# Sparse Sound Field Decomposition Using Group Sparse Bayesian Learning

Shoichi Koyama, Atsushi Matsubayashi, Naoki Murata, and Hiroshi Saruwatari*
* The University of Tokyo, Tokyo, Japan
E-mail: shoichi_koyama@ipc.i.u-tokyo.ac.jp Tel/Fax: +81-3-5841-6904

***Abstract*—A sparse sound field decomposition method is proposed. Sound field decomposition is the foundation of the various acoustic signal processing applications and enables the estimation of the entire sound field from pressure measurements. The plane wave decomposition, i.e., spatial Fourier analysis, of the sound field has been widely used; however, artifacts originating from spatial aliasing occur above the spatial Nyquist frequency. We have proposed a sparse sound field decomposition method based on a generative model as a sum of monopole source and plane wave components in the context of sound field recording and reproduction. For more accurate and robust decomposition, we propose three different group sparse signal models based on physical properties and a decomposition algorithm by extending sparse Bayesian learning. In simulation experiments, the accuracy of sparse decomposition was improved compared with that of current methods.**

## I. 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 the pressure measurements of multiple microphones. This makes it possible for the entire sound field to be estimated from the measurements. Plane wave decomposition, which corresponds to the spatial Fourier analysis of the sound field [1], has been commonly used because of its computational efficiency. In recent years, the sparse decomposition of the sound field has been proved to be effective in several applications [2]–[4] owing to the recent development of sparse decomposition algorithms in the context of compressed sensing [5], [6].

In sound field recording and reproduction 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. The driving signals of the loudspeakers used for reproduction must be calculated from the signals received by the microphones. This signal conversion can be achieved using the *wave field reconstruction (WFR) filtering* method, which is based on spatial Fourier analysis. We have derived the WFR filter for various array configurations [7]–[9]. Although this method enables stable and efficient signal conversion, artifacts originating from spatial aliasing notably occur, depending on the interelement spacing in the microphone or loudspeaker array. In the case of significant spatial aliasing artifacts, listeners may be 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*.

We previously developed a signal conversion method based on sparse sound field decomposition to reduce the effect of spatial aliasing artifacts [4], [10], in which the sound field is modeled as the sum of monopole source and plane wave components. Since only a few monopole components may exist inside a region near the microphones, it is possible to sparsely decompose the observed signals into basis functions, or dictionaries, consisting of Green's function. This method makes it possible to improve the reproduction accuracy 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.

For more accurate and robust sparse decomposition, prior information on the structure of the recording sound field may be useful. We have proposed three different group sparse signal models based on the physical properties of the sound field [10]. To address these signal models, an algorithm extended from the M-FOCUSS algorithm [11] was applied in [10]. In this paper, we extend the algorithm called sparse Bayesian learning (SBL) [12], [13] so that the proposed group sparse signal models can be addressed. Furthermore, we compare several algorithms in terms of their sparse sound field decomposition performance via numerical simulations.

## II. Generative model of sound field and its sparse decomposition

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 $\Omega$. When a sound pressure of temporal frequency $\omega$ 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 $\Omega$ and $k = \omega/c$ is the wave number obtained by setting the sound speed as $c$. Hereafter, $\omega$ 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. $p_{\mathrm{i}}(\mathbf{r})$ is represented as a convolution of $Q(\mathbf{r})$ and the three-dimensional

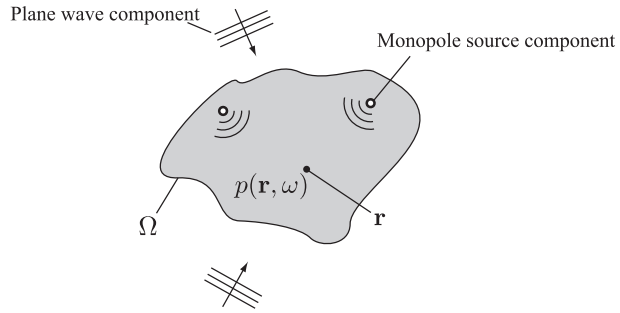



Fig. 1. Generative model of sound field.

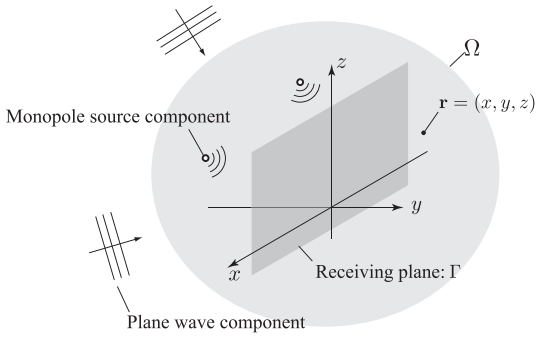



Fig. 2. Sound field modeled by sum of monopole source and plane wave components. The sound pressure is obtained on the receiving plane $\Gamma$.

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 confirmed by substituting it into (1) as

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

Since it is assumed that sound sources do not exist outside $\Omega$, the homogeneous term $p_{\mathrm{h}}(\mathbf{r})$ can be represented as the sum of plane waves.

Our objective is to decompose the sound field into $p_{\mathrm{i}}(\mathbf{r})$ and $p_{\mathrm{h}}(\mathbf{r})$ from sound pressure measurements inside $\Omega$. As shown in Fig. 2, we assume that the sound pressure distribution on the receiving plane $\Gamma$ is obtained. By spatial Fourier analysis, which corresponds to only the use of $p_{\mathrm{h}}(\mathbf{r})$, the signal energy may be spread over the basis functions even when a single monopole source exists; therefore, spatial aliasing artifacts cannot be avoided. On the other hand, when the decomposition into $p_{\mathrm{i}}(\mathbf{r})$ and $p_{\mathrm{h}}(\mathbf{r})$ is accurately performed, the dominant component of the observed signal can be represented by $p_{\mathrm{i}}(\mathbf{r})$ since the monopole components lie close to $\Gamma$. Furthermore, $Q(\mathbf{r})$ ($\mathbf{r}\in\Omega$) may become sparse because it can be considered that the monopole components exist only at a few locations in $\Omega$. When these assumptions are approximately satisfied, more accurate sound field estimation can be achieved, particularly above the spatial Nyquist frequency. For instance, in sound field recording and reproduction, the reproduction accuracy can be improved above the spatial Nyquist frequency by this representation [4].

In practice, (2) must be discretized to derive a computational algorithm for the decomposition. The region $\Omega$ is discretized as a set of grid points. Omnidirectional microphones are discretely aligned on $\Gamma$ 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 $\Omega$. The discrete form of (2) can be represented as

$$\mathbf{y} = \mathbf{D}\mathbf{x} + \mathbf{z}, \tag{5}$$

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, which corresponds to the ambient components, and $\mathbf{D}\in\mathbb{C}^{M\times N}$ is the dictionary matrix of the monopole components, whose elements consist of Green's functions between the grid points and the microphones. As discussed above, we assume that $\mathbf{x}$ is the dominant component of $\mathbf{y}$ and that $\mathbf{z}$ is a residual. Since it can be assumed that only a few monopole components exist in $\Omega$, a small number of elements in $\mathbf{x}$ may have nonzero values. Therefore, sparse decomposition algorithms [5], [6] can be applied to decompose $\mathbf{y}$ into $\mathbf{x}$ and $\mathbf{z}$.

## III. Group sparse signal models based on physical properties

A fundamental difficulty lies in the sparse decomposition of (5). On one hand, the grid points should cover $\Omega$ as densely as possible for an accurate signal representation. On the other hand, the dense distribution of the grid points leads to a high correlation between the columns of $\mathbf{D}$, which makes the sparse signal decomposition difficult [6]. Therefore, we exploit prior information on the structure of the sound field, i.e., the structure of the solution vector $\mathbf{x}$. We describe three different group sparse signal models based on the physical properties of the sound field.

### *Model 1: multiple time frames*

When multiple time frames of $\mathbf{y}$ are available and monopole components can be assumed to be static, each $\mathbf{x}$ may have the same sparsity pattern. The sparse decomposition problem using this model is known as the multiple measurement vector

(MMV) problem, for which several algorithms have been proposed [11], [13]–[15].

We denote the index of the time frame as $l \in \{1, \cdots, L\}$ and the signals of each $l$ as $\mathbf{y}_l \in \mathbb{C}^M$, $\mathbf{x}_l \in \mathbb{C}^N$, and $\mathbf{z}_l \in \mathbb{C}^M$. By concatenating them in vectors, we can represent (5) as

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_L \end{bmatrix} = \begin{bmatrix} \mathbf{D} & & & \mathbf{0} \\ & \mathbf{D} & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_L \end{bmatrix} + \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \vdots \\ \mathbf{z}_L \end{bmatrix}. \quad (6)$$

Each $\mathbf{x}_l$ is assumed to have nonzero values at the same positions.

*Model 2: temporal frequencies*

Many types of acoustic source signals have a broad frequency band. Therefore, each $\mathbf{x}$ in multiple frequency bins may have the same sparsity pattern. Similarly to model 1, using the index of the frequency bin $l \in \{1, \cdots, L\}$, we denote the signals of each $l$ as $\mathbf{y}_l \in \mathbb{C}^M$, $\mathbf{x}_l \in \mathbb{C}^N$, and $\mathbf{z}_l \in \mathbb{C}^M$. Since Green's function depends on the temporal frequency, the dictionary matrix of each $l$ is denoted as $\mathbf{D}_l \in \mathbb{C}^{M \times N}$. Therefore, (5) can be represented as

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_L \end{bmatrix} = \begin{bmatrix} \mathbf{D}_1 & & & \mathbf{0} \\ & \mathbf{D}_2 & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{D}_L \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_L \end{bmatrix} + \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \vdots \\ \mathbf{z}_L \end{bmatrix}. \quad (7)$$

Again, each $\mathbf{x}_l$ is assumed to have nonzero values at the same positions. Note that the dictionary matrices in (7) are different in each group, whereas those in (6) are the same.

*Model 3: image sources and multipole components*

Signals obtained in an ordinary room have reflections from walls in addition to direct sound. This phenomenon leads to the presence of monopole components at the reflective image source locations [16]. As another example, since the sound sources have a complex directivity pattern, multipole source components, such as dipole and quadrupole components, may exist at the same location as monopole components [1]. These properties can be represented by the same group sparse signal model.

When considering an image source model, using the index of the image sources $l \in \{1, \cdots, L\}$, we denote the signal of each $l$ as $\mathbf{x}_l \in \mathbb{C}^N$. Green's function between the $l$th image source location and the microphones is denoted as $\mathbf{D}_l \in \mathbb{C}^{M \times N}$. Therefore, (5) can be represented as

$$\mathbf{y} = \begin{bmatrix} \mathbf{D}_1, \mathbf{D}_2, \cdots, \mathbf{D}_L \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_L \end{bmatrix} + \mathbf{z}. \quad (8)$$

Again, each $\mathbf{x}_l$ is assumed to have nonzero values at the same positions. Note that the length of $\mathbf{y}$ is degenerated to $M$. Therefore, the structure of the dictionary matrix is different from that in (6) and (7). In this scenario, room geometry must be known to design $\mathbf{D}_l$.

In the case of multipole components, $\mathbf{D}_l$ in (8) becomes Green's function for each multipole.

*Combined models*

Models 1, 2, and 3 can be arbitrarily combined. For instance, in the case of combining models 1 and 2, each $\mathbf{x}_l$ in (6) is replaced by the solution vector in (7), and the dictionary matrix is designed accordingly. To combine $J$ groups, the sets of indexes of the groups are denoted as $\mathcal{G}_j$ ($j \in \{1, \cdots, J\}$), and the index of the $j$th group is denoted as $l_j \in \{1, \cdots, |\mathcal{G}_j|\}$. We redefine the signal vectors and dictionary matrix as $\mathbf{x} \in \mathbb{C}^{N|\mathcal{G}_1|\cdots|\mathcal{G}_J|}$, $\mathbf{y}$, $\mathbf{z}$, and $\mathbf{D}$. The sizes of $\mathbf{y}$, $\mathbf{z}$, and $\mathbf{D}$ depend on the types of combination of models. Each group has a nested structure in the solution vector $\mathbf{x}$. These variables can also be related as a linear equation as in (5). We previously proposed an algorithm for solving the group sparse decomposition problem by extending M-FOCUSS [10], [11].

## IV. Group sparse Bayesian learning

SBL was first proposed as a method of solving sparse representation problems in the context of regression and classification [17], [18]. Wipf and Rao applied SBL to signal processing [12] and extended it to solve the MMV problem [13]. Another extension to address the temporal correlation in the MMV problem was proposed by Zhang and Rao [19].

We now extend SBL to solve the group sparse signal representation problem. We assume the likelihood function $\rho(\mathbf{y}|\mathbf{x})$ as a circularly symmetric complex Gaussian distribution with the noise variance $\sigma^2$:

$$\rho\left(\mathbf{y}|\mathbf{x}\right) = \left(\pi\sigma^2\right)^{-N|\mathcal{G}_1|\cdots|\mathcal{G}_J|} \exp\left(-\frac{1}{\sigma^2}\|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2\right). \quad (9)$$

We assume that the prior distribution of each element of $\mathbf{x}$ is an independent complex Gaussian as follows:

$$\rho\left(\mathbf{x};\boldsymbol{\gamma}\right) = \prod_{n,l_1,\cdots,l_J} \left(\pi\gamma_n\right)^{-1} \exp\left(-\frac{|x_{n,l_1,\cdots,l_J}|^2}{\gamma_n}\right), \quad (10)$$

where $x_{n,l_1,\cdots,l_J}$ is the $n$th element of the $l_1, \cdots, l_J$th group of $\mathbf{x}$ and $\boldsymbol{\gamma} = [\gamma_1, \cdots, \gamma_N]^T \in \mathbb{R}_+^N$ is a positive variance parameter. Note that $\boldsymbol{\gamma}$ has the same value within each group. By combining (9) and (10), the posterior density of $\mathbf{x}$ also becomes the Gaussian

$$\rho\left(\mathbf{x}|\mathbf{y};\boldsymbol{\gamma}\right) = \frac{\rho\left(\mathbf{x},\mathbf{y};\boldsymbol{\gamma}\right)}{\int \rho\left(\mathbf{x},\mathbf{y};\boldsymbol{\gamma}\right) d\mathbf{x}} = \mathcal{N}\left(\boldsymbol{\mu},\boldsymbol{\Sigma}\right). \quad (11)$$

Here, the mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ are written as [20]

$$\boldsymbol{\mu} = \boldsymbol{\Gamma}\mathbf{D}^H\boldsymbol{\Sigma}_y^{-1}\mathbf{D}\boldsymbol{\Gamma} \quad (12)$$

$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma} - \boldsymbol{\Gamma}\mathbf{D}^H\boldsymbol{\Sigma}_y^{-1}\mathbf{D}\boldsymbol{\Gamma}, \quad (13)$$

where $\boldsymbol{\Gamma} = \mathrm{diag}\left(\boldsymbol{\gamma}, \cdots, \boldsymbol{\gamma}\right)$ and $\boldsymbol{\Sigma}_y = \sigma^2\mathbf{I} + \mathbf{D}\boldsymbol{\Gamma}\mathbf{D}^H$.

By optimizing $\boldsymbol{\gamma}$, appropriate grid points can be selected among $n$. In the empirical Bayesian strategy, the marginal

**Algorithm 1** Group sparse Bayesian learning algorithm.

Initialize $\boldsymbol{\gamma}$, $t = 1$
**while** loop $\neq 0$ **do**
  $\boldsymbol{\Sigma}_y \leftarrow \sigma^2 \mathbf{I} + \mathbf{D}\boldsymbol{\Gamma}\mathbf{D}^H$
  $\boldsymbol{\Sigma} \leftarrow \boldsymbol{\Gamma} - \boldsymbol{\Gamma}\mathbf{D}^H\boldsymbol{\Sigma}_y^{-1}\mathbf{D}\boldsymbol{\Gamma}$
  **for all** $n$ **do**
    $\gamma_n \leftarrow \frac{1}{|\mathcal{G}_1|\cdots|\mathcal{G}_J|}\sum_{l_1,\cdots,l_J}\left(\mu^2_{n,l_1,\cdots,l_J} + \mathrm{diag}\left(\boldsymbol{\Sigma}\right)_{n,l_1,\cdots,l_J}\right)$
  **end for**
  $t \leftarrow t + 1$
  **if** stopping condition is satisfied **then**
    loop $= 0$
  **end if**
**end while**

likelihood is used as the cost function,

$$\begin{aligned}\mathcal{L}(\boldsymbol{\gamma}) &= -\log\int \rho\left(\mathbf{y}|\mathbf{x}\right)\rho\left(\mathbf{x};\boldsymbol{\gamma}\right)d\mathbf{x} \\ &= -\log\rho\left(\mathbf{y};\boldsymbol{\gamma}\right) \\ &= \log|\boldsymbol{\Sigma}_y| + \mathbf{y}^H\boldsymbol{\Sigma}_y^{-1}\mathbf{y}, \end{aligned} \tag{14}$$

where the derivation of the last line can be found in [20]. Therefore, $\mathcal{L}(\boldsymbol{\gamma})$ must be minimized with respect to $\boldsymbol{\gamma}$.

To solve the minimization problem of (14), we apply the EM algorithm by treating $\mathbf{x}$ as hidden data [18]. For the E-step, the posterior moments are calculated using (13) as

$$\mathrm{E}_{\mathbf{x}|\mathbf{y},\boldsymbol{\gamma}^{(t)}}\left[|x_{n,l_1,\cdots,l_J}|^2\right] = \mu^2_{n,l_1,\cdots,l_J} + \mathrm{diag}(\boldsymbol{\Sigma})_{n,l_1,\cdots,l_J}, \tag{15}$$

where $\boldsymbol{\gamma}^{(t)}$ is the estimate of $\boldsymbol{\gamma}$ at the $t$th step. For the M-step, $\boldsymbol{\gamma}^{(t)}$ is updated as

$$\begin{aligned}\boldsymbol{\gamma}^{(t+1)} &= \arg\max_{\boldsymbol{\gamma}} \mathrm{E}_{\mathbf{x}|\mathbf{y},\boldsymbol{\gamma}^{(t)}}\left[\log p\left(\mathbf{y},\mathbf{x};\boldsymbol{\gamma}\right)\right] \\ &= \arg\max_{\boldsymbol{\gamma}} \mathrm{E}_{\mathbf{x}|\mathbf{y},\boldsymbol{\gamma}^{(t)}}\left[\log p\left(\mathbf{x};\boldsymbol{\gamma}\right)\right]. \end{aligned} \tag{16}$$

By setting the derivative of $\log p(\mathbf{x};\boldsymbol{\gamma})$ with respect to $\gamma_n$ as 0, the following update rule can be obtained for each $n$:

$$\begin{aligned}&\gamma_n^{(t+1)} \\ &= \mathrm{E}_{\mathbf{x}|\mathbf{y},\boldsymbol{\gamma}^{(t)}}\left[\frac{1}{|\mathcal{G}_1|\cdots|\mathcal{G}_J|}\sum_{l_1,\cdots,l_J}|x_{n,l_1,\cdots,l_J}|^2\right] \\ &= \frac{1}{|\mathcal{G}_1|\cdots|\mathcal{G}_J|}\sum_{l_1,\cdots,l_J}\left(\mu^2_{n,l_1,\cdots,l_J} + \mathrm{diag}\left(\boldsymbol{\Sigma}\right)_{n,l_1,\cdots,l_J}\right). \end{aligned} \tag{17}$$

The resulting algorithm is summarized in Algorithm 1.

Although the noise variance $\sigma^2$ in (9) is assumed to be known in the above derivation, it can be simultaneously estimated [18]. However, it is known that the estimate of $\sigma^2$ can be inaccurate in the SBL framework [12], [13].

## V. Experiments

Numerical simulations were conducted to evaluate the proposed method. First, we compared group SBL with several sparse decomposition algorithms using synthetic data. Second, we demonstrated a sparse sound field decomposition using the proposed method.

### A. Comparison of algorithms using synthetic data

We compared group SBL (G-SBL) with M-SBL [13], group FOCUSS (G-FOCUSS) [10], and M-FOCUSS [11] using randomly generated signals and dictionaries. Two different combinations models, models 1 and 2 and model 1 and 3, were investigated.

To evaluate the performance of sparse decomposition, the F-measure ($F_{\mathrm{msr}}$) and signal-to-distortion ratio (SDR) were used. The operator $\mathrm{supp}(\cdot)$ is used to extract a set of indexes such that the amplitude of each element of the estimated solution vector $\hat{\mathbf{x}}$ is larger than a threshold value $\epsilon$,

$$\mathrm{supp}(\hat{\mathbf{x}}) = \{n \in \{1,\cdots,N|\mathcal{G}_1||\mathcal{G}_2|\} \mid |\hat{x}_n| > \epsilon\}, \tag{18}$$

where $\hat{x}_n$ is the $n$th element of $\hat{\mathbf{x}}$. $F_{\mathrm{msr}}$ is defined as

$$F_{\mathrm{msr}} = 2\frac{|\mathrm{supp}(\hat{\mathbf{x}}) \cap \mathrm{supp}(\mathbf{x}_{\mathrm{true}})|}{|\mathrm{supp}(\hat{\mathbf{x}})| + |\mathrm{supp}(\mathbf{x}_{\mathrm{true}})|}, \tag{19}$$

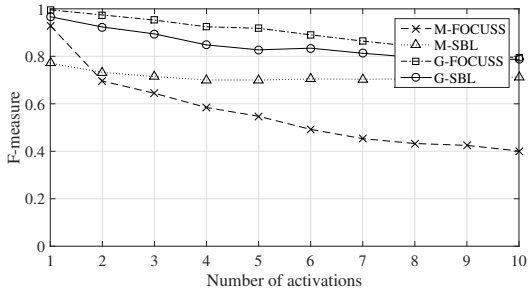
where $\mathbf{x}_{\mathrm{true}}$ is the true solution vector. Therefore, $F_{\mathrm{msr}}$ is equal to 1 when the activated indexes of these vectors are exactly the same. SDR is defined as

$$\mathrm{SDR} = 10\log_{10}\frac{\|\mathbf{x}_{\mathrm{true}}\|_2^2}{\|\mathbf{x}_{\mathrm{true}} - \hat{\mathbf{x}}\|_2^2}. \tag{20}$$

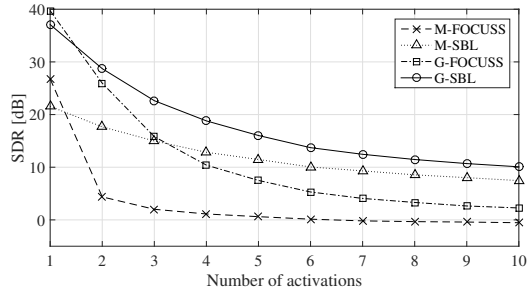
These values were averaged over 100 trials.

For the combination of models 1 and 2, the parameters were set to $M = 32$ and $N = 64$. The numbers of elements in both groups, $|\mathcal{G}_1|$ and $|\mathcal{G}_2|$, were 16. The activation locations of $\mathbf{x}$ were randomly selected from $n \in \{1,\cdots,N\}$ with the predetermined number of activations. The value of each element was generated by a Gaussian distribution with mean 0 and variance 1.0. The dictionary matrix $\mathbf{D} \in \mathbb{R}^{M|\mathcal{G}_1||\mathcal{G}_2| \times N|\mathcal{G}_1||\mathcal{G}_2|}$ was generated so that the columns of $\mathbf{D}$ were correlated [15]. First, each element of the matrix $\bar{\mathbf{D}} \in \mathbb{R}^{M|\mathcal{G}_1||\mathcal{G}_2| \times N|\mathcal{G}_1||\mathcal{G}_2|}$ was generated by a Gaussian distribution. Second, $\mathbf{D}$ was calculated as $\mathbf{D} = \bar{\mathbf{D}}\mathbf{W}^{1/2}$ by multiplying by the matrix $\mathbf{W} \in \mathbb{R}^{N|\mathcal{G}_1||\mathcal{G}_2| \times N|\mathcal{G}_1||\mathcal{G}_2|}$. The Toeplitz matrix $\bar{\mathbf{W}} \in \mathbb{R}^{N\times N}$, which is a block diagonal element of $\mathbf{W}$, was obtained so that the vector $\mathbf{v} \in \mathbb{R}^N$ is the first row of $\bar{\mathbf{W}}$. The $n$th element of $\mathbf{v}$, $v_n$, is $v_n = \xi^{n-1} + a_n$, where $a_n$ is from the vector comprising a uniformly random sequence sorted in decreasing order and $\xi \in [0,1]$ is a constant parameter used to determine the correlation strength. Here, $\xi$ was set to 0.9. In addition, white Gaussian noise was added to $\mathbf{y}$ to obtain an signal-to-noise ratio (SNR) of 40 dB.

Figure 3 shows the results of the sparse decomposition performance. The horizontal axis denotes the number of activations in both figures. The F-measure and SDR for G-FOCUSS were higher than those for M-FOCUSS. In the same manner, those for G-SBL were higher than those for M-SBL. In all the

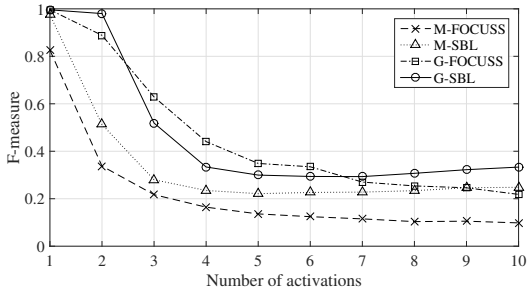



(a) F-measure

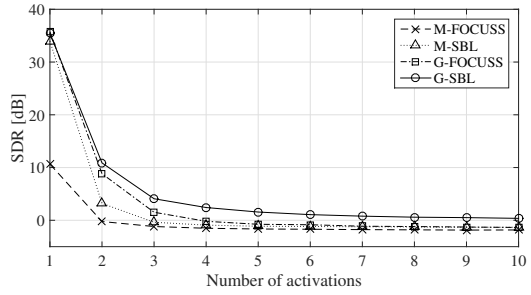



(b) SDR

Fig. 3. Results of sparse decomposition performance for combination of models 1 and 2.




(a) F-measure




(b) SDR

Fig. 4. Results of sparse decomposition performance for combination of models 1 and 3.

algorithms, the F-measure and SDR decreased as the number of activations was increased. When G-SBL and G-FOCUSS were compared, the F-measure for both methods was found to be almost the same. The SDR for G-SBL was higher than that for G-FOCUSS, particularly for a large number of activations.

For the combination of models 1 and 3, the parameters used to set the size of the signals were the same as before. Again, the dictionary matrix $\mathbf{D} \in \mathbb{R}^{M|\mathcal{G}_1| \times N|\mathcal{G}_1||\mathcal{G}_2|}$ was generated so that the columns of $\mathbf{D}$ were correlated. The parameter used to determine the correlation strength $\xi$ was 0.01.

The results for the combination of models 1 and 3 are shown in Fig. 4. The F-measure and SDR were relatively low compared with the results in Fig. 3. This is because the observation vector is degenerated, which makes the sparse decomposition difficult for a large number of activations. The differences between the four algorithms had similar tendency to those in Fig. 3. The F-measure and SDR for G-SBL and G-FOCUSS were higher than those for M-SBL and M-FOCUSS, respectively. The difference between G-SBL and G-FOCUSS was not apparent, in contrast to that in Fig. 3.

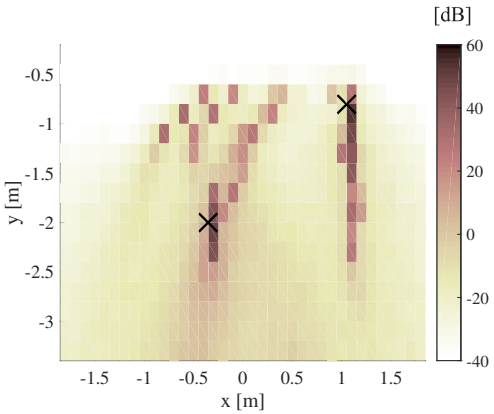
### *B. Evaluation of sound field decomposition performance*

Experiments to evaluate the performance for a sparse sound field decomposition were conducted under the free-field assumption. A linear array of omnidirectional microphones was set along the $x$-axis with the center at the origin. Thirty-two microphones were placed at intervals of 0.12 m. Static point sources were located at (1.05, –0.8, 0.0) m and (–0.35, –2.0, 0.0) m. The source signals were speech signals of male and female utterances. The grid points were aligned in a rectangular region of 3.8×3.4 m$^2$ on the $x$-$y$ plane at $z = 0$. The numbers of grid points were 38 in the $x$-direction and 17 in the $y$-direction with intervals of 0.1 m and 0.2 m, respectively. The sampling frequency was 16 kHz. The frame length of the short-term Fourier transform (STFT) was 512 samples and its shift length was 128 samples. We applied a combination of models 1 and 2 for multiple time frames and temporal frequencies. Sixteen time frames were clipped from the observed signals for use in model 1. White Gaussian noise was added to obtain an SNR of 20 dB.

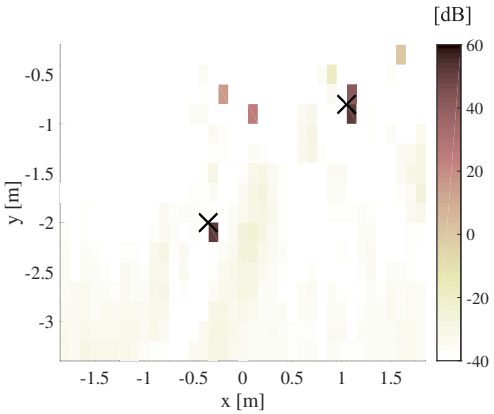
Figure 5 shows the distributions of $\mathbf{x}$ obtained by M-SBL and G-SBL. The amplitudes of $\mathbf{x}$ averaged over the frequencies and time frames are plotted in a dB scale. The crosses represent the true source locations. In M-SBL, the amplitude distribution was dispersed along the $y$-direction. This is because the power of the source signal decreases at several time frames and frequency bins. On the other hand, in G-SBL, the sparsity of the amplitude distribution was improved and the two true source locations were accurately detected, although small errors can be seen at a few grid points. Thus, the reconstruction accuracy may be improved using G-SBL.

## VI. Conclusion

A sparse sound field decomposition method using G-SBL was proposed. We formulated a generative model of a sound field as a sum of monopole source and plane wave components. In addition, three different group sparse signal models based on the physical properties of the sound field were proposed. The SBL algorithm was extended to address




(a) M-SBL




(b) G-SBL

Fig. 5. Results of sparse sound field decomposition. True source locations are denoted by crosses.

the proposed signal models and to perform group sparse decomposition. Numerical simulation results indicated that the methods using the group sparsity can outperform the methods used to solve the MMV problem. The performance of G-SBL was slightly better than that of G-FOCUSS, particularly for a large number of activations.

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

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

[7] 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.

[8] 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.

[9] S. Koyama, *Boundary Integral Approach to Sound Field Transform and Reproduction*, Ph.D. thesis, Graduate School of Information Science and Technology, the University of Tokyo, 2014.

[10] 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.

[11] 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.

[12] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, 2004.

[13] D. P. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3704–3716, 2007.

[14] J. Tropp, A. Gilbert, and M. Strauss, "Algorithms for simultaneous sparse approximation. Part I: greedy pursuit," *Signal Process.*, vol. 86, pp. 572–588, 2006.

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

[16] 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.

[17] M. E. Tipping, "The relevance vector machine," in *Neural Inform. Process. Syst.*, 2000, vol. 12, pp. 652–658.

[18] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Machine Learning Res.*, vol. 1, pp. 211–244, 2001.

[19] Z. Zhang and B. D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 912–926, 2011.

[20] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.