INVERSE WAVE PROPAGATION FOR REPRODUCING VIRTUAL SOURCES IN FRONT OF LOUDSPEAKER ARRAY

Shoichi Koyama, Yusuke Hiwasaki, Ken'ichi Furuya, and Yoichi Haneda

NTT Cyber Space Laboratories, NTT Corporation
3-9-11, Midori-Cho, Musashino-Shi, Tokyo 180-8585, Japan
phone: + (81) 422-59-3089, fax: + (81) 422-60-7811, email: koyama.shoichi@lab.ntt.co.jp
web: http://www.ntt.co.jp/cclab/e/

ABSTRACT

This paper presents a new combination of wave field synthesis with inverse wave propagation that can recreate virtual sound sources in front of a loudspeaker array. Wave field synthesis is a well-known sound field reproduction technique, based on the Rayleigh integral, that reconstructs sound pressure distribution by using a planar or linear loudspeaker array. We show that, using a holographic approach, the reconstruction position can be displaced towards the listener, in front of the secondary sources. As a result, virtual primary sources can be placed between the listener and the secondary sources. Numerical simulation results are presented to show the efficacy of the proposed method. We implemented an experimental system using linear microphone and loudspeaker arrays to reproduce the sound field in a real environment. Results of perceptual experiments showed that the proposed method can achieve sound localization accuracy for virtual sound sources equivalent to that for real sound sources.

1. INTRODUCTION

Wave field synthesis (WFS) is well known as a large-area sound-field reconstruction technique based on Kirchhoff-Helmholtz or Rayleigh integrals [1, 2, 3]. On the basis of WFS, a far-end sound field is recorded by using a planar or linear microphone array and reproduced at the near-end by using a similarly arranged loudspeaker array as *secondary sources*.

Generally, it has been possible to recreate a sound field created by primary sources behind secondary sources by using WFS. As a future application of WFS, it would be useful to create sound images in front of secondary sources to synchronize with 3D visual images. Recent studies have shown that it is possible to reproduce point sources within the target area by applying a focused source technique [4, 5, 6, 7]. However, this technique is only valid when the position, signal, or directivity of primary sources is known. This is because this technique is based on focusing sound pressure at a specified position. Decomposing a set of signals received by an array of microphones into parameters, i.e., the source position, direction, and original signal, is not a trivial task.

To recreate virtual primary sources in front of secondary sources without the above parameters, the sound pressure distribution at the receiving plane must be reconstructed within the target area. In other words, the reconstruction plane must be shifted forward while the secondary sources are fixed. That can be achieved by inversely obtaining the sound pressure distribution at a distance from the receiving plane in the opposite direction to wave propagation, which is an inverse problem similar to holography or inverse wave-
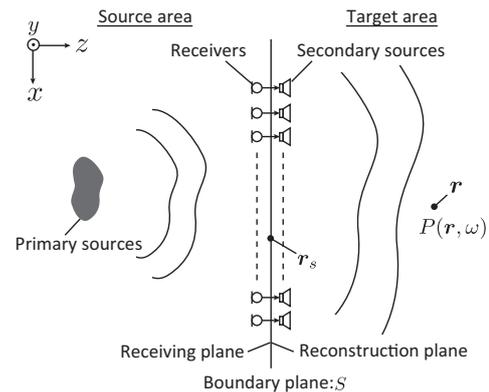

Figure 1: Sound field reproduction based on Rayleigh integrals [1].

field extrapolation [8, 9, 10]. We applied acoustical holography techniques for WFS to shift the reconstruction plane forward. As a result, primary sources were virtually recreated in front of secondary sources by using a filter determined by setting only a displacement parameter for shifting.

This paper is organized as follows. Section 2 deals with basic principles of WFS. Section 3 describes the application of inverse wave propagation to WFS for shifting the reconstruction plane. In Section 4, results of experiments are reported. Finally, Section 5 concludes this paper.

2. SOUND FIELD REPRODUCTION BASED ON WFS

Let us consider sound field reproduction based on the Rayleigh I integral [3]. This equation is a simplified representation of Kirchhoff-Helmholtz integrals, derived when the boundary geometry is spread out as an infinite plane as shown in Fig. 1 [8]:

$$P(\boldsymbol{r}, \omega) = -\iint_{-\infty}^{\infty} \frac{\partial P(\boldsymbol{r}_s, \omega)}{\partial z} \frac{\exp\left(-jk|\boldsymbol{r} - \boldsymbol{r}_s|\right)}{2\pi|\boldsymbol{r} - \boldsymbol{r}_s|} dx_s dy_s, (1)$$

where $\boldsymbol{r}$ denotes the vectorized position in the target area, $\boldsymbol{r}_s$ is the position vector at the boundary S, and $P(\boldsymbol{r}, \omega)$ and $P(\boldsymbol{r}_s, \omega)$ are the sound pressures of frequency ω at $\boldsymbol{r}$ and $\boldsymbol{r}_s$, respectively. $k = \omega/c$ is the wave number, and c is sound speed. The abbreviated notation $\partial/\partial z$ means the directional gradient in the direction of z at $\boldsymbol{r} = \boldsymbol{r}_s$. Equation (1) means that the sound field of the source area is reproduced by monopole secondary sources located at the reconstruction plane in the target area. Then, the driving signal of secondary
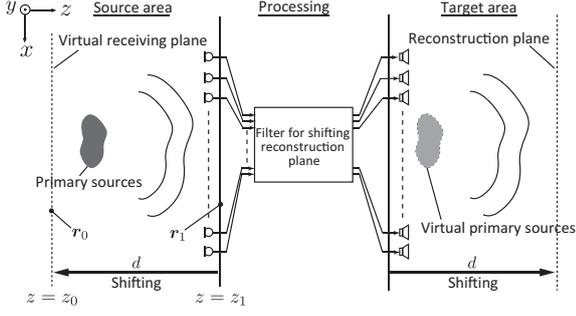
Figure 2: Shifting reconstruction plane forward. Virtual sources are recreated in front of secondary sources.

Figure 3: Block diagram of practical system for shifting reconstruction plane.

sources is represented as

$$D(\boldsymbol{r}_s, \omega) = -2\frac{\partial P(\boldsymbol{r}_s, \omega)}{\partial z}. \tag{2}$$

This equation is referred to as the *secondary source driving function* [3], which suggests the sound pressure gradient at the receiving plane is necessary for reproducing the sound field.

When considering only the horizontal ear-plane (at a constant y), an implementation of WFS can be reduced to a two-dimensional model, where the secondary source driving function of the linear loudspeaker array is represented as

$$D_{2.5D}(\boldsymbol{r}_s, \omega) = -\frac{g_c}{\sqrt{jk}}\frac{\partial P(\boldsymbol{r}_s, \omega)}{\partial z}. \tag{3}$$

Here, g_c is an amplitude correction factor [3, 11]. This approximation is known to result in artifacts in the target sound field, such as faster sound decay than in the real one [3].

3. INVERSE WAVE PROPAGATION FOR SHIFTING RECONSTRUCTION PLANE FORWARD

Shifting the reconstruction plane forward is identical to shifting the receiving plane backward. Therefore, it is necessary to obtain the gradient of the sound pressure in the opposite direction of wave propagation from the receiving plane. We approached this problem in a similar manner to holography and obtained a filter determined by setting only a displacement parameter. If the secondary sources reconstructed the sound pressure distribution of the virtual receiving plane, the sound field created by the secondary sources needs the sound pressure distribution of the actual receiving plane to be reconstructed at the shifted distance, i.e., reconstruction plane (Fig. 2).

3.1 Inverse Wave Propagation

Let us consider the direct direction of sound propagation from the virtual receiving plane at $z = z_0$ to the actual receiving plane at $z = z_1$. It is formulated as

$$P(\boldsymbol{r}_1, \omega) = \iint_{-\infty}^{\infty} P(\boldsymbol{r}_0, \omega)\frac{\partial}{\partial z}\frac{\exp(-jk|\boldsymbol{r}_1 - \boldsymbol{r}_0|)}{2\pi|\boldsymbol{r}_1 - \boldsymbol{r}_0|}dx_0 dy_0. \tag{4}$$

This equation is called the Rayleigh II integral [8].

If we neglect the evanescent components, the inversion formula of Eq. (4) is approximated as [9, 11]

$$P(\boldsymbol{r}_0, \omega) \simeq -\iint_{-\infty}^{\infty} P(\boldsymbol{r}_1, \omega)\frac{\partial}{\partial z}\frac{\exp(jk|\boldsymbol{r}_0 - \boldsymbol{r}_1|)}{2\pi|\boldsymbol{r}_0 - \boldsymbol{r}_1|}dx_1 dy_1. \tag{5}$$

Equation (5) provides the sound pressure at the virtual receiving plane obtained by inverse wave propagation when the sound pressure is known at the actual receiving plane. Note that sound pressure in the region behind of primary sources estimated by Eq. (5) is not ensured.

3.2 Application of inverse wave propagation to WFS

We propose applying inverse wave propagation to WFS. Sound pressure distribution at $z = z_1$ is obtained by using an equally distributed microphone array; the vectorized positions are denoted as $\boldsymbol{r}_n = (x_n, y_n, z_1)$, and signals are denoted as $P_n(\omega)$. By using inverse wave propagation, sound pressure distribution of discrete points at $z = z_0$ is obtained; the vectorized positions are denoted as $\boldsymbol{r}_m = (x_m, y_m, z_0)$, and signals are denoted as $P_m(\omega)$. When $P_m(\omega)$ is recreated by secondary sources, the reconstruction plane is shifted $d = |z_1 - z_0|$ in the target area. In this case, the secondary source driving function, denoted as $D_m(\omega)$, is obtained by substituting Eq. (5) into Eq. (2).

$$
\begin{aligned}
D_m(\omega) &= -2\frac{\partial}{\partial z}\left[\sum_n P_n(\omega)\right.\\
&\qquad \left.\frac{\partial}{\partial z}\left[\frac{\exp(jk|\boldsymbol{r}_m - \boldsymbol{r}_n|)}{2\pi|\boldsymbol{r}_m - \boldsymbol{r}_n|}\right]_{z=z_1}\right]_{z=z_0}\\
&= -\sum_n P_n(\omega)\left\{\frac{1 - jk|\boldsymbol{r}_m - \boldsymbol{r}_n|}{\pi|\boldsymbol{r}_m - \boldsymbol{r}_n|^3}\right.\\
&\qquad \left.+d^2\frac{3 - 3jk|\boldsymbol{r}_m - \boldsymbol{r}_n| - k^2|\boldsymbol{r}_m - \boldsymbol{r}_n|^2}{\pi|\boldsymbol{r}_m - \boldsymbol{r}_n|^5}\right\}\\
&\qquad \exp(jk|\boldsymbol{r}_m - \boldsymbol{r}_n|),
\end{aligned}
\tag{6}
$$

where

$$|\boldsymbol{r}_m - \boldsymbol{r}_n| = \sqrt{(x_m - x_n)^2 + (y_m - y_n)^2 + d^2}. \tag{7}$$

This equation provides the secondary source driving function to recreate the sound pressure distribution at the virtual receiving plane by using the sound pressure distribution at the actual receiving plane. Because the relative position of each

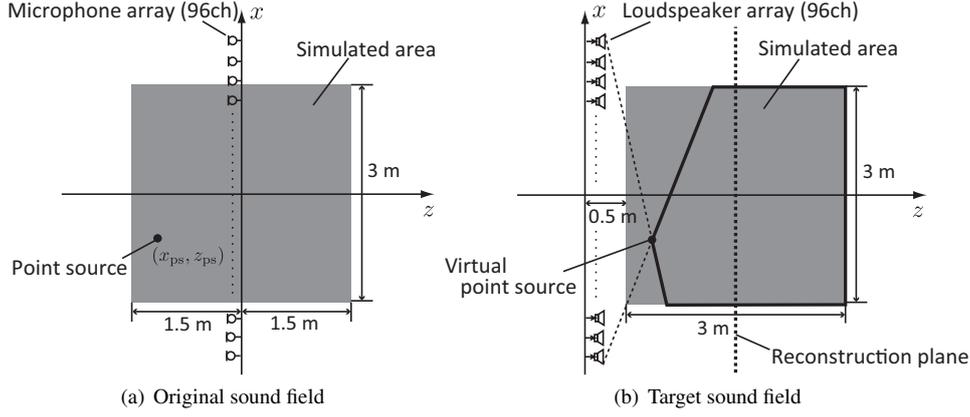(a) Original sound field   (b) Target sound field

Figure 4: Numerical simulation setup. When reconstruction plane is shifted 2 m forward, target sound field should ideally correspond with original sound field in simulated regions.

microphone $\boldsymbol{r}_n$ is known, $\boldsymbol{r}_m$ is geometrically determined by setting a displacement parameter d. Therefore, the filter for shifting the reconstruction plane is determined simply by setting d.

As previously stated, shifting the receiving plane backward is identical to shifting the reconstruction plane forward. If primary sources exist in $z_1 > z > z_0$, the primary sources are virtually reproduced in front of the secondary sources. Note that decomposition of the source parameters is unnecessary. However, the reproduction of the region behind of virtual primary sources is not accurate.

Two-dimensional formulation of Eq. (6) can be derived in a similar manner to the three-dimensional one. This is done using [11]

$$D_m(\omega) = -g_c \frac{\sqrt{jk}}{2} \sum_n P_n(\omega)$$
$$\left(\frac{|\boldsymbol{r}_m - \boldsymbol{r}_n|^2 + 2d^2}{|\boldsymbol{r}_m - \boldsymbol{r}_n|^3} H_1^{(1)} \left(k|\boldsymbol{r}_m - \boldsymbol{r}_n| \right) \right.$$
$$\left. - \frac{d^2}{|\boldsymbol{r}_m - \boldsymbol{r}_n|^2} k H_0^{(1)} \left(k|\boldsymbol{r}_m - \boldsymbol{r}_n| \right) \right) dx_1, \quad (8)$$

where

$$|\boldsymbol{r}_m - \boldsymbol{r}_n| = \sqrt{(x_m - x_n)^2 + d^2}. \quad (9)$$

Equation (8) provides the secondary source driving function to implement shifting of the reconstruction plane forward using linear arrays.

Figure 3 shows a block diagram of a practical system for shifting the reconstruction plane. Here, $p_n(t)$ and $d_m(t)$ are time domain representations of $P_n(\omega)$ and $D_m(\omega)$, respectively. $D_m(\omega)$ is derived by applying the filter for shifting reconstruction plane $G_{nm}(\omega)$ to $P_n(\omega)$.

$$D_m(\omega) = \sum_n G_{nm}(\omega) P_n(\omega), \quad (10)$$

where $G_{nm}(\omega)$ is derived from Eq. (6) or (8). In a two-dimensional case, $G_{nm}(\omega)$ is

$$G_{nm}(\omega) = -g_c \frac{\sqrt{jk}}{2} \left(\frac{|\boldsymbol{r}_m - \boldsymbol{r}_n|^2 + 2d^2}{|\boldsymbol{r}_m - \boldsymbol{r}_n|^3} H_1^{(1)} \left(k|\boldsymbol{r}_m - \boldsymbol{r}_n| \right) \right.$$
$$\left. - \frac{d^2}{|\boldsymbol{r}_m - \boldsymbol{r}_n|^2} k H_0^{(1)} \left(k|\boldsymbol{r}_m - \boldsymbol{r}_n| \right) \right). \quad (11)$$

The Hankel functions can be calculated numerically [12]. If the displacement parameter d is constant, G_{nm} is fixed as a matrix of filters. To reduce truncation errors caused by a finite array length, a tapering window is applied to the secondary source driving function [3].

4. EXPERIMENTS

4.1 Simulation results

Numerical simulation of 2D sound field reproduction using our proposed method for shifting the reconstruction plane was conducted. Figure 4 shows the simulation set up. Identical linear microphone and loudspeaker arrays were located at $z = 0$, with 96 channels in 1 array. The directivity of the array elements was considered as monopole. The elements of the arrays were equally spaced at 4 cm, so the lengths of the arrays became 3.8 m. The primary sound sources located at $z < 0$, i.e., the source area, were observed with a linear microphone array at $z = 0$. The original sound field created by primary sources in the source area was reproduced at $z \geq 0$ with a linear loudspeaker array at $z = 0$ as secondary sources. The original and target sound fields were simulated in 3×3 m regions (shaded regions in Fig. 4). When the reconstruction plane is shifted 2 m forward, the target sound field should ideally correspond with the original sound field in the simulated regions. The amplitude of reproducing signals, g_c in Eq. 8, was compensated to the proper reproduction level at the center of the simulated area, $(x, z) = (0.0 \text{ m}, 2.0 \text{ m})$.

Figure 5 shows the simulation results when the reconstruction plane was shifted 2 m forward and the sound source was a point source located at $(0.0 \text{ m}, -1.0 \text{ m})$. The source signal was a 1-kHz sinusoidal wave. Figure 5(a) shows the sound pressure distribution reproduced using the proposed method, and Fig. 5(b) shows the original sound pressure distribution in the source area. In this case, the virtual sound source appeared in the target area. When the target area was separated into two regions as shown in Fig. 5(c), the accuracy of reproduction was restricted in the bounded region. This is because it is impossible to perfectly reproduce divergence from a position in front of secondary sources by using a finite array. This restriction is similar to one in the conventional focused-source method [5]. We define the signal to distortion ratio (SDR) as the ratio of the original sound pressure distribution to the error of the reproduced sound pressure
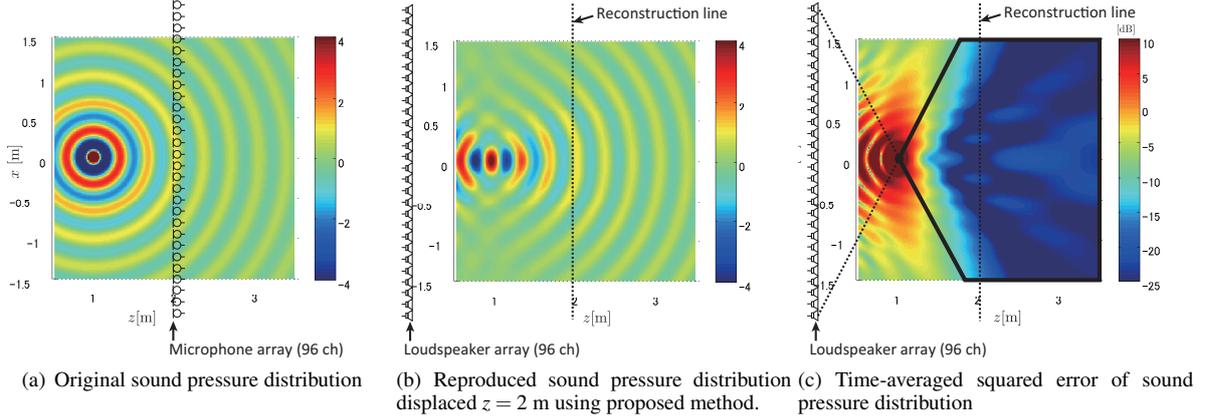
(a) Original sound pressure distribution

(b) Reproduced sound pressure distribution displaced $z = 2$ m using proposed method.

(c) Time-averaged squared error of sound pressure distribution

Figure 5: Simulation results when reconstruction plane was displaced 2 m.

(a) Original sound pressure distribution

(b) Reproduced sound pressure distribution displaced $z = 2$ m using proposed method.

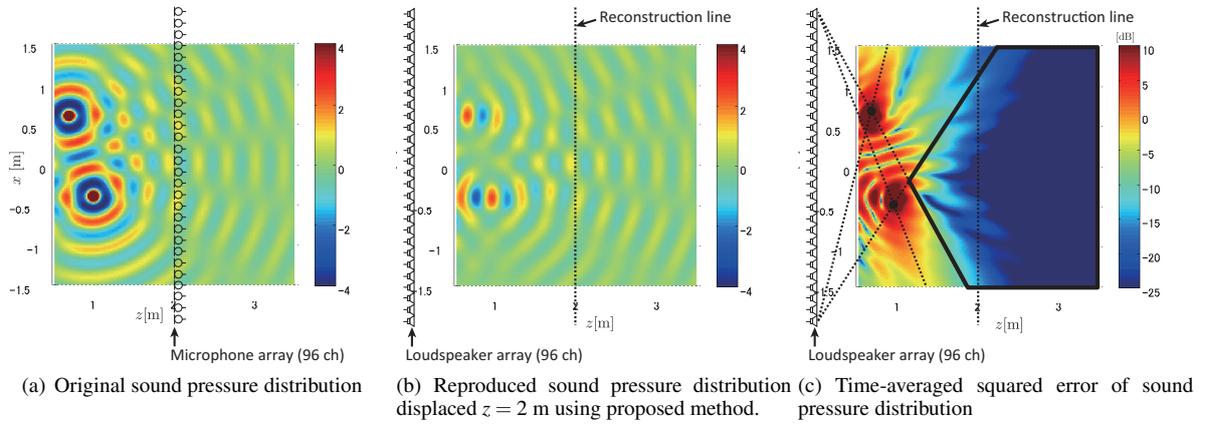(c) Time-averaged squared error of sound pressure distribution

Figure 6: Simulation results when two sources were used and reconstruction plane was displaced 2 m.

distribution. SDR can be written as

$$
\text{SDR} = -10\log_{10}\frac{1}{T}\sum_{t_k}\frac{\sum_{x_i}\sum_{z_j}|p(t_k,x_i,z_j) - p_{\text{org}}(t_k,x_i,z_j)|^2}{\sum_{x_i}\sum_{z_j}|p_{\text{org}}(t_k,x_i,z_j)|^2},
$$
(12)

where $p(t_k,x_i,z_j)$ and $p_{\text{org}}(t_k,x_i,z_j)$ are the reproduced and original sound pressure distributions in the time domain, respectively. T is the average time duration. Calculation of SDR was skipped in the region within 0.1 m of the primary sources. The SDR in the bounded region in Fig. 5(c) was 34.0 dB when $T = 10$ ms. Even when source parameters were known and the point source was reproduced by the conventional focused-source method, the SDR was almost the same.

Figure 6 shows the simulation results when the reconstruction plane was shifted 2 m forward, and the sound sources were two point sources located at $(-0.4 \text{ m}, -1.0 \text{ m})$ and $(0.6 \text{ m}, -0.8 \text{ m})$. The source signal was a 1-kHz sinusoidal wave. Figure 6(a) shows the sound pressure distribution reproduced using the proposed method, and Fig. 6(b) shows the original sound pressure distribution created by two point sources in the source area. The two sources were accurately reproduced inside the target area, while the exact region is restricted as shown in Fig. 6(c). The SDR in the bounded region in Fig. 6(c) was 36.5 dB when $T = 10$ ms. The filters used to produce the results in Figs. 5(b) and 6(b) were identical because the displacement parameters used in

the filters were the same. One advantage of the proposed method is that the same filter can be used for reproducing virtual sources in front of the loudspeaker array even when multiple sources are located in the source area.

4.2 Perceptual experiments of source localization

To perceptually evaluate the proposed method, we conducted a source localization listening test. Figure 7 shows the perceptual experiments setup. A linear microphone array and a loudspeaker array were set in the source room and in the reproduced room, respectively. Similar to the numerical simulation, there were 96 channels in an array, and the loudspeakers and microphones of the arrays were equally spaced at 4 cm. Both arrays were set 1.2 m above the floor. Spatial aliasing artifacts occured at frequencies above 4.3 kHz due to the finite intervals of the array elements. The loudspeaker elements were about 3.8 cm in diameter with enclosures and had almost monopole directivity up to 1.2 kHz. The speech signals of female and male utterances were used as source signals. The duration of each was 10 s, and the sampling frequency was 48 kHz. There were 7 listeners between the ages of 25 and 40. The amplitude of reproducing signals, g_c in Eq. (8), was compensated to the proper reproduction level at the listening position. The reverberation times of the source room and reproduced room were about 150 and 230 ms, respectively. The background noise level of the reproduced room was about 36.7 dBA.
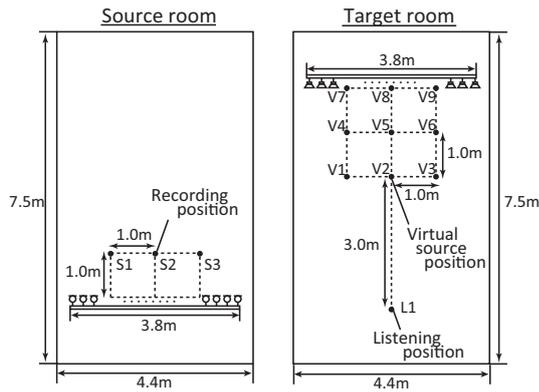
Figure 7: Perceptual experiments setup. S1–S3 indicate the recorded positions of primary sources, and V1–V9 indicate the reproduced positions of virtual sources.

As shown in Fig. 7, source signals were recorded at three positions, S1, S2, and S3, in the source room. A general closed loudspeaker was used as the primary sound source. Reproducing signals were created from signals recorded by the microphone array, and the reconstruction planes of these signals were shifted 1.0, 2.0, and 3.0 m by using the proposed method. Therefore, virtual sound sources were reproduced at V1–V9 by the loudspeaker array in the reproduced room. A listener was positioned at L1 and noted the reproduced positions of the virtual sound sources, V1–V9. A listener could see the signs of positions at V1–V9. The reproducing signals of V1–V9 were played randomly 27 times, so the reproducing signal of each virtual sound source was played 3 times.

Figure 8 shows the accuracy rate and its average including confidence intervals of listeners noting each virtual sound source location. The accuracy rate when real sound sources (loudspeakers) were located at V1–V9 and listeners noted the location of real sound sources is also shown. These results show that the accuracies of real and virtual source localization were almost the same. The accuracy rates of localization of real and virtual sources in line with the front-facing position of the listener, i.e., V2, V5, and V8, were relatively low. This is because differentiating the sound source distance is difficult for humans [13], so multiple sources located in the same direction caused some localization mistakes. For the remaining positions, the accuracy rate was found to be relatively high because the listener could distinguish the direction.

5. CONCLUSION

We presented a novel method that combines wave field synthesis with inverse wave propagation to recreate virtual sound sources in front of a loudspeaker array. This method shifts the reconstruction plane towards the listeners by inversely estimating the sound pressure at the virtual receiving plane. Compared to the conventional methods that require information on the primary sources at the far-end, this method enables us to recreate sound sources in front of secondary sources using a filter determined by a single displacement parameter. We also presented numerical simulation results showing that the sound field is accurately reproduced with a constraint due to limited secondary source length. The results of the perceptual experiments showed that this method can achieve a sound localization accuracy for virtual sound
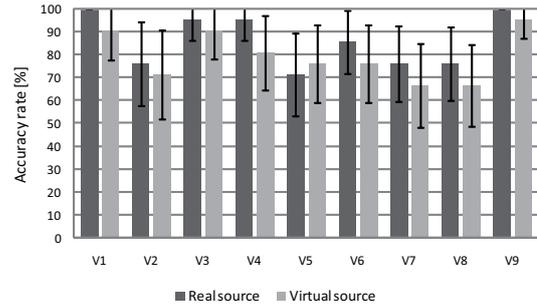

Figure 8: Results of perceptual experiments. Accuracy rate and its average including confidence intervals of listeners noting sound source location are shown.

sources equivalent to that for real sound sources.

REFERENCES

[1] A. J. Berkhout, D. de Vries, and P. Vogel. Acoustic control by wave field synthesis. *J. Acoust. Soc. Am.*, 93(5):2764–2778, 1993.

[2] D. Vries. *Wave Field Synthesis*. AES Monograph, 2009.

[3] S. Spors, R. Rabenstein, and J. Ahrens. The theory of wave field synthesis revisited. *124th AES Convention*, 2008.

[4] M. Boone, E. Verheijen, and G. Jansen. Virtual reality by sound reproduction based on wave field synthesis. *100th AES Convention*, 1996.

[5] S. Spors et al. Physical and perceptual properties of focused sources in wave field synthesis. *127th AES Convention*, 2009.

[6] T. Caulkins et al. Wave field synthesis interaction with the listening environment, improvements in the reproduction of virtual sources situated inside the listening room. *Proc. of the 6th Int. Conference on Digital Audio Effects (DAFx-03)*, 2003.

[7] S. Yon et al. Sound focusing in rooms: The time-reversal approach. *J. Acoust. Soc. Am.*, 113(3):1533–1543, 2003.

[8] E. G. Williams. *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. Academic Press, 1999.

[9] M. Nieto-Vesperinas. *Scattering and Diffraction in Physical Optics*. John Wiley & Sons, Inc, 1991.

[10] C. P. A. Wapenaar, G. L. Peeis, V. Budejicky, and A. J. Berkhout. Inverse extrapolation of primary seismic waves. *GEOPHYSICS*, 54(7):853–863, 1989.

[11] S. Koyama et al. Inverse wave propagation in wave field synthesis. *40th AES Conference*, 2010.

[12] D. E. Amos. A portable package for bessel functions of a complex argument and nonnegative order. *ACM Trans. Math. Software*, 12(3):265–273, 1986.

[13] J. Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localization*. The MIT Press, 1996.