Theoretical Study of Near 3D Sound Field Reproduction Based on Wave Field Synthesis

T. Kimura, Y. Yamakata and M. Katsumoto

National Institute of Information and Communications Technology, 4-2-1, Nukui-Kitamachi, Koganei, 184-8795 Tokyo, Japan
t-kimura@nict.go.jp

It is very important to develop near 3D sound field reproduction techniques in order to realize the ultrarealistic communication such as 3D TV and 3D tele-conference. In this report, the principle of the near 3D sound field reproduction technique using wave field synthesis is defined from Kirchhoff-Helmholtz integral equation and two methods (dipole control method and directional point control method) are proposed. The performance of two proposed methods is studied by computer simulation and it is shown that the dipole control method has good performance and that the directional point control method has good performance if the directivity of loudspeakers is unidirectional or shotgun.

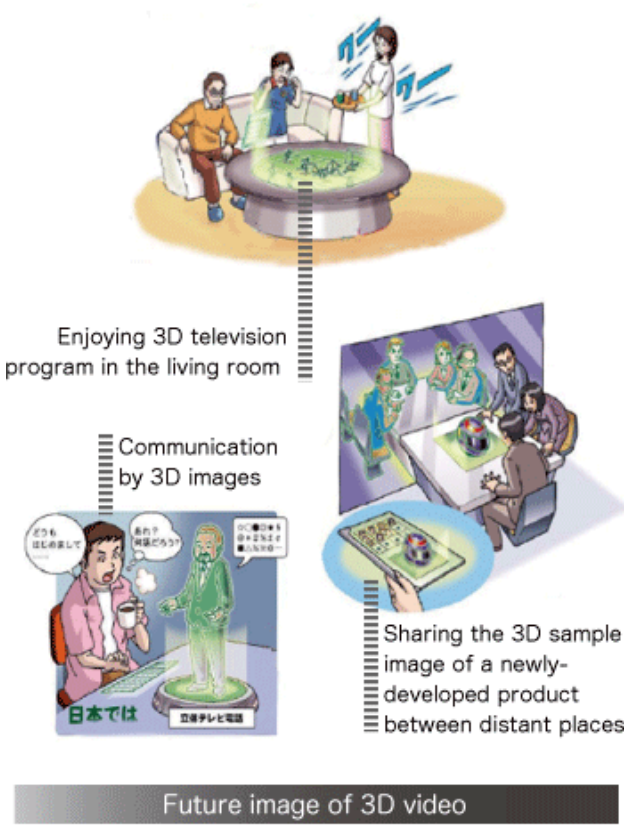

Figure 1: Future image of ultra-realistic communication using 3D video and 3D audio [1].

1 Introduction

We have been investigating ultra-realistic communication techniques as shown in Figure 1 [1]. If video and audio can be more realistically reproduced in a 3D space by applying 3D video and audio techniques, more realistic forms of communication (e.g. 3D television, 3D teleconferencing, etc.) will be possible than those currently provided by conventional video and audio techniques (HD video and 5.1 ch audio). 3D sound field reproduction techniques capable of providing the aural components of ultra-realistic communication include binaural [2], transaural [3], stereo dipole [4], wave field synthesis [5, 6, 7], and boundary surface control techniques [8]. In this paper, we focus on wave field synthesis and propose a 3D sound field reproduction technique based on wave field synthesis.

Wave field synthesis is a 3D sound field reproduction technique for reproducing wave fronts of a control area in a listening area based on Huygens' principle. Microphones placed on the boundary of a control area record the original sound and loudspeakers placed on the boundary of the listening area then play the recorded sound. The position of the loudspeakers is the same as that of the microphones. Multiple listeners can listen to the sound anywhere in the listening area without having to wear a device such as headphones because this technique reproduces the sound field of a 3D space rather than the sound field of binaural positions.

Conventional wave field synthesis systems have been constructed based on the Kirchhoff-Helmholtz integral equation, which mathematically defines Huygens' principle [9]. However, in these systems, the loudspeakers are placed around the listeners and the sound field of the inside of the loudspeaker array is reproduced, as shown in Figure 2(a). As a result, listeners cannot listen to sounds around the sound sources. In the proposed system, however, because the loudspeakers are placed around the sound sources and the sound field of the outside of the loudspeaker array is reproduced, as shown in Figure 2(b), listeners can listen to sounds around the sound sources. Thus, as shown in Figures 2(a) and 2(b), the direction of the microphones and loudspeakers in the proposed system is the opposite to that in the conventional system. Although the condition for reproducing wave fronts has been theoretically studied in the conventional system [10], it was not theoretically studied in the proposed system. In addition, while we assume that the sound field reproduced in the proposed system is the near sound field, where the distance between the sound sources and listeners is less than one meter, the condition for reproducing wave fronts in the near sound field was not investigated in the proposed system.

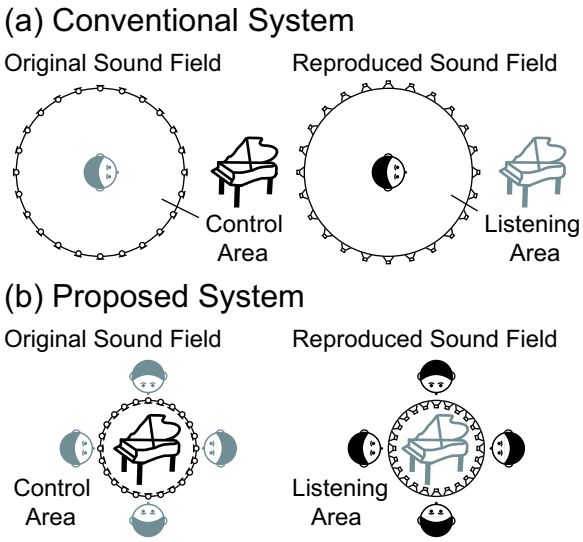

Figure 2: Original sound field and reproduced sound field in conventional and proposed systems.

In this paper, near 3D sound field reproduction techniques that enable listeners to listen to sounds around the sound sources are proposed based on wave field synthesis. The principle of the near 3D sound field reproduction technique is defined on the basis of the Kirchhoff-Helmholtz integral equation and two methods – dipole control and directional point control – are proposed in Section 2. The performance of the two methods was evaluated using computer simulation in Section 3.

2 Principle of near 3D sound field reproduction

2.1 Kirchhoff-Helmholtz integral equation

As shown in the left of Figure 3, if sound sources are surrounded by a continuous boundary surface S, and $\mathbf{r}_S$ and $\mathbf{r}$ are the position vectors on S and on space V (the outside of S), $P(\mathbf{r}, \omega)$ (sound pressure

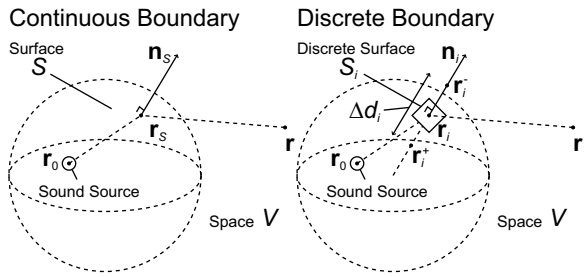

Figure 3: Coordinates in Kirchhoff-Helmholtz integral equation.

at $\mathbf{r}$) is denoted as follows:

$$P(\mathbf{r},\omega) = \frac{1}{4\pi}\oint_S \left\{ \frac{\partial P(\mathbf{r}_S,\omega)}{\partial \mathbf{n}_S}\frac{e^{-jk|\mathbf{r}-\mathbf{r}_S|}}{|\mathbf{r}-\mathbf{r}_S|} - P(\mathbf{r}_S,\omega)\frac{\partial}{\partial \mathbf{n}_S}\left(\frac{e^{-jk|\mathbf{r}-\mathbf{r}_S|}}{|\mathbf{r}-\mathbf{r}_S|}\right)\right\} dS, \quad (1)$$

where k is a wave number, and $\mathbf{n}_S$ is the normal unit vector toward the outside of the continuous boundary surface at $\mathbf{r}_S$. Eq. (1) shows that the sound pressure of the space V is reproduced if the monopole sources (amplitude $\partial P(\mathbf{r}_S,\omega)/\partial \mathbf{n}_S$) and dipole sources (amplitude $-P(\mathbf{r}_S,\omega)$) are played at $\mathbf{r}_S$.

However, to construct the system the boundary surface S must be discretized since the monopole and dipole sources are not placed continuously on S. As shown in the right of Figure 3, if $\mathbf{r}_i$ is the position vector on S_i (ith element of a discrete boundary surface), and $P(\mathbf{r}_i,\omega)$ (sound pressure at $\mathbf{r}_i$) and $\partial P(\mathbf{r}_i,\omega)/\partial \mathbf{n}_i$ (sound pressure gradient at $\mathbf{r}_i$) are constant in S_i, Eq. (1) can be converted as follows:

$$P(\mathbf{r},\omega) = \frac{1}{4\pi}\sum_{i=1}^{M}\left\{ \frac{\partial P(\mathbf{r}_i,\omega)}{\partial \mathbf{n}_i}\frac{e^{-jk|\mathbf{r}-\mathbf{r}_i|}}{|\mathbf{r}-\mathbf{r}_i|} - P(\mathbf{r}_i,\omega)\frac{\partial}{\partial \mathbf{n}_i}\left(\frac{e^{-jk|\mathbf{r}-\mathbf{r}_i|}}{|\mathbf{r}-\mathbf{r}_i|}\right)\right\}\Delta S_i, \quad (2)$$

where M is the total number of elements of the discrete boundary surface, ΔS_i is the area of S_i, and $\mathbf{n}_i$ is the normal unit vector toward the outside of the discrete boundary surface at $\mathbf{r}_i$. Eq. (2) shows that the sound pressure of the space V is reproduced if the monopole sources (amplitude $\partial P(\mathbf{r}_i,\omega)/\partial \mathbf{n}_i$) and dipole sources (amplitude $-P(\mathbf{r}_i,\omega)$) are played at M points (position $\mathbf{r}_i$).

2.2 Dipole control method

Sound pressure gradients $\partial P(\mathbf{r}_i,\omega)/\partial \mathbf{n}_i$ and sound pressures $P(\mathbf{r}_i,\omega)$ were approximated by the sound pressures at neighbor points as follows:

$$\frac{\partial P(\mathbf{r}_i,\omega)}{\partial \mathbf{n}_i} \approx \frac{P(\mathbf{r}_i^+,\omega) - P(\mathbf{r}_i^-,\omega)}{\Delta d_i}, \quad (3)$$

$$P(\mathbf{r}_i,\omega) \approx \frac{P(\mathbf{r}_i^+,\omega) + P(\mathbf{r}_i^-,\omega)}{2}, \quad (4)$$

where, as shown in the right of Figure 3, $\mathbf{r}_i^+$ and $\mathbf{r}_i^-$ are the position vector of the neighboring points at the inside and outside of $\mathbf{r}_i$, and $\Delta d_i (= |\mathbf{r}_i^+ - \mathbf{r}_i^-|)$ is the distance between the neighboring points. The dipole and monopole sources were also approximated by two monopole sources at neighboring points as follows:

$$\frac{\partial}{\partial \mathbf{n}_i}\left(\frac{e^{-jk|\mathbf{r}-\mathbf{r}_i|}}{|\mathbf{r}-\mathbf{r}_i|}\right) \approx \frac{1}{\Delta d_i}\left(\frac{e^{-jk|\mathbf{r}-\mathbf{r}_i^+|}}{|\mathbf{r}-\mathbf{r}_i^+|} - \frac{e^{-jk|\mathbf{r}-\mathbf{r}_i^-|}}{|\mathbf{r}-\mathbf{r}_i^-|}\right), \quad (5)$$

$$\frac{e^{-jk|\mathbf{r}-\mathbf{r}_i|}}{|\mathbf{r}-\mathbf{r}_i|} \approx \frac{1}{2}\left(\frac{e^{-jk|\mathbf{r}-\mathbf{r}_i^+|}}{|\mathbf{r}-\mathbf{r}_i^+|} + \frac{e^{-jk|\mathbf{r}-\mathbf{r}_i^-|}}{|\mathbf{r}-\mathbf{r}_i^-|}\right). \quad (6)$$

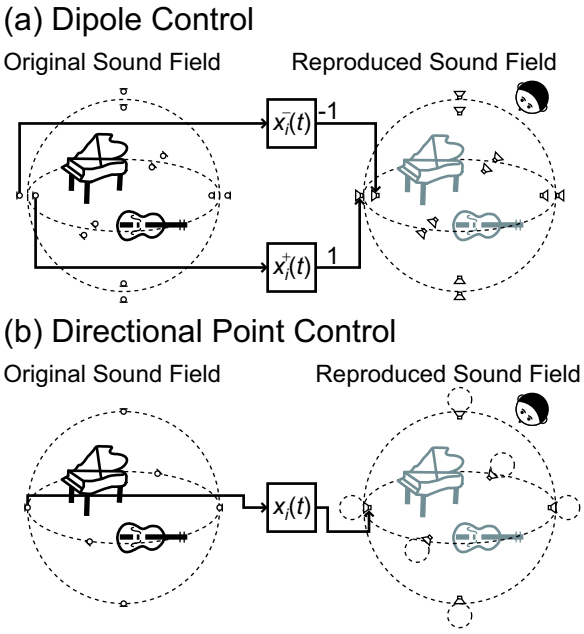

Figure 4: Near 3D sound field reproduction system based on dipole control and directional point control methods.

If Eqs. (3)-(6) are substituted for Eq. (2), the following equation is derived:

$$P(\mathbf{r},\omega) = \frac{1}{4\pi}\sum_{i=1}^{M}\left\{ P(\mathbf{r}_i^+,\omega)\frac{e^{-jk|\mathbf{r}-\mathbf{r}_i^-|}}{|\mathbf{r}-\mathbf{r}_i^-|} - P(\mathbf{r}_i^-,\omega)\frac{e^{-jk|\mathbf{r}-\mathbf{r}_i^+|}}{|\mathbf{r}-\mathbf{r}_i^+|}\right\}\frac{\Delta S_i}{\Delta d_i}. \quad (7)$$

Eq. (7) shows that the sound pressure of the space V is reproduced if the monopole sources (amplitude $P(\mathbf{r}_i^+,\omega)$) are played at M points (position $\mathbf{r}_i^-$) and the monopole sources (amplitude $-P(\mathbf{r}_i^-,\omega)$) are played at M points (position $\mathbf{r}_i^+$).

A diagram of the near 3D sound field reproduction system based on Eq. (7) (called the "dipole control method") is shown in Figure 4(a). Firstly, M pairs of omnidirectional microphones were placed on the boundary surface around the sound sources in the original sound field and audio signals ($x_i^+(t)$ and $x_i^-(t)$) were recorded. Secondly, M pairs of omnidirectional loudspeakers were placed at the same position as the microphone pairs in the reproduced sound field and the processed audio signals ($x_i^+(t)$ and $-x_i^-(t)$) were played. As a result, since wave fronts were reproduced on the outside of the boundary surface, listeners on the outside of the boundary surface experienced the feeling that the sound sources were being played on the inside of the boundary surface.

2.3 Directional point control method

The sound pressure gradients $\partial P(\mathbf{r}_i,\omega)/\partial \mathbf{n}_i$ were approximated as follows:

$$\begin{aligned}\frac{\partial P(\mathbf{r}_i,\omega)}{\partial \mathbf{n}_i} &= \frac{\partial}{\partial \mathbf{n}_i}\left(\frac{Ae^{-jk|\mathbf{r}_i-\mathbf{r}_0|}}{|\mathbf{r}_i-\mathbf{r}_0|}\right) \\ &= -\frac{Ae^{-jk|\mathbf{r}_i-\mathbf{r}_0|}}{|\mathbf{r}_i-\mathbf{r}_0|}\left(\frac{1}{|\mathbf{r}_i-\mathbf{r}_0|}+jk\right)\cos(\mathbf{n}_i,\mathbf{r}_i-\mathbf{r}_0) \\ &= -P(\mathbf{r}_i,\omega)\left(\frac{1}{|\mathbf{r}_i-\mathbf{r}_0|}+jk\right)\cos(\mathbf{n}_i,\mathbf{r}_i-\mathbf{r}_0) \\ &\approx -jkP(\mathbf{r}_i,\omega)\cos(\mathbf{n}_i,\mathbf{r}_i-\mathbf{r}_0)\left(\text{if } k \gg \frac{1}{|\mathbf{r}_i-\mathbf{r}_0|}\right),\end{aligned} \quad (8)$$

where $\mathbf{r}_0$ and A are the position vector and amplitude of the sound sources in an original sound field, and $(\mathbf{n}_i, \mathbf{r}_i - \mathbf{r}_0)$ denotes the angle

between the vector $\mathbf{n}_i$ and the vector $\mathbf{r}_i - \mathbf{r}_0$. The dipole sources were also approximated as follows:

$$\frac{\partial}{\partial \mathbf{n}_i}\left(\frac{e^{-jk|\mathbf{r}-\mathbf{r}_i|}}{|\mathbf{r}-\mathbf{r}_i|}\right) = -\frac{e^{-jk|\mathbf{r}-\mathbf{r}_i|}}{|\mathbf{r}-\mathbf{r}_i|}\left(\frac{1}{|\mathbf{r}-\mathbf{r}_i|} + jk\right)\cos(\mathbf{n}_i, \mathbf{r}_i - \mathbf{r}) \approx -jk\frac{e^{-jk|\mathbf{r}-\mathbf{r}_i|}}{|\mathbf{r}-\mathbf{r}_i|}\cos(\mathbf{n}_i, \mathbf{r}_i - \mathbf{r}) \left(\text{if } k \gg \frac{1}{|\mathbf{r}-\mathbf{r}_i|}\right), \quad (9)$$

where $(\mathbf{n}_i, \mathbf{r}_i - \mathbf{r})$ denotes the angle between the vector $\mathbf{n}_i$ and the vector $\mathbf{r}_i - \mathbf{r}$. If Eqs. (8)-(9) are substituted for Eq. (2), the following equation is derived:

$$P(\mathbf{r}, \omega) = \frac{jk}{4\pi}\sum_{i=1}^{M} P(\mathbf{r}_i, \omega)\frac{e^{-jk|\mathbf{r}-\mathbf{r}_i|}}{|\mathbf{r}-\mathbf{r}_i|} \{\cos(\mathbf{n}_i, \mathbf{r}_i - \mathbf{r}) - \cos(\mathbf{n}_i, \mathbf{r}_i - \mathbf{r}_0)\}\Delta S_i \quad (10)$$

This equation is known as the Fresnel-Kirchhoff diffraction formula [11]. In the proposed system, since the direction of vector $\mathbf{r}_i - \mathbf{r}_0$ is almost the same as that of vector $\mathbf{n}_i$, it can be approximated as $\cos(\mathbf{n}_i, \mathbf{r}_i - \mathbf{r}_0) \approx 1$. Since $\cos(\mathbf{n}_i, \mathbf{r}_i - \mathbf{r}) = \cos\{\pi - (\mathbf{n}_i, \mathbf{r} - \mathbf{r}_i)\}$, (10) can be converted as follows:

$$P(\mathbf{r}, \omega) \approx \frac{jk}{4\pi}\sum_{i=1}^{M} P(\mathbf{r}_i, \omega)\frac{e^{-jk|\mathbf{r}-\mathbf{r}_i|}}{|\mathbf{r}-\mathbf{r}_i|} \{-\cos(\mathbf{n}_i, \mathbf{r} - \mathbf{r}_i) - 1\}\Delta S_i \approx \frac{jk}{4\pi}\sum_{i=1}^{M} P(\mathbf{r}_i, \omega)D_i\frac{e^{-jk|\mathbf{r}-\mathbf{r}_i|}}{|\mathbf{r}-\mathbf{r}_i|}\Delta S_i, \quad (11)$$

where $D_i(\approx -\cos(\mathbf{n}_i, \mathbf{r} - \mathbf{r}_i) - 1)$ is the directivity of the monopole sources placed at $\mathbf{r}_i$. Eq. (11) shows that the sound pressure of the space V is reproduced if the directional monopole sources (amplitude $P(\mathbf{r}_i, \omega)$) are played at M points (position $\mathbf{r}_i$).

A diagram of the near 3D sound field reproduction system based on Eq. (11) (called the "directional point control method") is shown in Figure 4(b). Firstly, M omnidirectional microphones were placed on the boundary surface around the sound sources in the original sound field and audio signals ($x_i(t)$) were recorded. Secondly, M directional loudspeakers were placed at the same position as the microphones in the reproduced sound field and the recorded audio signals ($x_i(t)$) were played. The directivity of the loudspeakers was toward the outside of the boundary surface. As a result, because wave fronts were reproduced on the outside of the boundary surface, listeners on the outside of the boundary surface experienced the feeling that the sound sources were being played on the inside of the boundary surface.

3 Computer simulation

The performance of both the dipole control and directional point control methods was evaluated using computer simulation.

3.1 Simulation environment

As shown in Figure 5, 162 control points were placed on a sphere with a radius of 0.4 m and 162 synthesis points were placed on a sphere with a radius of 0.8 m. The position of 162 points corresponds to the vertex of a Class I Method 1 icosahedral geodesic dome following 4 frequencies [12].

The sound source signal $s(t)$ was a sine-wave signal with an amplitude of A and frequency of f as follows:

$$s(t) = A\sin 2\pi f t. \quad (12)$$

Let $\mathbf{r}$ be the position vector of a synthesis point. The $p_0(\mathbf{r}, f, t)$ (sound pressure at the synthesis point $\mathbf{r}$ in the original sound field)

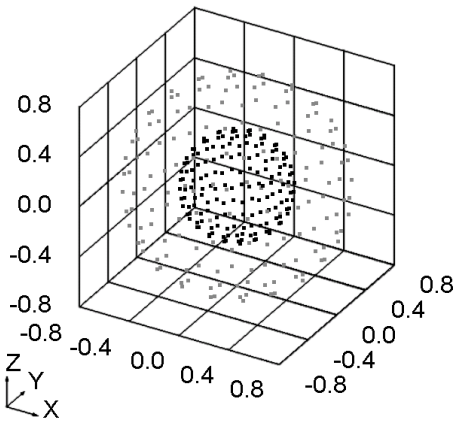

Figure 5: Position of control points (black) and synthesis points (gray) used in computer simulation.

is denoted as follows:

$$p_0(\mathbf{r}, f, t) = \frac{1}{d_0}s\left(t - \frac{d_0}{c}\right) = \frac{A}{d_0}\sin\left\{2\pi f\left(t - \frac{d_0}{c}\right)\right\}, \quad (13)$$

where $d_0(= |\mathbf{r} - \mathbf{r}_0|)$ is the distance between the sound source and the synthesis point, $\mathbf{r}_0$ is the position vector of the sound source, and c is the sound velocity.

In the case of the dipole control method, the $x_i^+(t)$ and $x_i^-(t)$ (recorded signals of the ith microphone pair) were denoted as follows:

$$x_i^+(t) = \frac{1}{d_{i0}^+}s\left(t - \frac{d_{i0}^+}{c}\right) = \frac{A}{d_{i0}^+}\sin\left\{2\pi f\left(t - \frac{d_{i0}^+}{c}\right)\right\}, \quad (14)$$

$$x_i^-(t) = \frac{1}{d_{i0}^-}s\left(t - \frac{d_{i0}^-}{c}\right) = \frac{A}{d_{i0}^-}\sin\left\{2\pi f\left(t - \frac{d_{i0}^-}{c}\right)\right\}, \quad (15)$$

where $d_{i0}^+(= |\mathbf{r}_i^+ - \mathbf{r}_0|)$ and $d_{i0}^-(= |\mathbf{r}_i^- - \mathbf{r}_0|)$ are the distance between the sound source and the ith microphone pair, and $\mathbf{r}_i^+$ and $\mathbf{r}_i^-$ are the position vector of the ith microphone pair defined as follows:

$$\mathbf{r}_i^+ = \mathbf{r}_i - \frac{\Delta d_i}{2}\mathbf{n}_i, \quad (16)$$

$$\mathbf{r}_i^- = \mathbf{r}_i + \frac{\Delta d_i}{2}\mathbf{n}_i, \quad (17)$$

where $\mathbf{r}_i$ is the position vector of the ith control point, $\mathbf{n}_i$ is the normal unit vector toward the outside of the discrete boundary surface at $\mathbf{r}_i$, and $\Delta d_i(= |\mathbf{r}_i^+ - \mathbf{r}_i^-|)$ is the distance between the microphones. The $p(\mathbf{r}, f, t)$ (sound pressure of the synthesis point $\mathbf{r}$ in the reproduced sound field) was calculated from $x_i^+(t)$ and $x_i^-(t)$ as follows:

$$p(\mathbf{r}, f, t) = \sum_{i=1}^{M}\left\{\frac{1}{d_i^-}x_i^+\left(t - \frac{d_i^-}{c}\right) - \frac{1}{d_i^+}x_i^-\left(t - \frac{d_i^+}{c}\right)\right\} = \sum_{i=1}^{M}\left[\frac{A}{d_i^- d_{i0}^+}\sin\left\{2\pi f\left(t - \frac{d_i^- + d_{i0}^+}{c}\right)\right\} - \frac{A}{d_i^+ d_{i0}^-}\sin\left\{2\pi f\left(t - \frac{d_i^+ + d_{i0}^-}{c}\right)\right\}\right], \quad (18)$$

where M is the total number of loudspeaker pairs, and $d_i^+(= |\mathbf{r} - \mathbf{r}_i^+|)$ and $d_i^-(= |\mathbf{r} - \mathbf{r}_i^-|)$ is the distance between the ith loudspeaker pair and the synthesis point.

In the case of the directional point control method, the $x_i(t)$ (recorded signal of the ith microphone) was denoted as follows:

$$x_i(t) = \frac{1}{d_{i0}}s\left(t - \frac{d_{i0}}{c}\right) = \frac{A}{d_{i0}}\sin\left\{2\pi f\left(t - \frac{d_{i0}}{c}\right)\right\}, \quad (19)$$

where $d_{i0}(= |\mathbf{r}_i - \mathbf{r}_0|)$ is the distance between the sound source and the ith microphone, and $\mathbf{r}_i$ is the position vector of the ith microphone. The $p(\mathbf{r}, f, t)$ (sound pressure of the synthesis point $\mathbf{r}$ in the

Table 1: Parametric conditions in computer simulation.

Amplitude of sound source (A)	1
Frequency of sound source (f)	125, 250, 500, 1000, 2000, 4000, 8000, 16000 Hz
Position vector of sound source ($\mathbf{r}_0$)	$(0,0,0)^T$ $(0.3,0,0)^T$ $(0,0.3,0)^T$ $(0,0,0.3)^T$
Sound velocity (c)	340 m/s
Total number of control points (M)	162
Radius of control points (r)	0.4 m
Total number of synthesis points (N)	162
Radius of synthesis points (R)	0.8 m
Normal unit vector ($\mathbf{n}_i$)	$\mathbf{r}_i/\lvert\mathbf{r}_i\rvert$
Neighbor distance (Δd_i)	0.002 m
Directivity of loudspeakers (D_i)	Omnidirectional, Bidirectional, Unidirectional, Shotgun

reproduced sound field) was calculated from $x_i(t)$ as follows:

$$p(\mathbf{r},f,t) = \sum_{i=1}^{M} \frac{D_i}{d_i} x_i\left(t-\frac{d_i}{c}\right) = \sum_{i=1}^{M} \frac{D_i A}{d_i d_{i0}} \sin\left\{2\pi f\left(t-\frac{d_i+d_{i0}}{c}\right)\right\} \quad (20)$$

where M is the total number of loudspeakers, $d_i(=|\mathbf{r}-\mathbf{r}_i|)$ is the distance between the ith loudspeaker and the synthesis point, and D_i is the directivity of the ith loudspeaker.

Parametric conditions are shown in Table 1. The maximum interval of the control points is about 13 cm, which is less than half of the wavelength of 1000 Hz sound ($=\frac{340\,\mathrm{m}}{1000\,\mathrm{Hz}}=34$ cm). Thus, the spatial sampling theorem for reproducing a wave front at sound frequencies under 1000 Hz is satisfied. The $\mathbf{r}_i$ and $\mathbf{r}$ (position vector of the control point and synthesis point) were set in a three-dimensional coordinate as follows:

$$\mathbf{r}_i = (r\cos\theta_i\cos\phi_i, r\sin\theta_i\cos\phi_i, r\sin\phi_i)^T (i=1...M), \quad (21)$$

$$\mathbf{r} = (R\cos\theta_j\cos\phi_j, R\sin\theta_j\cos\phi_j, R\sin\phi_j)^T (j=1...N), \quad (22)$$

where θ_i and ϕ_i are the azimuth and elevation angle of the ith control point, and θ_j and ϕ_j are the azimuth and elevation angle of the jth synthesis point.

The D_i (directivity of the ith loudspeaker) is defined as follows:

$$(\text{Omnidirectional})D_i = 1, \quad (23)$$

$$(\text{Bidirectional})D_i = \cos\theta_{is}, \quad (24)$$

$$(\text{Unidierctional})D_i = \frac{1}{2}(1+\cos\theta_{is}), \quad (25)$$

$$(\text{Shotgun})D_i = \begin{cases} \cos\theta_{is} & (|\theta_{is}| \le 90^\circ) \\ 0 & (|\theta_{is}| > 90^\circ) \end{cases}, \quad (26)$$

where $\cos\theta_{is} = \frac{\mathbf{n}_i\cdot(\mathbf{r}-\mathbf{r}_i)}{|\mathbf{n}_i||\mathbf{r}-\mathbf{r}_i|}$.

In addition, the sound intensity was calculated to evaluate the arrival direction of the sound at the synthesis point $\mathbf{r}$. The sound intensity was calculated using the cross-spectral method, as shown in Figure 6 [13]. Note that $I_x(\mathbf{r},f)$, $I_y(\mathbf{r},f)$ and $I_z(\mathbf{r},f)$ in Figure 6 are the x, y and z components of the sound intensity vectors, and $p(\mathbf{r}_x^+,f,t)$, $p(\mathbf{r}_x^-,f,t)$, $p(\mathbf{r}_y^+,f,t)$, $p(\mathbf{r}_y^-,f,t)$, $p(\mathbf{r}_z^+,f,t)$ and $p(\mathbf{r}_z^-,f,t)$ in Figure 6 are the sound pressure at six points ($\mathbf{r}_x^+$, $\mathbf{r}_x^-$, $\mathbf{r}_y^+$, $\mathbf{r}_y^-$, $\mathbf{r}_z^+$ and $\mathbf{r}_z^-$). The position vectors of the six points were set as

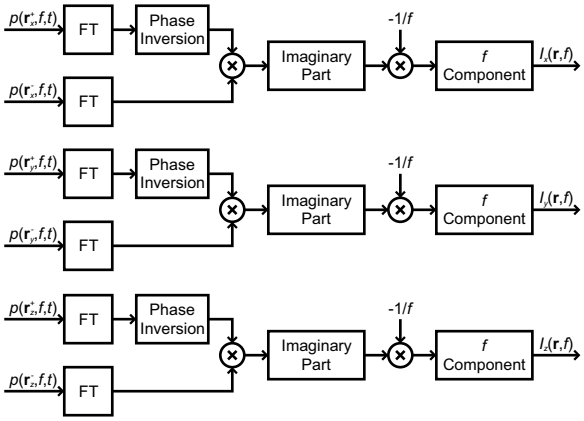

Figure 6: Block diagram of calculation of sound intensities in computer simulation (reproduced sound field).

follows:

$$\mathbf{r}_x^+ = \mathbf{r}+(\Delta,0,0)^T, \mathbf{r}_x^- = \mathbf{r}-(\Delta,0,0)^T, \quad (27)$$

$$\mathbf{r}_y^+ = \mathbf{r}+(0,\Delta,0)^T, \mathbf{r}_y^- = \mathbf{r}-(0,\Delta,0)^T, \quad (28)$$

$$\mathbf{r}_z^+ = \mathbf{r}+(0,0,\Delta)^T, \mathbf{r}_z^- = \mathbf{r}-(0,0,\Delta)^T, \quad (29)$$

where Δ is 0.001 m. The calculated sound intensity vector $\mathbf{I}(\mathbf{r},f) = \{I_x(\mathbf{r},f), I_y(\mathbf{r},f), I_z(\mathbf{r},f)\}^T$ was evaluated in section 3.2. The $\mathbf{I}_0(\mathbf{r},f) = \{I_{x0}(\mathbf{r},f), I_{y0}(\mathbf{r},f), I_{z0}(\mathbf{r},f)\}^T$ (sound intensity vector in the original sound field) was also calculated.

3.2 Simulation results

To quantitatively evaluate the performance at all the frequencies of the sound sources, two measures were applied. The first was the SNR of the RMSs of sound pressures. This measure indicates the difference in sound pressure distributions between the original and reproduced sound fields and is defined as follows:

$$\mathrm{SNR}(f)[\mathrm{dB}] = 10\log_{10}\frac{\sum_{\mathbf{r}}\{p_0(\mathbf{r},f)\}^2}{\sum_{\mathbf{r}}\{p(\mathbf{r},f)-p_0(\mathbf{r},f)\}^2}, \quad (30)$$

where $p_0(\mathbf{r},f)$ and $p(\mathbf{r},f)$ are the RMSs of the sound pressures in the original sound field and reproduced sound field defined as follows:

$$p_0(\mathbf{r},f) = \sqrt{\int_0^1 \{p_0(\mathbf{r},f,t)\}^2 dt}, \quad (31)$$

$$p(\mathbf{r},f) = \sqrt{\int_0^1 \{p(\mathbf{r},f,t)\}^2 dt}. \quad (32)$$

Note that $p_0(\mathbf{r},f)$ and $p(\mathbf{r},f)$ were normalized in the whole $\mathbf{r}$ before calculating the SNRs. The second measure used was the intensity direction error. This measure indicates the difference in the arrival direction of the sound between the original and reproduced sound fields and is defined as follows:

$$\theta(f)[\mathrm{degrees}] = \sqrt{\frac{1}{N}\sum_{\mathbf{r}}\left[\cos^{-1}\left\{\frac{\mathbf{I}(\mathbf{r},f)\cdot\mathbf{I}_0(\mathbf{r},f)}{|\mathbf{I}(\mathbf{r},f)||\mathbf{I}_0(\mathbf{r},f)|}\right\}\right]^2}, \quad (33)$$

where $\mathbf{I}_0(\mathbf{r},f)$ and $\mathbf{I}(\mathbf{r},f)$ are the sound intensity vectors in the original and reproduced sound fields, respectively, and $N(=162)$ is the total number of synthesis points.

The results of the SNRs and intensity direction errors for all the proposed methods are shown in Figures 7-8. In all the proposed methods, when the frequency of the sound sources was more than 2000 Hz, the SNRs were always less than 20 dB and the intensity direction errors were more than 20 degrees. This was due to the fact that the spatial sampling theorem to reproduce a wave front over 2000 Hz sound was not satisfied since the interval of the control points was more than half of the wavelength of over 2000 Hz sound.

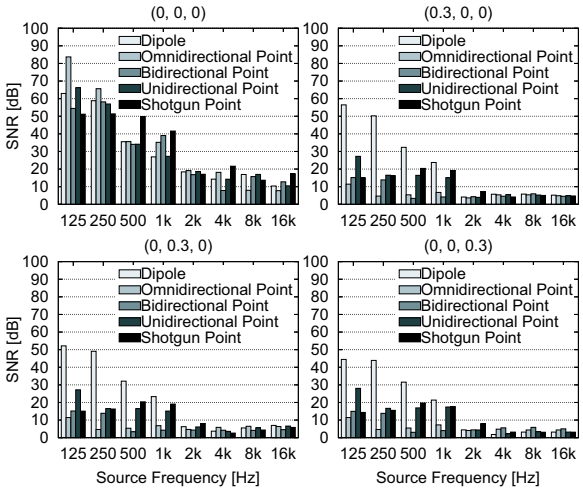

Figure 7: SNRs in proposed methods.

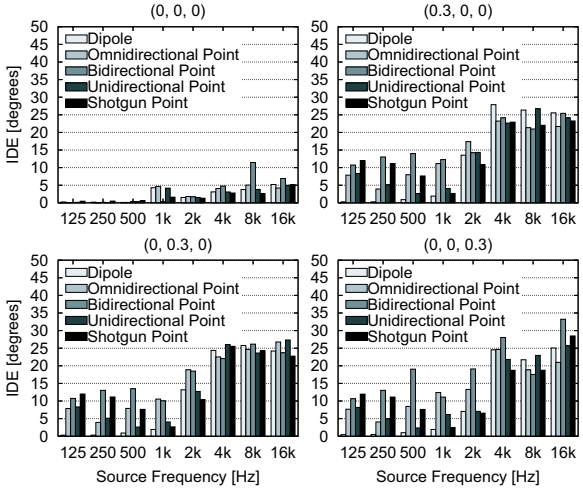

Figure 8: Intensity direction errors in proposed methods.

In the bidirectional point control method, when the position of the sound source was not central, the SNRs were less than 16 dB at frequencies of less than 1000 Hz and the intensity direction errors were more than 13 degrees. Thus, if systems were constructed using this method, listeners couldn't localize the direction of sound images since the sound intensities were not reproduced.

In the omnidirectional point control method, when the position of the sound source was not central, the SNRs were less than 12 dB at frequencies of less than 1000 Hz, although the intensity direction errors were less than 12.5 degrees. Thus, if systems were constructed using this method, listeners could localize the direction of sound images since the sound intensities were reproduced. However, they would not experience a realistic sensation when they moved around the loudspeaker array since the sound pressure distributions were not reproduced.

In the shotgun point control method, the SNRs were more than 14.3 dB and the intensity direction errors were less than 12.0 degrees at frequencies of less than 1000 Hz. In the unidirectional point control method, the SNRs were more than 15.0 dB and the intensity direction errors were less than 8.4 degrees at frequencies of less than 1000 Hz. Thus, if systems were constructed using these methods, listeners could localize the direction of sound images and experience a realistic sensation when they moved around the loudspeaker array since the sound pressure distributions and sound intensities were reproduced.

In the dipole control method, the SNRs were more than 21.3 dB and the intensity direction errors were less than 4.3 degrees at frequencies of less than 1000 Hz. Thus, if systems were constructed using this method, which requires twice number of microphones and loudspeakers, listeners could localize the direction of sound images and experience a more realistic sensation when they moved around the loudspeaker array since the sound pressure distributions and sound intensities were reproduced well.

4 Conclusion

In this paper, near 3D sound field reproduction techniques using wave field synthesis were proposed as methods of achieving ultra-realistic communication in applications such as 3D television and 3D teleconferencing. The principle was derived from the Kirchhoff-Helmholtz integral equation and two methods – dipole control and directional point control methods – were proposed. Computer simulation was used to evaluate the performance of the two methods. The results showed that the dipole control method performed very well, while the directional point control method performed satisfactorily if the directivity of the loudspeakers was unidirectional and shotgun.

Future work will include manufacturing microphone and loudspeaker arrays based on the proposed methods and evaluating the performance of the systems in a real environment using both acoustical measurement and subjective assessment.

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

[3] J. Bauck and D. H. Cooper, "Generalized transaural stereo and applications," *J. Audio Eng. Soc.*, vol. 44, no. 9, pp. 683–705, September 1996.

[4] O. Kirkeby, P. A. Nelson, and H. Hamada, "The 'Stereo Dipole': A virtual source imaging system using two closely spaced loudspeakers," *J. Audio Eng. Soc.*, vol. 46, no. 5, pp. 387–395, May 1998.

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

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

[7] 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.

[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] M. Born and E. Wolf, *Principles of Optics*, pp. 418–421, Cambridge University Press, Cambridge, UK, 7th(expanded) edition, 1999.

[10] 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 AES 13th Regional Conv.*, Tokyo, Japan, July 2007, number 0037, pp. 1–8.

[11] M. Born and E. Wolf, *Principles of Optics*, pp. 421–425, Cambridge University Press, Cambridge, UK, 7th(expanded) edition, 1999.

[12] H. Kenner, *Geodesic Math and How to Use It*, University of California Press, Berkeley, CA, second paperback edition, 2003.

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