PAPER *Special Section on Acoustic Signal Processing*

Direction of Arrival Estimation Using Nonlinear Microphone Array

Hidekazu KAMIYANAGIDA[†a)], **Hiroshi SARUWATARI**[†], **Kazuya TAKEDA**[††], **Fumitada ITAKURA**[†††], *and* **Kiyohiro SHIKANO**[†], *Regular Members*

SUMMARY This paper describes a new method for estimating the direction of arrival (DOA) using a nonlinear microphone array system based on complementary beamforming. Complementary beamforming is based on two types of beamformers designed to obtain complementary directivity patterns with respect to each other. In this system, since the resultant directivity pattern is proportional to the product of these directivity patterns, the proposed method can be used to estimate DOAs of $2(K-1)$ sound sources with K-element microphone array. First, DOA-estimation experiments are performed using both computer simulation and actual devices in real acoustic environments. The results clarify that DOA estimation for two sound sources can be accomplished by the proposed method with two microphones. Also, by comparing the resolutions of DOA estimation by the proposed method and by the conventional minimum variance method, we can show that the performance of the proposed method is superior to that of the minimum variance method under all reverberant conditions.

key words: *nonlinear array signal processing, microphone array, DOA estimation, complementary beamforming*

1. Introduction

The estimation of the direction of arrival (DOA) in array signal processing is a means of determining the direction of each source signal using the information contained in mixed signals observed in multiple input channels. In acoustic signal processing, DOA estimation plays an essential part in microphone-array technology as a preprocess in speech enhancement or noise-robust speech recognition [1]–[3]. For example, if we can obtain DOA information, the target speech and interfering noise can be separated spatially by appropriate signal processing.

The minimum variance (MV) method and the multiple signal classification (MUSIC) method are the conventional and popular DOA-estimation methods currently used in microphone array systems [1], [4], [5]. In the MV method, the array output power is minimized while maintaining constant gain along the look direction. When the sound source exists in the look direction, we obtain a high array output power because of the above-mentioned constraint for the look direction. In contrast, when no sound source exists in the look direction, since the contributions from all signals in other directions are minimized, we obtain a relatively low array output power. If there is no correlation among arriving signals, this minimization can be achieved by producing the directional nulls in acoustic space. In the MUSIC method, which is known to be an eigenvector-based technique, the correlation matrix of observed array data is decomposed into signal and noise subspaces by singular value decomposition. By using the modified correlation matrix which consists of only the noise subspace, we can estimate DOAs based on the fact that the noise subspace is orthogonal to the direction vectors corresponding to the actual angles of arrival.

In general, the DOA-estimation results of these methods are accurate, especially when dealing with a small number of sound sources with many microphones. However, the performances of these methods are greatly degraded when the number of sound sources exceeds that of microphones. Thus, it is impossible to estimate DOAs using the practical microphone array system with a small number of elements.

In order to resolve this problem, we newly propose to utilize a nonlinear microphone array system based on complementary beamforming which has been proposed by one of the authors for efficient speech enhancement [8]. Complementary beamforming is based on two types of beamformers designed to obtain complementary directivity patterns with respect to each other. In this system, since the resultant directivity pattern is proportional to the product of these directivity patterns, the proposed method can be used to estimate DOAs of $2(K-1)$ sound sources with K-element microphone array. The proposed method enables the estimation of DOAs with a fairly small and practical microphone array.

First, DOA-estimation experiments are performed using both computer simulation and actual devices in real acoustic environments. The results clarify that the DOA estimation for two sound sources can be achieved by the proposed method with two microphones. Also, by comparing the resolutions of DOA estimation by the

Manuscript received August 1, 2000.
Manuscript revised November 21, 2000.
[†]The authors are with Graduate School of Information Science, Nara Institute of Science and Technology, Ikoma-shi, 630–0101 Japan.
[††]The author is with Department of Information Electronics, Graduate School of Engineering, Nagoya University, Nagoya-shi, 464–8603 Japan.
[†††]The author is with Center for Information Media Studies, Nagoya University, Nagoya-shi, 464–8603 Japan.
a) E-mail: hide-ka@is.aist-nara.ac.jp

proposed method and by the conventional MV method, we can show that the performance of the proposed method is superior to that of the MV method under all reverberant conditions.

This paper is formatted as follows. In the following section, the conventional DOA-estimation methods are described. In Sect. 3, the proposed method based on the nonlinear microphone array is described. In Sects. 4 and 5, various experiments using computer simulations and actual devices are carried out to test the performance of the proposed method. In Sect. 6, to quantify the performance, the resolutions of DOA estimation are measured and reported. Following the discussion on the results of the experiments, we give conclusions in Sect. 7.

2. Conventional DOA-Estimation Methods and Their Problems

In this section, we describe typical examples of conventional DOA-estimation methods and their problems.

In this study, a straight-line array is assumed. The coordinates of the elements are designated as $x_k (k = 1, \cdots, K)$, and the directions of arrival of multiple signals are designated as θ_d $(d = 1, \cdots, D)$ (see Fig.1). In the following, it is assumed that each sound source is located in a sufficient distance away so that the plane-wave approximation holds for the arriving signals.

The most classical and simple DOA-estimation algorithm is the approach based on the delay and sum (DS) array. In this method, we obtain the resultant array output by adding the weighted output values of each element. Thus, the array output signal is described in the frequency domain as

$$S^{(\mathrm{BF})}(f) = \boldsymbol{g}\boldsymbol{o}(f), \tag{1}$$

$$\boldsymbol{g} \equiv [g_1, \cdots, g_k, \cdots, g_K], \tag{2}$$

$$\boldsymbol{o}(f) \equiv [O_1, \cdots, O_k, \cdots, O_K]^{\mathrm{T}}, \tag{3}$$

where $S^{(\mathrm{BF})}$(f) is the array output signal, $\boldsymbol{o}(f)$ is the observed signal vector specified by the coordinates of elements x_k, and $\boldsymbol{g}$ is the weight vector of the element. The superscript $^{\mathrm{T}}$ denotes transposition of the vector. By steering the look direction to all angles, we can observe the peaks of the array output power which correspond to the DOAs of sound sources. However, the resolution of the peaks is strictly limited by the array aperture, and a huge number of elements are required to achieve high DOA-estimation performance.

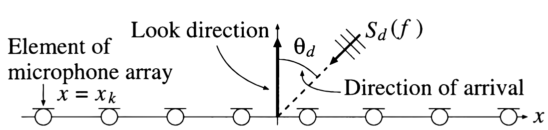


Fig. 1 Configuration of a microphone array and acoustic signals. x_k represents the coordinate of each element, and $S_d(f)$ represents the signal arriving from the direction θ_d.

In order to improve the performance, other approaches, e.g., the MV method [5] or MUSIC method [6] which are reffered to as “high resolution methods” have been proposed. The principle of the MV method is briefly described hereafter. The MV method is based on an adaptive beamformer technique in the adaptation procedure. The weight vector of element $\boldsymbol{g}$ is adapted for the signal so as to minimize the array output powers of signals arriving from beyond the look direction. This can be achieved by solving the following constrained minimization problem:

$$\min_{\boldsymbol{g}} \boldsymbol{g}\boldsymbol{R}(f)\boldsymbol{g}^{(\mathrm{H})}, \text{ subject to } \boldsymbol{g}\boldsymbol{a}_{d_0}(f) = 1, \tag{4}$$

$$\boldsymbol{R}(f) \equiv \mathrm{E}[\boldsymbol{o}(f)\boldsymbol{o}^{(\mathrm{H})}(f)], \tag{5}$$

$$\boldsymbol{a}_d(f) \equiv [a_{1,d}(f), \cdots, a_{k,d}(f), \cdots, a_{K,d}(f)]^{\mathrm{T}}, \tag{6}$$

$$a_{k,d}(f) \equiv \exp[j2\pi f \cdot x_k \cdot \sin(\theta)_d/c], \tag{7}$$

where $\boldsymbol{R}(f)$ is generally called the array correlation matrix, $\boldsymbol{g}\boldsymbol{R}(f)\boldsymbol{g}^{\mathrm{H}}$ is equal to the array output power $\mathrm{E}[|S^{(\mathrm{BF})}(f)|^2]$, and the superscript $^{\mathrm{H}}$ denotes the Hermitian transposition. Also, $a_{k,d}(f)$ corresponds to the phase difference in the signal coming from the direction of arrival θ_d at the coordinates of each element x_k; $\boldsymbol{a}_d(f)$ is generally called the steering vector, and c is the velocity of sound. In this study, d_0 corresponds to the look direction.

The solution of the constrained minimization problem given by Eq. (4) yields the optimal weight vector,

$$\boldsymbol{g}^{(\mathrm{opt})} = \frac{\boldsymbol{a}_{d_0}^{\mathrm{H}}(f)\boldsymbol{R}^{-1}(f)}{\boldsymbol{a}_{d_0}^{\mathrm{H}}(f)\boldsymbol{R}^{-1}(f)\boldsymbol{a}_{d_0}(f)}. \tag{8}$$

When the sound source exists in the look direction, we obtain a high array output power because of the constraint. In contrast, when no sound source exists in the look direction, since the contributions from all signals in other directions are minimized, we obtain a relatively low array output power. If there is no correlation among arriving signals, this minimization can be achieved by producing the directional nulls in acoustic space. However, since the maximum number of directional nulls is limited by $K-1$, the performance of the DOA estimation is greatly degraded when the number of sound sources exceeds that of microphones.

3. Proposed Algorithm

In this section, the nonlinear microphone array based on the complementary beamforming proposed in Ref. [7], [8] is briefly described, and the new DOA-estimation algorithm for this array is proposed.

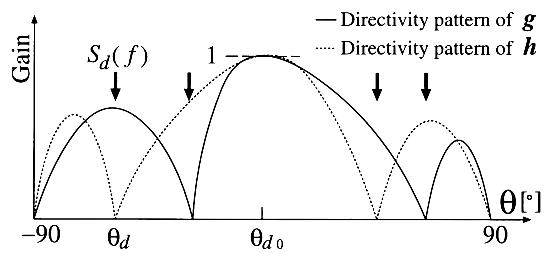


Fig. 2 Example of directivity patterns using complementary beamforming. The solid line represents the directivity pattern formed using the weight vector $\boldsymbol{g}$, and the dotted line represents the directivity pattern formed using the weight vector $\boldsymbol{h}$. The bold downnarrows indicate that the directional signal arrives from the corresponding direction θ_d.

3.1 Nonlinear Microphone Array and Its Optimization

First, using two types of complementary weight vectors of the element, $\boldsymbol{g} = [g_1, \cdots, g_K]$ and $\boldsymbol{h} = [h_1, \cdots, h_K]$ [9], we construct the signal spectra $S^{(g)}(f)$ and $S^{(h)}(f)$. The term, "complementary," implies one of the following conditions: "directivity pattern gain $|\boldsymbol{g}\boldsymbol{a}_d(f)| \gg$ directivity pattern gain $|\boldsymbol{h}\boldsymbol{a}_d(f)|$" or "directivity pattern gain $|\boldsymbol{g}\boldsymbol{a}_d(f)| \ll$ directivity pattern gain $|\boldsymbol{h}\boldsymbol{a}_d(f)|$" for an arbitrary direction. The exception is that the gain of both directivity patterns is unity with respect to the look direction (see Fig. 2).

For the signals $S^{(g)}(f)$ and $S^{(h)}(f)$, the following equations are applicable for the signals $S_d(f)$:

$$S^{(g)}(f) = \boldsymbol{g}\boldsymbol{o}(f) = \sum_{d=1}^{D} \boldsymbol{g}\boldsymbol{a}_d(f) \cdot S_d(f), \tag{9}$$

$$S^{(h)}(f) = \boldsymbol{h}\boldsymbol{o}(f) = \sum_{d=1}^{D} \boldsymbol{h}\boldsymbol{a}_d(f) \cdot S_d(f), \tag{10}$$

The sum of Eqs. (9) and (10) is defined as the primary signal, $S^{(\mathrm{p})}(f)$, and the difference is defined as the reference signal, $S^{(\mathrm{r})}(f)$. These can be given as

$$S^{(\mathrm{p})}(f) = \sum_{d=1}^{D} \{\boldsymbol{g}\boldsymbol{a}_d(f) + \boldsymbol{h}\boldsymbol{a}_d(f)\} \cdot S_d(f), \tag{11}$$

$$S^{(\mathrm{r})}(f) = \sum_{d=1}^{D} \{\boldsymbol{g}\boldsymbol{a}_d(f) - \boldsymbol{h}\boldsymbol{a}_d(f)\} \cdot S_d(f). \tag{12}$$

In the nonlinear microphone array, if there is no correlation among arriving signals, the output power spectrum is given by [7], [8]

$$\begin{aligned}
&|\hat{X}(f)|^2 \\
&= (1/4) \cdot \Big| \mathrm{E}\big[|S^{(\mathrm{p})}(f)|^2\big] - \mathrm{E}\big[|S^{(\mathrm{r})}(f)|^2\big] \Big| \\
&= (1/4) \cdot \Big| \sum_{d=1}^{D} \big\{ |\boldsymbol{g}\boldsymbol{a}_d(f) + \boldsymbol{h}\boldsymbol{a}_d(f)|^2 \\
&\quad - |\boldsymbol{g}\boldsymbol{a}_d(f) - \boldsymbol{h}\boldsymbol{a}_d(f)|^2 \big\} \cdot \mathrm{E}\big[|S_d(f)|^2\big] \Big| \\
&= (1/4) \cdot \Big| \sum_{d=1}^{D} 2\big\{ \boldsymbol{g}\boldsymbol{a}_d(f) \cdot (\boldsymbol{h}\boldsymbol{a}_d(f))^* \\
&\quad + (\boldsymbol{g}\boldsymbol{a}_d(f))^* \cdot \boldsymbol{h}\boldsymbol{a}_d(f) \big\} \cdot \mathrm{E}\big[|S_d(f)|^2\big] \Big| \\
&= \Big| \sum_{d=1}^{D} \mathrm{Re}\big[\boldsymbol{g}\boldsymbol{a}_d(f) \cdot (\boldsymbol{h}\boldsymbol{a}_d(f))^*\big] \cdot \mathrm{E}\big[|S_d(f)|^2\big] \Big|,
\end{aligned} \tag{13}$$

where $\mathrm{Re}[\cdot]$ represents the real part of the complex-valued argument and the superscript * denotes a complex conjugate.

Equation (13) indicates that the directional gain for signal $S_d(f)$ is $\mathrm{Re}\big[\boldsymbol{g}\boldsymbol{a}_d(f) \cdot (\boldsymbol{h}\boldsymbol{a}_d(f))^*\big]$, and, in geneal, this directional gain must be minimized for each direction except for the look direction. Accordingly, to reduce the signal component in Eq. (13), it is not necessary to produce small $|\boldsymbol{g}\boldsymbol{a}_d(f)|$ and $|\boldsymbol{h}\boldsymbol{a}_d(f)|$ individually, but to design them so as to obtain small $\mathrm{Re}\big[\boldsymbol{g}\boldsymbol{a}_d(f) \cdot (\boldsymbol{h}\boldsymbol{a}_d(f))^*\big]$ in the directivity patterns. Taking advantage of this complementary characteristic, even when the directional nulls of $K-1$ directions are produced in each directivity pattern $|\boldsymbol{g}\boldsymbol{a}_d(f)|$ and $|\boldsymbol{h}\boldsymbol{a}_d(f)|$, we can realize the directional nulls of $2(K-1)$ directions in the directivity pattern of the proposed array.

An adaptation for each signal direction can be achieved so as to minimize the array output $|\hat{X}(f)|^2$ by changing the weight vectors $\boldsymbol{g}$ and $\boldsymbol{h}$. However, since $\mathrm{Re}[\boldsymbol{g}\boldsymbol{a}_d(f) \cdot (\boldsymbol{h}\boldsymbol{a}_d(f))^*]$ may be both positive and negative in Eq. (13), the minimization of the array output $|\hat{X}(f)|^2$ is not a sufficient condition to minimize $\mathrm{Re}[\boldsymbol{g}\boldsymbol{a}_d(f) \cdot (\boldsymbol{h}\boldsymbol{a}_d(f))^*]$ of each signal component. For example, in the case that only two signals $S_1(f)$ and $S_2(f)$ arrive, the arbitrary directivity patterns which satisfy the following equation,

$$\begin{aligned}
&\mathrm{Re}[\boldsymbol{g}\boldsymbol{a}_1(f) \cdot (\boldsymbol{h}\boldsymbol{a}_1(f))^*] \cdot \mathrm{E}\big[|S_1(f)|^2\big] \\
&\quad = -\mathrm{Re}[\boldsymbol{g}\boldsymbol{a}_2(f) \cdot (\boldsymbol{h}\boldsymbol{a}_2(f))^*] \cdot \mathrm{E}\big[|S_2(f)|^2\big],
\end{aligned} \tag{14}$$

are always able to minimize the array output $|\hat{X}(f)|^2$.

To avoid this arbitrariness, an improved optimizing algorithm based on a block-averaged power spectrum, as described below, has been proposed [8].

First, power spectra $|S^{(\mathrm{p})}(f)|^2$ and $|S^{(\mathrm{r})}(f)|^2$ are calculated frame-by-frame along the time axis, and we obtain the approximated expectation value of each power spectrum in Eq. (13) by averaging the power spectra of each signal over several frames. A set of the frames used in the averaging process is regarded as a *block* in the time axis and this interframe-averaged power spectrum is designed to be $\langle |S^{(\mathrm{p})}(f)|^2 \rangle_b$ or $\langle |S^{(\mathrm{r})}(f)|^2 \rangle_b$, where the subscript b indicates that the interframe-averaged power spectra are obtained in the b th block.

Next, applying this block-averaging technique to Eq. (13), we calculate the array output in the b th block,

$\left(|\hat{X}(f)|_b^2\right)^2$, over some blocks ($b = 1, \ldots, B$), and define the squared sum of $\left(|\hat{X}(f)|_b^2\right)^2$ as a new minimized criterion. By assuming that the correlation among the signals is negligible in the averaged power spectra of each block, we can approximate the criterion as

$$\sum_{b=1}^{B}\left(|\hat{X}(f)|_b^2\right)^2$$
$$\equiv \sum_{b=1}^{B}(1/4)^2 \cdot \left|\langle|S^{(\mathrm{p})}(f)|^2\rangle_b - \langle|S^{(\mathrm{r})}(f)|^2\rangle_b\right|^2$$
$$\approx \sum_{b=1}^{B}\left|\sum_{d=1}^{D}\mathrm{Re}\left[\boldsymbol{g}\boldsymbol{a}_d(f)\cdot(\boldsymbol{h}\boldsymbol{a}_d(f))^*\right]\cdot\langle|S_d(f)|^2\rangle_b\right|^2. \tag{15}$$

If the mean power of each signal changes independently every block, Eq. (15) is minimized only when all directivity patterns $\mathrm{Re}[\boldsymbol{g}\boldsymbol{a}_d(f)\cdot(\boldsymbol{h}\boldsymbol{a}_d(f))^*]$ of each signal direction d are set to be zero. Thus, we can realize the directional nulls for each signal by minimizing $\Sigma_{b=1}^{B}\left(|\hat{X}(f)|_b^2\right)^2$ with respect to $\boldsymbol{g}$ and $\boldsymbol{h}$.

More practically, $|\hat{X}(f)|_b^2$ is calculated using the observation signal vector $\boldsymbol{o}(f)$, i.e., this can be given as

$$|\hat{X}(f)|_b^2 = (1/4)\cdot\big|(\boldsymbol{g}+\boldsymbol{h})\boldsymbol{R}_b(f)(\boldsymbol{g}+\boldsymbol{h})^{\mathrm{H}} - (\boldsymbol{g}-\boldsymbol{h})\boldsymbol{R}_b(f)(\boldsymbol{g}-\boldsymbol{h})^{\mathrm{H}}\big|, \tag{16}$$

$$\boldsymbol{R}_b(f) \equiv \langle\boldsymbol{o}(f)\boldsymbol{o}^{\mathrm{H}}(f)\rangle_b. \tag{17}$$

Based on Eqs. (15) and (16), the following constrained minimization problem is solved.

$$\min_{\boldsymbol{g},\boldsymbol{h}}\sum_{b=1}^{B}\big|(\boldsymbol{g}+\boldsymbol{h})\boldsymbol{R}_b(f)(\boldsymbol{g}+\boldsymbol{h})^{\mathrm{H}} - (\boldsymbol{g}-\boldsymbol{h})\boldsymbol{R}_b(f)(\boldsymbol{g}-\boldsymbol{h})^{\mathrm{H}}\big|^2 \tag{18}$$

$$\text{subject to}\,\boldsymbol{g}\boldsymbol{a}_{d_0}(f)=\boldsymbol{h}\boldsymbol{a}_{d_0}(f)=1 \tag{19}$$

Equation (19) is the constraint in which the gain of both directivity patterns is unity with respect to the look direction. Figure 3 shows the block diagram of this adaptation procedure.

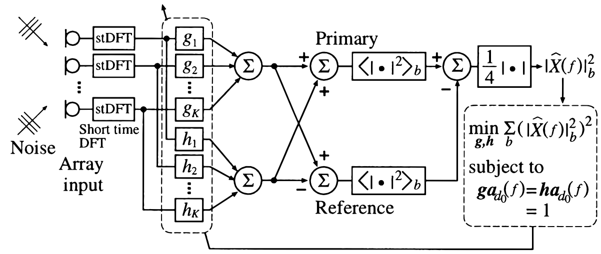


Fig. 3 Block diagram of adaptation procedure in nonlinear microphone array with complementary beamforming.

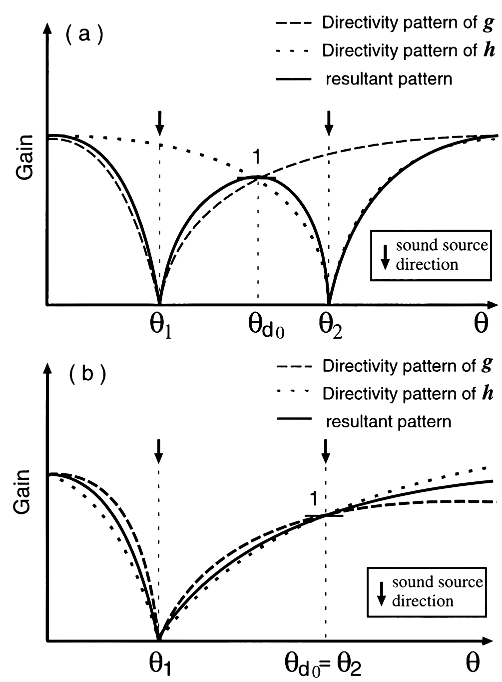


Fig. 4 Example of directivity patterns when (a) sound source does not exist in the look direction θ_{d_0}, and (b) sound source exists in the look direction θ_{d_0}. The dashed line represents the directivity pattern of $\boldsymbol{g}$, the dotted line represents the directivity pattern of $\boldsymbol{h}$, and the solid line represents the resultant pattern produced by the proposed method. The bold downnarrows indicate that the directional signals arrive from the corresponding directions, θ_1 and θ_2.

3.2 DOA Estimation by Nonlinear Microphone Array

In solving the constrained minimization problem with respect to each look direction θ_{d_0}, we can classify the behavior of the array output into the following four cases.

1. In the case that each signal changes dependently every block,
   - 1(a) when the sound source does not exist in θ_{d_0}, since $\mathrm{Re}[\boldsymbol{g}\boldsymbol{a}_d(f)\cdot(\boldsymbol{h}\boldsymbol{a}_d(f))^*]$ may be both positive and negative, $\boldsymbol{g}$ and $\boldsymbol{h}$ are adjusted so as to cancel out $\mathrm{Re}[\boldsymbol{g}\boldsymbol{a}_d(f)\cdot(\boldsymbol{h}\boldsymbol{a}_d(f))^*]\cdot\langle|S_d(f)|^2\rangle_b$ for all θ_d, or produce directional nulls to reduce the gains for the DOAs of sound sources. Thus, the array output becomes quite small.
   - 1(b) when the sound source exists in θ_{d_0}, the array output becomes quite small for the same reason as described in 1(a).
2. In the case that each signal changes independently every block,
   - 2(a) when the sound source does not exist in θ_{d_0}, $\boldsymbol{g}$ and $\boldsymbol{h}$ are able to produce directional nulls which are steered to DOAs of sound sources

complementarily to each other (see Fig. 4(a)). Thus, the array output becomes quite small.

2(b) when the sound source exists in θ_{d_0}, since the improved optimizing algorithm based on a block-averaged power spectrum is used, the cancellation which is described in case 1(a) is restricted. Thus, the resultant directivity patterns are produced as depicted in Fig. 4(b), and Eq. (15) yields a relatively large value.

Based on the above example, since Eq. (15) is extremely minimized in every angle, we cannot detect the existence of signals if the signals change every block dependently of each other. However, we can detect the existence of signals by steering the look direction θ_{d_0} to every direction and minimizing Eq. (15) if the signals change every block independently of each other, i.e., cases 2(a) and 2(b). Therefore, the DOA of $2(K-1)$ signals can be estimated using only K microphones in this case.

3.3 Optimal Solution of Constrained Minimization Problem

The constrained minimization problem given by Eqs. (18) and (19) can be solved using the method of Lagrange multipliers [4]. The Lagrangian is generally defined as a scalar-valued function which consists of the objective function to be minimized and the constraint function to be equal to 0. In this study, the Lagrangian L is given by

$$\begin{aligned} L=\sum_{b=1}^{B} E_b^2 + \boldsymbol{\lambda}[&\boldsymbol{g}^{\mathrm{R}}\boldsymbol{a}_{d_0}^{\mathrm{R}} - \boldsymbol{g}^{\mathrm{I}}\boldsymbol{a}_{d_0}^{\mathrm{I}} - 1,\ \boldsymbol{g}^{\mathrm{I}}\boldsymbol{a}_{d_0}^{\mathrm{R}} + \boldsymbol{g}^{\mathrm{R}}\boldsymbol{a}_{d_0}^{\mathrm{I}}, \\ &\boldsymbol{h}^{\mathrm{R}}\boldsymbol{a}_{d_0}^{\mathrm{R}} - \boldsymbol{h}^{\mathrm{I}}\boldsymbol{a}_{d_0}^{\mathrm{I}} - 1,\ \boldsymbol{h}^{\mathrm{I}}\boldsymbol{a}_{d_0}^{\mathrm{R}} + \boldsymbol{h}^{\mathrm{R}}\boldsymbol{a}_{d_0}^{\mathrm{I}}], \end{aligned} \tag{20}$$

$$\begin{aligned} E_b \equiv\ &(\boldsymbol{g}+\boldsymbol{h})\boldsymbol{R}_b(f)(\boldsymbol{g}+\boldsymbol{h})^{\mathrm{H}} \\ &- (\boldsymbol{g}-\boldsymbol{h})\boldsymbol{R}_b(f)(\boldsymbol{g}-\boldsymbol{h})^{\mathrm{H}}, \end{aligned} \tag{21}$$

where $\boldsymbol{a}_{d_0}^{\mathrm{R}}$ and $\boldsymbol{a}_{d_0}^{\mathrm{I}}$ are defined as the real and imaginary parts of $\boldsymbol{a}_{d_0}$, $\boldsymbol{\lambda} = [\lambda_1,\ \lambda_2,\ \lambda_3,\ \lambda_4]$ is the real-valued Lagrange multiplier vector, $\boldsymbol{g}^{\mathrm{R}}$ and $\boldsymbol{h}^{\mathrm{R}}$ are defined as the real parts of $\boldsymbol{g}$ and $\boldsymbol{h}$, and $\boldsymbol{g}^{\mathrm{I}}$ and $\boldsymbol{h}^{\mathrm{I}}$ are defined as their imaginary parts; they can be given as

$$\boldsymbol{g} = \boldsymbol{g}^{\mathrm{R}} + j\boldsymbol{g}^{\mathrm{I}} = [g_1^{\mathrm{R}}, \cdots, g_K^{\mathrm{R}}] + j[g_1^{\mathrm{I}}, \cdots, g_K^{\mathrm{I}}], \tag{22}$$

$$\boldsymbol{h} = \boldsymbol{h}^{\mathrm{R}} + j\boldsymbol{h}^{\mathrm{I}} = [h_1^{\mathrm{R}}, \cdots, h_K^{\mathrm{R}}] + j[h_1^{\mathrm{I}}, \cdots, h_K^{\mathrm{I}}]. \tag{23}$$

In this method, it is known that the stationary point of the Lagrangian L is the solution of Eqs. (18) and (19) if the stationary point gives the local minimizer of L. Thus, we can potentially obtain the optimal $\boldsymbol{g}$ and $\boldsymbol{h}$ to solve the following simultaneous equations:

$$\begin{aligned} \nabla_{g^{\mathrm{R}}} L =\ &4\sum_{b=1}^{B} E_b\big(\boldsymbol{R}_b(f)\boldsymbol{g}^{\mathrm{H}} + \boldsymbol{R}_b^{\mathrm{T}}(f)\boldsymbol{g}^{\mathrm{T}}\big) \\ &+ \lambda_1\boldsymbol{a}_{d_0}^{\mathrm{R}}(f) + \lambda_2\boldsymbol{a}_{d_0}^{\mathrm{I}}(f) \\ =\ &0 \end{aligned} \tag{24}$$

$$\begin{aligned} \nabla_{g^{\mathrm{I}}} L =\ &4j\sum_{b=1}^{B} E_b\big(\boldsymbol{R}_b(f)\boldsymbol{g}^{\mathrm{H}} + \boldsymbol{R}_b^{\mathrm{T}}(f)\boldsymbol{g}^{\mathrm{T}}\big) \\ &- \lambda_1\boldsymbol{a}_{d_0}^{\mathrm{I}}(f) + \lambda_2\boldsymbol{a}_{d_0}^{\mathrm{R}}(f) \\ =\ &0 \end{aligned} \tag{25}$$

$$\begin{aligned} \nabla_{h^{\mathrm{R}}} L =\ &4\sum_{b=1}^{B} E_b\big(\boldsymbol{R}_b(f)\boldsymbol{g}^{\mathrm{H}} + \boldsymbol{R}_b^{\mathrm{T}}(f)\boldsymbol{g}^{\mathrm{T}}\big) \\ &+ \lambda_3\boldsymbol{a}_{d_0}^{\mathrm{R}}(f) + \lambda_4\boldsymbol{a}_{d_0}^{\mathrm{I}}(f) \\ =\ &0 \end{aligned} \tag{26}$$

$$\begin{aligned} \nabla_{h^{\mathrm{I}}} L =\ &4j\sum_{b=1}^{B} E_b\big(\boldsymbol{R}_b(f)\boldsymbol{g}^{\mathrm{H}} + \boldsymbol{R}_b^{\mathrm{T}}(f)\boldsymbol{g}^{\mathrm{T}}\big) \\ &- \lambda_3\boldsymbol{a}_{d_0}^{\mathrm{I}}(f) + \lambda_4\boldsymbol{a}_{d_0}^{\mathrm{R}}(f) \\ =\ &0 \end{aligned} \tag{27}$$

$$\begin{aligned} \boldsymbol{g}^{\mathrm{R}}\boldsymbol{a}_{d_0}^{\mathrm{R}} - \boldsymbol{g}^{\mathrm{I}}\boldsymbol{a}_{d_0}^{\mathrm{I}} - 1 &= \boldsymbol{g}^{\mathrm{I}}\boldsymbol{a}_{d_0}^{\mathrm{R}} + \boldsymbol{g}^{\mathrm{R}}\boldsymbol{a}_{d_0}^{\mathrm{I}} \\ &= \boldsymbol{h}^{\mathrm{R}}\boldsymbol{a}_{d_0}^{\mathrm{R}} - \boldsymbol{h}^{\mathrm{I}}\boldsymbol{a}_{d_0}^{\mathrm{I}} - 1 \\ &= \boldsymbol{h}^{\mathrm{I}}\boldsymbol{a}_{d_0}^{\mathrm{R}} + \boldsymbol{h}^{\mathrm{R}}\boldsymbol{a}_{d_0}^{\mathrm{I}} \\ &= 0, \end{aligned} \tag{28}$$

where ∇_x ($x = \{g^{\mathrm{R}}, g^{\mathrm{I}}, h^{\mathrm{R}}, h^{\mathrm{I}}\}$) represents the following gradient operator with respect to the variable x.

$$\nabla_{g^{\mathrm{R}}} \equiv \big[\partial/\partial g_1^{\mathrm{R}},\ \partial/\partial g_2^{\mathrm{R}},\ \cdots,\ \partial/\partial g_K^{\mathrm{R}}\big]^{\mathrm{T}} \tag{29}$$

$$\nabla_{g^{\mathrm{I}}} \equiv \big[\partial/\partial g_1^{\mathrm{I}},\ \partial/\partial g_2^{\mathrm{I}},\ \cdots,\ \partial/\partial g_K^{\mathrm{I}}\big]^{\mathrm{T}} \tag{30}$$

$$\nabla_{h^{\mathrm{R}}} \equiv \big[\partial/\partial h_1^{\mathrm{R}},\ \partial/\partial h_2^{\mathrm{R}},\ \cdots,\ \partial/\partial h_K^{\mathrm{R}}\big]^{\mathrm{T}} \tag{31}$$

$$\nabla_{h^{\mathrm{I}}} \equiv \big[\partial/\partial h_1^{\mathrm{I}},\ \partial/\partial h_2^{\mathrm{I}},\ \cdots,\ \partial/\partial h_K^{\mathrm{I}}\big]^{\mathrm{T}} \tag{32}$$

Since Eqs. (24)–(28) are nonlinear simultaneous equations, they can be solved using an iterative method; Newton's method [10] is used in this work. To apply Newton's method to the solution of Eqs. (24)–(28), we define the Hessian matrix of L, the matrix of all the second partial derivatives of L, in the following block matrix form:

$$\boldsymbol{H} \equiv \begin{bmatrix} \boldsymbol{H}_{g^{\mathrm{R}},g^{\mathrm{R}}} & \boldsymbol{H}_{g^{\mathrm{R}},g^{\mathrm{I}}} & \boldsymbol{H}_{g^{\mathrm{R}},h^{\mathrm{R}}} & \boldsymbol{H}_{g^{\mathrm{R}},h^{\mathrm{I}}} & \boldsymbol{H}_{g^{\mathrm{R}},\lambda} \\ \boldsymbol{H}_{g^{\mathrm{I}},g^{\mathrm{R}}} & \boldsymbol{H}_{g^{\mathrm{I}},g^{\mathrm{I}}} & \boldsymbol{H}_{g^{\mathrm{I}},h^{\mathrm{R}}} & \boldsymbol{H}_{g^{\mathrm{I}},h^{\mathrm{I}}} & \boldsymbol{H}_{g^{\mathrm{I}},\lambda} \\ \boldsymbol{H}_{h^{\mathrm{R}},g^{\mathrm{R}}} & \boldsymbol{H}_{h^{\mathrm{R}},g^{\mathrm{I}}} & \boldsymbol{H}_{h^{\mathrm{R}},h^{\mathrm{R}}} & \boldsymbol{H}_{h^{\mathrm{R}},h^{\mathrm{I}}} & \boldsymbol{H}_{h^{\mathrm{R}},\lambda} \\ \boldsymbol{H}_{h^{\mathrm{I}},g^{\mathrm{R}}} & \boldsymbol{H}_{h^{\mathrm{I}},g^{\mathrm{I}}} & \boldsymbol{H}_{h^{\mathrm{I}},h^{\mathrm{R}}} & \boldsymbol{H}_{h^{\mathrm{I}},h^{\mathrm{I}}} & \boldsymbol{H}_{h^{\mathrm{I}},\lambda} \\ \boldsymbol{H}_{\lambda,g^{\mathrm{R}}} & \boldsymbol{H}_{\lambda,g^{\mathrm{I}}} & \boldsymbol{H}_{\lambda,h^{\mathrm{R}}} & \boldsymbol{H}_{\lambda,h^{\mathrm{I}}} & \boldsymbol{H}_{\lambda,\lambda} \end{bmatrix}, \tag{33}$$

where the block Hessian matrices $\boldsymbol{H}_{x,y}$ ($x, y = \{g^{\mathrm{R}}, g^{\mathrm{I}}, h^{\mathrm{R}}, h^{\mathrm{I}}, \lambda\}$) are defined as

$$\boldsymbol{H}_{x,y} \equiv \left[\frac{\partial}{\partial y_1}\nabla_x L,\ \cdots,\ \frac{\partial}{\partial y_K}\nabla_x L\right], \tag{34}$$

(for λ, we define $\nabla_\lambda \equiv [\partial/\partial\lambda_1, \partial/\partial\lambda_2, \partial/\partial\lambda_3, \partial/\partial\lambda_4]^{\mathrm{T}}$). Each block Hessian matrix in the upper triangular part of $\boldsymbol{H}$ is given as

$$\boldsymbol{H}_{g^{\mathrm{R}},g^{\mathrm{R}}} = 8\sum_{b=1}^{B}\left(\boldsymbol{R}_b(f)\boldsymbol{h}^{\mathrm{H}} + \boldsymbol{R}_b^{\mathrm{T}}(f)\boldsymbol{h}^{\mathrm{T}}\right) \cdot\left(\boldsymbol{R}_b(f)\boldsymbol{h}^{\mathrm{H}} + \boldsymbol{R}_b^{\mathrm{T}}(f)\boldsymbol{h}^{\mathrm{T}}\right)^{\mathrm{T}} \tag{35}$$

$$\boldsymbol{H}_{g^{\mathrm{R}},g^{\mathrm{I}}} = 8j\sum_{b=1}^{B}\left(\boldsymbol{R}_b(f)\boldsymbol{h}^{\mathrm{H}} + \boldsymbol{R}_b^{\mathrm{T}}(f)\boldsymbol{h}^{\mathrm{T}}\right) \cdot\left(\boldsymbol{R}_b(f)\boldsymbol{h}^{\mathrm{H}} - \boldsymbol{R}_b^{\mathrm{T}}(f)\boldsymbol{h}^{\mathrm{T}}\right)^{\mathrm{T}} \tag{36}$$

$$\boldsymbol{H}_{g^{\mathrm{R}},h^{\mathrm{R}}} = 8\sum_{b=1}^{B}\left(\boldsymbol{R}_b(f)\boldsymbol{h}^{\mathrm{H}} + \boldsymbol{R}_b^{\mathrm{T}}(f)\boldsymbol{h}^{\mathrm{T}}\right) \cdot\left(\boldsymbol{R}_b(f)\boldsymbol{g}^{\mathrm{H}} + \boldsymbol{R}_b^{\mathrm{T}}(f)\boldsymbol{g}^{\mathrm{T}}\right)^{\mathrm{T}} + 4\sum_{b=1}^{B}\left(\boldsymbol{R}_b(f) + \boldsymbol{R}_b^{\mathrm{T}}(f)\right)E_b \tag{37}$$

$$\boldsymbol{H}_{g^{\mathrm{R}},h^{\mathrm{I}}} = 8j\sum_{b=1}^{B}\left(\boldsymbol{R}_b(f)\boldsymbol{h}^{\mathrm{H}} + \boldsymbol{R}_b^{\mathrm{T}}(f)\boldsymbol{h}^{\mathrm{T}}\right) \cdot\left(\boldsymbol{R}_b(f)\boldsymbol{g}^{\mathrm{H}} - \boldsymbol{R}_b^{\mathrm{T}}(f)\boldsymbol{g}^{\mathrm{T}}\right)^{\mathrm{T}} + 4j\sum_{b=1}^{B}\left(\boldsymbol{R}_b^{\mathrm{T}}(f) - \boldsymbol{R}_b(f)\right)E_b \tag{38}$$

$$\boldsymbol{H}_{g^{\mathrm{R}},\lambda} = [\boldsymbol{a}_{d_0}^{\mathrm{R}}(f), \boldsymbol{a}_{d_0}^{\mathrm{I}}(f), \boldsymbol{0}, \boldsymbol{0}] \tag{39}$$

$$\boldsymbol{H}_{g^{\mathrm{I}},g^{\mathrm{I}}} = -8\sum_{b=1}^{B}\left(\boldsymbol{R}_b(f)\boldsymbol{h}^{\mathrm{H}} - \boldsymbol{R}_b^{\mathrm{T}}(f)\boldsymbol{h}^{\mathrm{T}}\right) \cdot\left(\boldsymbol{R}_b(f)\boldsymbol{h}^{\mathrm{H}} - \boldsymbol{R}_b^{\mathrm{T}}(f)\boldsymbol{h}^{\mathrm{T}}\right)^{\mathrm{T}} \tag{40}$$

$$\boldsymbol{H}_{g^{\mathrm{I}},h^{\mathrm{R}}} = 8j\sum_{b=1}^{B}\left(\boldsymbol{R}_b(f)\boldsymbol{h}^{\mathrm{H}} - \boldsymbol{R}_b^{\mathrm{T}}(f)\boldsymbol{h}^{\mathrm{T}}\right) \cdot\left(\boldsymbol{R}_b(f)\boldsymbol{g}^{\mathrm{H}} + \boldsymbol{R}_b^{\mathrm{T}}(f)\boldsymbol{g}^{\mathrm{T}}\right)^{\mathrm{T}} + 4j\sum_{b=1}^{B}\left(\boldsymbol{R}_b(f) - \boldsymbol{R}_b^{\mathrm{T}}(f)\right)E_b \tag{41}$$

$$\boldsymbol{H}_{g^{\mathrm{I}},h^{\mathrm{I}}} = -8\sum_{b=1}^{B}\left(\boldsymbol{R}_b(f)\boldsymbol{h}^{\mathrm{H}} - \boldsymbol{R}_b^{\mathrm{T}}(f)\boldsymbol{h}^{\mathrm{T}}\right) \cdot\left(\boldsymbol{R}_b(f)\boldsymbol{g}^{\mathrm{H}} - \boldsymbol{R}_b^{\mathrm{T}}(f)\boldsymbol{g}^{\mathrm{T}}\right)^{\mathrm{T}} + 4\sum_{b=1}^{B}\left(\boldsymbol{R}_b(f) + \boldsymbol{R}_b^{\mathrm{T}}(f)\right)E_b \tag{42}$$

$$\boldsymbol{H}_{g^{\mathrm{I}},\lambda} = [-\boldsymbol{a}_{d_0}^{\mathrm{I}}(f), \boldsymbol{a}_{d_0}^{\mathrm{R}}(f), \boldsymbol{0}, \boldsymbol{0}] \tag{43}$$

$$\boldsymbol{H}_{h^{\mathrm{R}},h^{\mathrm{R}}} = 8\sum_{b=1}^{B}\left(\boldsymbol{R}_b(f)\boldsymbol{g}^{\mathrm{H}} + \boldsymbol{R}_b^{\mathrm{T}}(f)\boldsymbol{g}^{\mathrm{T}}\right) \cdot\left(\boldsymbol{R}_b(f)\boldsymbol{g}^{\mathrm{H}} + \boldsymbol{R}_b^{\mathrm{T}}(f)\boldsymbol{g}^{\mathrm{T}}\right)^{\mathrm{T}} \tag{44}$$

$$\boldsymbol{H}_{h^{\mathrm{R}},h^{\mathrm{I}}} = 8j\sum_{b=1}^{B}\left(\boldsymbol{R}_b(f)\boldsymbol{g}^{\mathrm{H}} + \boldsymbol{R}_b^{\mathrm{T}}(f)\boldsymbol{g}^{\mathrm{T}}\right) \cdot\left(\boldsymbol{R}_b(f)\boldsymbol{g}^{\mathrm{H}} - \boldsymbol{R}_b^{\mathrm{T}}(f)\boldsymbol{g}^{\mathrm{T}}\right)^{\mathrm{T}} \tag{45}$$

$$\boldsymbol{H}_{h^{\mathrm{R}},\lambda} = [\boldsymbol{0}, \boldsymbol{0}, \boldsymbol{a}_{d_0}^{\mathrm{R}}(f), \boldsymbol{a}_{d_0}^{\mathrm{I}}(f)] \tag{46}$$

$$\boldsymbol{H}_{h^{\mathrm{I}},h^{\mathrm{I}}} = -8\sum_{b=1}^{B}\left(\boldsymbol{R}_b(f)\boldsymbol{g}^{\mathrm{H}} - \boldsymbol{R}_b^{\mathrm{T}}(f)\boldsymbol{g}^{\mathrm{T}}\right) \cdot\left(\boldsymbol{R}_b(f)\boldsymbol{g}^{\mathrm{H}} - \boldsymbol{R}_b^{\mathrm{T}}(f)\boldsymbol{g}^{\mathrm{T}}\right)^{\mathrm{T}} \tag{47}$$

$$\boldsymbol{H}_{h^{\mathrm{I}},\lambda} = [\boldsymbol{0}, \boldsymbol{0}, -\boldsymbol{a}_{d_0}^{\mathrm{I}}(f), \boldsymbol{a}_{d_0}^{\mathrm{R}}(f)] \tag{48}$$

$$\boldsymbol{H}_{\lambda,\lambda} = [\boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0}], \tag{49}$$

where $\boldsymbol{0} \equiv [0, \cdots, 0]^{\mathrm{T}}$. The remaining block matrices, in the lower triangular part of $\boldsymbol{H}$, can be easily calculated by transposing the upper triangular part. Using the Hessian matrix $\boldsymbol{H}$ given by Eqs. (33) and (35)–(49), we can apply Newton's method to the solution of Eqs. (24)–(28), i.e., $\mathrm{col}[\nabla_{g^{\mathrm{R}}}, \nabla_{g^{\mathrm{I}}}, \nabla_{h^{\mathrm{R}}}, \nabla_{h^{\mathrm{I}}}, \nabla_\lambda]L = [0, \cdots, 0]^{\mathrm{T}}$ where $\mathrm{col}[\cdot]$ denotes the column vector. By this method, we obtain the weight vectors of the element in the $(i+1)$ th step as

$$\begin{aligned}&\left[\boldsymbol{g}^{\mathrm{R}}, \boldsymbol{g}^{\mathrm{I}}, \boldsymbol{h}^{\mathrm{R}}, \boldsymbol{h}^{\mathrm{I}}, \boldsymbol{\lambda}\right]_{i+1}^{\mathrm{T}} \\ &= \left[\boldsymbol{g}^{\mathrm{R}}, \boldsymbol{g}^{\mathrm{I}}, \boldsymbol{h}^{\mathrm{R}}, \boldsymbol{h}^{\mathrm{I}}, \boldsymbol{\lambda}\right]_{i}^{\mathrm{T}} \\ &\quad - \alpha \cdot \boldsymbol{H}_i^{-1}\mathrm{col}\left[\nabla_{g^{\mathrm{R}}}, \nabla_{g^{\mathrm{I}}}, \nabla_{h^{\mathrm{R}}}, \nabla_{h^{\mathrm{I}}}, \nabla_\lambda\right]L_i,\end{aligned} \tag{50}$$

where the subscript i is used to explicitly express the value of the ith step in the iterations, and α ($0 < \alpha \leq 1$) is the step size parameter for iterations. Since the solutions led by Eq. (50) are inherently capable of giving both the local minimizer and the local maximizer of L, we should confirm whether or not L decreased following the iterations. If L did not decrease, we must repeat the iteration procedure to search for the local minimizer of L, changing the initial values of $\boldsymbol{g}$ and $\boldsymbol{h}$.

4. Experiments of DOA Estimation by Computer Simulation

In this section, DOA-estimation experiments are performed using a computer simulation to examine the applicability of the proposed method. The aim of this section is to show that DOAs of $2(K-1)$ sound sources can be estimated by the proposed method with K microphones, especially when confronted with two sound sources, i.e., $K = 2$. Also, we show the relation between the performance of DOA estimation and the degree of nonstationariness of each sound source.

4.1 Experimental Conditions

All sound data prepared in all experiments were sampled at 12 kHz with 16-bit resolution. A two-element array with the interelement spacing of 4 cm is assumed. Sound source signals are assumed to arrive from two directions, $-40°$ and $30°$. As described above, this experimental task corresponds to the case when the number of sound sources is equal to that of microphones. In this experiment, we assume the condition that there is no background noise.

Sound samples of 2 sec duration are used to perform the DOA estimation. In the proposed method, the optimization procedure shown in Eq. (50) is conducted under the following conditions: the frame length is 21.3 msec, the frame shift is half the frame length, the window function is rectangular, the block size is 50 frame lengths, and the block shift is 1 frame length. The step size parameter α for iterations in Newton's method is set to be 0.1.

We carry out the DOA-estimation procedure for each frequency independently. Hereafter we call the above-mentioned frequency as *analysis frequency*. In this experiment, the analysis frequencies are set to be 0.5, 1.0 and 2.0 kHz. We calculate and plot the resultant estimated DOA every $1°$ angle.

4.2 Human Speech-Like Noise

To evaluate the DOA-estimation ability of the proposed method for nonstationary signal, we use human speech-like noise (HSLN) [11] as a source signal. HSLN is a kind of bubble [12] noise generated by superimposing independent speech signals. By changing the number of superpositions, we can simulate various noise conditions. For example, HSLN of one or several superpositions can be considered as a nonstationary signal which sounds like a single speaker or an overlap of several speakers. When the number of superpositions is set to be some dozens, HSLN becomes a nonstationary signal which sounds like bubble noise. When the number of superpositions is greater than some hundreds, HSLN results in colored stationary noise while preserving the long-term spectrum of human speech. In this experiment, two HSLNs with the same superpositions are generated independently and assumed to arrive from sound source directions.

4.3 Initial Value of Weight Vector

The weight vectors $\boldsymbol{g}$ and $\boldsymbol{h}$ are optimized based on the criterion shown in Eq. (18) for each frequency independently. As the initial values in the iterative optimization, we design multiple weight vectors beforehand; each of them has a directional null at the direction of $(10 \cdot m)°$. where m is an integer ranging from -9 to 9. Taking two weight vectors as the initial values of $\boldsymbol{g}$ and $\boldsymbol{h}$, we perform the iteration procedure given by Eq. (50). By changing the pair of the initial weight vectors, we search for the optimal weight vectors $\boldsymbol{g}$ and $\boldsymbol{h}$ which minimize the criterion shown in Eq. (18).

4.4 Results of Experiments

Figure 5 shows the results of DOA estimation with each number of superpositions in HSLNs. These results are summarized as follows.

- In general, the conventional DOA-estimation method, such as the MV method or MUSIC method, cannot inherently identify DOAs of two sound sources with two microphones [4]. In contrast, our proposed method can be used to estimate DOAs of two sound sources with two microphones.
- The performance of DOA estimation of the proposed method degrades as the number of superpositions in HSLNs increases. Thus, the proposed method can be used to estimate the DOAs of nonstationary signals changing independently of each other. However it is difficult to use the proposed method to estimate the DOAs of stationary signals. The main reason for this phenomenon is that interframe-averaged power spectra of the stationary signals change negligibly among blocks and thus the criterion given by Eq. (15) could also be minimized. That is, in the case of arriving stationary signals, the array output decreases even when the sound source exists in the look direction because the array gain for each sound source is optimized to cancel each other.
- With lower analysis frequency, the performance of DOA estimation degrades.

These results clarify that the proposed method can be used to estimate DOAs with fewer microphones than in the conventional method, especially when the signals are nonstationary, such as speech, and independent of each other.

5. Experiments of DOA Estimation in Real Acoustic Environment

In this section, DOA-estimation results of the proposed method and the conventional MV method in a real acoustic environment are discussed.

5.1 Sound Recording in Reverberant Room

The experiment was carried out in the variable-reverberation-time room shown in schematically Fig. 6.

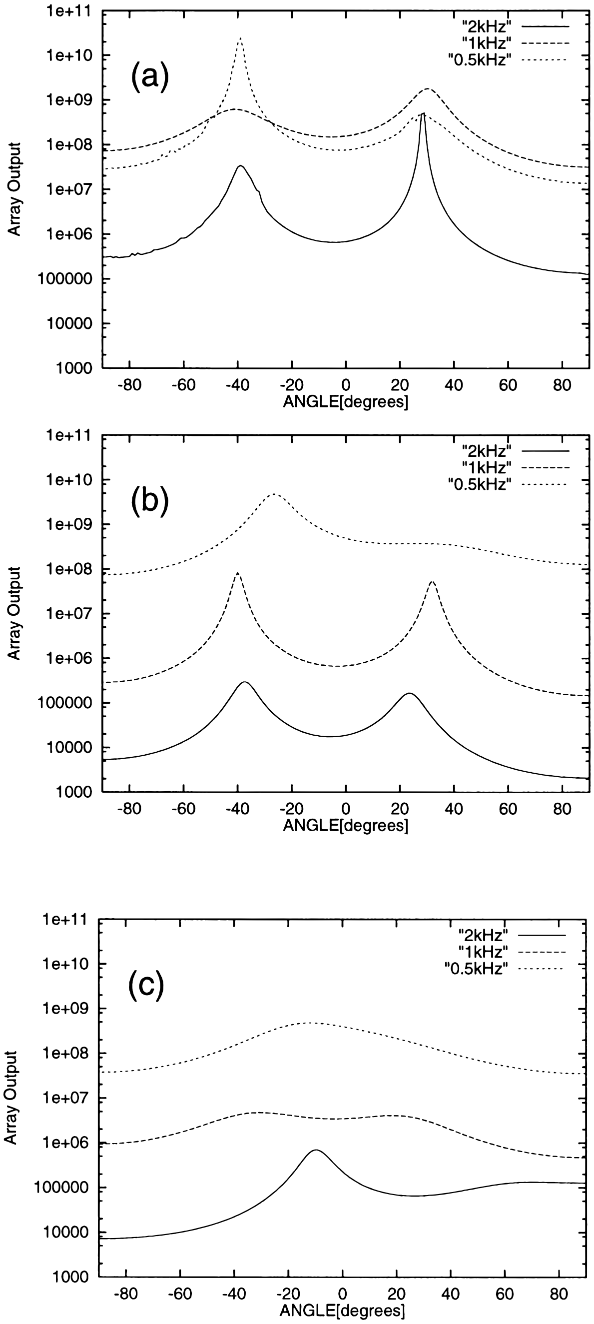


Fig. 5 The DOA-estimation results of the proposed method using a computer simulation (a) with 1 superposition in HSLNs, (b) with 10 superpositions in HSLNs, and (c) with 128 superpositions in HSLNs. The proposed method can be used to estimate the DOAs of two sound sources with two microphones. The solid lines represent the results for 2 kHz analysis frequency, the dashed lines represent the results for 1 kHz analysis frequency, and the dotted lines represent the results for 0.5 kHz analysis frequency.

Acoustic signals were recorded under the conditions that the reverberation times (RTs) were 150 and 300 msec. A loudspeaker was placed at each of the left side ($\theta = -40°$) and the right side ($\theta = 30°$) of the array. The background noise level and the target signal level measured at the array origin were 20 dB(A) and

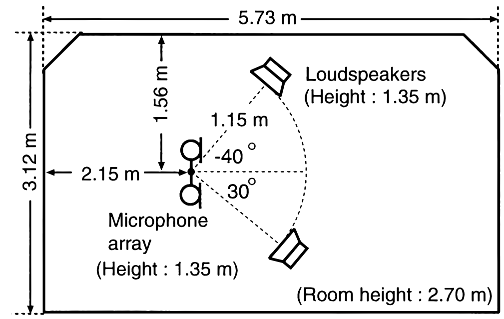


Fig. 6 Plan view of the layout of variable-reverberation-time room used in the experiments. The reverberation time is set to be 150 and 300 msec.

82 dB(A), respectively.

5.2 Experimental Setup

Two- and three-element arrays with the interelement spacing of 4 cm are used. In both the conventional MV method and the proposed method, we use sound samples of 2 sec duration to perform the DOA estimation. In the proposed method, the analysis conditions are the same as those in the computer simulation. The analysis conditions in the MV method are the same as those in the proposed method with respect to frame length, frame shift and window function. The analysis frequencies are set to be 1.0 and 3.0 kHz. We calculate and plot the resultant estimated DOA every 1° angle.

5.3 Experimental Results

First, the experimental results are shown in Fig. 7 for the case of a two-element array under two different reverberant conditions. In these figures, "Proposed" indicates the results of the proposed method, and "MV method" indicates those of the conventional MV method.

These figures show that the DOA-estimation results of the proposed method exhibit two peaks corresponding to the two sound sources, even when the conventional method fails. However, the positions of the two peaks deviate from each DOA as RT increases.

Next, the experimental results obtained with a three-microphone array under two different reverberant conditions are shown in Fig. 8.

These figures show that the MV method yields two peaks corresponding to the two sound sources under the condition that RT is relatively short because the number of microphones exceeds that of sound sources, however the method fails when RT is long. In contrast, to look at the envelops of the DOA results, the proposed method can produce two peaks at the correct directions under all reverberant conditions, while the raw DOA results are jagged. These jagged peaks are due to the existence of the local minimizers in solving Eqs. (18)

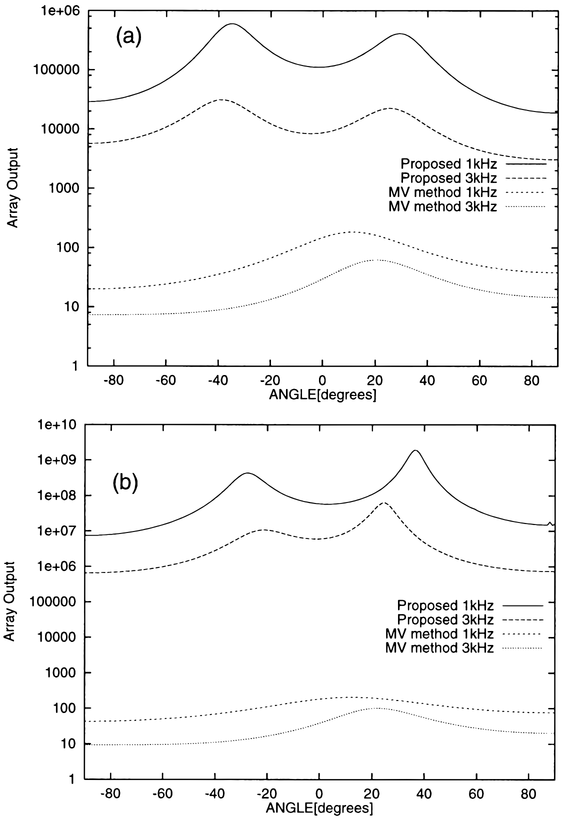


Fig. 7 The experimental results in the case of two-element array with (a) 150 msec reverberation time and (b) 300 msec reverberation time. The upper two lines represent the DOA-estimation results of the proposed method at 1 and 3 kHz analysis frequencies and the lower two lines represent the DOA-estimation results of the conventional MV method at 1 and 3 kHz analysis frequencies.

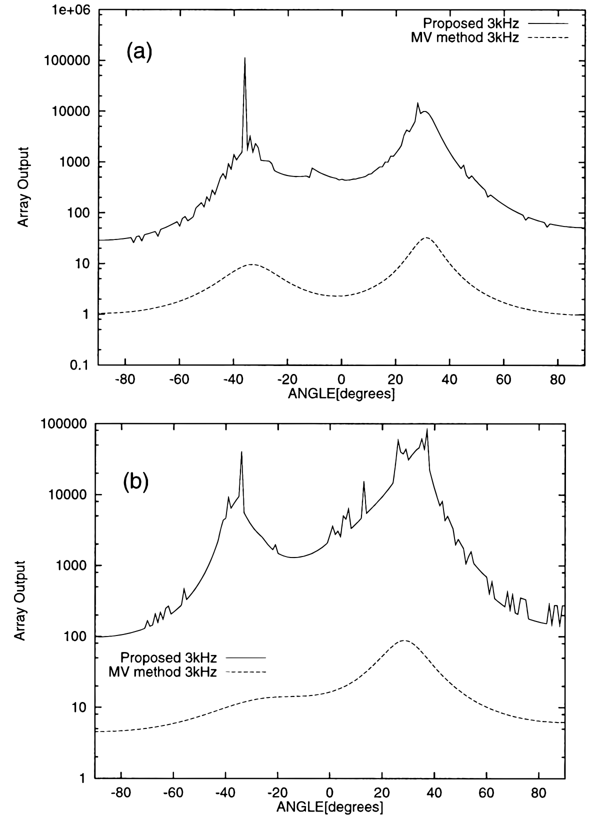


Fig. 8 The experimental results in the case of three-element array with (a) 150 msec reverberation time and (b) 300 msec reverberation time. The upper line represents the DOA-estimation result of the proposed method at 3 kHz analysis frequency and the lower line represents the DOA-estimation result of the conventional MV method at 3 kHz analysis frequency.

and (19), and we could not remove them automatically in this experiment; the removal of the jagged peaks is our future work.

6. Directional Resolution Measurement

In this section, the DOA-estimation performances of both the proposed and conventional methods are qualified under various reverberant conditions.

6.1 Definition of Directional Resolution

In this study, we use the directional resolution for the measurement of DOA-estimation quality, which is defined based on the following procedure (see Fig. 9). First, we set two loudspeakers with a narrow directional interval. Next, DOA estimations are performed for every additional two degrees with respect to the directional interval. Finally, when we obtain an accurate DOA of the loudspeakers, then the minimum directional interval Θ is defined as the directional resolution of our experiments.

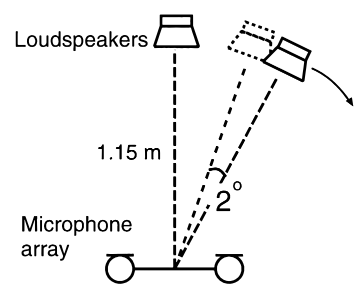


Fig. 9 Configuration of a microphone array and loudspeakers for measuring directional resolution. The angle of one sound source is fixed at 0°, and that of the other source is varied. DOA estimations are performed for every additional two degrees with respect to the directional interval.

6.2 Experimental Conditions

Acoustic signals prepared in this experiment are recorded in the variable-reverberation-time room. We use the following sound data: (1) artificial signals simulated by a computer, (2) acoustic signals recorded in the variable-reverberation-time room under the RT conditions of 150 and 300 msec. In the computer simulation,

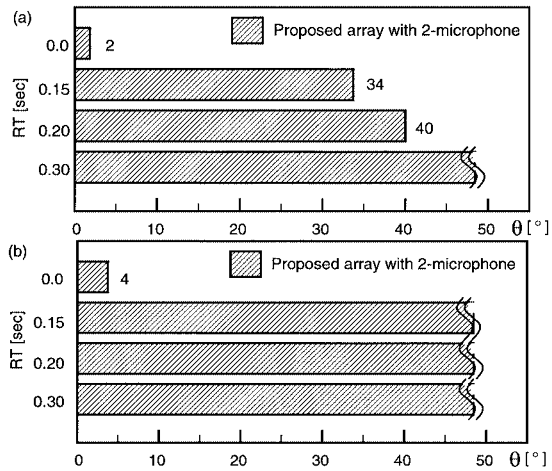


Fig. 10 The directional resolutions of the proposed method with two microphones at (a) 3 kHz analysis frequency and (b) 1 kHz analysis frequency.

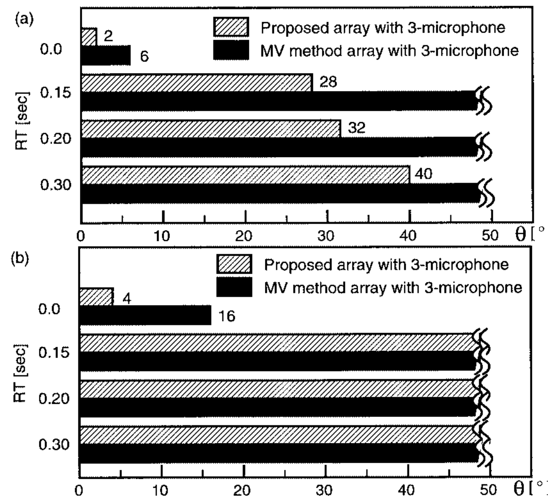


Fig. 11 The directional resolutions of proposed method and conventional MV method with three microphones at (a) 3 kHz analysis frequency and (b) 1 kHz analysis frequency.

the angle of one sound source is fixed at 0°, and that of the other source is changed from 2°. In the reverberant room, the angle of one sound source is fixed at 0°, and that of the other source is changed in the range from 10 to 48°.

Analysis frequencies are 1 and 3 kHz. We perform DOA estimation for HSLN of one superposition with both two- and three-element arrays with the interelement spacing of 4 cm. In both the conventional MV method and the proposed method, we use sound samples of 2 sec duration, and estimate DOA every 1° angle.

6.3 Experimental Results

Figure 10 shows the relation between the directional resolution of the proposed method and RT in the case of two microphones, where the analysis frequencies are set to be 1 and 3 kHz. We do not show the directional resolution of the MV method because the DOA estimation for two sound sources is impossible with only two microphones, as described in Sect. 6.2.

Figure 11 shows the relation between the directional resolution of the proposed method and RT in the case of three microphones. We can summarize the results as follows.

- The proposed method can be used to distinguish two DOAs with a directional interval of at least 4° at both analysis frequencies when two microphones are used and there is no reverberation. However, the directional resolution increases considerably, i.e., the accuracy of the DOA estimation deteriorates, as the RT increases.
- The directional resolution of the proposed method is superior to that of the conventional MV method under all reverberant conditions when three-element array are used. Also, the directional resolution of both the proposed method and the conventional method increases considerably as the RT increases.
- Increasing the number of elements in the proposed method prevents the deterioration of DOA estimation caused by reverberation.

7. Conclusion

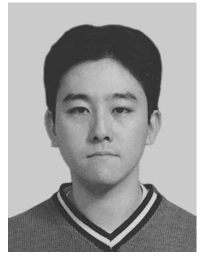
In this paper, a new method for direction of arrival (DOA) estimation using the nonlinear microphone array system based on complementary beamforming was described. To evaluate its effectiveness, DOA-estimation experiments were performed using both computer simulation and actual devices in real acoustic environments. From experiments of DOA estimation by computer simulation, it was shown that (1) the proposed method can be used to estimate DOAs of two sound sources with two microphones, and (2) the performance of DOA estimation of the proposed method degrades as the number of superpositions in HSLNs increases. Thus, the proposed method can be used to estimate the DOAs of nonstationary signals changing independently of each other, however it is difficult to use the proposed method to estimate the DOAs of stationary signals. From experiments of DOA estimation in real acoustic environments, comparing the conventional MV method and the proposed method, it was shown that the proposed method yields two peaks corresponding to the two sound sources with a two-element array, even when the conventional method fails. By comparing the resolutions of DOA estimation by the proposed method and by the conventional MV method, we could show that (1) the proposed method can distinguish two DOAs with a directional interval of at least 4° at analysis frequencies of 1 and 3 kHz when there is no reverberation. However, the accuracy of the DOA estimation deteriorates as the RT increases; (2) the directional resolution of the proposed method is superior to that of the conventional MV method under all reverberant conditions when a three-element array is used; (3) increasing in the number of elements in the proposed method prevents the deterioration of DOA estimation caused by reverberation.

Acknowledgement

This work was partly supported by Grant-in-Aid for COE Research (No. 11CE2005).

References

[1] S.U. Pillai, Array Signal Processing, Springer-Verlag, 1989.
[2] G.W. Elko, "Microphone array systems for hands-free telecommunication," Speech Communication, vol.20, pp.229–240, 1996.
[3] Y. Kaneda, "Microphone array technologies for speech recognition under noisy environment," J. Acoust. Soc. Jpn., vol.53, no.11, pp.872–876, 1997.
[4] D.H. Johnson and D.E. Dudgeon, Array Signal Processing: Concepts and Techniques, Prentice-Hall, New Jersey, 1993.
[5] J. Capon, "High resolution frequency-wavenumber spectrum analysis," Proc. IEEE, vol.57, pp.1408–1418, 1969.
[6] R.O. Schmidt, "Multiple emitter location and signal parameter estimation," Proc. RADC Spectral Est. Workshop, pp.243–258, Oct. 1979.
[7] H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Speech enhancement using nonlinear microphone array based on complementary beamforming," IEICE Trans. Fundamentals, vol.E82-A, no.8, pp.1501–1510, Aug. 1999.
[8] H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Speech enhancement using nonlinear microphone array based on noise adaptive complementary beamforming," IEICE Trans. Fundamentals, vol.E83-A, no.5, pp.866–876, May 2000.
[9] H. Saruwatari and M. Komura, "Synthetic aperture sonar in air medium using a nonlinear sidelobe canceller," IEICE Trans., vol.J81-A, no.5, pp.815–826, May 1998.
[10] A.L. Peressini, F.E. Sullivan, and J.J. Uhl, Jr., The Mathematics of Nonlinear Programming, Springer-Verlag, New York, 1988.
[11] S. Kajita, D. Kobayashi, K. Takeda, and F. Itakura, "Analysis of speech features included in human speech-like noise," J. Acoust. Soc. Jpn., vol.53, no.5, pp.337–345, 1997.
[12] IEICE, ed., Handbook for Electronics, Information and communication Engineers, p.2220, Ohmsha, Tokyo, 1988.



Hidekazu Kamiyanagida was born in Osaka, Japan on May 5, 1976. He received the B.E. degrees in electrical engineering from Nagoya University in 2000. He is currently a student of master cource of Graduate School of Information Science, Nara Institute of Science and Technology. His research interests include array signal processing and the DOA estimation. He is a member of the Acoustical Society of Japan.

Hiroshi Saruwatari was born in Nagoya, Japan, on July 27, 1967. He received the B.E., M.E. and Ph.D. degrees in electrical engineering from Nagoya University, Nagoya, Japan, in 1991, 1993 and 2000, respectively. He joined Intelligent Systems Laboratory, SECOM CO., LTD., Mitaka, Tokyo, Japan, in 1993, where he engaged in the research and development on the ultrasonic array system for the acoustic imaging. He is currently an associate professor of Graduate School of Information Science, Nara Institute of Science and Technology. He is a member of the IEEE and the Acoustical Society of Japan.

Kazuya Takeda was born in Sendai, Japan on September 1, 1960. He received B.E.E. and M.E.E. and Doctor of Engineering degrees all from Nagoya University in 1983, 1985 and 1994, respectively. In 1986, he joined ATR (Advanced Telecommunication Research Laboratories) where he involved in the two major projects of speech database construction and speech synthesis system development. In 1989, he moved to KDD R & D Laboratories and participated a project for constructing voice-activated telephone extension system. Since 1995, he has been working for Nagoya University as an Associate Prof.

Fumitada Itakura was born in Toyokawa, Japan on August 6 1940. He received the B.E.E., M.E.E. and Doctor of Engineering degrees all from Nagoya University in 1963, 1965 and 1972, respectively. In 1968, he joined the Electrical Communication Laboratory of NTT, Musashino, Tokyo and participated in the speech processing, including the maximum likelihood spectrum estimation, the PARCOR method, the line spectrum pair method, the composite sinusoidal method, and the APC-AB speech coding. In 1981, he was appointed to Head of Speech and Acoustics Research Section of the ECL, NTT. In 1984, he left NTT to become a professor of Nagoya University, where he teaches courses of communication theory and signal processing. In 1975, he received IEEE ASSP Senior Award for his paper on speech recognition based on the minimum prediction residual principle. He is a co-recipient with B.S. Atal of 1986 Morris N. Liebmann Award for contributions to linear predictive coding for speech processing. In 1997, he received IEEE Signal Processing Society Award.

Kiyohiro Shikano received the B.S., M.S., and Ph.D. degrees in electrical engineering from Nagoya University in 1970, 1972, and 1980, respectively. From 1972, he had been working at NTT Laboratories. During 1990–1993, he was the executive research scientist at NTT Human Interface Laboratories. During 1986–1990, he was the Head of Speech Processing Department at ATR Interpreting Telephony Research Laboratories. During 1984–1986, he was a visiting scientist in Carnegie Mellon University. He received the Yonezawa Prize from IEICE in 1975, the Signal Processing Society 1990 Senior Award from IEEE in 1991, and the Technical Development Award from ASJ in 1994. He is currently a professor of Nara Institute of Science and Technology (NAIST), where he is directing speech and acoustics laboratory. His major research areas are speech recognition, multi-modal dialog system, speech enhancement, adaptive microphone array, and acoustic field realization. He is a member of Information Processing Society of Japan, the Acoustical Society of Japan (ASJ), and the Institute of Electrical and Electronics, Engineers (IEEE).