SPARSE SOUND FIELD REPRESENTATION IN RECORDING AND REPRODUCTION FOR REDUCING SPATIAL ALIASING ARTIFACTS

Shoichi Koyama, Suehiro Shimauchi, Hitoshi Ohmuro

NTT Media Intelligence Laboratories, NTT Corporation, Tokyo, Japan

ABSTRACT

We propose a sound-pressure-to-driving-signal (SP-DS) conversion method for sound field reproduction based on sparse sound field representation. The most important problem in sound field reproduction is how to calculate driving signals of loudspeakers to reproduce desired sound fields. In common recording and reproduction systems, sound pressures at multiple positions obtained in a recording area are only known as the desired sound field; therefore, SP-DS conversion algorithms are necessary. Current SP-DS conversion methods do not take into account sound sources to be reproduced, which results in severe spatial aliasing artifacts. Our proposed method decomposes the received sound pressure distribution based on the generative model of the sound field. Numerical simulation results indicate that the proposed method can achieve higher reproduction accuracy compared to the current methods, especially in higher frequencies above the spatial Nyquist frequency.

Index Terms— Sound field reproduction, sparse signal representation, wave field synthesis, wave field reconstruction filter, super-resolution

1. INTRODUCTION

Physical sound field reproduction makes it possible for high-fidelity audio systems. In practical recording and reproduction systems, the sound field to be reproduced may be obtained with microphones in a recording area; therefore, sound pressures at multiple positions in the desired sound field are only known. This means that driving signals of loudspeakers for reproducing sound field needs to be calculated from the received signals of the microphones. We call this type of signal transformation sound-pressure-to-driving-signal (SP-DS) conversion. We focus on the SP-DS conversion problem when the array configuration of the microphones and loudspeakers are planar or linear.

Wave field synthesis (WFS) [1] is a sound field reproduction method based on the Kirchhoff-Helmholtz or Rayleigh integrals. By using the Rayleigh integral of the first kind, the driving signals of the WFS for a planar or linear loudspeaker array needs to be equivalent to the distribution of the sound pressure gradient of the desired sound field at the position of the array [2]. Because it is difficult to obtain the distribution of sound pressure gradient in practice, WFS cannot be directly applied as SP-DS conversion.

We have proposed an SP-DS conversion method for planar or linear arrays of microphones and loudspeakers based on the *wave field reconstruction (WFR) filter* [3]. In the WFR filtering method, SP-DS conversion is achieved by decomposing the received sound pressure distribution into spatial Fourier basis functions that correspond to uniformly-sampled plane waves. This representation makes it possible for stable and efficient signal conversion. However, in this procedure, artifacts originating from spatial aliasing notably occur, which depends on the intervals of microphones and loudspeakers. As an example, let us assume a system in which loudspeakers are densely arranged compared to microphones. Even though this system has potential to reproduce a sound field in the frequency bands up to the spatial Nyquist frequency determined by the loudspeaker intervals, it can be hardly expected to reproduce it in the frequency bands above the spatial Nyquist frequency determined by the microphone intervals by using the WFR filtering method. Under the significant effect of the spatial aliasing artifacts, listeners may not clearly localize reproduced sound images. Additionally, frequency characteristics of the reproduced direct sound are greatly affected, which is called the *coloration effect* [4].

Now, the question is how to reproduce the sound field in the frequency bands above the spatial Nyquist frequency determined by the microphone intervals in the above-mentioned setup. This so-called super-resolution of sound field reproduction may be possible if the received sound pressure distribution can be represented using a small number of basis functions because the signals can be interpolated more precisely. We used basis functions that depend on sound sources to be reproduced, *primary sources*, for sparse representation of the sound field. We formulated a generative model of the sound field, which is represented by the sum of monopole source and plane wave components. Such decomposition is not a trivial task, but recently developed sparse decomposition algorithms [5] can be applied to this problem. These decomposed signals are separately converted, and then the driving signals of the loudspeakers are obtained as the sum of them. We demonstrate that the proposed method is robust against spatial aliasing artifacts as well as modeling errors.

Sound pressure control methods based on inverse filtering can be applied as SP-DS conversion [6, 7]. Even though these methods are aimed to minimize errors between the synthesized and desired sound pressures at control points, for example, in a least square error sense, the above-mentioned spatial aliasing artifacts cannot be avoided. In a spherical array, Wabnitz *et al.* [8] proposed an upscaling method for Ambisonics order based on sparse plane-wave decomposition. However, plane-wave decomposition of the received sound pressure distribution can be hardly sparse in our planar or linear array case. In [9], the optimized basis functions were obtained using eigenvalue decomposition based on MAP estimation with the prior knowledge of the primary sources. Therefore, these locations need to be given *a priori*. Our new perspective in the generative model and sparse-representation-based algorithm for sound field recording and reproduction are the main contributions of this paper.

2. GENERATIVE MODEL OF SOUND FIELD

We divide a sound field in the recording area into two regions, internal and external, of a closed surface. The internal region is denoted as Ω. We assume that components approximated as monopole sources exist only inside Ω. When the sound pressure of the tempo-

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

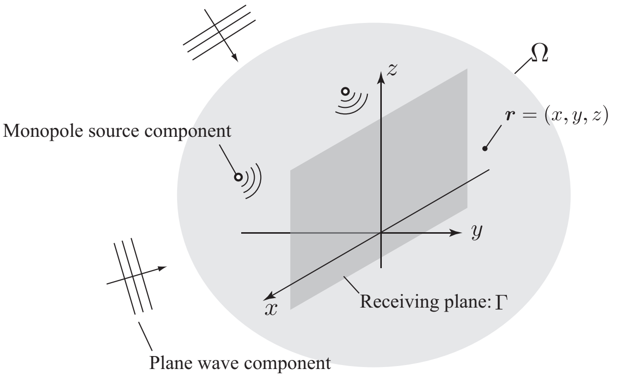


Fig. 1. Sound field in recording area is modeled by sum of monopole source and plane wave components. Sound pressure distribution is obtained by planar distribution of receivers.

ral frequency ω at the position $\boldsymbol{r}$ is denoted as $p(\boldsymbol{r},\omega)$, the following equation should be satisfied:

$$\left(\nabla^2+k^2\right)p(\boldsymbol{r},\omega)=\begin{cases}-Q(\boldsymbol{r},\omega), & \boldsymbol{r}\in\Omega\\ 0, & \boldsymbol{r}\notin\Omega\end{cases}, \tag{1}$$

where $Q(\boldsymbol{r},\omega)$ is the distribution of the monopole source components inside Ω, $k=\omega/c$ is the wave number, and c is the sound speed. Hereafter, ω is omitted for notational simplicity. Equation (1) means that $p(\boldsymbol{r})$ follows the inhomogeneous and homogeneous Helmholtz equations at $\boldsymbol{r}\in\Omega$ and $\boldsymbol{r}\notin\Omega$, respectively. Therefore, the solution of (1) can be represented as the sum of inhomogeneous and homogeneous terms, $p_{\rm i}(\boldsymbol{r})$ and $p_{\rm h}(\boldsymbol{r})$. Additionally, $p_{\rm i}(\boldsymbol{r})$ is represented as a convolution of $Q(\boldsymbol{r})$ and the three-dimensional free-field Green's function $G(\boldsymbol{r}|\boldsymbol{r}')$ as [10]:

$$\begin{aligned}p(\boldsymbol{r}) &= p_{\rm i}(\boldsymbol{r})+p_{\rm h}(\boldsymbol{r})\\ &= \int_{\boldsymbol{r}'\in\Omega}Q(\boldsymbol{r}')G(\boldsymbol{r}|\boldsymbol{r}')d\boldsymbol{r}'+p_{\rm h}(\boldsymbol{r}),\end{aligned} \tag{2}$$

where

$$G(\boldsymbol{r}|\boldsymbol{r}')=\frac{e^{jk|\boldsymbol{r}-\boldsymbol{r}'|}}{4\pi|\boldsymbol{r}-\boldsymbol{r}'|}. \tag{3}$$

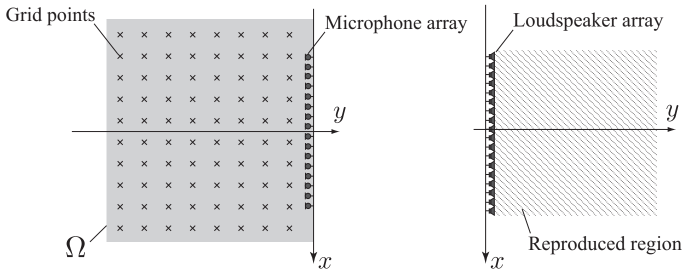
Here, $G(\boldsymbol{r}|\boldsymbol{r}')$ corresponds to the transfer function of the monopole source. Equation (2) can be confirmed by substituting it into (1) as:

$$\begin{aligned}&(\nabla^2+k^2)\left\{\int_{\boldsymbol{r}'\in\Omega}Q(\boldsymbol{r}')G(\boldsymbol{r}|\boldsymbol{r}')d\boldsymbol{r}'+p_{\rm h}(\boldsymbol{r})\right\}\\ &\quad=-\int_{\boldsymbol{r}'\in\Omega}\delta(\boldsymbol{r}-\boldsymbol{r}')Q(\boldsymbol{r}')d\boldsymbol{r}'\\ &\quad=\begin{cases}-Q(\boldsymbol{r}), & \boldsymbol{r}\in\Omega\\ 0, & \boldsymbol{r}\notin\Omega\end{cases}.\end{aligned} \tag{4}$$

The homogeneous term $p_{\rm h}(\boldsymbol{r})$ can be represented as the sum of plane waves; therefore, $p_{\rm h}(\boldsymbol{r})$ is described using the spatial frequency spectrum $P_{\rm h}(k_x,k_z)$ as [10]:

$$p_{\rm h}(\boldsymbol{r})=\frac{1}{4\pi^2}\int_{-\infty}^{\infty}dk_x\int_{-\infty}^{\infty}dk_z P_{\rm h}(k_x,k_z)e^{j(k_x x+k_y y+k_z z)}, \tag{5}$$

where k_x and k_z respectively denote the spatial frequencies with respect to x and z, and $k_y=\pm\sqrt{k^2-k_x^2-k_z^2}$.



(a) Recording area (b) Target area

Fig. 2. Microphone and loudspeaker arrays are respectively set in recording and target areas. Region Ω is discretized as set of grid points.

As shown in Fig. 1, we assume that the sound pressure distribution on the receiving plane Γ is obtained in the recording area. All the monopole source and plane wave components are assumed to be in the region of $y<0$. The secondary sources are also assumed to be planarly distributed in the target area. The driving signals of the secondary sources for reproducing the captured sound field should be equivalent to the sound pressure gradient on Γ [2]. Therefore, if the received sound pressure distribution $p(\boldsymbol{r})(\boldsymbol{r}\in\Gamma)$ can be decomposed into $p_{\rm i}(\boldsymbol{r})$ and $p_{\rm h}(\boldsymbol{r})$, the driving signals of the secondary sources $d(\boldsymbol{r})$ can be uniquely calculated as:

$$\begin{aligned}d(\boldsymbol{r}) &= \left.\frac{\partial}{\partial y}p(\boldsymbol{r})\right|_{y=0}\\ &= \int_{\boldsymbol{r}'\in\Omega}Q(\boldsymbol{r}')\left.\frac{\partial G(\boldsymbol{r}|\boldsymbol{r}')}{\partial y}\right|_{y=0}d\boldsymbol{r}'\\ &\quad+\frac{1}{4\pi^2}\int_{-\infty}^{\infty}dk_x\int_{-\infty}^{\infty}dk_z jk_y P_{\rm h}(k_x,k_z)e^{j(k_x x+k_z z)}.\end{aligned} \tag{6}$$

In the WFR filtering method, all the captured sound field is treated as the homogeneous term $p_{\rm h}(\boldsymbol{r})$, and $p(\boldsymbol{r})$ $(\boldsymbol{r}\in\Gamma)$ is decomposed into the spatial Fourier basis functions as in (5). Even when the captured sound field is generated by a single point source, the energy of the received sound pressure distribution spreads over on the basis functions in this decomposition; therefore, it is difficult to avoid folding noise. On the other hand, if the decomposition into $p_{\rm i}(\boldsymbol{r})$ and $p_{\rm h}(\boldsymbol{r})$ is achieved, dominant components of the received sound pressure distribution may lie in $p_{\rm i}(\boldsymbol{r})$ because $G(\boldsymbol{r}|\boldsymbol{r}')$ correspond to the basis functions of the monopole sources that are located adjacent to Γ. Additionally, $Q(\boldsymbol{r})$ $(\boldsymbol{r}\in\Omega)$ may become sparse because it can be considered that the monopole source components exist only at a few locations in Ω. When these assumptions are nearly satisfied, more accurate calculation of $d(\boldsymbol{r})$ can be achieved.

3. SP-DS CONVERSION ALGORITHM BASED ON SPARSE DECOMPOSITION

We describe the SP-DS conversion algorithm based on (2) and (6). As shown in Fig. 2, arrays of microphones and loudspeakers are respectively used as the distributions of receivers and secondary sources in the recording and target areas. The region Ω is discretized as a set of grid points. The numbers of microphones, loudspeakers, and grid points are denoted as M, L, and N, respectively. Here, $N\gg M$ and $L>M$ are assumed. Vectors $\mathbf{p}^{(i)}\in\mathbb{C}^M$ and $\mathbf{d}^{(i)}\in\mathbb{C}^L$ respectively denote the received signals of the

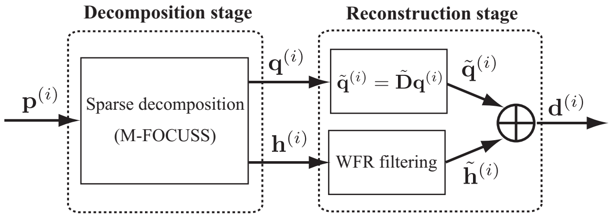


Fig. 3. Block diagram of proposed method.

microphones and driving signals of the loudspeakers in the temporal frequency domain. The superscript i is the index of the time frame. Therefore, SP-DS conversion must obtain $\mathbf{d}^{(i)}$ from $\mathbf{p}^{(i)}$. As shown in Fig. 3, the SP-DS conversion algorithm has two stages: decomposition and reconstruction.

In the decomposition stage, $\mathbf{p}^{(i)}$ is decomposed into two components as (2). When the dictionary matrix of the monopole components, which has the Green's function (3) between the grid points and microphones in each element, is denoted as $\mathbf{D} \in \mathbb{C}^{M \times N}$, the discrete form of (2) is described as:

$$\mathbf{p}^{(i)} = \mathbf{D}\mathbf{q}^{(i)} + \mathbf{h}^{(i)}, \tag{7}$$

where $\mathbf{q}^{(i)} \in \mathbb{C}^N$ is the distribution of the monopole components at the grid points, and $\mathbf{h}^{(i)} \in \mathbb{C}^M$ is the homogeneous term of the received signals. When the multiple time frames of $\mathbf{p}^{(i)}$, $\mathbf{q}^{(i)}$, and $\mathbf{h}^{(i)}$ $(i \in \{1, \cdots, I\})$ are represented by single matrices as $\mathbf{P} = [\mathbf{p}^{(1)}, \cdots, \mathbf{p}^{(I)}]$, $\mathbf{Q} = [\mathbf{q}^{(1)}, \cdots, \mathbf{q}^{(I)}]$, and $\mathbf{H} = [\mathbf{h}^{(1)}, \cdots, \mathbf{h}^{(I)}]$, (7) is rewritten as:

$$\mathbf{P} = \mathbf{D}\mathbf{Q} + \mathbf{H}. \tag{8}$$

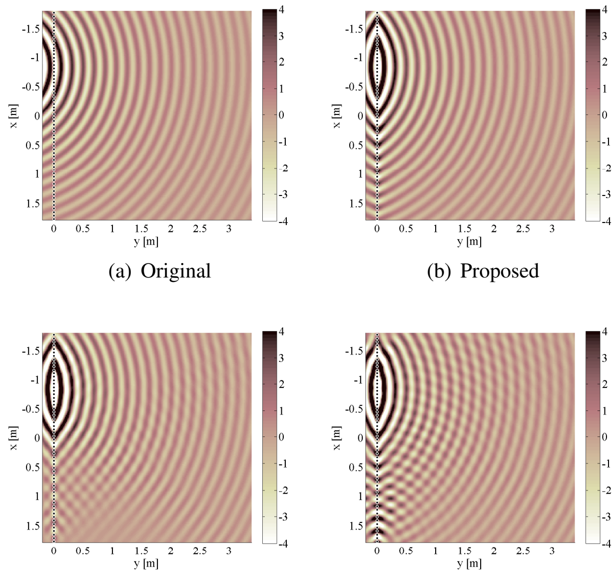
If Ω includes all the direct primary sources, $\mathbf{q}^{(i)}$ is a dominant component of $\mathbf{p}^{(i)}$ compared to $\mathbf{h}^{(i)}$. In contrast, $\mathbf{h}^{(i)}$ corresponds to ambient components such as reverberation and diffused noise. Additionally, only a few element of $\mathbf{q}^{(i)}$ may have non-zero values. Therefore, the sparse decomposition algorithm [5] can be applied to decompose $\mathbf{p}^{(i)}$ into $\mathbf{q}^{(i)}$ and $\mathbf{h}^{(i)}$. Practically, the directivity of the primary sources is different from monopole characteristics and the source locations may not be precisely on the grid points. However, these modeling errors can also be included in $\mathbf{h}^{(i)}$. The sampling interval and size of Ω affect the size of $\mathbf{D}$ as well as the restricted isometry property (RIP) condition [11]; therefore, these parameters need to be designed based on approximation error and computational complexity. In the context of source localization, similar algorithms to this decomposition stage have been proposed [12, 13].

For sparse decomposition of $\mathbf{p}^{(i)}$, it can be assumed that the locations of the primary sources, i.e., the positions of non-zero elements of $\mathbf{q}^{(i)}$, are constant at each i in a short period. Therefore, the sparse decomposition algorithm should take into account that multiple measurements of $\mathbf{p}^{(i)}$ have the same sparsity structure. The diversity measure of $\mathbf{Q}$ for this multiple measurement vectors (MMV) problem is generally defined as [14]:

$$J^{(p,q)}(\mathbf{Q}) = \sum_{n=1}^{N} \left(\|\mathbf{Q}[n]\|_q\right)^p, \ \ 0 \leq p \leq 1, \ q \geq 1, \tag{9}$$

where $\mathbf{Q}[n]$ denotes the n-th row of $\mathbf{Q}$. The optimization criteria is described as:

$$\arg\min_{\mathbf{Q}} \left\{ \|\mathbf{P} - \mathbf{D}\mathbf{Q}\|_F^2 + \lambda J^{(p,q)}(\mathbf{Q}) \right\}, \tag{10}$$



(a) Original (b) Proposed

(c) WFR (d) SPC

Fig. 4. Original and reproduced sound pressure distributions when source signal was 1.8 kHz. Spatial aliasing artifacts can be seen in WFR and SPC.

where λ is a predefined parameter that balances the approximation error and the sparsity-inducing penalty $J^{(p,q)}(\mathbf{Q})$. We applied the M-FOCUSS algorithm [15] to solve (10). Then, $\mathbf{H}$ can be simply obtained as $\mathbf{H} = \mathbf{P} - \mathbf{D}\mathbf{Q}$.

In the reconstruction stage, $\mathbf{d}^{(i)}$ is obtained from the estimates of $\mathbf{q}^{(i)}$ and $\mathbf{h}^{(i)}$. These two components $\mathbf{q}^{(i)}$ and $\mathbf{h}^{(i)}$ are separately converted into the driving signals, then $\mathbf{d}^{(i)}$ is obtained as the sum of them:

$$\mathbf{d}^{(i)} = \tilde{\mathbf{q}}^{(i)} + \tilde{\mathbf{h}}^{(i)}, \tag{11}$$

where $\tilde{\mathbf{q}}^{(i)} \in \mathbb{C}^L$ and $\tilde{\mathbf{h}}^{(i)} \in \mathbb{C}^L$ are the driving signals calculated from $\mathbf{q}^{(i)}$ and $\mathbf{h}^{(i)}$, respectively. We define $\tilde{\mathbf{D}} \in \mathbb{C}^{L \times N}$ as the matrix that converts $\mathbf{q}^{(i)}$ into $\tilde{\mathbf{q}}^{(i)}$ as $\tilde{\mathbf{q}}^{(i)} = \tilde{\mathbf{D}}\mathbf{q}^{(i)}$. Based on (6), each element of $\tilde{\mathbf{D}}$ has the differential of the Green's function (3) with respect to y. An alternative method for obtaining $\tilde{\mathbf{q}}^{(i)}$ from $\mathbf{q}^{(i)}$ is the spectral division method (SDM) [16]. Although the SDM is based on the signal representation in the spatial frequency domain, it is possible to calculate $\tilde{\mathbf{q}}^{(i)}$ at the resolution determined by the loudspeaker intervals because the position and strength of the monopole components, $\mathbf{q}^{(i)}$, can be given. The other component $\tilde{\mathbf{h}}^{(i)}$ can be obtained by applying the WFR filter to $\mathbf{h}^{(i)}$ [3]. To obtain the L signal of $\tilde{\mathbf{h}}^{(i)}$ from the M signal of $\mathbf{h}^{(i)}$, spatial signal interpolation is necessary.

4. EXPERIMENTS

Numerical simulations were conducted under the free-field assumption to compare three methods; the proposed method based on sparse representation (Proposed), WFR filtering method (WFR) [3], and sound pressure control method (SPC) [6]. Although the proposed method was derived in the case of the planar arrays of microphones and loudspeakers, we assumed that these arrays are linear in the experiments. The proposed method can be straightforwardly extended to the linear case.

As in Fig. 2, the microphones and loudspeakers were aligned along the x-axis with the center at the origin in the recording and target areas, respectively. The numbers of microphones and loudspeakers were respectively 32 and 64. The intervals of the microphones were 12 cm and those of the loudspeakers were 6 cm; therefore, both the array lengths of the microphones and loudspeakers were 3.84 m. The directivity of the array elements was assumed to be omni-directional. A single static sound source as a primary source was located in the recording area. The directivity of the primary source was assumed to be uni-directional. The primary source was located at $(-0.84, -0.98, 0.0)$ m and directed to $y > 0$. With a linear loudspeaker array, the amplitude decay of the reproduced source becomes faster than desired even though this artifact may not be critical to listeners' perception [4]. Therefore, we eliminated this effect for feasible evaluation by modifying the amplitude decay of the primary source in the original sound field from $1/r$ to $1/r\sqrt{r}$, where r is the distance from the source location. The sound pressure distributions were simulated in a 3.6×3.6 m region at intervals of 1.5 cm on the x-y-plane at $z = 0$. The amplitudes were normalized using the averaged squared amplitude in the region of $y \geq 0.5$ m in the simulated region. The sampling frequency was 48 kHz and the length of the single time frame was 2048 samples.

In Proposed, the grid points were aligned in a rectangular region of 4.0×3.0 m centered at $(0.0, -1.5, 0.0)$ m. The interval of the grid points was 0.1 m; therefore, the number of the grid points was 40×30. In M-FOCUSS of the proposed method, p and q in (9) were set as $p = 0.8$ and $q = 2$, λ in (10) was set as 1.0×10^{-4}, and 40 time frames were used for the decomposition. The SDM was applied to obtain $\tilde{\mathbf{q}}^{(i)}$. In Proposed and WFR, sinc interpolation was used for signal up-sampling to obtain signals of 64 channels from those of 32 channels. A spatial Tukey window function was applied to Proposed and WFR. In SPC, control points were set along the line of $y = 0.5$ m. The inverse of the transfer function matrix was calculated as the minimum norm solution. The microphone array was set at $y = 0.5$ m in the recording area to match the target sound field with that of Proposed and WFR.

The general reproduction accuracy was evaluated using the signal to distortion ratio (SDR) defined as:

$$\mathrm{SDR} = 10\log_{10}\frac{\sum_i\sum_j\sum_k|\bar{p}_{\mathrm{org}}(x_i,y_j,t_k)|^2}{\sum_i\sum_j\sum_k|\bar{p}_{\mathrm{rep}}(x_i,y_j,t_k)-\bar{p}_{\mathrm{org}}(x_i,y_j,t_k)|^2}, \quad (12)$$

where $\bar{p}_{\mathrm{rep}}(x_i, y_j, t_k)$ and $\bar{p}_{\mathrm{org}}(x_i, y_j, t_k)$ are the reproduced and original sound pressure distributions in the time domain, (x_i, y_j) denotes discrete positions in the simulated region, and t_k denotes discrete time. The total number of time samples was set at 10 ms, i.e., 480 samples. The SDR was calculated in the region of $y \geq 0.5$ m.

Fig. 4 shows the simulation results when the source signal was a 1.8-kHz sinusoidal wave, which is above the spatial Nyquist frequency determined by the interval of the microphones, 1.4 kHz. The black dots indicate the locations of the loudspeakers. Even though spatial aliasing artifacts arose in the WFR and SPC (Figs. 4c and d), they did not arise in Proposed (Fig. 4b). The SDRs of Proposed, WFR, and SPC were 23.6, 11.5, and 10.3 dB, respectively.

Fig. 5 plots the relation between SDRs and the frequency of the source signal. The spatial Nyquist frequency determined by the interval of the microphones is indicated by the dashed line. The SDRs of Proposed and WFR at low frequencies were relatively smaller than those of SPC. This error originated from that of the sinc interpolation. The SDRs of Proposed, at a higher frequency than the spatial Nyquist frequency, were maintained compared to WFR and SPC. Therefore, Proposed can be considered as being more robust against spatial aliasing artifacts. It should be noted that the directivity of the primary source was different from the monopole characteristics and the source location was not exactly on the grid points.

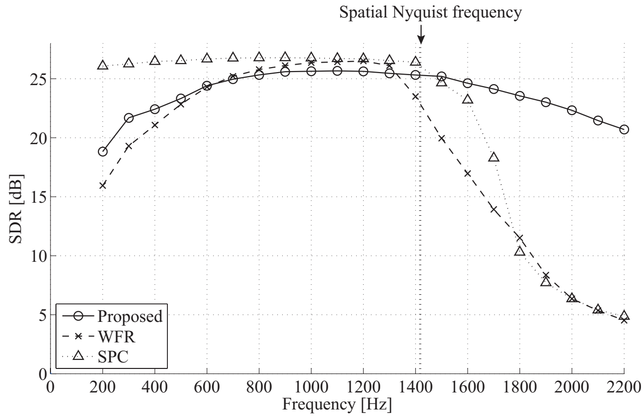


Fig. 5. Relation between SDR and frequency of source signal.

5. CONCLUSION

We proposed the SP-DS conversion method based on the sparse decomposition of the received sound pressure distribution. We formulated the generative model of the sound field as the sum of the monopole source and plane wave components. The received signals of the microphones are decomposed into these two components by using the sparse decomposition algorithm. These components are then separately converted, and the driving signals of the loudspeakers are obtained as the sum of them. Numerical simulations were conducted to compare the proposed method with the WFR filtering method and sound pressure control method based on the minimum norm solution. The reproduction accuracy of the proposed method was better than that of the other methods, especially at frequencies above the spatial Nyquist frequency. This result indicates that the proposed method is robust against spatial aliasing artifacts as well as modeling errors. Although the proposed method aims at super-resolution of sound field reproduction, the generative model and decomposition algorithm of the proposed method may also be applied to sound field analysis.

6. REFERENCES

[1] 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.

[2] S. Spors, R. Rabenstein, and J. Ahrens, "The theory of wave field synthesis revisited," in *Proc. 124th Conv. AES*, Amsterdam, The Netherlands, Oct. 2008.

[3] S. Koyama, K. Furuya, Y. Hiwasaki, and Y. Haneda, "Analytical approach to wave field reconstruction filtering in spatio-temporal frequency domain," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 4, pp. 685–696, 2013.

[4] D. de Vries, *Wave Field Synthesis*, AES Monograph. Audio Eng. Soc., 2009.

[5] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, 2010.

[6] P. A. Gauthier and A. Berry, "Sound-field reproduction in-room using optimal control techniques: Simulations in the frequency domain," *J. Acoust. Soc. Am.*, vol. 117, no. 2, pp. 662–678, 2005.

[7] M. Kolundžija, C. Faller, and M. Vitterli, "Reproducing sound fields using MIMO acoustic channel inversion," *J. Audio Eng. Soc.*, vol. 59, no. 10, pp. 721–734, 2011.

[8] A. Wabnitz, N. Epain, A. McEwan, and C. Jin, "Upscaling ambisonics sound scenes using compressed sensing techniques," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, 2011, pp. 1–4.

[9] S. Koyama, K. Furuya, Y. Hiwasaki, and Y. Haneda, "MAP estimation of driving signals of loudspeakers for sound field reproduction from pressure measurements," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, 2013.

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

[11] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[12] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, 2005.

[13] J. Le Roux, P. T. Boufounos, K. Kang, and J. R. Hershey, "Source localization in reverberant environments using sparse optimization," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, 2013, pp. 4310–4314.

[14] A. Rakotomamonjy, "Surveying and computing simultaneous sparse approximation (or group-lasso) algorithms," *Signal Processing*, vol. 91, pp. 1505–1526, 2011.

[15] S. F. Cotter, D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, 2005.

[16] J. Ahrens and S. Spors, "Sound field reproduction using planar and linear arrays of loudspeakers," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 8, pp. 2038–2050, 2010.