captured using a microphone array and it is then reproduced in a listening area using a loudspeaker array. The arrays are placed at the boundaries of their respective areas. The positions of the microphones and the loudspeakers are identical in their respective areas. In this technique, information on the position of listeners and the direction faced by them in the listening area is not required; in contrast, this information is required in other sound field reproduction techniques such as binaural [4] and transaural [5] techniques.

The mathematical description of Huygens' principle is as follows: sound pressures in a listening area can be accurately reproduced if sound pressures and their gradients on the boundary surface are controlled using the Kirchhoff-Helmholtz integral equation [6]. If there are sound sources outside a space V and $\mathbf{r}$ denotes the position vector inside the space as shown on the left side in Fig. 1, $P(\mathbf{r},\omega)$ (the sound pressure at $\mathbf{r}$) can be expressed using the Kirchhoff-Helmholtz integral equation as follows:

$$P(\mathbf{r},\omega) = \oint_S \left\{ \frac{\partial P(\mathbf{r}_S,\omega)}{\partial \mathbf{n}_S} G(\mathbf{r}_S|\mathbf{r},\omega) - P(\mathbf{r}_S,\omega)\frac{\partial G(\mathbf{r}_S|\mathbf{r},\omega)}{\partial \mathbf{n}_S} \right\} dS \quad (\mathbf{r}\in V), \qquad (1)$$

where S is the continuous boundary surface of V, $\mathbf{r}_S$ is the position vector on S, and $\mathbf{n}_S$ is the normal unit vector directed toward the outside of V at $\mathbf{r}_S$. $G(\mathbf{r}_S|\mathbf{r},\omega)(=\frac{\exp(-jk|\mathbf{r}_S-\mathbf{r}|)}{4\pi|\mathbf{r}_S-\mathbf{r}|})$ corresponds to the acoustic transfer func-

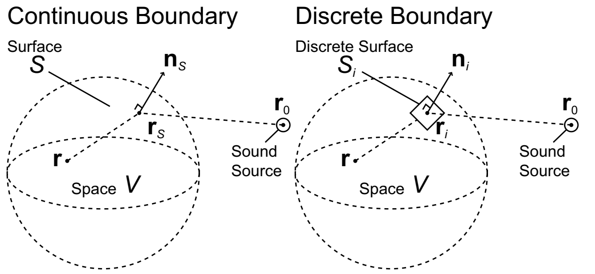

Fig. 1: Coordinates used in the Kirchhoff-Helmholtz integral equation.

tion from $\mathbf{r}_S$ to $\mathbf{r}$. Note that $k(=\omega/c)$ is the wave number, and c is the sound velocity. Eq. (1) shows that the sound pressure in V can be reproduced if sound pressures $P(\mathbf{r}_S,\omega)$ and their gradients $\partial P(\mathbf{r}_S,\omega)/\partial \mathbf{n}_S$ at $\mathbf{r}_S$ are recorded in the original sound field and monopole sound sources with an amplitude $\partial P(\mathbf{r}_S,\omega)/\partial \mathbf{n}_S$ and dipole sound sources with an amplitude $-P(\mathbf{r}_S,\omega)$ are played at $\mathbf{r}_S$ in the reproduced sound field.

In order to develop a sound field reproduction system, S must be discretized since the monopole and dipole sources are not placed continuously along S. As shown on the right side in Fig. 1, if $\mathbf{r}_i$ is the position vector of S_i (the ith element of a discrete boundary surface) and if $P(\mathbf{r}_i,\omega)$ (the sound pressure at $\mathbf{r}_i$) and $\partial P(\mathbf{r}_i,\omega)/\partial \mathbf{n}_i$ (the sound pressure gradient at $\mathbf{r}_i$) are constant for S_i, Eq. (1) can be rewritten as follows:

$$P(\mathbf{r},\omega)=\sum_{i=1}^{M}\left\{\frac{\partial P(\mathbf{r}_i,\omega)}{\partial \mathbf{n}_i}G(\mathbf{r}_i|\mathbf{r},\omega)-P(\mathbf{r}_i,\omega)\frac{\partial G(\mathbf{r}_i|\mathbf{r},\omega)}{\partial \mathbf{n}_i}\right\}\Delta S_i \quad (\mathbf{r}\in V), \qquad (2)$$

where M is the total number of elements in the discrete boundary surface, ΔS_i is the area of S_i, and $\mathbf{n}_i$ is the normal unit vector directed toward the outside of the discrete boundary surface at $\mathbf{r}_i$. Eq. (2) shows that the sound pressure in V can be reproduced if sound pressures $P(\mathbf{r}_i,\omega)$ and their gradients $\partial P(\mathbf{r}_i,\omega)/\partial \mathbf{n}_i$ at M points are recorded in the original sound field and monopole sound sources with an amplitude $\partial P(\mathbf{r}_i,\omega)/\partial \mathbf{n}_i$ and dipole sound sources with an amplitude $-P(\mathbf{r}_i,\omega)$ are played at M points in the reproduced sound field.

From a practical point of view, it is very difficult to realize sound field reproduction systems based on Eq. (2) because the monopole sound source and the dipole sound source cannot be placed at identical positions corresponding to the M points in the reproduced sound field. On the other hand, if approximations are introduced in Eq. (2), the following equation (known as the Fresnel-Kirchhoff diffraction formula [7]) can be derived:

$$P(\mathbf{r},\omega)=jk\sum_{i=1}^{M}P(\mathbf{r}_i,\omega)G(\mathbf{r}_i|\mathbf{r},\omega)(\cos\theta_i-\cos\theta_{i0})\Delta S_i \quad (\mathbf{r}\in V), \qquad (3)$$

where $\cos\theta_i(=\frac{\mathbf{n}_i\cdot(\mathbf{r}_i-\mathbf{r})}{|\mathbf{n}_i||\mathbf{r}_i-\mathbf{r}|})$ denotes the cosine of the angle between the vector $\mathbf{n}_i$ and the vector $\mathbf{r}_i-\mathbf{r}$ and $\cos\theta_{i0}(=\frac{\mathbf{n}_i\cdot(\mathbf{r}_i-\mathbf{r}_0)}{|\mathbf{n}_i||\mathbf{r}_i-\mathbf{r}_0|})$ represents the cosine of the angle between the vector $\mathbf{n}_i$ and the vector $\mathbf{r}_i-\mathbf{r}_0$; $\mathbf{r}_0$ denote the position vectors of the sound sources. Further, since the direction of the vector $\mathbf{r}_i-\mathbf{r}$ is almost the same as that of the vector $\mathbf{n}_i$, we can use the approximation $\cos\theta_i\approx 1$. Thus, Eq. (3) can also be written as

$$P(\mathbf{r},\omega)=jk\sum_{i=1}^{M}D_m(\mathbf{r}_0|\mathbf{r}_i)P(\mathbf{r}_i,\omega)G(\mathbf{r}_i|\mathbf{r},\omega)\Delta S_i \quad (\mathbf{r}\in V), \qquad (4)$$

where $D_m(\mathbf{r}_0|\mathbf{r}_i)(=1-\cos\theta_{i0})$ corresponds to the directivities of the microphones placed at $\mathbf{r}_i$. Eq. (4) shows that the sound pressure in V can be reproduced if the sound pressures $P(\mathbf{r}_i,\omega)$ at the M points are recorded by directional microphones in the original sound field and monopole sound sources with an amplitude $jkD_m(\mathbf{r}_0|\mathbf{r}_i)P(\mathbf{r}_i,\omega)$ are played at M points in the reproduced sound field. Since the sound pressure gradients can be controlled by the directivity of the microphones, only M microphones and loudspeakers are required to control sound pressures on the boundary surface in the system based on Eq. (4); in contrast, $2M$ microphones and loudspeakers are needed to control both sound pressures and their gradients on the boundary surface in the system based on Eq. (2).

Sound field reproduction systems, in which the microphones and loudspeakers are placed in a line or on a plane and sound pressures on the boundary are controlled, have been proposed [1, 2]. In such a system, the sound field in front of the listeners' position in the original area is presented to the listeners by playing the recording sound through the loudspeakers placed in front of the listeners; this leads to the listeners perceiving the original

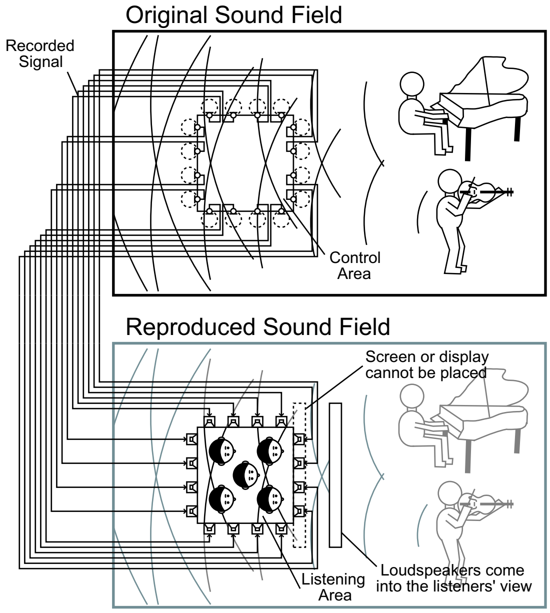

Fig. 2: Conventional 3D sound field reproduction system; the system involves the use of directional microphones and wave field synthesis.

frontal sound field. However, in the case of sound arriving from all directions, the original spatial sound field cannot be reproduced using this system. On the other hand, as shown in Fig. 2, a sound field reproduction system in which directional microphones and loudspeakers are placed around the listeners according to Eq. (4) to reproduce the original spatial sound field has been proposed, and the conditions under which wave fronts are accurately reproduced have been investigated [3].

In the conventional system shown in Fig. 2, the loudspeakers are placed on the boundary surface of the listening area. Thus, when visual systems are introduced in the same listening area for an audio-visual presentation, the screen or display of the visual system should be placed on or outside the boundary surface, as shown in the lower side of Fig. 2. However, the screen or display cannot be placed on the boundary surface, and the loudspeakers obstruct the audiences' field of view if the screen or display is placed outside the boundary surface. Thus, in order to solve these problems, it is necessary to develop a technique that does not require the positioning of loudspeakers on the boundary surface of the listening area for reproducing the 3D sound field.

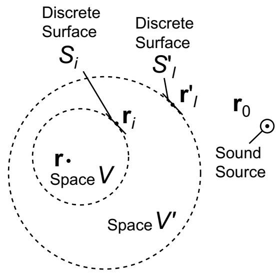

Fig. 3: Coordinates in the mathematical derivation of the principle of the proposed system.

In this study, a novel 3D sound field reproduction system based on directional microphones and boundary surface control [8] is proposed to reproduce the 3D sound field in the listening area without requiring loudspeakers to be placed at the boundary of the area. In Section 2, the theoretical basis of the proposed system is discussed. The 3D sound field in the listening area can be accurately reproduced by carrying out inverse filtering based on acoustic transfer functions, even when the loudspeakers are not placed at the boundary of the listening area. In Section 3, to evaluate the validity of the proposed system, a computer simulation of the implementation of the proposed system in a listening room is performed and the 3D sound field reproduced in the simulation is numerically analyzed.

2. THEORETICAL STUDY

First, the theoretical basis of the proposed system is presented. The 3D sound field in the listening area can be accurately reproduced by introducing a inverse filtering to Eq. (4), even when loudspeakers are not placed at the boundary of the listening area.

As shown in Fig. 3, a space V' is considered outside the space V (i.e., $V \in V'$). Let S'_l ($l = 1...N$) be the lth discrete boundary surface of V'. $\mathbf{r}'_l$ and N are the position vector and total number of discrete surfaces, respectively. To reproduce the 3D sound field in V', the following equation, derived from Eq. (4), is used:

$$P(\mathbf{r},\omega) = jk \sum_{l=1}^{N} D_m(\mathbf{r}_0|\mathbf{r}'_l) P(\mathbf{r}'_l,\omega) G(\mathbf{r}'_l|\mathbf{r},\omega) \Delta S'_l \quad (\mathbf{r} \in V'), \qquad (5)$$

where $\Delta S'_l$ is the area of S'_l. Since $\mathbf{r}_i$ is always in V', the following equation can be derived from Eq. (5):

$$P(\mathbf{r}_i,\omega) = jk\sum_{l=1}^{N} D_m(\mathbf{r}_0|\mathbf{r}'_l)P(\mathbf{r}'_l,\omega)G(\mathbf{r}'_l|\mathbf{r}_i,\omega)\Delta S'_l \quad (\mathbf{r}_i \in V'). \tag{6}$$

On the other hand, if the 3D sound field in which there are point sources at $\mathbf{r}'_l$ is reproduced in V, the following equation can be derived using Eq. (4):

$$P(\mathbf{r},\omega) = jk\sum_{i=1}^{M} D_m(\mathbf{r}'_l|\mathbf{r}_i)P(\mathbf{r}_i,\omega)G(\mathbf{r}_i|\mathbf{r},\omega)\Delta S_i \quad (\mathbf{r} \in V). \tag{7}$$

If Eq. (6) is substituted in Eq. (7), the following equation is obtained:

$$P(\mathbf{r},\omega) = jk\sum_{l=1}^{N} D_m(\mathbf{r}_0|\mathbf{r}'_l)P(\mathbf{r}'_l,\omega) \left\{ jk\sum_{i=1}^{M} D_m(\mathbf{r}'_l|\mathbf{r}_i)G(\mathbf{r}'_l|\mathbf{r}_i,\omega)G(\mathbf{r}_i|\mathbf{r},\omega)\Delta S_i \right\}\Delta S'_l \quad (\mathbf{r} \in V, \mathbf{r}_i \in V'). \tag{8}$$

Thus, by comparing Eq. (8) with Eq. (5), the following equation is derived:

$$G(\mathbf{r}'_l|\mathbf{r},\omega) = jk\sum_{i=1}^{M} D_m(\mathbf{r}'_l|\mathbf{r}_i)G(\mathbf{r}'_l|\mathbf{r}_i,\omega) G(\mathbf{r}_i|\mathbf{r},\omega)\Delta S_i \quad (\mathbf{r} \in V, \mathbf{r}_i \in V'). \tag{9}$$

The following scheme is considered: the sound is recorded by M directional microphones placed at $\mathbf{r}_i$, the recorded signals are processed by M-input N-output filters, and the filtered signals are played by the N loudspeakers placed at $\mathbf{r}'_l$. If all the M recorded signals are denoted as $D_m(\mathbf{r}_0|\mathbf{r}_i)P(\mathbf{r}_i,\omega)$, the N filtered signals $P'(\mathbf{r}'_l,\omega)$ ($l = 1...N$) are expressed as follows:

$$P'(\mathbf{r}'_l,\omega) = \sum_{i=1}^{M} H_{li}(\omega)D_m(\mathbf{r}_0|\mathbf{r}_i)P(\mathbf{r}_i,\omega), \tag{10}$$

where $H_{li}(\omega)$ are the coefficients of the M-input N-output filters. Thus, the reproduced 3D sound field $P'(\mathbf{r},\omega)$ is written as

$$P'(\mathbf{r},\omega) = \sum_{l=1}^{N} P'(\mathbf{r}'_l,\omega)G(\mathbf{r}'_l|\mathbf{r},\omega)\Delta S'_l$$
$$= \sum_{i=1}^{M} D_m(\mathbf{r}_0|\mathbf{r}_i)P(\mathbf{r}_i,\omega) \left\{ \sum_{l=1}^{N} H_{li}(\omega)G(\mathbf{r}'_l|\mathbf{r},\omega)\Delta S'_l \right\}. \tag{11}$$

If Eq. (9) is substituted in Eq. (11), the following equation is obtained:

$$P'(\mathbf{r},\omega) = jk\sum_{i=1}^{M} D_m(\mathbf{r}_0|\mathbf{r}_i)P(\mathbf{r}_i,\omega) \left[\sum_{n=1}^{M} G(\mathbf{r}_n|\mathbf{r},\omega) \left\{ \sum_{l=1}^{N} H_{li}(\omega) D_m(\mathbf{r}'_l|\mathbf{r}_n)G(\mathbf{r}'_l|\mathbf{r}_n,\omega)\Delta S'_l \right\}\Delta S_n \right]. \tag{12}$$

If the M-input N-output filters are defined as

$$\sum_{l=1}^{N} H_{li}(\omega)D_m(\mathbf{r}'_l|\mathbf{r}_n)G(\mathbf{r}'_l|\mathbf{r}_n,\omega)\Delta S'_l = \begin{cases} 1 & n = i \\ 0 & n \neq i \end{cases}, \tag{13}$$

the 3D sound field is reproduced in V; this is indicated by the following equation derived from Eq. (12):

$$P'(\mathbf{r},\omega) = jk\sum_{i=1}^{M} D_m(\mathbf{r}_0|\mathbf{r}_i)P(\mathbf{r}_i,\omega)G(\mathbf{r}_i|\mathbf{r},\omega)\Delta S_i$$
$$= P(\mathbf{r},\omega) \quad (\mathbf{r} \in V, \mathbf{r}_i \in V'). \tag{14}$$

Thus, if the recorded signals are processed by the filters defined as in Eq. (13), the 3D sound field can be reproduced in the listening area even if the loudspeakers are placed at the position ($\mathbf{r}'_l$), which is not the same as the recorded position ($\mathbf{r}_i$). Since $\mathbf{r}_i$ and $\mathbf{r}'_l$ are points on the boundary surfaces S and S', respectively, it is necessary for the loudspeakers to be placed on the boundary surface (S') which envelops the boundary surface used for recording (S).

Let the matrix form of Eq. (13) be denoted as follows:

$$\mathbf{G}(\omega)\mathbf{H}(\omega) = \mathbf{I} \tag{15}$$

$$\mathbf{G}(\omega) = \begin{pmatrix} D_{m11}G_{11} & \dots & D_{mN1}G_{N1} \\ \vdots & \ddots & \vdots \\ D_{m1M}G_{1M} & \dots & D_{mNM}G_{NM} \end{pmatrix} \tag{16}$$

$$\mathbf{H}(\omega) = \begin{pmatrix} H_{11}(\omega) & \dots & H_{1M}(\omega) \\ \vdots & \ddots & \vdots \\ H_{N1}(\omega) & \dots & H_{NM}(\omega) \end{pmatrix} \tag{17}$$

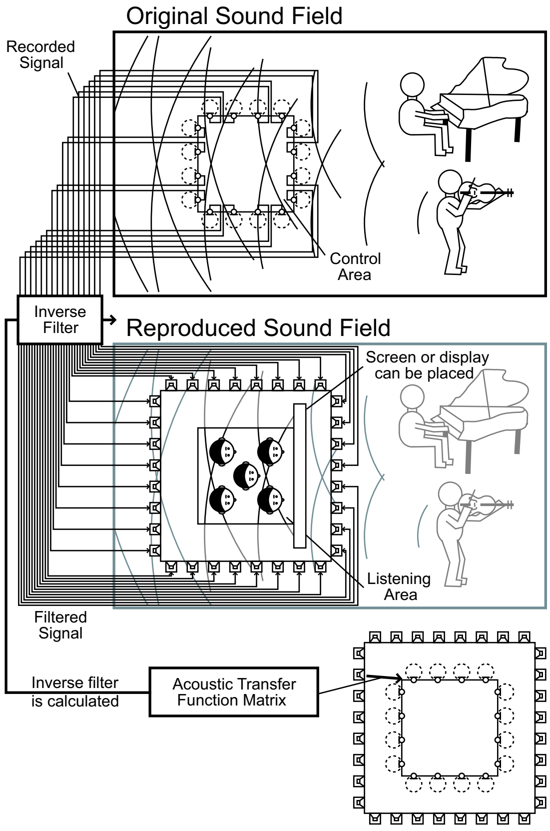

Fig. 4: Proposed 3D sound field reproduction system based on directional microphones and boundary surface control.

$$\mathbf{I} = \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix}, \tag{18}$$

where $D_{mli}G_{li}$ denotes $D_m(\mathbf{r}'_l|\mathbf{r}_i)G(\mathbf{r}'_l|\mathbf{r}_i,\omega)\Delta S'_l$. Thus, the filters $\mathbf{H}(\omega)$ are calculated according to the following equation:

$$\mathbf{H}(\omega) = \mathbf{G}^+(\omega), \tag{19}$$

where $\mathbf{G}^+(\omega)$ is the Moore-Penrose pseudo inverse matrix of $\mathbf{G}(\omega)$. Since $\mathbf{G}(\omega)$ is the matrix consisting of the acoustic transfer functions from $\mathbf{r}'_l$ to $\mathbf{r}_i$, the calculated filters $\mathbf{H}(\omega)$ are inverse filters of the acoustic transfer functions.

The diagram of the proposed 3D sound field reproduction system based on directional microphones and boundary surface control is shown in Fig. 4. First, in the original sound field, M directional microphones are placed at the points $\mathbf{r}_i$ on the boundary surface of the control area and the sound $D_m(\mathbf{r}_0|\mathbf{r}_i)P(\mathbf{r}_i,\omega)$ is recorded. The directional microphones are then directed toward the outside of the control area. Second, in the reproduced sound field, M directional microphones are placed at the points $\mathbf{r}_i$ on the boundary surface of the listening area, and N loudspeakers are placed at the points $\mathbf{r}'_l$ on a boundary surface enveloping the listening area. The position and direction of the M directional microphones are the same as those of the microphones during recording. Third, the acoustic transfer functions from the N loudspeakers to the M directional microphones $D_m(\mathbf{r}'_l|\mathbf{r}_i)G(\mathbf{r}'_l|\mathbf{r}_i,\omega)\Delta S'_l$ are measured, and the inverse filters $H_{li}(\omega)$ are calculated from the measured acoustic transfer functions. Finally, the recorded signals are filtered by the inverse filters, and the filtered signals $P'(\mathbf{r}'_l,\omega)$ are played by the N loudspeakers. As a result, since the 3D sound field is reproduced in the listening area, listeners feel as if they are listening to the sound in the original sound field. Since it is not necessary for the directional microphones to be placed in the reproduced sound field when the sound is played, it is possible to construct an audio-visual system in which the screen or display of the visual system is placed on or outside the boundary surface of the listening area. Since the sound pressure at $\mathbf{r}_i$ is the same as the recorded sound $D_m(\mathbf{r}_0|\mathbf{r}_i)P(\mathbf{r}_i,\omega)$ in the reproduced sound field, the sound sources can be placed at arbitrary positions outside the control area. Thus, if the sound sources are placed outside the control area and within the loudspeaker array, listeners can be made to feel as if there are sound images before the loudspeaker array.

When the 3D sound field reproduction system is constructed in a real environment, since the acoustic transfer functions include an initial delay, filters that do not satisfy the causality are obtained if the inverse filters are calculated using Eq. (19). In order to obtain the filters satisfying the causality, the inverse filters are calculated using the following equations:

$$\mathbf{H}(\omega) = \mathbf{G}^+(\omega)\mathbf{I}(\omega) \tag{20}$$

$$\mathbf{I}(\omega) = \begin{pmatrix} e^{-j\omega T} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & e^{-j\omega T} \end{pmatrix}, \tag{21}$$

where T is the delay time required for the calculation of the inverse filters satisfying the causality. The total delay of the system is T and the reproduced 3D sound field can be expressed as follows:

$$P'(\mathbf{r},\omega) = e^{-j\omega T} jk \sum_{i=1}^{M} D_m(\mathbf{r}_0|\mathbf{r}_i) P(\mathbf{r}_i,\omega) G(\mathbf{r}_i|\mathbf{r},\omega)\Delta S_i$$
$$= e^{-j\omega T} P(\mathbf{r},\omega) \quad (\mathbf{r} \in V, \mathbf{r}_i \in V'). \qquad (22)$$

3. NUMERICAL ANALYSIS

3.1. Simulation Environment

The original sound field was formed in free space where there were no reflection sounds. The control area in the original sound field was a square prism with dimensions of 2 m (width) × 2 m (depth) × 1 m (height). The directional microphone array was placed on the surface of the square prism. The reproduced sound field was considered to be formed in free space in order to simplify the computer simulation. The listening area in the reproduced sound field corresponded to the area occupied by the square prism, whose size was the same as that of the control area. Further, another square prism with dimensions of 4 m (width) × 4 m (depth) × 2 m (height) was set outside the listening area. The loudspeaker array was placed on the surface of this square prism.

The directional microphone array in the original sound field and the loudspeaker array in the reproduced sound field are shown in Fig. 5. The gray shaded area in the reproduced sound field in Fig. 5 denotes the listening area. As shown in Fig. 5, directional microphones were placed on six planes of the square prism, with a lattice spacing of Δr_m, and they were directed toward the outside of the control area. Loudspeakers were placed on six planes of the other square prism, with a lattice spacing of Δr_s. The origin of the three-dimensional coordinates was the center of the control and listening areas.

The sound source signal $s(t)$ was a sinusoidal signal with frequency f and amplitude A (i.e., $s(t) = A\sin 2\pi f t$). Let $\mathbf{r}$ be the position vector of an arbitrary point in the control area. Then, $p_o(\mathbf{r},t)$ (the sound pressure at $\mathbf{r}$ in the original sound field) is denoted as follows:

$$p_o(\mathbf{r},t) = \frac{1}{|\mathbf{r}-\mathbf{r}_0|} s\left(t - \frac{|\mathbf{r}-\mathbf{r}_0|}{c}\right)$$

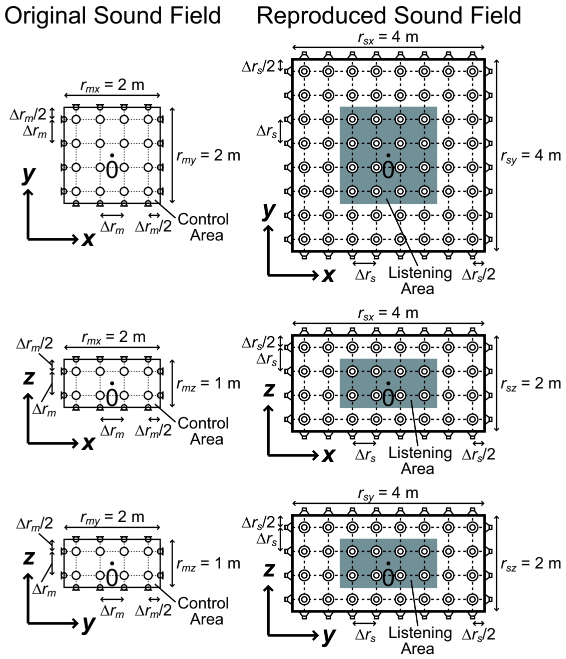

Fig. 5: Directional microphone array in the original sound field and loudspeaker array in the reproduced sound field in the computer simulation.

$$= \frac{A}{|\mathbf{r}-\mathbf{r}_0|} \sin\left\{2\pi f\left(t - \frac{|\mathbf{r}-\mathbf{r}_0|}{c}\right)\right\}, \qquad (23)$$

where $\mathbf{r}_0$ is the position vector of the sound source and c is the sound velocity.

The 3D sound field is reproduced by the proposed system according to the following steps. First, $x_i(t)$ ($i = 1...M$) (the signal recorded by the ith directional microphone at $\mathbf{r}_i$) is expressed as follows:

$$x_i(t) = \frac{D_m(\mathbf{r}_0|\mathbf{r}_i)}{|\mathbf{r}_i-\mathbf{r}_0|} s\left(t - \frac{|\mathbf{r}_i-\mathbf{r}_0|}{c}\right)$$
$$= \frac{D_m(\mathbf{r}_0|\mathbf{r}_i)A}{|\mathbf{r}_i-\mathbf{r}_0|} \sin\left\{2\pi f\left(t - \frac{|\mathbf{r}_i-\mathbf{r}_0|}{c}\right)\right\}, \qquad (24)$$

where $D_m(\mathbf{r}_0|\mathbf{r}_i)$ and $\mathbf{r}_i$ are the directivity and the position vector of the ith directional microphone, respectively. Note that M is the total number of directional microphones. Second, $y_l(t)$ ($l = 1...N$) (the signal obtained by filtering the recorded signal using the inverse filter) is denoted as

$$y_l(t) = \sum_{i=1}^{M} \Xi_{li} x_i\left(t - \frac{\Theta_{li}}{2\pi f}\right)$$

$$= \sum_{i=1}^{M} \frac{\Xi_{li} D_m(\mathbf{r}_0|\mathbf{r}_i) A}{|\mathbf{r}_i - \mathbf{r}_0|} \sin\left\{2\pi f\left(t - \frac{|\mathbf{r}_i - \mathbf{r}_0|}{c}\right) - \Theta_{li}\right\}, \quad (25)$$

where $\Xi_{li}(=|H_{li}(\omega)|)$ and $\Theta_{li}(=\arg H_{li}(\omega))$ denote the absolute value and the angle of the inverse filters. Since a sinusoidal signal is used as the sound source signal, Ξ_{li} and Θ_{li} can be regarded as constants that do not depend on ω in this computer simulation. Since the time length of $s(t)$ is infinite and since it is not necessary to calculate inverse filters satisfying the causality, the inverse filters are obtained using Eq. (19) in this computer simulation. Finally, $p_\mathrm{p}(\mathbf{r}, t)$ (the sound pressure at $\mathbf{r}$ in the reproduced sound field) is expressed using the played signal ($y_l(t)$) by the N loudspeakers placed at $\mathbf{r}'_l$ as follows:

$$p_\mathrm{p}(\mathbf{r}, t) = \sum_{l=1}^{N} \frac{1}{|\mathbf{r} - \mathbf{r}'_l|} y_l\left(t - \frac{|\mathbf{r} - \mathbf{r}'_l|}{c}\right)$$
$$= \sum_{l=1}^{N} \sum_{i=1}^{M} \frac{\Xi_{li} D_m(\mathbf{r}_0|\mathbf{r}_i) A}{|\mathbf{r} - \mathbf{r}'_l||\mathbf{r}_i - \mathbf{r}_0|} \sin\left\{2\pi f\left(t - \frac{|\mathbf{r} - \mathbf{r}'_l| + |\mathbf{r}_i - \mathbf{r}_0|}{c}\right) - \Theta_{li}\right\}. \quad (26)$$

The parametric conditions are listed in Table 1. The microphone interval is 16.67 cm, 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.

$\mathbf{r}_0$, $\mathbf{r}$, $\mathbf{r}_i$, and $\mathbf{r}'_l$ were defined in three-dimensional coordinates as follows:

$$\mathbf{r}_0 = d\mathbf{u} \quad (27)$$

$$\mathbf{r} = \begin{pmatrix} r_x & r_y & r_y \end{pmatrix}^T \quad (|r_x|, |r_y| < 2, |r_z| < 1) \quad (28)$$

$$\mathbf{r}_i = \begin{cases} \begin{pmatrix} \frac{r_{mx}}{2}(-1)^i \\ \Delta r_m R(\frac{i}{2}, \frac{r_{my}}{\Delta r_m}) + \frac{\Delta r_m - r_{my}}{2} \\ \Delta r_m Q(\frac{i}{2}, \frac{r_{my}}{\Delta r_m}) + \frac{\Delta r_m - r_{mz}}{2} \end{pmatrix} & (i = 1\sim144) \\ \begin{pmatrix} \Delta r_m R(\frac{i-144}{2}, \frac{r_{mx}}{\Delta r_m}) + \frac{\Delta r_m - r_{mx}}{2} \\ \frac{r_{my}}{2}(-1)^{i-144} \\ \Delta r_m Q(\frac{i-144}{2}, \frac{r_{mx}}{\Delta r_m}) + \frac{\Delta r_m - r_{mz}}{2} \end{pmatrix} & (i = 145\sim288) \\ \begin{pmatrix} \Delta r_m R(\frac{i-288}{2}, \frac{r_{mx}}{\Delta r_m}) + \frac{\Delta r_m - r_{mx}}{2} \\ \Delta r_m Q(\frac{i-288}{2}, \frac{r_{mx}}{\Delta r_m}) + \frac{\Delta r_m - r_{my}}{2} \\ \frac{r_{mz}}{2}(-1)^{i-288} \end{pmatrix} & (i = 289\sim576) \end{cases} \quad (29)$$

Table 1: Parametric conditions used in the computer simulation.

Amplitude (A)	1
Frequency (f)	63, 125, 250, 500, 1000 Hz
Distance (d)	2, 10, 50 m
Direction vector ($\mathbf{u}$)	$(1,0,0)^T$ $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0)^T$ $(\frac{2}{3}, \frac{2}{3}, \frac{1}{3})^T$
Sound velocity (c)	340 m/s
Microphone number (M)	576
Microphone interval (Δr_m)	0.1667 m
Microphone directivity ($D_m(\mathbf{r}_0\|\mathbf{r}_i)$)	Shotgun
Microphone array size (r_{mx}, r_{my}, r_{mz})	2×2×1 m
Loudspeaker number (N)	2304
Loudspeaker interval (Δr_s)	0.1667 m
Loudspeaker array size (r_{sx}, r_{sy}, r_{sz})	4×4×2 m

$$\mathbf{r}'_l = \begin{cases} \begin{pmatrix} \frac{r_{sx}}{2}(-1)^l \\ \Delta r_s R(\frac{l}{2}, \frac{r_{sy}}{\Delta r_s}) + \frac{\Delta r_s - r_{sy}}{2} \\ \Delta r_s Q(\frac{l}{2}, \frac{r_{sy}}{\Delta r_s}) + \frac{\Delta r_s - r_{sz}}{2} \end{pmatrix} & (l = 1\sim576) \\ \begin{pmatrix} \Delta r_s R(\frac{l-576}{2}, \frac{r_{sx}}{\Delta r_s}) + \frac{\Delta r_s - r_{sx}}{2} \\ \frac{r_{sy}}{2}(-1)^{l-576} \\ \Delta r_s Q(\frac{l-576}{2}, \frac{r_{sx}}{\Delta r_s}) + \frac{\Delta r_s - r_{sz}}{2} \end{pmatrix} & (l = 577\sim1152) \\ \begin{pmatrix} \Delta r_s R(\frac{l-1152}{2}, \frac{r_{sx}}{\Delta r_s}) + \frac{\Delta r_s - r_{sx}}{2} \\ \Delta r_s Q(\frac{l-1152}{2}, \frac{r_{sx}}{\Delta r_s}) + \frac{\Delta r_s - r_{sy}}{2} \\ \frac{r_{sz}}{2}(-1)^{l-1152} \end{pmatrix} & (l = 1153\sim2304) \end{cases}, \quad (30)$$

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

$D_m(\mathbf{r}_0|\mathbf{r}_i)$ (directivity of the ith directional microphone) is the shotgun directivity, using which wave fronts have been accurately reproduced in a past study [3], defined as follows:

$$D_m(\mathbf{r}_0|\mathbf{r}_i) = \begin{cases} \cos\theta_{im} & (|\theta_{im}| \le 90°) \\ 0 & (|\theta_{im}| > 90°) \end{cases}, \quad (31)$$

where $\cos\theta_{im}(= \frac{(\mathbf{r}_0 - \mathbf{r}_i)\cdot\mathbf{n}_{im}}{|\mathbf{r}_0 - \mathbf{r}_i||\mathbf{n}_{im}|})$ is the cosine of the angle between the vector $\mathbf{r}_0 - \mathbf{r}_i$ and the vector $\mathbf{n}_{im}$, and $\mathbf{n}_{im}$ (the directional vector of the ith directional microphone) is defined as follows:

$$\mathbf{n}_{im} = \begin{cases} \left((-1)^i \quad 0 \quad 0\right)^T & (i = 1\sim 144) \\ \left(0 \quad (-1)^{i-144} \quad 0\right)^T & (i = 145\sim 288) \\ \left(0 \quad 0 \quad (-1)^{i-288}\right)^T & (i = 289\sim 576) \end{cases}. \quad (32)$$

On the other hand, in the case of the conventional 3D sound field reproduction system based on Eq. (4), since the 3D sound field is reproduced by playing the recorded sound from the M loudspeakers placed at $\mathbf{r}_i$ in the reproduced sound field, $p_c(\mathbf{r},t)$ (the sound pressure at $\mathbf{r}$ in the reproduced sound field) is denoted as follows:

$$p_c(\mathbf{r},t) = \sum_{i=1}^{M} \frac{D_m(\mathbf{r}_0|\mathbf{r}_i)A}{|\mathbf{r}-\mathbf{r}_i||\mathbf{r}_i-\mathbf{r}_0|} \sin\left\{2\pi f\left(t - \frac{|\mathbf{r}-\mathbf{r}_i|+|\mathbf{r}_i-\mathbf{r}_0|}{c}\right)\right\}. \quad (33)$$

3.2. Simulation Result

Wave fronts of the original sound field, synthesized wave fronts, and the differences between them at $t = 0$ s and $f = 500$ Hz are shown in Figs. 6-8. In these figures, only the XY plane ($z = 0$), XZ plane ($y = 0$), and YZ plane ($x = 0$) in the space surrounded by the loudspeaker array (4 m × 4 m × 2 m) are plotted, and the dashed lines denote the boundary of the listening area (2 m × 2 m × 1 m) surrounded by the loudspeaker array. In these figures, absolute values of sound pressures ($p_o(\mathbf{r},t)$, $p_c(\mathbf{r},t)$, $p_p(\mathbf{r},t)$) and the differences ($p_c(\mathbf{r},t) - p_o(\mathbf{r},t)$, $p_p(\mathbf{r},t) - p_o(\mathbf{r},t)$) are plotted; color bars are shown on the right side of the figures. The wave fronts in the listening area are accurately synthesized if the differences are white. Note that the differences are calculated after normalizing the sound pressures ($p_o(\mathbf{r},t)$, $p_c(\mathbf{r},t)$, $p_p(\mathbf{r},t)$) in a space with dimensions of 1 m (width) × 1 m (depth) × 0.5 m (height) in the listening area (i.e., $|r_x| < 0.5$, $|r_y| < 0.5$, $|r_z| < 0.25$).

When the conventional system is used, the wave fronts are not accurately reproduced, as is evident from the differences not being white. This is due to the fact that the acoustic transfer functions from the loudspeakers to the listening position tend to diverge to infinity since the distance between the loudspeakers and the listening position is short in this computer simulation (e.g., the shortest distance is 0.5 m if the listening position is at the center of the listening area). On the other hand, when the proposed system is used, the wave fronts are accurately reproduced, as is apparent from the differences being white. This is due to the fact that the acoustic transfer functions from the loudspeakers to the listening position do not diverge to infinity since the distance between the loudspeakers and the listening position is long when the loudspeakers are positioned outside the listening area (e.g., the shortest distance is 1 m if the listening position is at the center of the listening area). Thus, the proposed system can reproduce wave fronts more accurately than the conventional system.

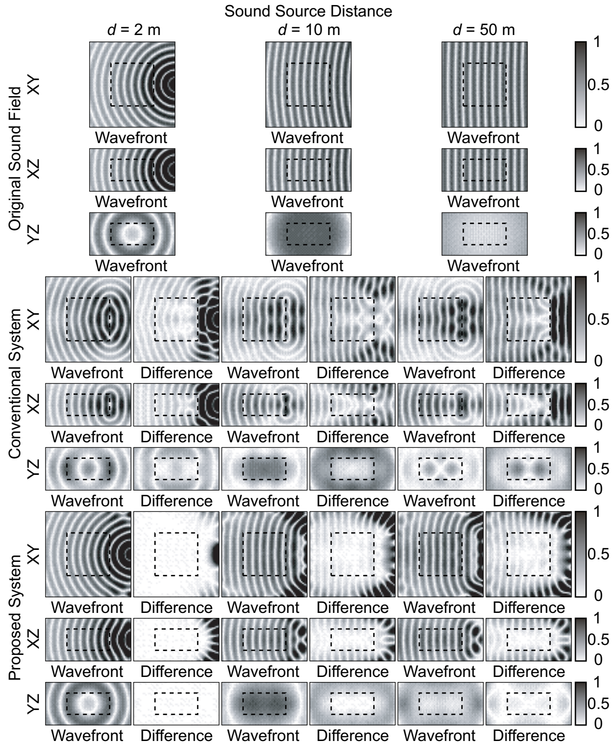

Fig. 6: Wave fronts of the original sound field, synthesized wave fronts, and the differences between them in the computer simulation ($t = 0$ s, $f = 500$ Hz, $\mathbf{u} = (1,0,0)^T$).

In order to quantitatively evaluate the reproduced sound field, SNRs (signal-to-noise ratios) were calculated as follows:

$$\mathrm{SNR} = 10\log_{10} \frac{\sum_f \sum_{\mathbf{r}} \{p_o(\mathbf{r},0)\}^2}{\sum_f \sum_{\mathbf{r}} \{p_{c(p)}(\mathbf{r},0) - p_o(\mathbf{r},0)\}^2}, \quad (34)$$

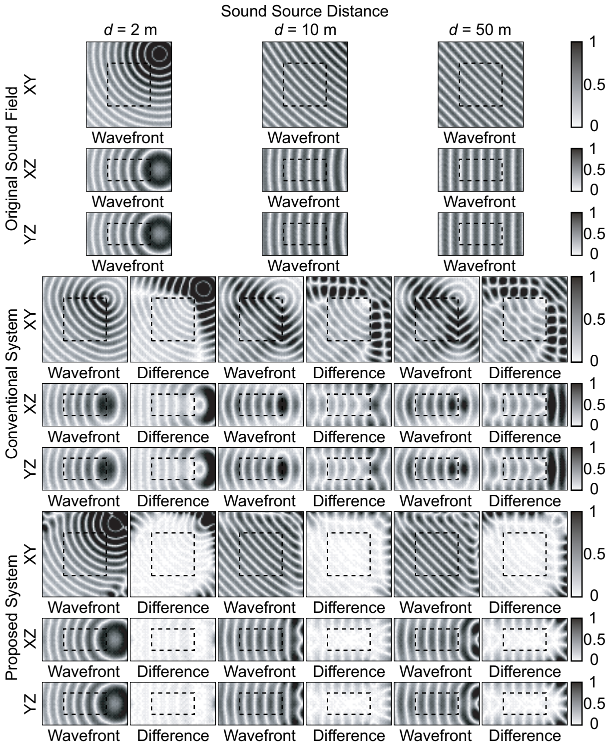

Fig. 7: Wave fronts of the original sound field, synthesized wave fronts, and the differences between them in the computer simulation (t = 0 s, f = 500 Hz, $\mathbf{u} = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0)^T$).

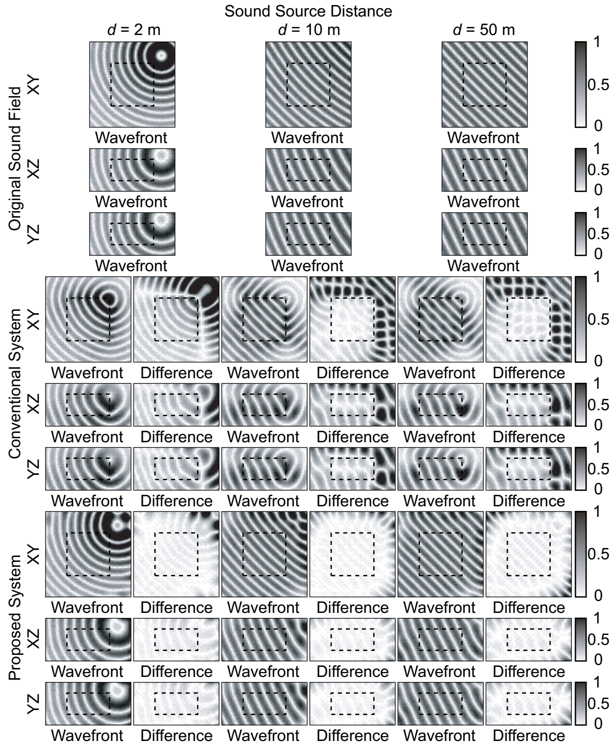

Fig. 8: Wave fronts of the original sound field, synthesized wave fronts, and the differences between them in the computer simulation (t = 0 s, f = 500 Hz, $\mathbf{u} = (\frac{2}{3}, \frac{2}{3}, \frac{1}{3})^T$).

where the range of $\mathbf{r}$ was limited to the space with dimensions of 2 m × 2 m × 1 m (i.e., $|r_x| < 1$, $|r_y| < 1$, $|r_z| < 0.5$). Note that $p_o(\mathbf{r}, 0)$, $p_c(\mathbf{r}, 0)$, and $p_p(\mathbf{r}, 0)$ were normalized in the range described above before calculating the SNRs.

The SNR values for each system are shown in Fig. 9. The values for the conventional system are always lesser than 5 dB for all source distances and directions. On the other hand, in the proposed system, the SNRs are always greater than 10 dB and exceed the values for the conventional system by more than 7 dB for all source distances and directions. It is thus inferred that wave fronts can be accurately reproduced in the proposed system.

The SNR values for the proposed system are lower than the SNR value (about 30 dB) obtained in a previous study on the basis of a boundary surface control technique [8, 9]. However, while both the sound pressures and their gradients on the boundary of the control and listening areas were controlled in the previous study, only the sound pressures on the boundary of these areas are controlled in the proposed system. Thus, since the number of microphones and loudspeakers employed in the proposed system is half the number used in the previous study, the calculation amount of inverse filters is lesser than that in the previous study. From the discussions above, the proposed system is expected to be effective in practice.

4. CONCLUSION

In this study, we proposed a 3D sound field reproduction system based on directional microphones and the boundary surface control; the system was constructed by using the inverse filters in the conventional system. The proposed system can reproduce a 3D sound field in a listening area even if loudspeakers are not placed on the

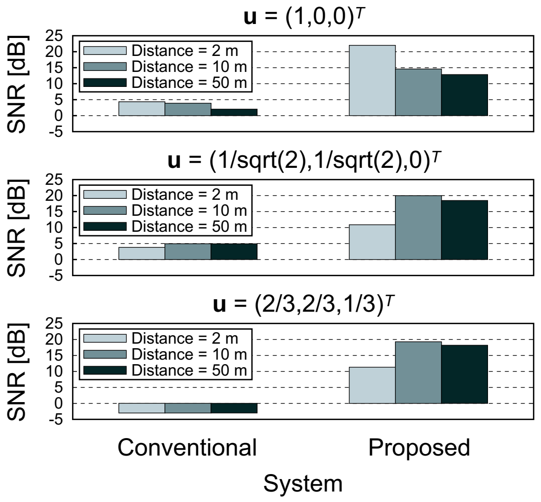

Fig. 9: SNRs of the conventional and proposed systems in the computer simulation.

boundary surface. We performed a computer simulation in order to numerically analyze the reproduced 3D sound field. The results of the simulation showed that the proposed system can reproduce wave fronts more accurately than the conventional system.

In this study, although the computer simulation was performed under conditions where there were no reflected sounds in the reproduced sound field, there will be reflected and reverberant sounds in the reproduced sound field when the system is constructed in an actual room. Further, when the screen or display of the visual system is placed on or outside the boundary surface of the listening area, the acoustic transfer functions from the loudspeakers to the directional microphones may vary with the screen or display. Thus, in a future study, the accuracy of the reproduced 3D sound field when the loudspeaker array and the screen or display are placed in an actual room should be studied; such a study will contribute to the development of an audio-visual system using the proposed system.

Although the number of microphones and loudspeakers considered in this computer simulation is 576 and 2304, respectively, so as to satisfy the spatial sampling theorem, these numbers are too large for a real system; therefore, it would be difficult to use the result of this computer simulation for developing a real system. On the other hand, if the performance of the reproduced 3D sound field is evaluated by considering the human auditory system, the number of microphones and loudspeakers required to construct a real system can be reduced [10]. Thus, it is necessary to evaluate the performance of the constructed system by carrying out listening tests in order to reduce the number of microphones and loudspeakers.

5. REFERENCES

[1] H. Fletcher, "Symposium on wire transmission of symphonic music and its reproduction on auditory perspective: Basic requirement," *Bell System Technical Journal*, vol. 13, no. 2, pp. 239–244, April 1934.

[2] 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.

[3] T. kimura and K. Kakehi, "Effects of directivity of microphones and loudspeakers on accuracy of synthesized wave fronts in sound field reproduction based on wave field synthesis," in *Papers of the AES 13th Regional Convention*, July 2007, number 0037, pp. 1–8.

[4] J. Blauert, *Spatial Hearing*, pp. 372–392, MIT Press, Cambridge, Mass, revised edition, 1997.

[5] 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, October 1974.

[6] B. B. Baker and E. T. Copson, *The Mathematical Theory of Huygens' Principle*, pp. 23–26, Oxford University Press, London, UK, second edition, 1950.

[7] B. B. Baker and E. T. Copson, *The Mathematical Theory of Huygens' Principle*, pp. 72–74, Oxford University Press, London, UK, second edition, 1950.

[8] S. Ise, "A principle of sound field control based on the Kirchhoff-Helmholtz integral equation and the theory of inverse systems," *ACUSTICA - Acta Acustica*, vol. 85, no. 1, pp. 78–87, January/February 1999.

[9] S. Takane, Y. Suzuki, and T. Sone, "A new method for global sound field reproduction based on Kirchhoff's integral equation," *ACUSTICA - Acta Acustica*, vol. 85, no. 2, pp. 250–257, March/April 1999.

[10] T. Kimura, K. Kakehi, K. Takeda, and F. Itakura, "Subjective assessments for the effect of the number of channel signals on the sound field reproduction used in wavefield synthesis," in *Proc. ICA*, April 2004, number Th.P1.18 in IV, pp. 3159–3162.