SOUND FIELD DECOMPOSITION IN REVERBERANT ENVIRONMENT USING SPARSE AND LOW-RANK SIGNAL MODELS

Shoichi Koyama and Hiroshi Saruwatari

Graduate School of Information Science and Technology, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

ABSTRACT

A sound field decomposition method for a reverberant environment is proposed. Sound field decomposition is the foundation of various acoustic signal processing applications and enables the estimation of the entire sound field from pressure measurements. Although spatial Fourier analysis of the sound field has been widely used, sparse decomposition of the sound field has recently been proved to be effective in several applications. However, in current methods, no constraints are imposed on ambiance components, whereas source components are assumed to be sparsely distributed in the space. This results in inaccurate decomposition in a reverberant environment. The proposed method is based on sparse and low-rank signal models, which are used for simultaneous decomposition of the observed signals into source and ambiance components. Numerical simulation results indicated that the decomposition accuracy is superior to that of current methods.

Index Terms— Sound field decomposition, sound field analysis, super-resolution, sparse representation, convex optimization

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 wave equation (or Helmholtz equation) from pressure measurements. This makes it possible for the entire sound field to be estimated from the signals received by multiple microphones. Plane wave decomposition, which corresponds to spatial Fourier analysis of the sound field [1], has been commonly used because of its computational efficiency. In recent years, sparse decomposition of the sound field has been proved to be effective in several applications, such as acoustic holography, source localization, and sound field recording and reproduction [2–5], owing to the recent development of sparse decomposition algorithms in the context of compressed sensing [6, 7].

Acoustic holography is used to measure a pressure or velocity distribution on a surface close to acoustic sources using a limited number of pressure measurements [8]. In near-field acoustic holography (NAH) [1, 8], the pressure distribution measured by microphones is decomposed into spatial Fourier basis functions, such as plane wave functions, cylindrical harmonics, and spherical harmonics, to reconstruct the pressure or velocity distribution in the inverse direction of sound propagation.

Sound field recording and reproduction is targeted at high-fidelity audio systems. Sound pressures at multiple positions in a recording area are obtained with microphones and are then reproduced with loudspeakers in a target area. This conversion from the signals received by microphones into the driving signals of the loudspeakers implicitly includes an estimation of pressure and velocity distributions; therefore, this problem is closely related to acoustic holography [9–11]. The *wave field reconstruction (WFR) filtering* method, which is based on the spatial Fourier analysis, enables efficient and stable signal conversion for recording and reproduction [11, 12].

A critical issue in sound field decomposition based on spatial Fourier analysis is artifacts originating from spatial aliasing, depending on the interelement spacing in the microphone array. For example, in sound field recording and reproduction, listeners are unable to clearly localize the reproduced sound images. Furthermore, the frequency characteristics of the reproduced sound are adversely affected, which is referred to as the *coloration effect* [13]. We previously proposed a sparse sound field decomposition method to reduce these artifacts [4, 14], in which the sound field is modeled as the sum of monopole sources inside a predefined *source region* and plane waves from outside the region, i.e., source and ambiance components. Since only a few monopole components are assumed to exist inside the source region, it is possible to sparsely decompose the observed signals into basis functions, or dictionaries, consisting of Green's functions. This method makes it possible to improve the reproduction accuracy above the spatial Nyquist frequency, which can be regarded as super-resolution in recording and reproduction.

In current methods of sparse sound field decomposition, no constraints are imposed on ambiance components, whereas source components are assumed to be sparse. For example, in [4, 14], plane wave components are treated as residuals. This assumption is valid when the plane wave components include only spatially uncorrelated signals. However, when there are intense monopole components outside the source region, such as reflections, this assumption does not hold. As a result, the decomposition accuracy can be degraded. Although several methods for sparse sound field decomposition have been proposed in various contexts [2, 3, 5], the above-mentioned problem cannot be avoided since the ambiance components are not explicitly defined or are assumed to be residuals.

We propose a sound field decomposition method based on sparse and low-rank signal models. In addition to the assumption of a sparse distribution of monopole components, we assume that signals derived from plane wave components have a low-rank structure. This assumption means that the signal components outside the source region, which are mainly reflections, can be approximated as the direct product of the source signals and their steering vectors. Therefore, these signal models can lead to more accurate sound field decomposition even in a reverberant environment. We derive a decomposition algorithm based on these signal models by using the alternating direction method of multipliers (ADMM) [15, 16]. Numerical simulations are conducted to evaluate the proposed method for sound field decomposition and reconstruction.

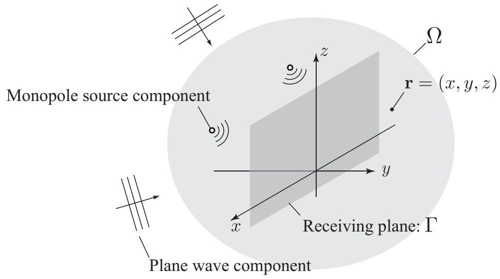


Fig. 1. Generative model of sound field.

2. GENERATIVE MODEL OF SOUND FIELD AND ITS SPARSE DECOMPOSITION

A generative model of a sound field, which was first proposed in [4], is briefly revisited. As shown in Fig. 1, a sound field is divided into two regions, internal and external, of a closed surface. The internal region is denoted as Ω, i.e., the source region. When a sound pressure of temporal frequency ω at position $\mathbf{r}$ is denoted as $p(\mathbf{r}, \omega)$, the following equation should be satisfied:

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

where $Q(\mathbf{r}, \omega)$ is the distribution of the monopole 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. Therefore, the solution of (1) can be represented as the sum of the inhomogeneous and homogeneous terms, $p_{\mathrm{i}}(\mathbf{r})$ and $p_{\mathrm{h}}(\mathbf{r})$, respectively. The inhomogeneous term $p_{\mathrm{i}}(\mathbf{r})$ is represented as a convolution of $Q(\mathbf{r})$ and the three-dimensional free-field Green's function $G(\mathbf{r}|\mathbf{r}')$ as [1]

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

where

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

Here, $G(\mathbf{r}|\mathbf{r}')$ corresponds to the transfer function between the monopole source at $\mathbf{r}'$ and the position $\mathbf{r}$. Equation (2) can be verified by substituting it into (1). Since it is assumed that sound sources do not exist outside Ω, the homogeneous term $p_{\mathrm{h}}(\mathbf{r})$ can be represented as the sum of plane waves. Our objective is to decompose $p(\mathbf{r})$ into $p_{\mathrm{i}}(\mathbf{r})$ and $p_{\mathrm{h}}(\mathbf{r})$ from sound pressure measurements inside Ω. Here, we assume that the sound pressure distribution on the receiving plane Γ is obtained as shown in Fig. 1.

To address the sound field decomposition problem described above as a sparse representation problem, the region Ω is discretized as a set of grid points. Omnidirectional microphones are discretely aligned on Γ to capture the sound pressure distribution. 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 (2) can be represented as

$$\mathbf{y} = \mathbf{D}\mathbf{x} + \mathbf{z}, \tag{4}$$

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 monopole components at the grid points, $\mathbf{z} \in \mathbb{C}^M$ is the homogeneous term, and $\mathbf{D} \in \mathbb{C}^{M \times N}$ is the dictionary matrix of the monopole components, whose elements consists of Green's functions between the grid points and the microphones. Since it can be assumed that only a few monopole components exist in Ω, a small number of elements of $\mathbf{x}$ have nonzero values. Therefore, when it can be assumed that $\mathbf{x}$ is a dominant component of $\mathbf{y}$ and $\mathbf{z}$ is a residual, $\mathbf{y}$ can be decomposed into $\mathbf{x}$ and $\mathbf{z}$ using sparse decomposition algorithms [7].

Although (4) represents the signal model of a single frequency bin and single time frame, it is possible to exploit several group sparse signal models arising from the physical properties of the sound field [14]. For simplicity, we here consider only the group-sparsity derived from the measurements of multiple time frames, which is referred to as the multiple-measurement-vector (MMV) problem [17]. When multiple time frames of $\mathbf{y}$ are available and the monopole components are assumed to be static, each $\mathbf{x}$ has the same sparsity pattern. We denote the index of the time frame as $l \in \{1, \cdots, L\}$ and the signals of each l as $\mathbf{y}_l$, $\mathbf{x}_l$, and $\mathbf{z}_l$. Matrices $\mathbf{Y} \in \mathbb{C}^{M \times L}$, $\mathbf{X} \in \mathbb{C}^{N \times L}$, and $\mathbf{Z} \in \mathbb{C}^{M \times L}$ are defined with each column consisting of $\mathbf{y}_l$, $\mathbf{x}_l$, and $\mathbf{z}_l$, respectively. Therefore, (4) can be rewritten as

$$\mathbf{Y} = \mathbf{D}\mathbf{X} + \mathbf{Z}. \tag{5}$$

When the monopole components are static at each l, the row of $\mathbf{X}$ becomes sparse. Therefore, this sparse decomposition with respect to the row can be achieved by solving the following optimization problem:

$$\underset{\mathbf{X}}{\text{minimize}} \, \|\mathbf{X}\|_{p/q} \text{ subject to } \mathbf{Y} = \mathbf{D}\mathbf{X}, \tag{6}$$

where $\|\cdot\|_{p/q}$ is the $\ell_{p/q}$-norm defined as

$$\|\mathbf{X}\|_{p/q} = \sum_{n=1}^{N} \|X_{n,\cdot}\|_q^p. \tag{7}$$

Here, $X_{n,\cdot}$ represents the nth row of $\mathbf{X}$ and $0 \leq p \leq 1$ and $q \geq 1$ are parameters for inducing the row sparsity of $\mathbf{X}$. Several algorithms for solving (6) have been proposed [18]. Then, $\mathbf{Z}$ can be simply obtained as $\mathbf{Y} - \mathbf{D}\mathbf{X}$.

3. SOUND FIELD DECOMPOSITION USING SPARSE AND LOW-RANK SIGNAL MODELS

In the optimization criterion (6), the homogeneous term $\mathbf{Z}$ is not constrained and is treated as a residual by assuming that it has a complex Gaussian distribution. This assumption is valid when $\mathbf{Z}$ includes only spatially uncorrelated signals; however, it does not hold in a reverberant environment. In [14], by setting a large region Ω, the group sparsity with respect to direct and image source locations is introduced to overcome this problem. However, it is difficult to apply this method when the accurate shape and size of the room are unknown. Therefore, it is necessary to develop a method for sound field decomposition in a reverberant environment without prior information about its size and shape.

3.1. Signal model of sparse and low-rank components

In addition to the assumption of the row sparsity of $\mathbf{X}$, we assume that $\mathbf{Z}$ has a low-rank structure. More specifically, the signal components outside Ω, which are mainly reflections, can be approximated as the direct product of the source signals and their steering vectors. When the source signals are mutually uncorrelated, the rank of the spatial covariance matrix of $\mathbf{Z}$, i.e., $\mathbf{Z}\mathbf{Z}^H$, can be limited by the number of sound sources. By introducing this model, it is possible to accurately decompose $\mathbf{X}$ so that the reflective components are included in $\mathbf{Z}$.

The signal decomposition discussed above can be formulated as follows:

$$\underset{\mathbf{X},\mathbf{Z}}{\text{minimize}}\ \lambda\|\mathbf{X}\|_{1/2} + \|\mathbf{Z}\|_* \ \text{ subject to } \ \mathbf{Y} = \mathbf{D}\mathbf{X} + \mathbf{Z}, \tag{8}$$

where $\|\cdot\|_*$ represents the nuclear norm, which is the tightest convex lower bound of the rank function [19]. Additionally, we set $p = 1$ and $q = 2$ so that $\|\mathbf{X}\|_{p/q}$ becomes a convex function. The parameter λ determines the balance between them. Therefore, (8) can be solved as a convex optimization problem.

A problem related to (8) is robust principal component analysis [20]. In this problem, $\mathbf{X}$ itself is assumed to be sparse and the constraint condition is $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$. This model is used to decompose $\mathbf{Y}$ into pulsive and low-rank components.

3.2. Decomposition algorithm based on ADMM

We derive an algorithm based on ADMM for solving (8) [15, 16]. First, we define the augmented Lagrangian function L_ρ as

$$\begin{aligned} L_\rho(\mathbf{X},\mathbf{Z},\mathbf{W}) &= \lambda\|\mathbf{X}\|_{1/2} + \|\mathbf{Z}\|_* \\ &+\langle \mathbf{W}, \mathbf{D}\mathbf{X}+\mathbf{Z}-\mathbf{Y}\rangle + \frac{1}{2\rho}\|\mathbf{D}\mathbf{X}+\mathbf{Z}-\mathbf{Y}\|_F^2, \end{aligned} \tag{9}$$

where $\langle\cdot,\cdot\rangle$ represents the inner product, $\mathbf{W}$ is the Lagrangian multiplier, and $\rho > 0$ is a constant parameter. In ADMM, each variable is alternately updated, starting with arbitrary initial values, $\mathbf{X}^1$, $\mathbf{Z}^1$, and $\mathbf{W}^1$, as

$$\begin{cases} \mathbf{X}^{k+1} = \arg\min_{\mathbf{X}} L_\rho(\mathbf{X},\mathbf{Z}^k,\mathbf{W}^k) \\ \mathbf{Z}^{k+1} = \arg\min_{\mathbf{Z}} L_\rho(\mathbf{X}^{k+1},\mathbf{Z},\mathbf{W}^k) \\ \mathbf{W}^{k+1} = \mathbf{W}^k + \left(\mathbf{D}\mathbf{X}^{k+1}+\mathbf{Z}^{k+1}-\mathbf{Y}\right)/\rho \end{cases}, \tag{10}$$

where k is the index of the iteration. The Lagrangian function is minimized for one variable while fixing the other variables. At each update, $\mathbf{X}$ and $\mathbf{Z}$ can be efficiently updated by evaluating proximal operators [21,22].

The $\mathbf{X}$-update can be formulated as

$$\begin{aligned} \mathbf{X}^{k+1} &= \arg\min_{\mathbf{X}} \lambda\|\mathbf{X}\|_{1/2} + \frac{1}{2\rho}\|\mathbf{P}^k - \mathbf{D}\mathbf{X}\|_F^2 \\ &= \mathcal{T}_{\lambda\eta}\left(\mathbf{X} - \frac{\eta}{\rho}\mathbf{D}^H(\mathbf{D}\mathbf{X}-\mathbf{P}^k)\right), \end{aligned} \tag{11}$$

where $\mathbf{P}^k = \mathbf{Y} - \mathbf{Z}^k - \rho\mathbf{W}^k$, η is a stepsize parameter, the (n,l)th element of the operator $\mathcal{T}_{\lambda\rho}(\cdot)$ is defined as

$$\mathcal{T}_\alpha(\mathbf{A})_{nl} = \max\{\|A_{n,\cdot}\|_2 - \alpha, 0\}\frac{A_{nl}}{\|A_{n,\cdot}\|_2}, \tag{12}$$

and A_{nl} represents the (n,l)th element of $\mathbf{A}$. In a similar manner, the $\mathbf{Z}$-update can be formulated as

$$\begin{aligned} \mathbf{Z}^{k+1} &= \arg\min_{\mathbf{Z}} \|\mathbf{Z}\|_* + \frac{1}{2\rho}\|\mathbf{Q}^k - \mathbf{Z}\|_F^2 \\ &= \mathbf{U}^k \mathrm{diag}\left(\max\{\sigma_i^k - \rho, 0\}\right)\left(\mathbf{V}^k\right)^H, \end{aligned} \tag{13}$$

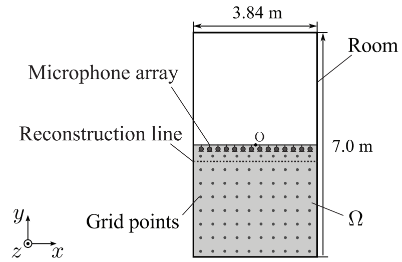


Fig. 2. Simulation setup.

where $\mathbf{Q}^k = \mathbf{Y} - \mathbf{D}\mathbf{X}^{k+1} - \rho\mathbf{W}$, $\mathrm{diag}(\cdot)$ represents the diagonal matrix with the arguments as elements, and $\mathbf{U}^k$, $\mathbf{V}^k$, and σ_i^k are derived by the singular value decomposition of $\mathbf{Q}^k$ as

$$\mathbf{Q}^k = \mathbf{U}^k \mathrm{diag}\left(\sigma_1^k, \cdots, \sigma_r^k\right)\left(\mathbf{V}^k\right)^H. \tag{14}$$

Here, r is the rank of $\mathbf{Q}^k$. Since (11) and (13) respectively amount to soft thresholding for the matrix element and the singular values, each iteration can be efficiently computed.

Although the group sparsity of multiple time frames is only assumed in (8), it is straightforward to use other group-sparse signal models, such as multiple frequency bins [14].

4. EXPERIMENTS

Numerical simulations were conducted to evaluate the proposed method in a two-dimensional sound field. First, the sparse decomposition performances of the proposed method and a current method are compared. Second, we demonstrate super-resolution in sound field reconstruction using the proposed method.

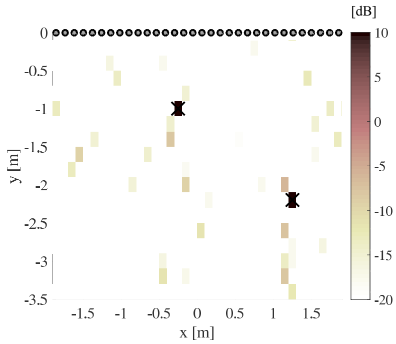
4.1. Evaluation of sound field decomposition

We compared the proposed method (**Proposed**) with iterative shrinkage-thresholding for the MMV problem (**M-IST**) [23]. **M-IST** is an algorithm for solving (6), in which $\mathbf{X}$ is iteratively updated by soft-thresholding as in (11). The sound sources were in a rectangular room of size 3.84×7.0 m^2 as shown in Fig. 2. The origin of the coordinate system was set at the center of the room. A linear microphone array was set along the x-axis with its center at the origin. The number of microphones was 32 and they were set at intervals of 0.12 m; therefore, the spatial Nyquist frequency was 1.4 kHz. The directivity of the microphones was assumed to be omnidirectional. The room reverberation was simulated by the image method [24]. The reflection coefficients were set as 0.84, which corresponds to a reverberation time (T_{60}) of about 500 ms.

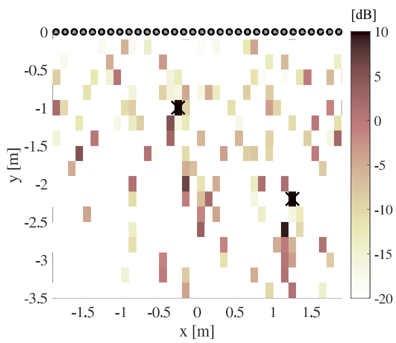
The size of the two-dimensional region Ω was set to be half of the room size, i.e., 3.84×3.5 m^2 on the x-y plane. The number of grid points was 38×17. Their intervals were 0.1 m in the x direction and 0.2 m in the y direction. The center of the grid points was at (0.0, –1.8) m.

Two point sources were located at (–0.25, –1.0) and (1.25, –2.2) m. The source signal was a single-frequency sinusoidal wave. The amplitudes of the source signals were independently generated by a complex Gaussian distribution with a mean of 0.0 and a variance of 10.0. Sensor noise was added so that the signal-to-noise ratio was 40 dB.

In **Proposed**, λ, ρ, and η in (11) and (13) were set as 0.06, 0.4, and 0.1, respectively. **M-IST** also requires a parameter corresponding to ρ in (11) to be set, which was set as 3.7×10^{-2}. The number of



(a) Proposed



(b) M-IST

Fig. 3. Results of sound field decomposition. Crosses indicate true source locations.

time frames of the observed signal was 100. The maximum number of iterations in both methods was 5000.

To evaluate the performance of sparse decomposition, we defined an F-measure (F_{msr}) and a signal-to-distortion ratio (SDR). An operator $\mathrm{supp}(\cdot)$ extracted a set of row indexes such that the squared amplitude of each row of the solution matrix $\mathbf{X}$ was larger than a threshold value μ,

$$\mathrm{supp}(\mathbf{X}) = \left\{ n \in \{1, \cdots, N\} : \|X_{n,\cdot}\|_2^2 > \mu \right\}, \qquad (15)$$

where μ was set as 0.5. F_{msr} is defined as

$$F_{\mathrm{msr}} = 2\frac{|\mathrm{supp}(\mathbf{X}_{\mathrm{est}}) \cap \mathrm{supp}(\mathbf{X}_{\mathrm{true}})|}{|\mathrm{supp}(\mathbf{X}_{\mathrm{est}})| + |\mathrm{supp}(\mathbf{X}_{\mathrm{true}})|}, \qquad (16)$$

where $\mathbf{X}_{\mathrm{est}}$ and $\mathbf{X}_{\mathrm{true}}$ are the estimated and true solution matrices, respectively. Therefore, F_{msr} is equal to 1 when the sets of activated indexes of these matrices are exactly the same. SDR is defined as

$$\mathrm{SDR} = 10 \log_{10} \frac{\|\mathbf{X}_{\mathrm{true}}\|_F^2}{\|\mathbf{X}_{\mathrm{true}} - \mathbf{X}_{\mathrm{est}}\|_F^2}. \qquad (17)$$

Figure 3 shows the amplitude distribution of $\mathbf{X}$ when the frequency of the source signal was 2.0 kHz. Note that this frequency is above the spatial Nyquist frequency. The two crosses indicate the true source locations. In **Proposed**, the observed signal was accurately decomposed, although several grid points other than the true source locations had small amplitudes. In contrast, in **M-IST**, the amplitudes of the decomposed signals were dispersed over the grid points. This is because the signal components outside Ω cannot be treated as residuals. The values of F_{msr} for **Proposed** and **M-IST** were 1.0 and 0.09, and the SDR values were 6.8 and 5.5 dB, respectively.

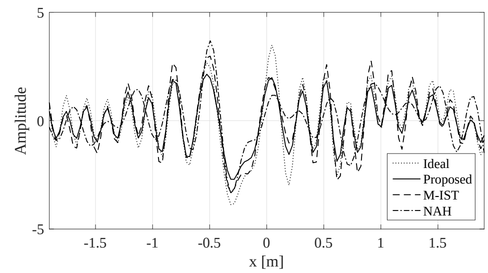


Fig. 4. Reconstructed sound pressure distribution on the line $y = -0.5$ m. $\mathrm{SDR}_{\mathrm{rec}}$ for **Proposed**, **M-IST**, and **NAH** was 5.27, 0.66, and 0.59 dB, respectively.

4.2. Evaluation of sound field reconstruction

We applied the proposed method to an acoustic holography problem. The goal was to reconstruct the sound pressure distribution of direct sounds on a reconstruction line by using the observed signals. **NAH** [1] was also used for comparison in addition to **M-IST**.

The simulation setup is the same as that in Section 4.1. Two point sources were located at (–0.25, –1.0) and (1.25, –2.2) m. The frequency of the source signal was 2.0 kHz. The sound pressure distribution on the line $y = -0.5$ m was estimated at 128 points at intervals of 0.03 m. Here, the SDR for reconstruction, $\mathrm{SDR}_{\mathrm{rec}}$, is defined as

$$\mathrm{SDR}_{\mathrm{rec}} = 10 \log_{10} \frac{\sum_x \sum_t |p_{\mathrm{ideal}}(x,t)|^2}{\sum_x \sum_t |p_{\mathrm{ideal}}(x,t) - p_{\mathrm{est}}(x,t)|^2}, \qquad (18)$$

where $p_{\mathrm{ideal}}(\cdot)$ and $p_{\mathrm{est}}(\cdot)$ are the ideal and estimated sound pressure distributions on the reconstruction line, respectively.

Figure 4 shows the ideal and estimated distributions of the instantaneous sound pressures. The distribution estimated by **NAH** included significant errors due to spatial aliasing artifacts. In **M-IST**, the estimated distribution still contained errors because its signal decomposition was not accurate. The distribution estimated by **Proposed** exhibited relatively high accuracy. $\mathrm{SDR}_{\mathrm{rec}}$ for **Proposed**, **M-IST**, and **NAH** was 7.44, 5.75, and 1.17 dB, respectively. Therefore, the accurate decomposition of **Proposed** enables high reconstruction accuracy above the spatial Nyquist frequency.

5. CONCLUSION

A sound field decomposition method based on sparse and low-rank signal models was proposed. In current methods, no constraints are imposed on the ambiance components, whereas a spatially sparse distribution of source components is assumed. We assumed that the ambiance components have a low-rank signal structure. An algorithm for decomposition based on ADMM was derived. In the numerical simulations, the performance of signal decomposition was first evaluated. Then, the reconstruction accuracy of the sound pressure distribution was compared with that of other methods. The results indicated that the proposed method enables more accurate sparse sound field decomposition in a reverberant environment.

6. REFERENCES

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

[2] G. Chardon, L. Daudet, A. Peillot, F. Ollivier, N. Bertin, and R. Gribonval, "Near-field acoustic holography using sparsity and compressive sampling principles," *J. Acoust. Soc. Am.*, vol. 132, no. 3, pp. 1521–1534, 2012.

[3] A. Asaei, H. Bourlard, M. Taghizadeh, and V. Cevher, "Model-based sparse component analysis for reverberant speech localization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Florence, May 2014, pp. 1453–1457.

[4] S. Koyama, S. Shimauchi, and H. Ohmuro, "Sparse sound field representation in recording and reproduction for reducing spatial aliasing artifacts," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Florence, May 2014, pp. 4476–4480.

[5] T. Noohi, N. Epain, and C. T. Jin, "Super-resolution in acoustic imaging using sparse recovery with spatial aliasing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Brisbane, 2015, pp. 2414–2418.

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

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

[8] J. Maynard, E. G. Williams, and Y. Lee, "Nearfield acoustic holography: I. theory of generalized holography and the development of NAH," *J. Acoust. Soc. Am.*, vol. 78, no. 4, pp. 1395–1413, 1985.

[9] F. M. Fazi and P. A. Nelson, "The relationship between sound field reproduction and near-field acoustical holography," in *Proc. 129th AES Conv.*, 2010.

[10] S. Koyama, K. Furuya, Y. Hiwasaki, and Y. Haneda, "Reproducing virtual sound sources in front of a loudspeaker array using inverse wave propagator," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 6, pp. 1746–1758, 2012.

[11] 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.

[12] S. Koyama, K. Furuya, Y. Hiwasaki, Y. Haneda, and Y. Suzuki, "Wave field reconstruction filtering in cylindrical harmonic domain for with-height recording and reproduction," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 10, pp. 1546–1557, 2014.

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

[14] S. Koyama, N. Murata, and H. Saruwatari, "Structured sparse signal models and decomposition algorithm for super-resolution in sound field recording and reproduction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Brisbane, Apr. 2015, pp. 619–623.

[15] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *J. Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[16] M. B. McCoy, V. Cevher, Q. T. Dinh, A. Asaei, and L. Baldassarre, "Convexity in source separation: Models, geometry, and algorithms," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 87–95, 2014.

[17] 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.

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

[19] B. Recht, M. Fazel, and P. Parrilo, "Guaranteed minimum-rank solutions to linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.

[20] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *J. ACM*, vol. 52, no. 3, 2011.

[21] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. Advances in Neural Inf. Process. Systems (NIPS)*, Granada, 2011, pp. 612–620.

[22] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM Multiscale Model. Simul.*, vol. 4, no. 4, pp. 1168–1200, 2005.

[23] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[24] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, 1979.