

EFFECT OF MULTIPOLE DICTIONARY IN SPARSE SOUND FIELD DECOMPOSITION FOR SUPER-RESOLUTION IN RECORDING AND REPRODUCTION

Shoichi Koyama, Naoki Murata, and Hiroshi Saruwatari

The University of Tokyo, Graduate School of Information Science and Technology, Bunkyo-ku, Tokyo, Japan
email: koyama.shoichi@ieee.org

A sound field decomposition method using a multipole dictionary for super-resolution in recording and reproduction is proposed. Conventional methods are based on plane-wave decomposition; however, this procedure leads to severe spatial aliasing artifacts at frequencies above the spatial Nyquist frequency, determined by the intervals between microphones. We have proposed a sparse sound field decomposition method based on a model consisting of near-field monopole source and far-field plane-wave components. Although this method enables super-resolution in recording and reproduction, the use of only a monopole source dictionary is not appropriate in several cases because the region of the near-field source components is discretized as grid points. We propose the use of a multipole dictionary to accurately represent acoustically compact sources with complex directivity. We prove that this multipole dictionary is also effective for representing source components off the grid. Numerical simulations are performed to evaluate the proposed method.

Keywords: Sound field reproduction, sound field decomposition, sparse representation, group sparsity, super-resolution

1. Introduction

Sound field decomposition is a fundamental problem in sound field analysis, reconstruction, and visualization. The objective of sound field decomposition is to represent a sound field as a linear combination of fundamental solutions of the Helmholtz equation from pressure measurements. In sound field recording and reproduction, this makes it possible to calculate the driving signals of loudspeakers required for reproduction from the signals received by microphones, which is referred to as *sound-pressure-to-driving-signal (SP-DS) conversion*. Plane-wave decomposition, which corresponds to spatial Fourier analysis of a sound field [1], has been commonly used because of its computational efficiency. In recent years, the sparse representation of a sound field has been proved to be effective in several applications, such as acoustic holography [2], source localization [3], and sound field recording and reproduction [4], owing to the recent development of sparse decomposition algorithms in the context of compressed sensing [5]. We here consider a sparse sound field decomposition problem for recording and reproduction.

Sound field recording and reproduction is targeted at high-fidelity audio systems. Wave field synthesis (WFS) [6] is a sound field synthesis method based on the Kirchhoff–Helmholtz or Rayleigh integrals, which is generally applied to a planar or linear array of loudspeakers. Since it is necessary to obtain the distribution of the sound pressure gradient of the desired sound field on the secondary sources, WFS cannot be directly applied to SP-DS conversion. However, the *wave field reconstruction (WFR) filtering* method enables efficient SP-DS conversion [7], which has been extended to several

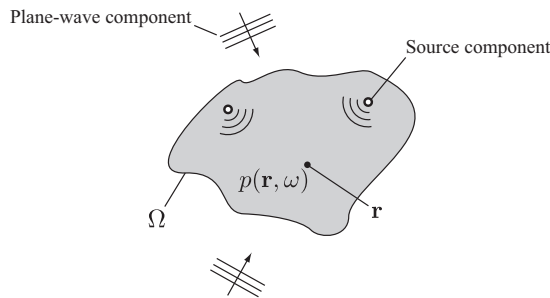


Figure 1: Generative model of sound field consisting of near-field source and far-field plane-wave components.

array geometries [8, 9]. Higher-order ambisonics (HOA) is based on a sound field representation in a spherical or circular harmonic domain [10, 11]. In HOA, a sound field is generally analyzed and synthesized using spherical or circular arrays of microphones and loudspeakers, and it can be applied to SP-DS conversion through encoding and decoding processes [10]. These methods can be classified as analytical approaches to recording and reproduction.

Another type of sound field reproduction method involves a numerical approach, which is typically based on controlling sound pressures at discrete positions in a target area [12]. We define this method as the *sound pressure control (SPC)* method. The inverse of a known transfer function matrix between loudspeakers and control points is numerically calculated. Since only the desired sound pressures at the control points are required to obtain the driving signals of the loudspeakers, these methods can also be applied to SP-DS conversion.

As described above, the analytical approach to recording and reproduction is generally based on spatial Fourier analysis of the captured sound field. Although this procedure makes stable and efficient signal conversion possible, artifacts originating from spatial aliasing notably occur, depending on the interelement spacing between the microphones and loudspeakers. These artifacts cannot be avoided even when a numerical approach is applied. Owing to the significant effect of the spatial aliasing artifacts, listeners will not clearly localize reproduced sound images. Furthermore, the frequency characteristics of the reproduced direct sound are greatly affected.

We previously proposed a sound field decomposition method based on a model consisting of near-field monopole source and far-field plane-wave components [4, 13]. Under the assumption that the source distribution inside the predefined near-field region of the microphones is spatially sparse, the captured sound field is decomposed into these two components using the sparse decomposition algorithm. This method makes it possible to improve the reproduction accuracy at frequencies above the spatial Nyquist frequency when the number of microphones is smaller than that of loudspeakers, which can be regarded as super-resolution in recording and reproduction. However, the use of only monopole source components as a dictionary matrix will not be sufficient to represent acoustically compact sources with complex directivity because the near-field region must be discretized as grid points. We propose the use of a multipole dictionary for sparse sound field decomposition. We prove that this multipole dictionary is also effective for representing source components off the grid. Numerical simulations are conducted to compare the methods using monopole and multipole dictionaries.

2. Generative model of sound field and its sparse decomposition

2.1 Sound field consisting of source and plane-wave components

We briefly revisit the generative model of the sound field proposed in Ref. [4]. As shown in Fig. 1, we divide a sound field in the recording area into internal and external regions, i.e., near-field and far-

field regions, of a closed surface. The internal region is denoted as Ω . Sound pressures are measured inside Ω using multiple microphones. It is assumed that source components only exist inside Ω . When the sound pressure of temporal frequency ω at position \mathbf{r} is denoted as $p(\mathbf{r}, \omega)$, the following equation should be satisfied:

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

where $Q(\mathbf{r}, \omega)$ is the distribution of the source components inside Ω and $k = \omega/c$ is the wave number obtained by setting the sound speed as c . Hereafter, ω is omitted for notational simplicity. Equation (1) indicates that $p(\mathbf{r})$ satisfies the inhomogeneous and homogeneous Helmholtz equations at $\mathbf{r} \in \Omega$ and $\mathbf{r} \notin \Omega$, respectively. The solution of Eq. (1) can be represented as the sum of the inhomogeneous and homogeneous terms, $p_i(\mathbf{r})$ and $p_h(\mathbf{r})$, respectively. Here, $p_i(\mathbf{r})$ is represented as the convolution of $Q(\mathbf{r})$ and the three-dimensional free-field Green's function $G(\mathbf{r}|\mathbf{r}')$ as [1]

$$\begin{aligned} p(\mathbf{r}) &= p_i(\mathbf{r}) + p_h(\mathbf{r}) \\ &= \int_{\mathbf{r}' \in \Omega} Q(\mathbf{r}') G(\mathbf{r}|\mathbf{r}') d\mathbf{r}' + p_h(\mathbf{r}), \end{aligned} \quad (2)$$

where

$$G(\mathbf{r}|\mathbf{r}') = \frac{e^{jk\|\mathbf{r}-\mathbf{r}'\|_2}}{4\pi\|\mathbf{r}-\mathbf{r}'\|_2}. \quad (3)$$

Here, $G(\mathbf{r}|\mathbf{r}')$ corresponds to the transfer function between the monopole source at \mathbf{r}' and the position \mathbf{r} . Since it is assumed that sound sources do not exist outside Ω , the homogeneous term $p_h(\mathbf{r})$ can be represented as the sum of plane waves [1]. Therefore, we define these two terms $p_i(\mathbf{r})$ and $p_h(\mathbf{r})$ as source and plane-wave components, respectively.

If the captured sound field can be represented as Eq. (2), the driving signals of the secondary sources used to reproduce the sound field can be uniquely obtained [4, 13]. The source component $p_i(\mathbf{r})$ is reproduced with model-based sound field synthesis methods [14, 15]. The WFR filtering method can be applied to reproduce the plane-wave component $p_h(\mathbf{r})$ [7, 8, 9]. As an example, we here consider planar distributions of receivers and secondary sources. The sound pressure distribution is obtained on the x - z plane at $y = 0$. The plane-wave component $p_h(\mathbf{r})$ can be described using the spatial frequency spectrum on the receiving plane, $P_h(k_x, k_y)$, as

$$p_h(\mathbf{r}) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_z P_h(k_x, k_z) e^{j(k_x x + k_y y + k_z z)}, \quad (4)$$

where k_x and k_z respectively denote the spatial frequencies with respect to x and z , and $k_y = \pm\sqrt{k^2 - k_x^2 - k_z^2}$. All the source and plane-wave components are assumed to be in the region of $y < 0$. When secondary sources are also arranged on a plane and are assumed to be monopole sources, the driving signals of the secondary sources can be obtained as [4]

$$\begin{aligned} d(\mathbf{r}) &= 2 \frac{\partial}{\partial y} p(\mathbf{r}) \Big|_{y=0} \\ &= 2 \int_{\mathbf{r}' \in \Omega} Q(\mathbf{r}') \frac{\partial G(\mathbf{r}|\mathbf{r}')}{\partial y} \Big|_{y=0} d\mathbf{r}' + \frac{1}{2\pi^2} \int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_z j k_y P_h(k_x, k_z) e^{j(k_x x + k_z z)}. \end{aligned} \quad (5)$$

This equation can be regarded as a combination of the synthesis of the source components $Q(\mathbf{r})$ by WFS [14] and the WFR filtering of $p_h(\mathbf{r})$ [7]. If the decomposition of the captured sound field into source and plane-wave components is achieved, the dominant component will lie in $p_i(\mathbf{r})$; therefore, $d(\mathbf{r})$ obtained by Eq. (5) will become more accurate than that obtained on the basis of plane-wave decomposition, particularly at high frequencies [4].

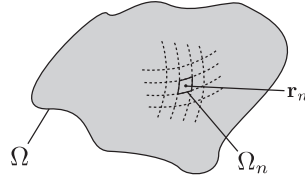


Figure 2: Discretization of Ω to treat sound field decomposition as sparse representation problem.

2.2 Sparse decomposition based on discrete model

A major problem is how to decompose the observed sound pressures into the source and plane-wave components as in Eq. (2). We apply sparse decomposition algorithms under the assumption that the distribution of the source components inside Ω is spatially sparse.

To address the sound field decomposition problem (Eq. (2)) as a sparse representation problem, the region Ω is discretized as a set of grid points. Omnidirectional microphones are discretely arranged inside Ω to capture the sound pressures. The numbers of microphones and grid points are denoted as M and N , respectively. We assume $N \gg M$ because the grid points should entirely and densely cover the region Ω . The discrete form of Eq. (2) can be represented as

$$\mathbf{y} = \mathbf{D}\mathbf{x} + \mathbf{z}, \quad (6)$$

where $\mathbf{y} \in \mathbb{C}^M$ and $\mathbf{x} \in \mathbb{C}^N$ respectively denote the signals received by the microphones and the distribution of the source component at the grid points, $\mathbf{z} \in \mathbb{C}^M$ is the plane-wave component, and $\mathbf{D} \in \mathbb{C}^{M \times N}$ is the dictionary matrix of the source components whose elements consist of Green's functions (Eq. (3)) between the grid points and the microphones. We assumed that \mathbf{x} is the dominant component of \mathbf{y} and \mathbf{z} is the residual. Since it can be assumed that only a few source components exist in Ω , a small number of elements in \mathbf{x} have nonzero values. Therefore, the decomposition of \mathbf{y} into \mathbf{x} and \mathbf{z} can be achieved by solving the following optimization problem:

$$\underset{\mathbf{x}}{\text{minimize}} \frac{1}{2} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_p^p, \quad (7)$$

where $\|\mathbf{x}\|_p$ ($0 < p \leq 1$) is the ℓ_p -norm of \mathbf{x} , which corresponds to the penalty term used to induce sparsity of the solution \mathbf{x} , and λ is a parameter that balances the approximation error and the sparsity-inducing penalty. Several algorithms for solving (7) have been proposed [16]. Although Eq. (6) represents the signal model of a single frequency bin and single time frame, it is possible to exploit several group-sparse models arising from the physical properties of the sound field [13].

3. Multipole dictionary to reduce discretization error

3.1 Theoretical interpretation of the use of multipole dictionary

By discretizing Ω into grid points, the dictionary of monopole sources will not be sufficient to represent acoustically compact sources with complex directivity, particularly when the true source location is not exactly on the grid. We show that the use of a discrete monopole dictionary can be regarded as the approximation of the Green's function with a zeroth-order Taylor expansion. Therefore, the approximation error of the discretization of Ω can be reduced by using higher orders. We propose a sparse sound field decomposition method using a dictionary matrix consisting of multipole sources, which corresponds to the use of higher orders of the Taylor expansion.

We separate the region Ω into a set of small regions Ω_n ($n \in \{1, \dots, N\}$) as in Fig. 2. The location of a representative point of Ω_n is denoted as \mathbf{r}_n . Then, the source component $p_i(\mathbf{r})$ is given

by

$$p_i(\mathbf{r}) = \int_{\mathbf{r}' \in \Omega} Q(\mathbf{r}') G(\mathbf{r}|\mathbf{r}') d\mathbf{r}' = \sum_{n=1}^N \int_{\mathbf{r}_n \in \Omega_n} Q(\mathbf{r}') G(\mathbf{r}|\mathbf{r}') d\mathbf{r}'. \quad (8)$$

We here consider the approximation of $G(\mathbf{r}|\mathbf{r}')$ by a Taylor expansion with respect to \mathbf{r}_n as follows:

$$G(\mathbf{r}|\mathbf{r}') = \sum_{l_x=0}^{\infty} \sum_{l_y=0}^{\infty} \sum_{l_z=0}^{\infty} \frac{(x' - x_n)^{l_x} (y' - y_n)^{l_y} (z' - z_n)^{l_z}}{l_x! l_y! l_z!} \left(\frac{\partial}{\partial x} \right)^{l_x} \left(\frac{\partial}{\partial y} \right)^{l_y} \left(\frac{\partial}{\partial z} \right)^{l_z} G(\mathbf{r}|\mathbf{r}') \Big|_{\mathbf{r}'=\mathbf{r}_n}, \quad (9)$$

where l_x , l_y , and l_z are referred to as the expansion order. This equation corresponds to the multipole expansion of the Green's function. By truncating the expansion order of $G(\mathbf{r}|\mathbf{r}')$ up to L_x , L_y , and L_z , Eq. (8) can be approximated as

$$p_i(\mathbf{r}) \simeq \sum_{n=1}^N \sum_{l_x=0}^{L_x} \sum_{l_y=0}^{L_y} \sum_{l_z=0}^{L_z} \left(\frac{\partial}{\partial x} \right)^{l_x} \left(\frac{\partial}{\partial y} \right)^{l_y} \left(\frac{\partial}{\partial z} \right)^{l_z} G(\mathbf{r}|\mathbf{r}') \Big|_{\mathbf{r}'=\mathbf{r}_n} \cdot \int_{\mathbf{r}' \in \Omega_n} Q(\mathbf{r}') \frac{(x' - x_n)^{l_x} (y' - y_n)^{l_y} (z' - z_n)^{l_z}}{l_x! l_y! l_z!} d\mathbf{r}'. \quad (10)$$

The expansion coefficients for Ω_n including the sources will only have non-zero values. Since $Q(\mathbf{r})$ is assumed to be sparse, this leads to a group-sparse structure in the discrete model as described in the next section. When L_x , L_y , and L_z are set as 0, $p_i(\mathbf{r})$ becomes a linear combination of monopole sources, as in Eq. (6), by using the grid points at \mathbf{r}_n .

3.2 Sparse decomposition with multipole dictionary

We describe a discrete model based on Eq. (10) including group sparsity with respect to multiple time frames and frequencies as in Ref. [13]. Using the subscripts indicating the indexes of time frame $t \in \{1, \dots, T\}$ and frequency $f \in \{1, \dots, F\}$, the following equation can be obtained:

$$\mathbf{y}_{t,f} = \sum_{l=1}^L \mathbf{D}_{f,l} \mathbf{x}_{t,f,l} + \mathbf{z}_{t,f}, \quad (11)$$

where $l \in \{1, \dots, L\}$ is the index of the multipole components combining l_x , l_y , and l_z . The dictionary matrix $\mathbf{D}_{f,l}$ consists of the l th pole of the Green's function between \mathbf{r}_m and \mathbf{r}_n of the f th frequency as

$$D_{m,n,f,l} = \left(\frac{\partial}{\partial x} \right)^{l_x} \left(\frac{\partial}{\partial y} \right)^{l_y} \left(\frac{\partial}{\partial z} \right)^{l_z} G_f(\mathbf{r}_m|\mathbf{r}') \Big|_{\mathbf{r}'=\mathbf{r}_n}, \quad (12)$$

where $G_f(\cdot)$ is the Green's function (Eq. (3)) of the f th frequency.

By combining the signals and redefining the variables as $\mathbf{y} \in \mathbb{C}^{MTF}$, $\mathbf{D} \in \mathbb{C}^{MTF \times NTFL}$, $\mathbf{x} \in \mathbb{C}^{NTFL}$, and $\mathbf{z} \in \mathbb{C}^{MTF}$, they can be still related by a linear equation as in Eq. (6). Each $\mathbf{x}_{t,f,l}$ should have the same sparse structure. Decomposition by exploiting this group-sparse structure can be achieved by solving the following optimization problem:

$$\underset{\mathbf{x}}{\text{minimize}} \frac{1}{2} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda J_{p,2}(\mathbf{x}). \quad (13)$$

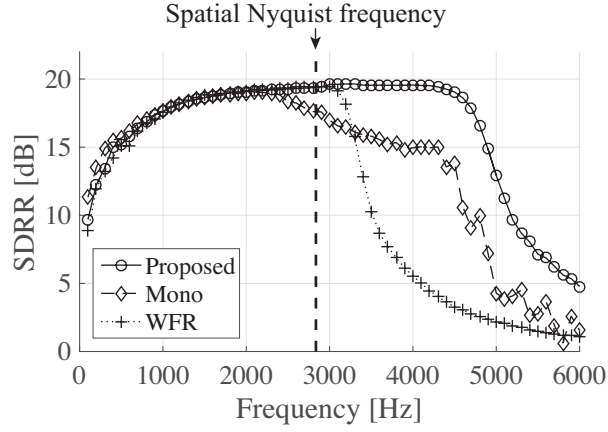


Figure 3: Relationship between frequency and SDRR.

Here, $J_{p,2}(\mathbf{x})$ is a groupwise diversity measure of \mathbf{x} defined as

$$J_{p,2}(\mathbf{x}) = \sum_{n=1}^N \|\mathbf{x}_n\|_2^p = \sum_{n=1}^N \left(\sum_{t=1}^T \sum_{f=1}^F \sum_{l=1}^L |x_{n,t,f,l}|^2 \right)^{p/2}, \quad (14)$$

where $\mathbf{x} \in \mathbb{C}^{TFL}$ is the n th vector component of each group and $x_{n,t,f,l}$ is an element of \mathbf{x}_n .

The optimization problem in Eq. (13) can be solved by using the iteratively reweighted least-squares algorithm proposed in [13], which can be regarded as an extension of the FOCUSS algorithm [16, 17]. The source and plane-wave components are separately converted into the driving signals of the loudspeakers as in Eq. (5). The sum of these signals is the final output of the loudspeakers used for reproduction.

4. Experiments

Numerical simulations were performed using linear arrays of microphones and loudspeakers. We compared the proposed method, the method only using the monopole dictionary, and the WFR filtering method, which are denoted as “**Proposed**”, “**Mono**”, and “**WFR**”, respectively.

A linear microphone and loudspeaker arrays were set along the x -axis with the center at the origin in the recording and target areas, respectively. The number of microphones M was 32 and the intervals between the microphones were 0.06 m. The number of loudspeakers was 48 and their intervals were 0.04 m. Therefore, the spatial Nyquist frequency of the microphone array was about 2.8 kHz and that of the loudspeaker array was about 4.3 kHz. The array lengths of both the microphones and loudspeakers were 1.92 m. The directivity of the microphones and loudspeakers was assumed to be omnidirectional.

In **Proposed** and **Mono**, the two-dimensional region Ω was set to be a square of $2.4 \times 2.4 \text{ m}^2$ centered at $(0.0, -2.0, 0.0) \text{ m}$ on the x - y plane at $z = 0$. The number of grid points was 25×13 and they were set at intervals of 0.1 m in the x direction and 0.2 m in the y direction. The number of time frames T was 8. The number of positive-frequency bins F was set as 128 and the sampling frequency was 16 kHz. For the multipole components, a monopole and a dipole for x and y were used for the dictionary \mathbf{D} ; therefore, L was set to 3. In **Mono**, only the dictionary of the monopole was used.

The sound pressure distributions in the target area were simulated in a $2.1 \times 2.4 \text{ m}^2$ region on the x - y plane at $z = 0$ with the center at $(0.0, 1.0, 0.0) \text{ m}$. The intervals between the simulated positions were 0.015 m. The amplitudes were normalized using the averaged squared amplitude in the region $y \geq 0.5 \text{ m}$. The general reproduction accuracy was evaluated using the signal-to-distortion ratio for

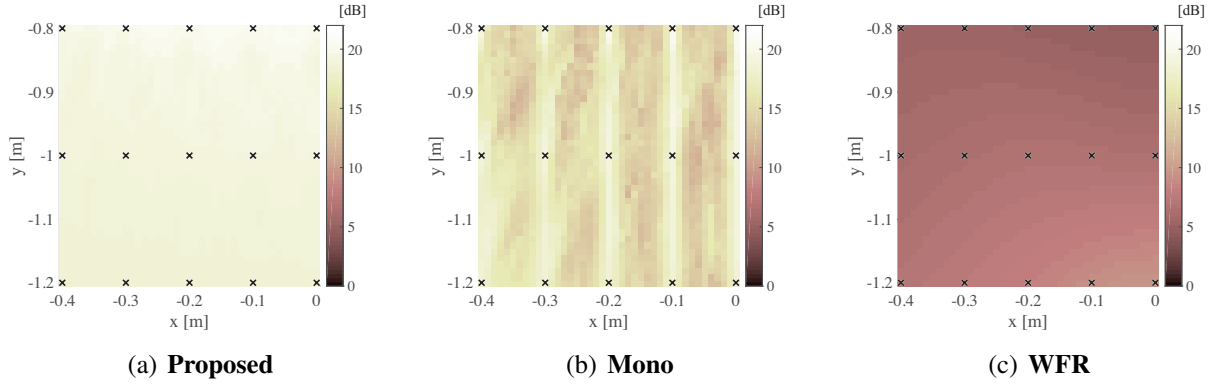


Figure 4: Relationship between source location and SDRR. Cross marks indicate grid points.

reproduction (SDRR) defined as

$$\text{SDRR} = 10 \log_{10} \frac{\sum_i \sum_j \sum_k |\bar{p}_{\text{org}}(x_i, y_j, t_k)|^2}{\sum_i \sum_j \sum_k |\bar{p}_{\text{rep}}(x_i, y_j, t_k) - \bar{p}_{\text{org}}(x_i, y_j, t_k)|^2}, \quad (15)$$

where $\bar{p}_{\text{org}}(\cdot)$ and $\bar{p}_{\text{rep}}(\cdot)$ are the original and reproduced sound pressure distributions in the time domain, respectively, (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 to 160, i.e., the duration was 10 ms.

Figure 3 shows the relationship between the frequency of the source signal and SDRR when a point source was located at $(-0.32, -0.84, 0.0)$ m. Note that the primary source was not exactly on a grid point. To choose λ for **Proposed** and **Mono**, we calculated SDRR from 0.001 to 0.01 at intervals of 0.001 for λ at 4.0 kHz. Then, λ that gives the highest SDRR was chosen for each method. The SDRRs of the three methods remained high up to the spatial Nyquist frequency of the microphone array. In **WFR**, the SDRR sharply decreased above this frequency. The SDRR of **Mono** deteriorated but was maintained to some extent above the spatial Nyquist frequency. On the other hand, the SDRR of **Proposed** remained high up to 4.3 kHz, which is approximately the spatial Nyquist frequency of the loudspeaker array.

To demonstrate the effect of the model of multipole components for the off-grid case, the SDRR when a single primary source was inside a square region of $0.4 \times 0.4 \text{ m}^2$ with the center at $(-0.2, -1.0, 0.0)$ m is shown in Fig. 4. The SDRRs were calculated at intervals of 0.01 m inside the region by the three methods and the frequency of the source signal was 4.0 kHz. The cross marks indicate the locations of the grid points. The SDRRs of **WFR** were low regardless of the location of the primary source. Although the SDRRs of **Mono** were high when the primary source was located around a grid point, they decreased between the grid points. High SDRRs were maintained over the entire region for **Proposed**; therefore, the multipole dictionary is effective for off-grid cases.

5. Conclusion

A sparse sound field decomposition method using a multipole dictionary was proposed. In the method previously proposed by the authors, the near-field region was discretized as grid points and a sparse decomposition algorithm was applied, where the dictionary of the source components consisted of the transfer function of the monopole sources between the grid points and the microphones. However, the monopole dictionary is not sufficient to represent acoustically compact sources with complex directivity. We proposed the use of a multipole dictionary for sparse sound field decomposition. This multipole dictionary is effective for representing source components off the grid, which can be interpreted as the use of higher orders of the Taylor expansion of the Green's function. Numerical experiments were also conducted to show the effectiveness of the multipole dictionary.

REFERENCES

1. Williams, E. G., *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*, Academic Press, New York (1999).
2. Chardon, G., Daudet, L., Peillot, A., Ollivier, F., Bertin, N. and Gribonval, R. Near-field acoustic holography using sparsity and compressive sampling principles, *J. Acoust. Soc. Am.*, **132** (3), 1521–1534, (2012).
3. Malioutov, D., Cetin, M. and Willsky, A. S. A sparse signal reconstruction perspective for source localization with sensor arrays, *IEEE Trans. Signal Process.*, **53** (8), 3010–3022, (2005).
4. Koyama, S., Shimauchi, S. and Ohmuro, H. Sparse sound field representation in recording and reproduction for reducing spatial aliasing artifacts, *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Florence, May, pp. 4443–4447, (2014).
5. Elad, M., *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, New York (2010).
6. Berkhout, A. J., de Vries, D. and Vogel, P. Acoustic control by wave field synthesis, *J. Acoust. Soc. Am.*, **93** (5), 2764–2778, (1993).
7. Koyama, S., Furuya, K., Hiwasaki, Y. and Haneda, Y. Analytical approach to wave field reconstruction filtering in spatio-temporal frequency domain, *IEEE Trans. Audio, Speech, Lang. Process.*, **21** (4), 685–696, (2013).
8. Koyama, S., Furuya, K., Hiwasaki, Y., Haneda, Y. and Suzuki, Y. Wave field reconstruction filtering in cylindrical harmonic domain for with-height recording and reproduction, *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, **22** (10), 1546–1557, (2014).
9. Koyama, S., Furuya, K., Wakayama, K., Shimauchi, S. and Saruwatari, H. Analytical approach to transforming filter design for sound field recording and reproduction using circular arrays with a spherical baffle, *J. Acoust. Soc. Am.*, **139** (3), 1024–1036, (2016).
10. Daniel, J. Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonics format, *Proc. 23rd AES Int. Conf.*, Copenhagen, May, (2003).
11. Poletti, M. Three-dimensional surround sound systems based on spherical harmonics, *J. Audio Eng. Soc.*, **53** (11), 1004–1025, (2005).
12. Gauthier, P.-A. and Berry, A. Sound-field reproduction in-room using optimal control techniques: Simulations in the frequency domain, *J. Acoust. Soc. Am.*, **117** (2), 662–678, (2005).
13. Koyama, S., Murata, N. and Saruwatari, H. Structured sparse signal models and decomposition algorithm for super-resolution in sound field recording and reproduction, *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Brisbane, Apr., pp. 619–623, (2015).
14. Spors, S., Rabenstein, R. and Ahrens, J. The theory of wave field synthesis revisited, *Proc. 124th AES Conv.*, Amsterdam, Oct., (2008).
15. Ahrens, J. and Spors, S. An analytical approach to sound field reproduction using circular and spherical loudspeaker distributions, *Acta Acustica United with Acustica*, **94**, 988–999, (2008).
16. Gorodnitsky, I. F. and Rao, B. D. Sparse signal reconstruction from limited data using FOCUSS: a reweighted minimum norm algorithm, *IEEE Trans. Signal Process.*, **45** (3), 600–616, (1997).
17. Cotter, S. F., Rao, D., Engan, K. and Kreutz-Delgado, K. Sparse solutions to linear inverse problems with multiple measurement vectors, *IEEE Trans. Signal Process.*, **53** (7), 2477–2488, (2005).