Exterior and Interior Sound Field Separation Using Convex Optimization: Comparison of Signal Models

Yuhta Takida[†], Shoichi Koyama[†,††], Hiroshi Saruwatari[†]
[†]The University of Tokyo, Graduate School of Information Science and Technology,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
[††]Institute Langevin, ESPCI, Paris Diderot University, CNRS UMR 7587,
1 rue Jussieu, Paris 75005, France

***Abstract*—An exterior (direct source) and interior (reverberant) sound field separation method using a convex optimization algorithm is proposed. Extracting the exterior sound field from mixed observations using multiple microphones can be an effective preprocessing approach to analyzing the sound field inside a region including sources in a reverberant environment. We formulate signal models of the exterior and interior sound fields by exploiting the signal characteristics of each sound field. The interior sound field is sparsely represented using overcomplete plane-wave functions. Two models using harmonic functions and a low-rank structure are proposed for the exterior sound field. The separation algorithms for each model are derived by the alternating direction method of multipliers. Numerical simulation results indicate that higher separation accuracy than that for existing methods can be achieved by the proposed method with a small number of microphones and a flexible microphone arrangement.**

I. Introduction

When analyzing a sound field inside a region of interest including sound sources, the interference of sound waves incoming from outside of the region, such as reverberation, can significantly affect the analysis. For example, the measurement of loudspeaker characteristics [1], the visualization of sound field (i.e., acoustic holography) [2], [3], and source identification (i.e., source localization and/or reconstruction of source field) [4]–[6] become extremely difficult tasks in a reverberant environment. Extracting a direct source field, i.e., the exterior sound field of a region, from mixed observations can be an effective preprocessing approach for these analyses. We address the separation problem of exterior (direct source) and interior (reverberant) sound fields using multiple microphones.

Previous studies of sound field separation were mainly based on the discretization of integral equations of sound propagation. Bi et al. [1] proposed a method based on the equivalent source method (ESM) [7]–[9], which can be regarded as the discretized form of the single layer potential [10]. By using microphones on two-layer surfaces adjacent to the boundary of a region, the exterior and interior sound fields are approximated by the sum of imaginary monopoles (equivalent sources) placed on an imaginary layer. By estimating the amplitudes of the equivalent sources from the observed signals, exterior and interior sound fields can be separately obtained. The method proposed in [11] is based on sound field representation in a harmonic domain. By arranging microphones on two spherical layers, the harmonic coefficients of each sound field are estimated by calculating a surface integral using the sound pressures obtained by the microphones. These methods basically require a large number of microphones for accurate calculation of the integral equations. Furthermore, the microphones have to be uniformly arranged on the two layers. These limitations on the number of microphones and microphone geometries significantly decrease their practical applicability.

We propose an exterior and interior sound field separation method using a convex optimization algorithm. Signal models incorporating a convex relaxation are formulated by exploiting the signal characteristics of each sound field. The interior sound field is sparsely represented by overcomplete plane-wave functions. A representation using harmonic functions and a low-rank structure is proposed for the exterior sound field. Then, two types of convex optimization problem are formulated and algorithms based on the alternating direction method of multipliers (ADMM) [12]–[14] are derived. Numerical simulations are conducted to evaluate and compare the proposed signal models and algorithms with the existing methods.

II. Statement of Problem

Suppose that a closed region Ω including sound sources is set inside a measurement room. The sound pressure of frequency ω at position $\mathbf{r}$ is denoted as $u(\mathbf{r},\omega)$. Inside a region adjacent to $\partial\Omega$ (the boundary of Ω) not including any sources, $u(\mathbf{r},\omega)$ is represented as the sum of the exterior and interior sound fields of Ω, $u_{\mathrm{E}}(\cdot)$ and $u_{\mathrm{I}}(\cdot)$, respectively, as

$$u(\mathbf{r},\omega) = u_{\mathrm{E}}(\mathbf{r},\omega) + u_{\mathrm{I}}(\mathbf{r},\omega). \tag{1}$$

We hereafter omit ω for notational simplicity.

We assume that M microphones are set adjacent to $\partial\Omega$, for example, on two-layer surfaces as in Fig. 1. The vector of the observed signals consisting of $u(\mathbf{r}_m)$ with the mth microphone location $\mathbf{r}_m$ ($m \in \{1,\dots,M\}$) is denoted as $\mathbf{y} \in \mathbb{C}^M$. From (1), $\mathbf{y}$ can also be represented as the sum of the exterior and interior sound field components, $\mathbf{x} \in \mathbb{C}^M$ and $\mathbf{z} \in \mathbb{C}^M$, as

$$\mathbf{y} = \mathbf{x} + \mathbf{z}. \tag{2}$$

We here assume that T time frames of microphone observations are available. By denoting the index of the time frame as $t \in \{1,\dots,T\}$, the signal vectors of the tth time

ISBN 978-90-827970-1-5

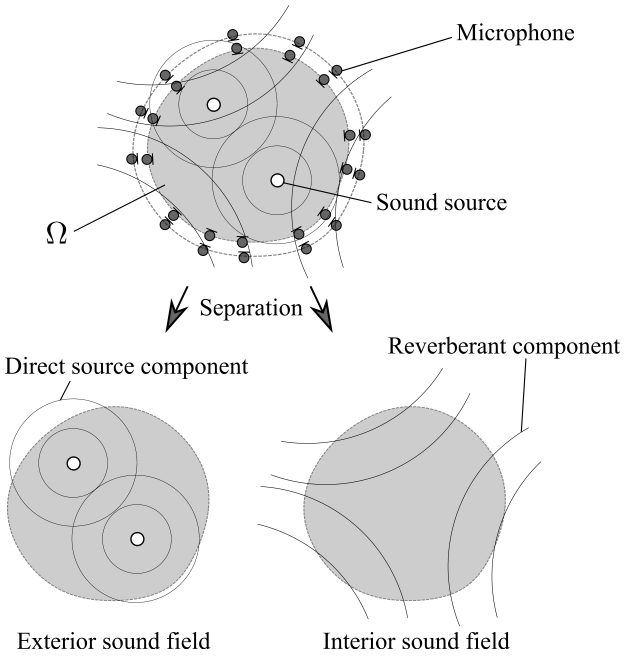

Fig. 1: Separation of exterior and interior sound fields using microphones adjacent to boundary of Ω.

frame are represented by subscript t as $\mathbf{y}_t$, $\mathbf{x}_t$, and $\mathbf{z}_t$. When the matrices consisting of each signal vector are defined as $\mathbf{Y} := [\mathbf{y}_1, \ldots, \mathbf{y}_T] \in \mathbb{C}^{M\times T}$, $\mathbf{X} := [\mathbf{x}_1, \ldots, \mathbf{x}_T] \in \mathbb{C}^{M\times T}$, and $\mathbf{Z} := [\mathbf{z}_1, \ldots, \mathbf{z}_T] \in \mathbb{C}^{M\times T}$, (2) can be reformulated as

$$\mathbf{Y} = \mathbf{X} + \mathbf{Z}. \tag{3}$$

Our objective is to separate the observed signals $\mathbf{Y}$ into the exterior and interior sound field components $\mathbf{X}$ and $\mathbf{Z}$, respectively, as in (3).

In the ESM [1], two imaginary layers are set inside and outside $\partial\Omega$, and $u_{\mathrm{E}}(\cdot)$ and $u_{\mathrm{I}}(\cdot)$ are approximated by the sum of imaginary monopoles on the inner and outer imaginary layers, respectively. The amplitudes of these equivalent sources are determined by matching them with the microphone observations. The theoretical basis of this technique can be regarded as the discretization of the single-layer potential [10], which is an integral equation for representing a homogeneous sound field. However, a large number of microphones is generally required for accurate separation [1]. The array geometry of the microphones must be two layers on $\partial\Omega$. The method proposed in [11] is based on the approximation of $u_{\mathrm{E}}(\cdot)$ and $u_{\mathrm{I}}(\cdot)$ by spherical harmonic functions. The harmonic coefficients are estimated by calculating the surface integral of $u(\cdot)$. This method also requires a large number of microphones since the accuracy of the surface integral strongly depends on the number of microphones and the frequency. A two-layer spherical microphone arrangement is required to calculate this integral.

III. Signal Models for Sound Field Separation

We formulate the separation problem of $\mathbf{Y}$ into $\mathbf{X}$ and $\mathbf{Z}$ as a convex optimization problem exploiting the signal characteristics of the exterior and interior sound fields. This strategy enables a more flexible geometry of the microphones and reduces the number of microphones.

We introduce signal models for the interior and exterior sound fields. The interior sound field is assumed to be sparse in the plane-wave domain. Two models are proposed for the exterior sound field: representation by harmonic functions and by low-rank signals. For simplicity, a two-dimensional (2D) sound field is assumed; however, the proposed method can be extended to the three-dimensional (3D) case with several modifications.

A. Signal Models for Interior Sound Field

We assume that the interior sound field can be sparsely represented in a plane-wave domain. According to [15], a homogeneous sound field in a certain star-shaped region can be well approximated by a limited number of plane waves, which has been successfully applied in various contexts [3], [16], [17]. Therefore, the interior sound field $u_{\mathrm{I}}(\mathbf{r})$ can be sparsely decomposed by using overcomplete plane-wave basis functions $e^{i\mathbf{k}_l\cdot\mathbf{r}}$ ($l \in \{1, \ldots, L\}$) as

$$u_{\mathrm{I}}(\mathbf{r}) \approx \sum_{l=1}^{L} s_l e^{i\mathbf{k}_l\cdot\mathbf{r}}, \tag{4}$$

where the wave-vectors $\mathbf{k}_l$ are chosen uniformly on the circle of radius k in the wave-number domain, and a limited number of expansion coefficients s_l have nonzero values. By denoting the plane wave dictionary matrix as $\mathbf{W} \in \mathbb{C}^{M\times L}$, whose elements consist of plane-wave basis functions, the interior sound field component $\mathbf{Z}$ is represented as

$$\mathbf{Z} = \mathbf{W}\mathbf{S}, \tag{5}$$

where $\mathbf{S} \in \mathbb{C}^{M\times L}$ will have a sparse structure. When the source positions and acoustic characteristics of the room are assumed to be static during T time frames, each column of $\mathbf{S}$ will have nonzero values at the same indexes. Then, $\mathbf{S}$ can be assumed to be row-sparse.

B. Signal Models for Exterior Sound Field

We introduce two types of signal model for the exterior sound field. The first model is a representation using circular harmonic functions. By denoting the position vector in polar coordinates as $\mathbf{r} = (r, \phi)$, $u_{\mathrm{E}}(\mathbf{r})$ can be expanded as

$$u_{\mathrm{E}}(\mathbf{r}) = \sum_{n=-\infty}^{\infty} g_n H_n(kr) e^{in\phi}, \tag{6}$$

where $H_n(\cdot)$ is the nth-order Hankel function of the first kind, k ($= \omega/c$) is the wave number, and c is the sound velocity. By truncating order n at an appropriate number N, $u_{\mathrm{E}}(\mathbf{r})$ will be well approximated [18]. Then, $\mathbf{X}$ can be decomposed by using the dictionary matrix of circular harmonic functions $\mathbf{D} \in \mathbb{C}^{M\times(2N+1)}$. From (5) and (6), (3) can be represented as

$$\mathbf{Y} = \mathbf{D}\mathbf{G} + \mathbf{W}\mathbf{S}, \tag{7}$$

ISBN 978-90-827970-1-5 © EURASIP 2018

where $\mathbf{G} \in \mathbb{C}^{(2N+1)\times T}$ consists of the expansion coefficients of the circular harmonics. The signal decomposition based on model (7) can be achieved by solving the following optimization problem:

$$\begin{aligned} &\underset{\mathbf{G},\mathbf{S}}{\text{minimize}} \ \|\mathbf{G}\|_F + \lambda\|\mathbf{S}\|_{1,2} \\ &\text{subject to } \|\mathbf{D}\mathbf{G} + \mathbf{W}\mathbf{S} - \mathbf{Y}\|_F^2 = 0, \end{aligned} \tag{8}$$

where $\|\cdot\|_F$ is the Frobenius norm and $\|\cdot\|_{1,2}$ is the $\ell_{1,2}$-norm defined as

$$\|\mathbf{S}\|_{1,2} = \sum_{l=1}^{L} \|\mathbf{S}_{l,\cdot}\|_2. \tag{9}$$

Here, $\mathbf{S}_{l,\cdot}$ represents the lth row of $\mathbf{S}$. The $\ell_{1,2}$-norm is typically used to induce row-sparsity of the matrix because the optimization problem becomes convex [19].

The second model is based on the assumption that the exterior sound field components have a low-rank structure. The exterior sound field component $\mathbf{X}$ is represented as the direct product of the source signals and their transfer functions; therefore, when the source signals are mutually uncorrelated, the rank of the covariance matrix, $\mathbf{X}\mathbf{X}^{\mathsf{H}}$, will approximately correspond to the number of direct sources inside Ω. The signal decomposition based on this model can be achieved by solving the following optimization problem:

$$\begin{aligned} &\underset{\mathbf{X},\mathbf{S}}{\text{minimize}} \ \|\mathbf{X}\|_* + \lambda\|\mathbf{S}\|_{1,2} \\ &\text{subject to } \|\mathbf{X} + \mathbf{W}\mathbf{S} - \mathbf{Y}\|_F^2 = 0, \end{aligned} \tag{10}$$

where $\|\cdot\|_*$ represents the nuclear norm, which is the tightest convex lower bound of the rank function [20].

IV. Signal Decomposition Algorithm Using ADMM

The optimization problems for the two models, which are hereafter referred to as the Frobenius–sparse (FS) and Low-rank–sparse (LS) models, are formulated as convex problems. Therefore, ADMM can be applied to solve these problems with high computational efficiency. We here describe the derivation of the decomposition algorithms for the FS and LS models using ADMM.

A. Decomposition Algorithm for FS model

For the FS model (8), we define the augmented Lagrangian function $\mathcal{L}_{\mathrm{FS}}$ as

$$\begin{aligned} &\mathcal{L}_{\mathrm{FS}}(\mathbf{G},\mathbf{S},\mathbf{\Theta}) = \|\mathbf{G}\|_F + \lambda\|\mathbf{S}\|_{1,2} \\ &+ \langle\mathbf{\Theta}, \mathbf{D}\mathbf{G} + \mathbf{W}\mathbf{S} - \mathbf{Y}\rangle + \frac{1}{2\rho}\|\mathbf{D}\mathbf{G} + \mathbf{W}\mathbf{S} - \mathbf{Y}\|_F, \end{aligned} \tag{11}$$

where $\langle\cdot,\cdot\rangle$ represents the inner product, $\mathbf{\Theta}$ is the Lagrangian multiplier, and $\rho > 0$ is a constant parameter. In ADMM, each variable is alternately updated, starting with an arbitrary initial value, as follows:

$$\begin{cases} \mathbf{G}^{(i+1)} = \underset{\mathbf{G}}{\arg\min}\, \mathcal{L}_{\mathrm{FS}}(\mathbf{G}, \mathbf{S}^{(i)}, \mathbf{\Theta}^{(i)}) \\ \mathbf{S}^{(i+1)} = \underset{\mathbf{S}}{\arg\min}\, \mathcal{L}_{\mathrm{FS}}(\mathbf{G}^{(i+1)}, \mathbf{S}, \mathbf{\Theta}^{(i)}) \\ \mathbf{\Theta}^{(i+1)} = \mathbf{\Theta}^{(i)} + \dfrac{1}{\rho}(\mathbf{D}\mathbf{G}^{(i+1)} + \mathbf{W}\mathbf{S}^{(i+1)} - \mathbf{Y}) \end{cases}, \tag{12}$$

where (i) is the iteration index. The Lagrangian function for each update is minimized for one variable while fixing the other variables, which can be efficiently computed using proximal operators [21].

The update of $\mathbf{G}$ can be derived as

$$\begin{aligned} \mathbf{G}^{(i+1)} &= \underset{\mathbf{G}}{\arg\min}\left[\|\mathbf{G}\|_F + \frac{1}{2\rho}\left\|\mathbf{P}_{\mathrm{FS}}^{(i)}(\mathbf{G}) + \rho\mathbf{\Theta}^{(i)}\right\|_F^2\right] \\ &= \mathcal{T}_{\rho/\eta_D}^F\left(\mathbf{G}^{(i)} - \frac{\rho}{\eta_D}\mathbf{D}^{\mathsf{H}}\left(\mathbf{\Theta}^{(i)} + \frac{1}{\rho}\mathbf{P}_{\mathrm{FS}}^{(i)}(\mathbf{G}^{(i)})\right)\right), \end{aligned} \tag{13}$$

where $\mathbf{P}_{\mathrm{FS}}^{(i)}(\mathbf{G}) := \mathbf{D}\mathbf{G} + \mathbf{W}\mathbf{S}^{(i)} - \mathbf{Y}$ and $\eta_D > \sigma_{\max}^2(\mathbf{D})$ is a constant. Here, $\sigma_{\max}^2(\mathbf{D})$ represents the maximum eigenvalue of $\mathbf{D}\mathbf{D}^{\mathsf{H}}$. The proximal operator $\mathcal{T}_\alpha^F(\cdot)$ is defined as

$$\mathcal{T}_\alpha^F(\mathbf{A}) := \mathbf{U}_{\mathbf{A}} \max\left\{1 - \frac{\alpha}{\|\mathrm{diag}(\mathbf{\Sigma}_{\mathbf{A}})\|_2}, 0\right\} \mathbf{\Sigma}_{\mathbf{A}}\mathbf{V}_{\mathbf{A}}^{\mathsf{H}}, \tag{14}$$

where $\mathbf{U}_{\mathbf{A}}$, $\mathbf{\Sigma}_{\mathbf{A}}$, and $\mathbf{V}_{\mathbf{A}}$ are obtained by the singular value decomposition of $\mathbf{A}$ as $\mathbf{A} = \mathbf{U}_{\mathbf{A}}\mathbf{\Sigma}_{\mathbf{A}}\mathbf{V}_{\mathbf{A}}^{\mathsf{H}}$. This update rule is obtained by linearization of the second term of the first line of (13) at $\mathbf{G}^{(i)}$ with a positive constant parameter η_D [13].

The update of $\mathbf{S}$ can be derived by a similar procedure as

$$\begin{aligned} \mathbf{S}^{(i+1)} &= \underset{\mathbf{S}}{\arg\min}\left[\lambda\|\mathbf{S}\|_{1,2} + \frac{1}{2\rho}\left\|\mathbf{Q}_{\mathrm{FS}}^{(i)}(\mathbf{S}) + \rho\mathbf{\Theta}^{(i)}\right\|_F^2\right] \\ &= \mathcal{T}_{\lambda\rho/\eta_W}^{\ell_{1,2}}\left(\mathbf{S}^{(i)} - \frac{\rho}{\eta_W}\mathbf{W}^{\mathsf{H}}\left(\mathbf{\Theta}^{(i)} + \frac{1}{\rho}\mathbf{Q}_{\mathrm{FS}}^{(i)}(\mathbf{S}^{(i)})\right)\right), \end{aligned} \tag{15}$$

where $\mathbf{Q}_{\mathrm{FS}}^{(i)}(\mathbf{S}) := \mathbf{D}\mathbf{G}^{(i+1)} + \mathbf{W}\mathbf{S} - \mathbf{Y}$ and $\eta_W > \sigma_{\max}^2(\mathbf{W})$ is a constant. The (n,t)th element of $\mathcal{T}_\beta^{\ell_{1,2}}(\cdot)$ is defined as

$$\{\mathcal{T}_\beta^{\ell_{1,2}}(\mathbf{B})\}_{n,t} := \max\{\|\mathbf{B}_{n,\cdot}\|_2 - \beta, 0\}\frac{b_{n,t}}{\|\mathbf{B}_{n,\cdot}\|_2}, \tag{16}$$

where $b_{n,t}$ is the (n,t)th element of $\mathbf{B}$. The stopping rule of (12) can be obtained from the Karush–Kuhn–Tucker (KKT) condition as [13]

$$\begin{aligned} &\|\mathbf{D}\mathbf{G}^{(i+1)} + \mathbf{W}\mathbf{S}^{(i+1)} - \mathbf{Y}\|_F / \|\mathbf{Y}\|_F \le \xi_1 \\ &\max(\sqrt{\eta_D}\|\mathbf{G}^{(i+1)} - \mathbf{G}^{(i)}\|_F, \\ &\qquad \sqrt{\eta_W}\|\mathbf{S}^{(i+1)} - \mathbf{S}^{(i)}\|_F)/\rho\|\mathbf{Y}\|_F \le \xi_2, \end{aligned}$$

where ξ_1 and ξ_2 are sufficiently small constants.

B. Decomposition Algorithm for LS Model

In a similar manner, the decomposition algorithm for the LS model based on ADMM (10) can be derived. The augmented Lagrangian function $\mathcal{L}_{\mathrm{LS}}$ for (10) is defined as

$$\begin{aligned} &\mathcal{L}_{\mathrm{LS}}(\mathbf{X},\mathbf{S},\mathbf{\Theta}) = \|\mathbf{X}\|_* + \lambda\|\mathbf{S}\|_{1,2} \\ &+ \langle\mathbf{\Theta}, \mathbf{X} + \mathbf{W}\mathbf{S} - \mathbf{Y}\rangle + \frac{1}{2\rho}\|\mathbf{X} + \mathbf{W}\mathbf{S} - \mathbf{Y}\|_F. \end{aligned} \tag{17}$$

Again, each variable is alternately updated as

$$\begin{cases} \mathbf{X}^{(i+1)} = \underset{\mathbf{X}}{\arg\min}\, \mathcal{L}_{\mathrm{LS}}(\mathbf{X}, \mathbf{S}^{(i)}, \mathbf{\Theta}^{(i)}) \\ \mathbf{S}^{(i+1)} = \underset{\mathbf{S}}{\arg\min}\, \mathcal{L}_{\mathrm{LS}}(\mathbf{X}^{(i+1)}, \mathbf{S}, \mathbf{\Theta}^{(i)}) \\ \mathbf{\Theta}^{(i+1)} = \mathbf{\Theta}^{(i)} + \dfrac{1}{\rho}(\mathbf{X}^{(i+1)} + \mathbf{W}\mathbf{S}^{(i+1)} - \mathbf{Y}). \end{cases} \tag{18}$$

ISBN 978-90-827970-1-5 © EURASIP 2018

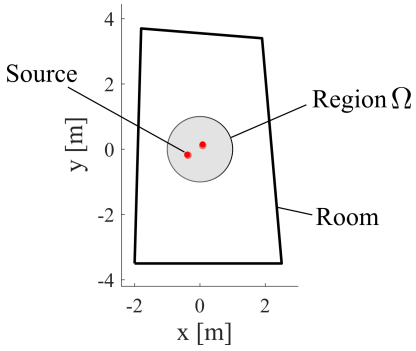

Fig. 2: Experimental setup.

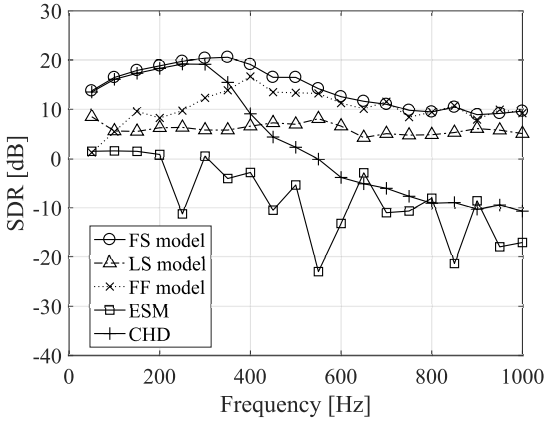

Fig. 3: Relationship between frequency and SDR for uniform arrangement case.

By applying the proximal operator to the update of $\mathbf{X}$ and $\mathbf{S}$, the following update rule can be derived:

$$\begin{aligned}\mathbf{X}^{(i+1)} &= \arg\min_{\mathbf{X}} \left[\|\mathbf{X}\|_* + \frac{1}{2\rho} \left\| \mathbf{P}_{\mathrm{LS}}^{(i)}(\mathbf{X}) + \rho\boldsymbol{\Theta}^{(i)} \right\|_F^2 \right] \\ &= \mathcal{T}_{\rho}^{\mathrm{tr}} \left(\mathbf{Y} - \mathbf{W}\mathbf{S}^{(i)} - \rho\boldsymbol{\Theta}^{(i)} \right) \end{aligned} \tag{19}$$

$$\begin{aligned}\mathbf{S}^{(i+1)} &= \arg\min_{\mathbf{S}} \left[\lambda\|\mathbf{S}\|_{1,2} + \frac{1}{2\rho} \left\| \mathbf{Q}_{\mathrm{LS}}^{(i)}(\mathbf{S}) + \rho\boldsymbol{\Theta}^{(i)} \right\|_F^2 \right] \\ &= \mathcal{T}_{\lambda\rho/\eta_W}^{\ell_{1,2}} \left(\mathbf{S}^{(i)} - \frac{\rho}{\eta_W} \mathbf{W}^{\mathsf{H}} (\boldsymbol{\Theta}^{(i)} + \frac{1}{\rho} \mathbf{Q}_{\mathrm{LS}}^{(i)}(\mathbf{S}^{(i)})) \right), \end{aligned} \tag{20}$$

where $\mathbf{P}_{\mathrm{LS}}^{(i)}(\mathbf{X}) := \mathbf{X} + \mathbf{W}\mathbf{S}^{(i)} - \mathbf{Y}$ and $\mathbf{Q}_{\mathrm{LS}}^{(i)}(\mathbf{S}) := \mathbf{X}^{(i+1)} + \mathbf{W}\mathbf{S} - \mathbf{Y}$. The operator $\mathcal{T}_{\gamma}^{\mathrm{tr}}(\cdot)$ is defined as

$$\mathcal{T}_{\gamma}^{\mathrm{tr}}(\mathbf{C}) = \mathbf{U}_{\mathbf{C}} \max\{\boldsymbol{\Sigma}_{\mathbf{C}} - \gamma\mathbf{I}, 0\} \mathbf{V}_{\mathbf{C}}^{\mathsf{H}}, \tag{21}$$

where $\mathbf{U}_{\mathbf{C}}$, $\boldsymbol{\Sigma}_{\mathbf{C}}$, and $\mathbf{V}_{\mathbf{C}}$ are obtained by the singular value decomposition of $\mathbf{C}$ as $\mathbf{C} = \mathbf{U}_{\mathbf{C}}\boldsymbol{\Sigma}_{\mathbf{C}}\mathbf{V}_{\mathbf{C}}^{\mathsf{H}}$. Again, the stopping rule of (18) can be obtained from the KKT condition as in (17).

V. Experiments

Numerical simulations were conducted to evaluate the proposed method in a 2D sound field. We compared the **FS** and **LS** models with methods based on the equivalent source method (**ESM**) [1] and circular harmonic decomposition (**CHD**) [11]. For reference, the Frobenius–Frobenius (**FF**) model, in which the Frobenius norm is imposed on $\mathbf{S}$ in (7), was also used for comparison to validate the effect of the sparsity assumption.

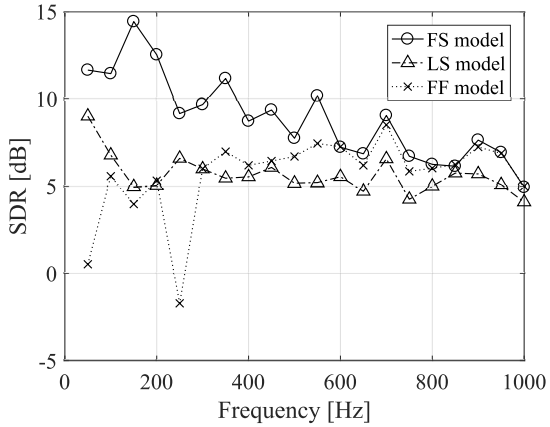

Fig. 4: Relationship between frequency and SDR for random arrangement.

The room geometry was set as shown in Fig. 2 with the absorption coefficient of the walls 0.6. The finite element method (FEM) was used to simulate reverberation in the room [22]. In FEM, the second-degree polynomial interpolation was used to obtain pressure values between the meshes. The region Ω was circular with radius $R = 1.0$ m and the coordinate origin was set at the center of Ω. Omnidirectional microphones were arranged on two layers of circular lines on $\partial\Omega$ with their centers at the origin. The radii of the inner and outer layers were 1.0 and 1.075 m, respectively. Sixteen microphones were set on each layer. Two types of microphone arrangement, uniform and random arrangements in the angular direction, were investigated. Two point sources were located at $(0.35, -0.20)$ m and $(0.10, 0.10)$ m inside Ω. The source signals were single-frequency sine waves, whose amplitudes were generated by a circularly symmetric complex Gaussian distribution. The number of time frames used for separation T was 100. Gaussian noise was also added to the observations so that the signal-to-noise ratio (SNR) became 30 dB.

To evaluate the separation accuracy, we define the signal-to-distortion ratio (SDR) as

$$\mathrm{SDR} = 10\log_{10} \frac{\|\mathbf{X}_{\mathrm{true}}\|^2}{\|\mathbf{X}_{\mathrm{true}} - \mathbf{X}_{\mathrm{est}}\|^2}, \tag{22}$$

where $\mathbf{X}_{\mathrm{true}}$ and $\mathbf{X}_{\mathrm{est}}$ are the true and estimated exterior sound field components, respectively.

The constant parameters η_D and η_W were determined as $\eta_D = 1.02\sigma_{\max}^2(\mathbf{D})$ and $\eta_W = 1.02\sigma_{\max}^2(\mathbf{W})$. The parameter ρ was adaptively changed at each iteration and had an initial value of with 4.8×10^{-3} [13]. The order N in (6) was truncated as $N = \lceil kR \rceil$, where $\lceil\cdot\rceil$ is the ceiling function [23]. Uniformly sampled plane waves from 0 to 2π rad were used for the plane-wave dictionary $\mathbf{W}$. In **FF**, the number of plane waves was set as $2\lceil kR \rceil + 1$ to make the dictionary size

ISBN 978-90-827970-1-5 © EURASIP 2018

of $\mathbf{W}$ equal to that of $\mathbf{D}$. An overcomplete dictionary was constructed for **FS** by using $8\lceil kR \rceil$ plane waves. The balancing parameter λ was determined so that the SDR becomes the highest value using the golden section search algorithm [24]. Initial values were set as matrices of all ones. The maximum number of iterations was 3000.

Figs. 3 and 4 show the SDR with respect to the frequency of the source signal for uniform and random arrangements, respectively. For the uniform arrangement, although the SDR of **CHD** was relatively high below 400 Hz, it sharply decreased as the frequency increased. The SDR of **ESM** was significantly low at all frequencies. On the other hand, **FS** achieved the highest SDR at most of the frequencies and maintained relatively high SDRs even at high frequencies. The difference between **FS** and **FF** was large at low frequencies although their SDRs were almost the same above 500 Hz. The SDR of **LS** was lower than that of **FS** and **FF** at all frequencies. For the random arrangement, the results of **ESM** and **CHD** are not shown because their SDRs were lower than –20 dB at all frequencies. On the other hand, relatively high SDRs were still maintained for the proposed method. In particular, the SDR of **FS** was higher than that of **FF** and **LS**. These results indicate that high separation accuracy and a more flexible microphone arrangement can be achieved by the proposed method, particularly **FS**.

VI. Conclusion

We proposed an exterior and interior sound field separation method using a convex optimization algorithm. Signal models for exterior and interior sound fields are formulated by exploiting the signal characteristics of each sound field. Two models are proposed for the exterior sound field: representation by harmonic functions and by low-rank signals. The interior sound field is assumed to be sparse in the plane-wave domain. The two optimization problems become convex; therefore, we derived the algorithm based on ADMM. The numerical simulation results indicated that high separation accuracy compared with that for the existing methods can be achieved by the proposed method with flexible microphone arrangement. In particular, the FS model achieved the highest SDR. Therefore, the signal models using harmonic functions and overcomplete plane-wave functions are suitable for representing exterior and interior sound fields, respectively.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP15H05312 and SECOM Science and Technology Foundation.

References

[1] C. Bi, X. Chen, and J. Chen, "Sound field separation technique based on equivalent source method and its application in nearfield acoustic holography," *J. Acoust. Soc. Am.*, vol. 123, no. 3, pp. 1472–1478, 2008.

[2] J. D. 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.

[3] 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.

[4] 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.

[5] A. Peillot, F. Ollivier, G. Chardon, and L. Daudet, "Localization and identification of sound sources using "compressive sampling" techniques," in *Proc. Int. Congr. Sound, Vibr. (ICSV)*, Rio de Janeiro, July 2011, pp. 1–8.

[6] 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. 4443–4447.

[7] G. H. Koopmann, L. Song, and J. B. Fahnline, "A method for computing acoustic fields based on the principle of wave superposition," *J. Acoust. Soc. Am.*, vol. 86, no. 5, pp. 2433–2438, 1989.

[8] L. Song, G. H. Koopmann, and J. B. Fahnline, "Numerical errors associated with the method of superposition for computing acoustic fields," *J. Acoust. Soc. Am.*, vol. 89, no. 6, pp. 2625–2633, 1991.

[9] M. E. Johnson, S. J. Elliot, K. H. Baek, and J. Garcia-Bonito, "An equivalent source technique for calculating the sound field inside an enclosure containing scattering objects," *J. Acoust. Soc. Am.*, vol. 104, no. 3, pp. 1221–1231, 1998.

[10] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*. New York: Springer, 2013.

[11] A. Fahim, P. N. Samarasinghe, and T. D. Abhayapala, "Extraction of exterior field from a mixed sound field for 2D height-invariant sound propagation," in *Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC)*, Xi'an, Sept. 2016, pp. 1–5.

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

[13] 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.

[14] 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.

[15] A. Moiola, R. Hiptmair, and I. Perugia, "Plane wave approximation of homogeneous Helmholtz solutions," *Z. Angew. Math. Phys.*, vol. 62, pp. 809–837, 2011.

[16] W. Jin and W. B. Kleijin, "Theory and design of multizone soundfield reproduction using sparse methods," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 12, pp. 2343–2355, 2015.

[17] S. Koyama and L. Daudet, "Comparison of reverberation models for sparse sound field decomposition," in *Proc. IEEE Int. Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, New Paltz, Oct. 2017, pp. 214–218.

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

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

[20] 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.

[21] 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.

[22] F. Hecht, "New development in freefem++," *J. Numer. Math.*, vol. 20, no. 3–4, pp. 251–266, 2012.

[23] D. B. Ward and T. D. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 6, pp. 697–707, 2001.

[24] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*. Cambridge: Cambridge University Press, 1988.

ISBN 978-90-827970-1-5 © EURASIP 2018