Underdetermined Sound Source Separation Using Power Spectrum Density Estimated by Combination of Directivity Gain

Yusuke Hioka, *Senior Member, IEEE,* Ken'ichi Furuya, *Senior Member, IEEE,* Kazunori Kobayashi, Kenta Niwa, *Member, IEEE,* and Yoichi Haneda, *Senior Member, IEEE*

***Abstract*—A method for separating underdetermined sound sources based on a novel power spectral density (PSD) estimation is proposed. The method enables up to $M(M-1)+1$ sources to be separated when we use a microphone array of M sensors and a Wiener post-filter calculated by the estimated PSDs. The PSD of a beamformer's output is modelled by a mixture of source PSDs multiplied by the beamformer's directivity gain in the particular angle where each source is located. Based on this model, the PSD of each sound source is estimated from the PSD of multiple fixed beamformers' outputs using the difference in the combination of directivity gains. Simulation results proved that the proposed method effectively separated up to $M(M-1)+1$ sound sources if the fixed beamformers were appropriately selected. Experiments were also conducted in a reverberant chamber to ensure the proposed method was also effective in practical use.**

***Index Terms*—microphone array, underdetermined sound source separation, power spectral density estimation, Wiener post-filter, fixed beamformer.**

I. Introduction

HANDS-FREE microphones are utilised as sound interfaces in voice conference or automatic speech recognition (ASR) systems because of their convenience. The common disadvantage of hands-free microphones compared to conventional close-talking microphones is their low signal-to-interference ratio (SIR). Thus, various source separation or noise reduction techniques have been studied [1] in order to improve this ratio. Among the various techniques, the use of a microphone array [2] has been well studied because it can spatially distinguish and separate sound sources located in different directions/positions. In recent years, several commercial products using microphone arrays to improve the quality of the target source have also been introduced. [3]

Generally, sound source separation problems can be categorised according to two different aspects: whether the problem is blind/non-blind, and (over)determined/underdetermined. A blind problem should be solved without (or with very little) prior knowledge of the mixing process and the sound sources. Such unknown knowledge normally includes the spatial information, i.e. the position of the sound sources and the microphones. In contrast, the problem is underdetermined if the number of sound sources exceeds the number of available microphones.

Of the various microphone array technologies, beamforming is one of the most basic approaches for sound source separation. Because beamforming realises a spatial filter whose spatial degree of freedom is restricted by the number of microphones, it can only solve the *determined* problem. Fixed beamforming, which includes the conventional delay-and-sum beamforming, emphasises the target signal by pointing its main beam to the target source or by pointing directivity nulls toward the interferences. On the other hand, an adaptive beamformer optimises a cost function to tailor its directivity to the received signals. It can effectively reduce undesired interference by pointing directivity nulls while steering the main beam to the target source. For example, the linearly constrained minimum variance (LCMV) beamformer [4] optimises itself by reducing the total power of the output while maintaining the target signal without distortion. The well-known generalised sidelobe canceller (GSC) [5] is an extension of LCMV that implements the optimisation by employing an open-loop process. As these conventional beamforming methods require spatial information of sound sources and microphones, they are only effective in dealing with the non-blind problem. To address the blind problem, the use of independent component analysis (ICA) [6] has been extensively studied in the last decade [1]. However, previous blind source separation methods based on ICA are also affected by the underdetermined problem since the resultant source separation filter performs as an adaptive beamforming technique [7].

Various attempts have been made to solve *underdetermined* problems [8] recently. One of the most well-known approaches is the exploitation of W-disjoint orthogonality [9], [10], [11] or the sparseness of source signals. This framework was first established by Jourjine *et al.* [9] as the DUET algorithm and was later explored in detail by Yilmaz *et al.* in [11].

Y. Hioka was with the NTT Cyber Space Laboratories (current NTT Media Intelligence Laboratories), NTT Corporation. He is now with the Department of Electrical and Computer Engineering, University of Canterbury, Christchurch, 8140 New Zealand (e-mail: yusuke.hioka@ieee.org).

K. Furuya was with the NTT Media Intelligence Laboratories, NTT Corporation. He is now with the Department of Computer Science and Intelligent Systems, Oita University, Oita, 870-1192 Japan (e-mail: furuya-kenichi@oita-u.ac.jp).

K. Niwa and K. Kobayashi are with the NTT Media Intelligence Laboratories, NTT Corporation, Musashino, Tokyo, 180-8585 Japan (niwa.kenta@lab.ntt.co.jp, kobayashi.kazunori@lab.ntt.co.jp).

Y. Haneda was with NTT Media Intelligence Laboratories. He is now with the Graduate School of Informatics and Engineering at The University of Electro-Communications, Tokyo, 182-8585 Japan.

Manuscript received March 17, 2012; revised November 26, 2012; accepted February 12, 2013.

Several extensions of this framework have also been studied [8], [12], [13]. Although these previous methods are normally classified as blind methods since they do not require any spatial information, they actually need to adopt some small prior assumptions about the source signals such as sparseness. Thus, the existing techniques can only be applied to separate specific sound sources that follow the assumed source models.

We focused on a *non-blind underdetermined* problem in this study. In return for the restriction in which knowledge of the angle of sound sources and the microphone arrangement are required a priori, the proposed method can separate any arbitrary types of sound sources. Needless to say, the underdetermined source separation problem is usually achieved by applying some nonlinear techniques [8]. In the nonlinear methods, the application of an adaptive post-filter to a beamformer's output is a common approach [14], [15], [16] to enhance the noise reduction performance. The post-filter consists of the estimated power spectral density (PSD) of the target and interferences. For the PSD estimation, Zelinski proposed a method using the auto- and cross-power spectral density of the received signals [14]. The drawback of Zelinski's method is that it is based on an assumption that the interferences are completely incoherent, i.e., they are uncorrelated between microphones. As this assumption seldom holds in a real acoustic environment, McCowan *et al.* later proposed a method using the diffuse noise model, which is a more reasonable assumption in a real world environment [15]. Nevertheless, various modifications and improvements of the post-filter calculation have been reported recently, [17], [18], [19], [20], although to the best of our knowledge, none of the previous works have applied a post-filter to reduce coherent signals because the beamformer is normally used for that function [16] unless the problem is underdetermined.

This paper proposes a novel technique to estimate the PSD of coherent signals in order to calculate a post-filter. The remarkable advantages of the proposed method are:

- It makes it possible to separate up to $M(M-1)+1$ sources when the number of microphones is M,
- No models or restrictions are assumed for the source signals.

The key strategy of the proposed method is the use of fixed beamformers, not to straightforwardly separate sources by controlling the directivity but to estimate the PSD of each source by exploiting the difference in the combination of directivity gains in each source angle. When there are multiple beamformers, the set of gains of the beamformers to a particular angle differs from the sets of other angles. The proposed method uses this difference as a clue to estimate the PSD of each sound source mixed in the output of a beamformer. A similar idea was proposed by Saruwatari *et al.* that uses complementary beamformers to estimate the PSD of sound sources [21], [22]. Although this method also succeeded in solving the underdetermined problem without any assumptions on the source signals, it allows us to separate only $2M-1$ sources, as the complementary beamformers also steer nulls to interferences. The proposed method avoids the null steering, so it succeeded in increasing the maximum number of separable sources to $M(M-1)+1$.

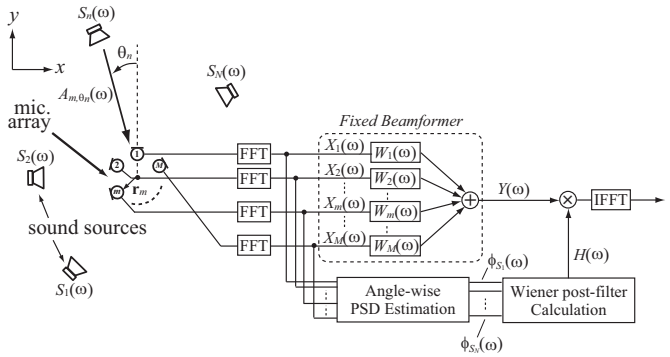

Fig. 1. Problem definition and flow chart of proposed method.

This paper is organised as follows. In Sec. II, we propose our novel PSD estimation technique using a combination of directivity patterns achieved by multiple fixed beamformers; then we discuss the maximum number of separable sound sources in Sec. III. Simulation results that reveal the effectiveness of the proposed method are explained in Sec. IV. Then some results of an experimental evaluation conducted in a real acoustic field are described in Sec. V. Finally, we give some concluding remarks in Sec. VI.

II. Power Spectrum Estimation Using Combination of Directivity Gains

A. Problem Definition

Assume N different sound sources located at angle θ_n are observed by a microphone array with M microphones, as shown in Fig. 1. Here, signals observed by the m-th microphone are described by

$$X_m(\omega) = \sum_{n=1}^{N} A_{m,\theta_n}(\omega) S_n(\omega), \tag{1}$$

where $S_n(\omega)$ is the source n, and $A_{m,\theta_n}(\omega)$ is the transfer function from the source n to the microphone m. Note, for simplicity, that we describe the signals in the frequency domain so that the convolution is expressed by the multiplication. Several different models for sound propagation could replace the transfer function. The simplest model assumes that only direct sound regarded as a plane wave arrives from angle θ_n, allowing us to substitute $A_{m,\theta_n}(\omega)$ as follows

$$A_{m,\theta_n}(\omega) := \exp(j\mathbf{k}_{\theta_n}^T \cdot \mathbf{r}_m), \tag{2}$$

where $\mathbf{k}_{\theta_n}$ is the wave number vector for the sound source n in angle θ_n, i.e. $\mathbf{k}_{\theta_n} = [\frac{2\pi \sin\theta_n}{c}, \frac{2\pi \cos\theta_n}{c}]^T$, and $\mathbf{r}_m$ is the coordinate of the microphone m defined in Fig. 1. Our problem is to estimate source PSDs from this observation.

B. Modelling PSD of Fixed Beamformer Output

Now as shown in Fig. 2, we define N angles $\check{\theta}_n$ around a microphone array, then apply L fixed beamformers, each with a different directivity pattern, to the array input signal. In the rest of this paper, we call θ_n and $\check{\theta}_n$ the *actual source angle*

and the *assumed source angle*, respectively. Although the assumed source angle is prerequisite prior knowledge, it could be estimated by any existing source localisation techniques [2]. Note, without loss of generality, that the number of source angles is always less than the number of fixed beamformers, i.e. $L \geq N$. If the assumed source angle coincides with the actual source angle, i.e. $\breve{\theta}_n = \theta_n$, the output of the l-th beamformer is described by

$$Y_l(\omega) = \sum_{m=1}^{M} W_{l,m}(\omega) X_m(\omega). \tag{3}$$

where $W_{l,m}(\omega)$ is the filter coefficient of mic m for beamformer l. Substituting $X_m(\omega)$ in Eq.(1) into Eq.(3) will derive

$$\begin{aligned} Y_l(\omega) &= \sum_{m=1}^{M}\sum_{n=1}^{N} W_{l,m}(\omega) A_{m,\theta_n}(\omega) S_n(\omega) \\ &= \sum_{n=1}^{N} D_{l,\theta_n}(\omega) S_n(\omega), \end{aligned} \tag{4}$$

where the beamformer's response to the i-th source is defined by

$$\begin{aligned} D_{l,\theta_n}(\omega) &= \sum_{m=1}^{M} W_{l,m}(\omega) A_{m,\theta_n}(\omega) && (5) \\ &= \mathbf{w}_l^T(\omega)\mathbf{a}_{\theta_n}(\omega), && (6) \end{aligned}$$

where $\mathbf{w}_l(\omega) = [W_{l,1}(\omega), \cdots, W_{l,M}(\omega)]^T$ and $\mathbf{a}_{\theta_n}(\omega) = [A_{1,\theta_n}(\omega), \cdots, A_{M,\theta_n}(\omega)]^T$ are respectively the vector of filter coefficients and the array manifold vector for angle θ_n.

The sound sources are assumed to be mutually uncorrelated, and therefore, the PSD of a beamformer's output is given by

$$\begin{aligned} \phi_{Y_l}(\omega) &= E[Y_l(\omega)Y_l^*(\omega)] && (7) \\ &= E\left[\sum_n\sum_{n'} D_{\theta_n}(\omega)S_n(\omega)D^*_{\theta_{n'}}(\omega)S^*_{n'}(\omega)\right] && (8) \\ &= \sum_n |D_{\theta_n}(\omega)|^2 E[|S_n(\omega)|^2] \\ &\quad + \sum_n\sum_{n'\neq n}\Big(D_{\theta_n}(\omega)D^*_{\theta_{n'}}(\omega)E[S_n(\omega)S^*_{n'}(\omega)] \\ &\quad D_{\theta_{n'}}(\omega)D^*_{\theta_n}(\omega)E[S_{n'}(\omega)S^*_n(\omega)]\Big) && (9) \\ &\approx \sum_n |D_{l,\theta_n}(\omega)|^2 E[|S_n(\omega)|^2] && (10) \\ &:= \sum_{n=1}^{N} |D_{l,\theta_n}(\omega)|^2 \phi_{S_n}(\omega). && (11) \end{aligned}$$

where we applied the uncorrelatedness assumption to ignore the second term of Eq. (9). Thus, Eq. (10) shows that the PSD of the beamformer's output can be approximated by the sum of source PSDs multiplied by the squared amplitude of the beam pattern (we refer to this as *directivity gain* hereafter).

C. Angle-wise PSD Estimation

Now, by stacking up the PSD of the output signal of L different fixed beamformers to become a vector form, we obtain

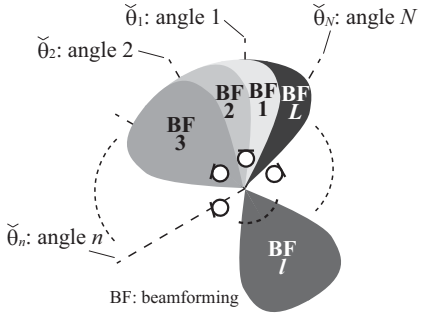

Fig. 2. Definition of N assumed source angles and introduction of L fixed beamformers of different directivity patterns.

$$\underbrace{\begin{bmatrix}\phi_{Y_1}\\ \phi_{Y_2}\\ \vdots\\ \phi_{Y_L}\end{bmatrix}}_{\Phi_Y(\omega)} = \underbrace{\begin{bmatrix}|D_{1,\theta_1}|^2 & |D_{1,\theta_2}|^2 & \cdots & |D_{1,\theta_N}|^2\\ |D_{2,\theta_1}|^2 & & \cdots & \\ \vdots & \vdots & \ddots & \vdots\\ |D_{L,\theta_1}|^2 & & \cdots & |D_{L,\theta_N}|^2\end{bmatrix}}_{\mathbf{D}(\omega)} \underbrace{\begin{bmatrix}\phi_{S_1}\\ \phi_{S_2}\\ \vdots\\ \phi_{S_N}\end{bmatrix}}_{\Phi_S(\omega)}. \tag{12}$$

Note that we omitted ω for simplicity. Because the PSD of the beamformer's output is calculated from the PSD of the observed signals, and the directivity gains can be given a priori, the PSD of sound sources of each angle is estimated by solving the simultaneous equation in Eq. (12)

$$\begin{aligned} \widehat{\Phi}_S(\omega) &= \begin{bmatrix}\widehat{\phi}_{S_1}(\omega) & \widehat{\phi}_{S_2}(\omega) & \cdots & \widehat{\phi}_{S_N}(\omega)\end{bmatrix}^T \\ &\approx \begin{cases}\mathbf{D}^{-1}(\omega)\Phi_Y(\omega) & \text{if } L = N\\ \mathbf{D}^{+}(\omega)\Phi_Y(\omega) & \text{if } L > N\end{cases} \end{aligned} \tag{13}$$

where $^+$ and $\widehat{\ }$ denote a Moore-Penrose pseudo-inverse and an estimated value, respectively. Each component of the estimated vector corresponds to the PSD estimate of a source located in θ_n. Note that the estimated PSD $\widehat{\Phi}_S(\omega)$ might include some negative values because there are no constraints on the values appearing in $\widehat{\Phi}_S(\omega)$. A discussion on applying a suitable constraint on the estimated PSD remains as a future problem. Within this paper, such an unrealistically negative PSD is substituted by another value. In the simulation and experiments explained later in Secs. IV and V, we applied the following rule; negative estimates are substituted by the absolute value of the original estimate defined by Eq. (14).

$$\widehat{\phi}_{|S_n|}(\omega) = |\widehat{\phi}_{S_n}(\omega)| \tag{14}$$

D. Wiener Post-filter based on Estimated PSD

Now we apply the estimated PSD of each angle and calculate the Wiener post-filter given by

$$H(\omega) = \frac{\sum_{n\in\Theta_S}\widehat{\phi}_{|S_n|}(\omega)}{\sum_{n\in\Theta_S}\widehat{\phi}_{|S_n|}(\omega) + \sum_{n\in\Theta_N}\widehat{\phi}_{|S_n|}(\omega)} \tag{15}$$

where Θ_S and Θ_N are the sets of indices of angles where we want to receive sources or reduce sources. Applying $H(\omega)$ to the output of a fixed beamformer will emphasise the sound sources located in the receiving angles. Fig. 1 depicts the entire flow of the proposed sound source separation method.

III. Number of separable sound sources

In general, the spatial resolution of a microphone array is determined by the number of microphones (M); thus, conventional beamformers are capable of separating up to M different sources. However, since the proposed PSD estimation defined by Eq. (12) is independent from the number of microphones, one would expect that the proposed method could estimate more source PSDs in different angles than the number of microphones. If this is true, we are also interested in the maximum number of source PSDs that can be estimated correctly when the number of microphones is fixed. In this section, we attempt to clarify the answer to this question.

A. Maximum Number of Separable Sound Sources

The accuracy of the proposed PSD estimation defined in Eq. (12) depends on the condition of the matrix $\mathbf{D}(\omega)$. If $\mathbf{D}(\omega)$ is ill-conditioned, the simultaneous equation becomes indefinite, which may give an unstable estimate. Because the ill-conditioned matrix is usually rank deficient, the maximum number of sources that the proposed method can separate will be found by pursuing N where $\mathbf{D}(\omega)$ starts to become rank deficient.

From Eq. (6), each component of $\mathbf{D}(\omega)$ is derived by

$$|D_{l,\theta_n}(\omega)|^2 = \mathbf{w}_l^T(\omega)\mathbf{a}_{\theta_n}(\omega)\mathbf{a}_{\theta_n}^H(\omega)\mathbf{w}_l(\omega) \quad (16)$$

$$= \mathbf{w}_l^T(\omega)R_{\theta_n}(\omega)\mathbf{w}_l(\omega) \quad (17)$$

where $R_{\theta_n}(\omega) = \mathbf{a}_{\theta_n}(\omega)\mathbf{a}_{\theta_n}^H(\omega)$. Because $L \geq N$, we first start by looking into the case where $\mathbf{D}(\omega)$ is column-rank deficient, i.e. the column vectors of $\mathbf{D}(\omega)$ are linear dependent. Namely, the following Eq. (18) holds for all l given $\beta_n(\omega)(\in \mathbb{R}) \neq 0\ \forall n$.

$$\sum_{n=1}^{N}\beta_n(\omega)|D_{l,\theta_n}(\omega)|^2 = 0 \quad (18)$$

Substituting $|D_{l,\theta_n}(\omega)|^2$ defined in Eq. (17) into Eq. (18) results in

$$\sum_{n=1}^{N}\beta_n(\omega)\mathbf{w}_l^T(\omega)R_{\theta_n}(\omega)\mathbf{w}_l(\omega) = \mathbf{w}_l^T\left[\sum_{i=1}^{N}\beta_n(\omega)R_{\theta_n}(\omega)\right]\mathbf{w}_l(\omega) = 0. \quad (19)$$

Because $R_{\theta_n}(\omega)$ is a non-zero matrix and $\beta_n(\omega) \neq 0\ \forall n$, Eq. (19) holds if either

$$[\text{case 1}] \quad \mathbf{w}_l^T(\omega)R_{\theta_n}(\omega)\mathbf{w}_l(\omega) = 0 \qquad \forall l, n \quad (20)$$

or

$$[\text{case 2}] \quad \sum_{n=1}^{N}\beta_n(\omega)R_{\theta_n}(\omega) = O_M \quad (21)$$

is satisfied, where O_M is a $M \times M$ zero matrix.

Case 1 appears when $\mathbf{w}_l^T(\omega)\mathbf{a}_{\theta_n}(\omega) = 0$ holds for all l and θ_n. This can be realised only if we have L different fixed beamformers whose directivity gain is 0 for all θ_N. Such fixed beamformers could exist only if $M > N$. Hence, the rank deficiency of $\mathbf{D}(\omega)$ due to case 1 never arises if $N \geq M$. Because the proposed method is aimed at the case where $N \geq M$, we recognise that this case has no adverse affects on the proposed method.

To further look into case 2, we reformulate components in the matrix $R_{\theta_n}(\omega)$ to arrange them in column vector form and then combine the vectors of different $R_{\theta_n}(\omega)$ to be a matrix $\mathcal{R}(\omega)$. Now we have

$$\underbrace{\begin{bmatrix} r_{\theta_1}^{(11)}(\omega) & \cdots & r_{\theta_N}^{(11)}(\omega) \\ r_{\theta_1}^{(12)}(\omega) & \cdots & r_{\theta_N}^{(12)}(\omega) \\ \vdots & & \vdots \\ r_{\theta_1}^{(ij)}(\omega) & \cdots & r_{\theta_N}^{(ij)}(\omega) \\ \vdots & & \vdots \\ r_{\theta_1}^{(MM)}(\omega) & \cdots & r_{\theta_N}^{(MM)}(\omega) \end{bmatrix}}_{\mathcal{R}(\omega)} \underbrace{\begin{bmatrix} \beta_1(\omega) \\ \beta_2(\omega) \\ \vdots \\ \beta_N(\omega) \end{bmatrix}}_{\mathcal{B}(\omega)} = \mathbf{0} \quad (22)$$

where $r_{\theta_n}^{(ij)}(\omega)$ denotes the ij component of $R_{\theta_n}(\omega)$ and $\mathbf{0}$ is a zero vector. Because $\beta_n(\omega)$ is a non-zero real value, Eq. (22) is satisfied only if the column vectors of $\mathcal{R}(\omega)$ are linear dependent. Now we see that the number of rows of $\mathcal{R}(\omega)$ is fixed once M is determined, whereas the number of columns can be varied by setting different N. This implies that the maximum N where Eq. (22) does not hold is determined by the row-rank of $\mathcal{R}(\omega)$. In other words, the maximum number of separable sound sources can be found by searching the maximum N where the rank of $\mathcal{R}(\omega)$ is equal to N.

With regard to the row-rank of $\mathcal{R}(\omega)$, although there are M^2 rows in $\mathcal{R}(\omega)$, M of them are all the same vectors consisting of N unity since $r_{\theta_n}^{(ii)}(\omega) = 1$ for all i. Therefore, the maximum row-rank of $\mathcal{R}(\omega)$ is $M^2 - (M-1) = M(M-1)+1$. In other words, $\mathbf{D}(\omega)$ is rank deficient if $N > M(M-1)+1$. Thus, the maximum number of sound sources that the proposed method can separate is

$$M(M-1)+1 \quad (23)$$

if we have M microphones.

B. Effect of Array Configuration on Maximum Number of Separable Sound Sources

We note here that in fact, there is a specific case where the row-rank of $\mathcal{R}(\omega)$ is further reduced. It appears when the orientation and inter-microphone distance of more than one pair of microphones in the microphone array are exactly the same. For instance, a rectangular microphone array (see Fig. 3(a)) has two sets of pairs of microphones whose orientation and inter-microphone distance are equal. In this case, the phase differences between the microphones in these sets are equal, i.e. $r_{\theta_n}^{(12)}(\omega) = r_{\theta_n}^{(43)}(\omega)$ and $r_{\theta_n}^{(23)}(\omega) = r_{\theta_n}^{(14)}(\omega)$. As the equality also holds for the conjugate, i.e., $r_{\theta_n}^{(21)}(\omega) = r_{\theta_n}^{(34)}(\omega)$ and $r_{\theta_n}^{(32)}(\omega) = r_{\theta_n}^{(41)}(\omega)$, the row-rank of $\mathcal{R}(\omega)$ will be reduced to $4 \times (4-1) + 1 - 4 = 9$. Table I lists the number of row-rank reductions for some common array configurations shown in Fig. 3. The worst case scenario arises when every microphone pair has the same inter-microphone distance and

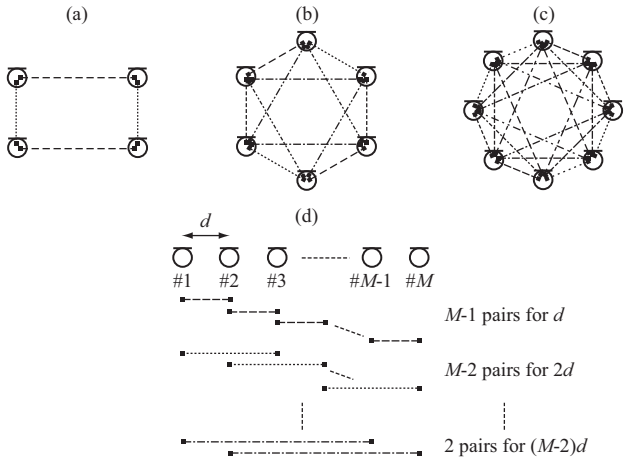

Fig. 3. Example where the rank of $\mathcal{R}(\omega)$ is reduced because of the configuration of the microphone array. There are multiple pairs whose inter-microphone phase differences are the same.

TABLE I
NUMBER OF ROW-RANK REDUCTIONS WITH SPECIFIC ARRAY CONFIGURATION.

Configuration	Number of row-rank reductions
Square/Rectangle	4
Regular Hexagon	12
Regular Octagon	24
Linear	2×number of pairs with equal distance

is in the same orientation, i.e. a uniformly spaced linear configuration. In this case, we have $(M-1)(M-2)$[1] rows whose components are equal to that of the other row; thus, the maximum number of sources is reduced to $M(M-1)+1-(M-1)(M-2)=2M-1$.

Naturally, even if $\mathcal{R}(\omega)$ is full row-rank, Eq. (22) holds if $r_{\theta_n}^{(ij)}(\omega)=r_{\theta_n'}^{(ij)}(\omega)$ for all i and j. This occurs in two different cases: (1) when the line/plane of the microphone array is the axis/plane of symmetry for the angles θ_n and θ_n', (2) when the phase difference between microphones for angle θ_n coincides with that of angle θ_n' due to the spatial aliasing. Generally, these cases should be ignored as both of them are abnormal situations for the microphone array.

In conclusion, the maximum possible number of source angles where the proposed method can estimate PSD may drop depending on the array configuration. Because $\mathcal{R}(\omega)$ is determined once the array configuration and assumed source angles are given, we can check the decline of N by calculating the rank of $\mathcal{R}(\omega)$ in advance.

C. Appropriate Choice of Fixed Beamformers

We also need to discuss the case when $\mathbf{D}(\omega)$ is row-rank deficient. Notice that each row of $\mathbf{D}(\omega)$ consists of the directivity gain of a fixed beamformer to N source angles. The fixed beamformers should be chosen carefully to avoid $\mathbf{D}(\omega)$ being row-rank deficient. The solution to this problem is not actually unique; there are various options for the preferred

[1]See Appendix for further details on the derivation.

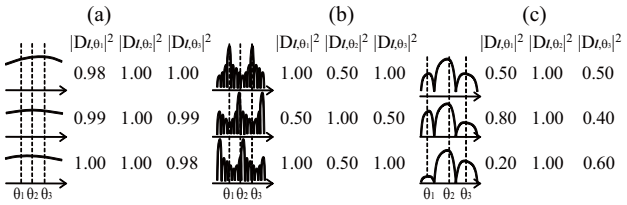

Fig. 4. Shape of directivity gain where $\mathbf{D}(\omega)$ becomes row-rank deficient. (a) Low frequency, (b) spatial aliasing, and (c) inappropriate combination of beamformers.

design of beamformers that meet the requirements. Thus, in this paper, we do not go too far into specifying the best design of fixed beamformers, but instead, we refer to the following cases where the shape of directivity gain causes the row-rank deficiency of $\mathbf{D}(\omega)$.

1) Low frequency
2) Spatial aliasing
3) Inappropriate combination of beamformers

As we see in the example in Fig. 4(a), the shape of directivity gain generally broadens in low frequency bands. The directivity gain of the source angles exhibits a similar value, which results in the rank deficiency of $\mathbf{D}(\omega)$. On the contrary, in the second case, folding occurs in the directivity gain due to the spatial aliasing. In this case, the directivity gain of two different angles may become equal, as shown in Fig. 4(b), which also causes rank deficiency of $\mathbf{D}(\omega)$. Finally, the last case occurs more accidentally. With some specific combinations of a beamformer and source angles, a row of $\mathbf{D}(\omega)$ can be expressed by the linear combination of other rows; an example is shown in Fig. 4(c). Because this case can be avoided by applying different beamformers by checking the rank of $\mathbf{D}(\omega)$ in advance, we would like to stress that the appropriate choice of beamformers is crucial to the proposed method. Further investigation into the most appropriate choice of fixed beamformers remains as a future research subject.

IV. SIMULATION RESULTS

A computer simulation was conducted to evaluate the performance of the proposed method. A microphone array observation was simulated under the far-field propagation model in an ideal anechoic environment; thus, the observed signal was simulated by calculating the signal defined in Eq. (1) with the transfer function Eq. (2) based on the plane wave model. Cylindrically isotropic noise was added as background noise where the signal-to-noise ratio (SNR) was varied from -6 dB to 20 dB. A uniform circular microphone array with M microphones was used, and its radius was set to d cm. Basic settings of parameters are given in Table II. For the multiple fixed beamformers used in the proposed PSD estimation, delay-and-sum beamformers whose main lobes pointed to each of the assumed source angles were applied. The source signals used in the simulations were: zero mean white noise, music, and speech signals, which are summarised in Table III.

TABLE II
BASIC PARAMETER SET USED IN THE SIMULATION.

Parameter	Value
d [cm]	4
N	3
L	3
Sampling frequency [Hz]	16000
FFT point	512
Frame shift	256
Frame length	512
Assumed source angle θ_n [deg]	{3, 123, 243}
SNR [dB]	20

TABLE III
SOUND SOURCES USED IN SIMULATION.

ID	Signal	ID	Signal
SP1	speech (female/JP)	MU1	music (classical)
SP2	speech (male/JP)	MU2	music (vocal)
SP3	speech (female/EN)	WN	white noise

* JP: Japanese, EN: English

A. Evaluation Metrics

The metrics we used to evaluate the effectiveness of sound source separation were *Signal to Interference Ratio improvement* (SIR) [23] and *Signal to Distortion Ratio* (SDR) [24]. The SIR indicates the amount of noise reduced by the source separation defined by

$$\triangle \mathrm{SIR} = \mathrm{SIR}_{\mathrm{O},i} - \mathrm{SIR}_{\mathrm{I},i} \quad (24)$$

where

$$\mathrm{SIR}_{\mathrm{O},i} = 10\log_{10}\frac{\sum_t z_{is_i}^2(t)}{\sum_t \left(\sum_{j\neq i} z_{is_j}(t)\right)^2} \quad (25)$$

$$\mathrm{SIR}_{\mathrm{I},i} = 10\log_{10}\frac{\sum_t x_{ms_i}^2(t)}{\sum_t \left(\sum_{j\neq i} x_{ms_j}(t)\right)^2} \quad (26)$$

and t is the sample index. Signal $z_{is_i}(t)$ is the output of the whole system when only the source $s_i(t)$ is active and the other sources are silent, whereas signal $x_{ms_i}(t)$ is the input signal of source $s_i(t)$ observed by microphone m. SDR is a metric used to observe the amount of distortion in the output of sound source separation, given by

$$\mathrm{SDR} = 10\log_{10}\left(\frac{\sum_t |\xi s_{\mathrm{in}}(t-\delta)|^2}{\sum_t |s_{\mathrm{out}}(t) - \xi s_{\mathrm{in}}(t-\delta)|^2}\right)[\mathrm{dB}], \quad (27)$$

where ξ and δ are the parameters for aligning the amplitude and delay of the signal, and $s_{\mathrm{in}}(t)$ and $s_{\mathrm{out}}(t)$ are the input and output of the target source signal, respectively. For equal evaluation, the performance of the proposed method was evaluated by only the effect of a Wiener post-filter.

Because the performance of the proposed method depends on the frequency band, we also used *band-limited Signal to Interference Ratio improvement* (bSIR) to measure the SIR improvement in the particular frequency band Ω defined by

$$\triangle \mathrm{bSIR}(\Omega) = \mathrm{bSIR}_{\mathrm{O},i}(\Omega) - \mathrm{bSIR}_{\mathrm{I},i}(\Omega) \quad (28)$$

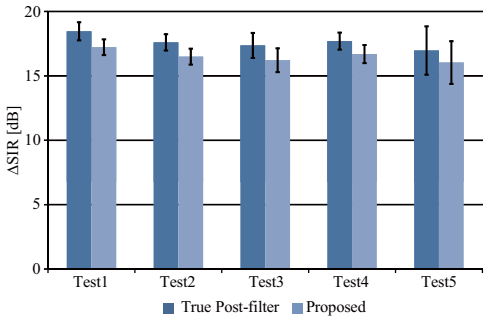

Fig. 5. Average SIR for each test in Table IV.

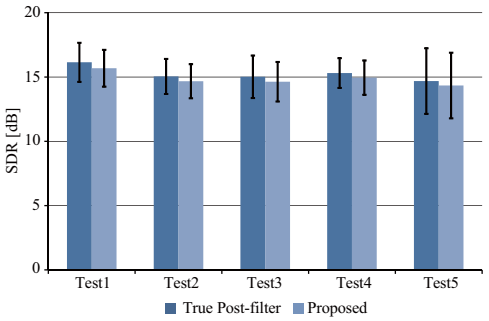

Fig. 6. Average SDR for each test in Table IV.

where

$$\mathrm{bSIR}_{\mathrm{O},i}(\Omega) = 10\log_{10}\frac{\sum_{\omega\in\Omega}|Z_{is_i}(\omega)|^2}{\sum_{\omega\in\Omega}\left|\sum_{j\neq i} Z_{is_j}(\omega)\right|^2}, \quad (29)$$

$$\mathrm{bSIR}_{\mathrm{I},i}(\Omega) = 10\log_{10}\frac{\sum_{\omega\in\Omega}|X_{ms_i}(\omega)|^2}{\sum_{\omega\in\Omega}\left|\sum_{j\neq i} X_{ms_j}(\omega)\right|^2}. \quad (30)$$

Here, $Z_{is_i}(\omega) = \mathcal{F}\{z_{is_i}^2(t)\}$ and $X_{ms_i}(\omega) = \mathcal{F}\{x_{ms_i}^2(t)\}$, where $\mathcal{F}\{\cdot\}$ denotes Fourier transform.

B. Performance Evaluation with Basic Experimental Setting

First of all, we compare the performance of the proposed method with that of the Wiener filter calculated from the true PSD of the target and the interferences (this method is referred to as "true post-filter" hereafter) under the basic parameter set given in Table II. Because the proposed method recovers only the spectral amplitude of the target, the performance of this true post-filter is recognised as the upper boundary

TABLE IV
TYPE AND ANGLE OF SOUND SOURCES

Test	Source1		Source2		Source3	
	ID	Actual	ID	Actual	ID	Actual
1	SP1	3°	SP2	123°	SP3	243°
2	SP1	3°	SP2	123°	MU1	243°
3	SP1	3°	SP2	123°	WN	243°
4	SP1	3°	MU1	123°	MU2	243°
5	SP1	3°	MU1	123°	WN	243°

* Actual: Actual source angle θ_i

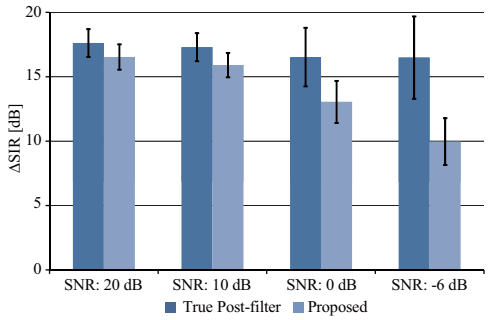

Fig. 7. Average SIR for different input SNRs.

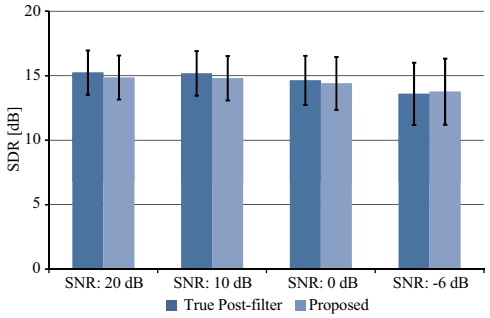

Fig. 8. Average SDR for different input SNRs.

of the performance achieved by recovering only the spectral amplitude. In the basic set, the number of sources is the same as the number of microphones, and the true source angles coincide with the assumed source angles. The input sources and their actual source angles are summarised in Table IV.

Because speech and music signals are nonstationary, we applied Eq. (31) to calculate the expectation where τ is the frame index and α is the forgetting factor set to 0.1.

$$\phi_{Y_l}^{(\tau)}(\omega) \quad = \quad \alpha\phi_{Y_l}^{(\tau-1)}(\omega) + (1-\alpha)|Y_l^{(\tau)}(\omega)|^2 \qquad (31)$$

The Wiener post-filter was calculated by Eq. (15), where one of the assumed angles in Table II was set to be the angle of target Θ_S , and the others were the angles of interferences Θ_N.

Fig. 5 and Fig. 6 show the average SIR and SDR for each test summarised in Table IV. Ten trials (the cylindrically uniform noise was changed in each trial) were performed for each test, and the average was calculated over both trials and sound sources. The error bars show the standard deviation of the estimated value from the average. The proposed method demonstrated performance close to that of the true post-filter in both SIR and SDR. Nevertheless, there is generally a trade-off between the amount of noise reduction (i.e. SIR) and the amount of distortion in the emphasised signal (i.e. SDR); the proposed method achieved high marks for both due to its accurate PSD estimation.

Finally, Figs. 7 and 8 show the average SIR and SDR for different input SNRs. Because the proposed method is only effective for the coherent interference, the performance in reducing interference is degraded when the input SNR is decreased. On the other hand, the degradation in terms of SDR caused by noise is relatively small; the proposed method maintains its performance close to that of the true post-filter. These results lead us to conclude that the proposed method is robust against incoherent noise in terms of target source distortion, whereas the performance in reducing coherent interferences degrades when the input SNR is very low.

C. Performance with Underdetermined Problems

As has been discussed so far, there is a theoretical upper limit to the number of sound sources that the proposed method can separate properly. In this section, we look into the performance of the proposed method regarding the number of sources from an experimental point of view. Because the PSD estimation degrades when $\mathbf{D}(\omega)$ is ill-conditioned, we first measured the condition number [25] of $\mathbf{D}(\omega)$ for different numbers of microphones and sources. The larger the condition number of a matrix is, the more ill-conditioned the matrix becomes. Fig. 9(a) shows the condition number of $\mathbf{D}(\omega)$ for the circular array of $M = 3$ microphones when N is changed. Note that the vertical axis of the graph is expressed in the log-scale, and the set of assumed angles for each N is summarised in Table V. We can read some important findings about the condition of $\mathbf{D}(\omega)$ from this figure.

A) The condition number blows up when N exceeds $M(M-1)+1$. As we theoretically analysed in Sec. III-A, the matrix $\mathbf{D}(\omega)$ is rank deficient without loss of generality if $N > M(M-1)+1$. In Fig. 9(a), the condition number diverged when $N = 8$ or 9 because of $M(M-1)+1 = 7$.

B) The condition number increases in the lower frequency band as the number of sound sources increases. This is due to the effect of the broad directivity gain of the beamformer in the low frequency band, as discussed in Sec. III-C. Since neither the number of microphones nor the radius of the microphone array is changed, directivity gains used to compose $\mathbf{D}(\omega)$ exhibit similar values when N is increased. As Fig. 10(a) shows, the condition number in the lower frequency band was reduced when the radius of the array was doubled.

C) Some local peaks of the condition number are found in the middle to high frequency bands. As we also discussed in Sec. III-C, the spatial aliasing is another cause of the rank deficiency of $\mathbf{D}(\omega)$. When the radius of the array was widened, the number of peaks increased, and the peaks appeared in the lower frequency band. Thus, spatial aliasing is responsible for the peaks of the condition numbers. Next, we confirmed these characteristics of the proposed method by evaluating its performance in sound source separation. Both Fig. 9(b) and

TABLE V
ASSUMED SOURCE ANGLES FOR EACH NUMBER OF SOUND SOURCES

N	Assumed source angles								
3	3	123	243						
4	3	93	183	273					
5	3	75	147	219	291				
6	3	63	123	183	243	303			
7	3	53	93	133	183	233	283		
8	3	48	93	138	183	228	273	318	
9	3	43	83	123	163	203	243	283	323

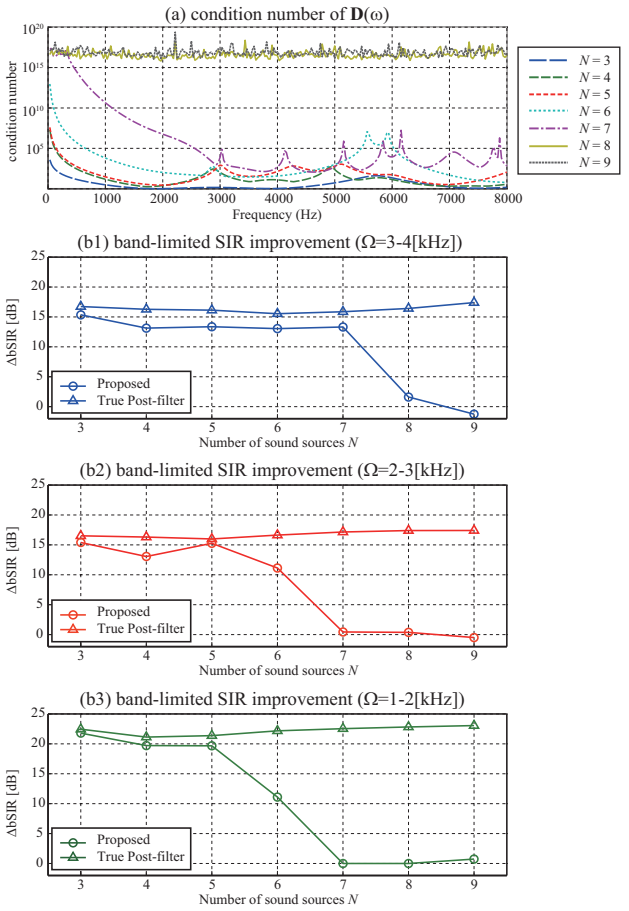

Fig. 9. Effect of number of sound sources on condition number of $\mathbf{D}(\omega)$ and performance of sound source separation: (a) condition number of $\mathbf{D}(\omega)$ at different frequencies, (b1)–(b3) band-limited SIR improvement of different frequency bands when number of sound sources was changed. The radius of microphone array was 4 cm. Note that the vertical axis of graph (a) is expressed in the log-scale.

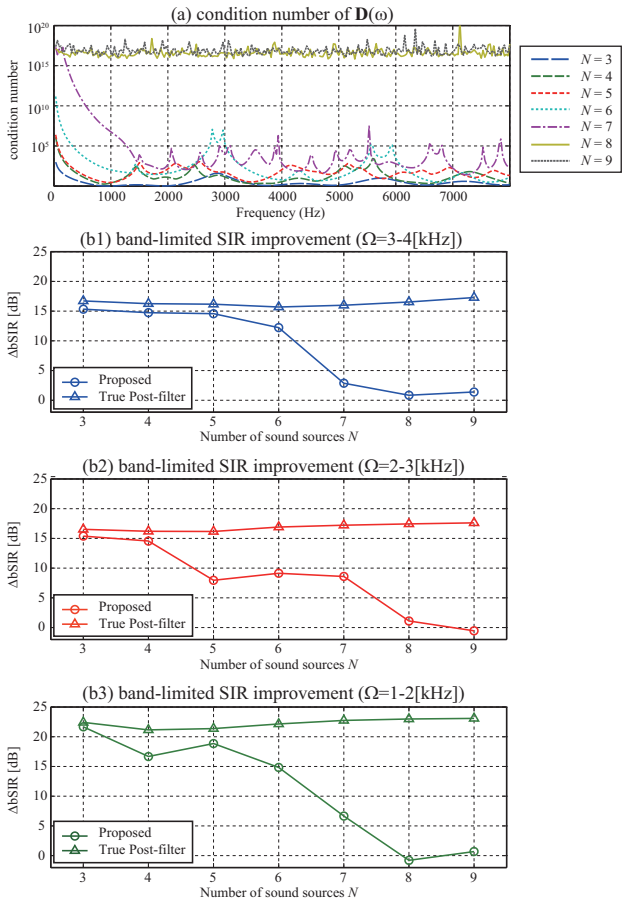

Fig. 10. Effect of number of sound sources on condition number of $\mathbf{D}(\omega)$ and performance of sound source separation: (a) condition number of $\mathbf{D}(\omega)$ at different frequencies, (b1)–(b3) band-limited SIR improvement of different frequency bands when number of sound sources was changed. The radius of the microphone array was 8 cm. Note that the vertical axis of graph (a) is expressed in the log-scale.

Fig. 10(b) show the bSIR of $\Omega = 3-4$ kHz for the source located at $\theta = 3°$ when the number of sound sources was changed. As can be seen in Fig. 9(b), the proposed method maintains its performance up to $N = 7$ but drastically degrades when N exceeds 7, which exactly meets the theoretical upper limit of the number of sound sources (A) stated above. In the lower frequency band, the performance of the method decreases before N reaches the upper limit due to the effect of either/both the broad directivity gain in the low frequency band (B) and/or the spatial aliasing (C). A similar trend can be seen in Fig. 10(b), but the situation is more complicated as the detrimental effect of aliasing arises above 1.5 kHz for $N = 5$ or 6, which prevents the method from maintaining its performance up to $N = 7$.

We further investigated the source separation performance when more microphones were used. Fig. 11 shows the bSIR of $\Omega = 3-4$ kHz for the target source when the number of sound sources was increased where the radius of the microphone array was fixed at $d = 16$ cm. The upper limits of the number of sound sources are indicated by the thicker vertical dashed lines. The proposed method maintained its performance until the number of sound sources reached nearly the upper limit

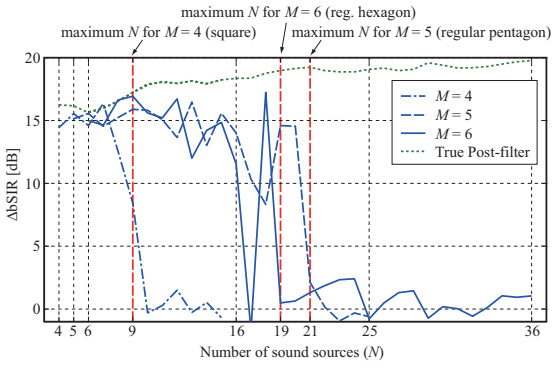

Fig. 11. Band-limited SIR improvement when number of sound sources was changed for different numbers of microphones.

of each M. Actually, we can find some drops just before the upper limit. Such degradation is thought to be the result of using a simple delay-and-sum beamformer (DS) as the fixed beamformer; thus, the performance in this region will be improved if a more appropriate fixed beamformer is used.

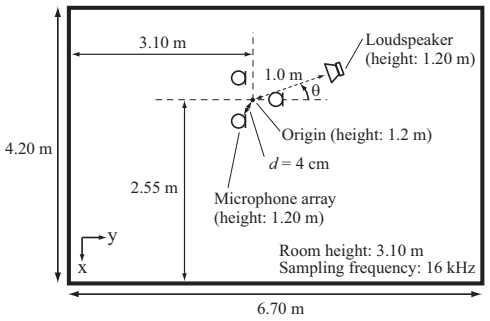

Fig. 12. Setup of microphone array and loudspeaker for experiment in reverberant chamber.

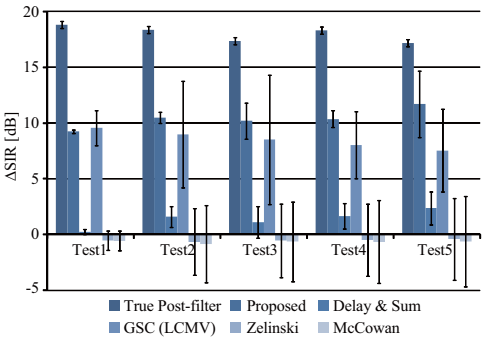

Fig. 13. Average SIR for each test in Table IV measured in experiments.

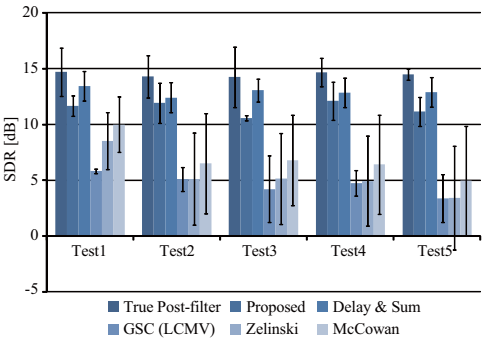

Fig. 14. Average SDR for each test in Table IV measured in experiments.

V. Experimental Results in Practical Environment

Finally, we conducted some experiments in a real acoustic environment to make sure the proposed method was effective in practical use and was superior to the conventional methods. For the conventional beamformings that were compared with the proposed method, we tested a DS and a GSC (equivalent to LCMV). We also examined a Wiener postfilter designed using the techniques of Zelinski *et al.* [14] and McCowan *et al.* [15]. A microphone array with the same arrangement as that used in the simulation with the basic parameter set was installed in a moderately reverberant chamber whose reverberation time was approximately 300 ms. The details of the experimental setup are described in Fig. 12. The same sets of sound sources summarised in Table IV were used in the experiment.

The separation performance for each sound source is summarised in Figs. 13 and 14. Because GSC is an alternative realisation of LCMV [4], we found that the performance of GSC was equivalent to that of LCMV. We can see that the proposed method outperformed GSC for every case in the real acoustic environment due to GSC's sensitivity to errors. As the proposed method is relatively robust against errors, the method maintained its effectiveness in the real acoustic chamber. In contrast, neither the Zelinski nor McCowan postfilters performed well because the model of interferences that these previous techniques assumed was far from the characteristics of coherent interferences.

Finally, Fig. 15 plots the average band-limited SIR for $\Omega = 3 - 4$ kHz when the number of sound sources was changed. Although some fluctuations can be seen in the performance, the proposed method outperformed the conventional methods in noise suppression performance up to $N = 7$.

VI. Concluding Remarks

We have proposed a method for underdetermined sound source separation based on post-filtering. In order to calculate the post-filter, a novel PSD estimation algorithm that exploits the combination of directivity gain to each source angle was introduced. The most prominent advantage of the proposed method is its ability to separate more sound sources than the number of microphones used. Through a theoretical discussion, we found that the maximum number of separable sound sources by the proposed method is $M(M-1)+1$ when the number of microphones is M. However, there are various practical cases where the maximum number of sound sources is reduced. Thus, the beamformer for the PSD estimation and array configuration should be carefully selected. A more theoretical investigation to pursue the best design of the beamformer remains as future work.

In the simulation, we used a simple delay-and-sum beamformer and confirmed that the proposed method performed comparably to the performance of the true post-filter. Although the frequency band was limited, the proposed method was effective for separating up to $M(M-1)+1$ sound sources using M microphones. In a comparison with conventional methods, the proposed method outperformed DS in its noise suppression performance while keeping its robustness against perturbations in practical acoustic environment to which GSC was very sensitive. The experimental results also proved that the proposed method is effective in reducing coherent interferences where the conventional postfiltering techniques do not work well.

Finally, we remark that an attempt to constrain the estimated PSD to be a positive value remains as another future topic of research.

VII. Acknowledgements

The authors would like to express their sincere appreciation to Dr. Tomohiro Nakatani, Dr. Hiroshi Sawada, Dr. Shoko Araki, Dr. Keisuke Kinoshita, and Dr. Mehrez Souden at NTT Communication Science Laboratories for their valuable comments.

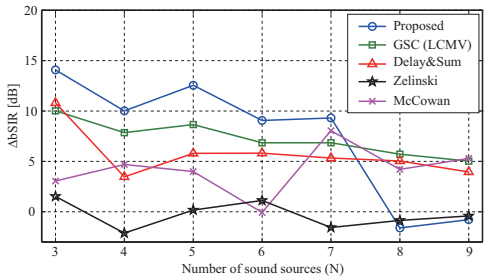

Fig. 15. Effect of number of angles on bSIR in actual reverberant environment.

APPENDIX

When the microphone array is a uniformly spaced linear array, we find several microphone pairs whose orientation and the inter-microphone distance are equal. As described in Fig. 3(d), we have $M-1$ pairs with their distance equal to d, $M-2$ pairs with their distance equal to $2d$, continuing up to 2 pairs with their distance equal to $(M-2)d$. If one component for each distance is left and the number for the conjugates is doubled, the total number of rows that are equal to other rows is

$$2 \times (1 + 2 + \cdots + (M-2)) = (M-2)(1 + (M-2)) = (M-2)(M-1) \quad (32)$$

REFERENCES

[1] J. Benesty, S. Makino, and J. Chen, *Speech enhancement*, ser. Signals and communication technology. Springer, 2005.

[2] M. Brandstein and D. Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*, 1st ed. Springer, Jun. 2001.

[3] K. Kobayashi, Y. Haneda, K. Furuya, and A. Kataoka, "A hands-free unit with noise reduction by using adaptive beamformer," *Consumer Electronics, IEEE Transactions on*, vol. 54, no. 1, pp. 116 –122, february 2008.

[4] D. H. Johnson and D. E. Dudgeon, *Array signal processing : concepts and techniques*. P T R Prentice Hall, Englewood Cliffs, NJ :, 1993.

[5] L. Griffiths and C. Jim, "An alternative approach to linearly constrained adaptive beamforming," *Antennas and Propagation, IEEE Transactions on*, vol. 30, no. 1, pp. 27 – 34, jan 1982.

[6] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, ser. Adaptive and Learning Systems for Signal Processing, Communications and Control Series. Wiley, 2004.

[7] S. Araki, S. Makino, Y. Hinamoto, R. Mukai, T. Nishikawa, and H. Saruwatari, "Equivalence between frequency-domain blind source separation and frequency-domain adaptive beamforming for convolutive mixtures," *EURASIP J. Appl. Signal Process.*, vol. 2003, pp. 1157–1166, Jan. 2003.

[8] P. O'Grady, B. Pearlmutter, and S. Rickard, "Survey of sparse and non-sparse methods in source separation," *International Journal of Imaging Systems and Technology, special issue on Blind Source Separation and Deconvolution in Imaging and Image Processing, Vol. 15, Issue 1, pages 18-33*, July 2005.

[9] A. Jourjine, S. Rickard, and O. Yilmaz, "Blind separation of disjoint orthogonal signals: demixing n sources from 2 mixtures," in *Acoustics, Speech, and Signal Processing, 2000. ICASSP '00. Proceedings. 2000 IEEE International Conference on*, vol. 5, 2000, pp. 2985 –2988.

[10] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Processing*, vol. 81, no. 11, pp. 2353 – 2362, 2001.

[11] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *Signal Processing, IEEE Transactions on*, vol. 52, no. 7, pp. 1830 – 1847, 2004.

[12] S. Araki, H. Sawada, R. Mukai, and S. Makino, "Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors," *Signal Processing*, vol. 87, no. 8, pp. 1833–1847, 2007.

[13] R. J. Weiss, M. I. M, and D. P. W. Ellis, "Source separation based on binaural cues and source model constraints," in *in Proc. Interspeech*, 2008, pp. 419–422.

[14] R. Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms," in *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on*, apr 1988, pp. 2578 –2581 vol.5.

[15] I. McCowan and H. Bourlard, "Microphone array post-filter based on noise field coherence," *Speech and Audio Processing, IEEE Transactions on*, vol. 11, no. 6, pp. 709 – 716, nov. 2003.

[16] S. Fischer and K. U. Simmer, "Beamforming microphone arrays for speech acquisition in noisy environments," *Speech Communication*, vol. 20, no. 3-4, pp. 215 – 227, 1996.

[17] S. Lefkimmiatis, D. Dimitriadis, and P. Maragos, "An optimum microphone array post-filter for speech applications." in *INTERSPEECH*. ISCA, 2006.

[18] T. Yoshioka and T. Nakatani, "A microphone array system integrating beamforming, feature enhancement, and spectral mask-based noise estimation," in *Hands-free Speech Communication and Microphone Arrays (HSCMA), 2011 Joint Workshop on*, 30 2011-june 1 2011, pp. 219 –224.

[19] A. Kamkar-Parsi and M. Bouchard, "Improved noise power spectrum density estimation for binaural hearing aids operating in a diffuse noise field environment," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 17, no. 4, pp. 521 –533, may 2009.

[20] H. Dam, S. Nordholm, H. Dam, and S. Low, "Postfiltering using multichannel spectral estimation in multispeaker environments," *EURASIP Journal on Advances in Signal Processing*, 2008.

[21] H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Speech enhancement using nonlinear microphone array based on complementary beamforming (special section on digital signal processing)," *IEICE transactions on fundamentals of electronics, communications and computer sciences*, vol. 82, no. 8, pp. 1501–1510, 1999.

[22] ——, "Speech enhancement using nonlinear microphone array based on noise adaptive complementary beamforming," *IEICE transactions on fundamentals of electronics, communications and computer sciences*, vol. 83, no. 5, pp. 866–876, 2000.

[23] S. Araki and S. Makino, *Speech Enhancement*, 1st ed. Springer, 2005, ch. 14, pp. 329–352.

[24] S. Araki, H. Sawada, R. Mukai, and S. Makino, "Underdetermined sparse source separation of convolutive mixtures with observation vector clustering," *Proceedings of ISCAS 2006*, pp. 21–24, 2006.

[25] G. Strang, *Linear Algebra and Its Applications*. Brooks Cole, February 1988.

Yusuke Hioka (S'04-M'05-SM'12) received his B.E., M.E., and Ph.D. degrees in electrical engineering in 2000, 2002, and 2005 from Keio University, Yokohama, Japan. From 2005 to 2012, he was with the NTT Cyber Space Laboratories, Nippon Telegraph and Telephone Corporation (NTT). From 2010 to 2011, he was also a Visiting Research Fellow at Massey University and a visiting researcher at Victoria University of Wellington, both located in Wellington, New Zealand. In 2013 he joined the Department of Electrical and Computer Engineering at the University of Canterbury, Christchurch, New Zealand where he is currently a Lecturer. His research interests include microphone array signal processing and room acoustics. He is also a member of the IEICE and the Acoustical Society of Japan (ASJ).

Ken'ichi Furuya (M'96-SM'10) received his B.E. and M.E. degrees in acoustic design from Kyushu Institute of Design, Fukuoka, Japan, in 1985 and 1987, and his Ph.D. degree from Kyushu University, Japan, in 2005. From 1987 to 2012, he was with the laboratories of Nippon Telegraph and Telephone Corporation (NTT), Tokyo, Japan. In 2012, he joined the Department of Computer Science and Intelligent Systems of Oita University, Oita, Japan, where he is currently a Professor. His current research interests include signal processing in acoustic engineering. Dr. Furuya was awarded the Sato Prize by the ASJ in 1991. He is a member of the ASJ, the Acoustical Society of America, and IEICE.

Kazunori Kobayashi received the B.E., M.E., and Ph.D. degrees in electrical and electronic system engineering from Nagaoka University of Technology in 1997, 1999, and 2003. Since joining Nippon Telegraph and Telephone Corporation (NTT) in 1999, he has been engaged in research on microphone arrays, acoustic echo cancellers and hands-free systems. He is now Senior Research Engineer of NTT Media Intelligence Laboratories. He is a member of the IEICE and ASJ.

Kenta Niwa received his B.E. and M.E. degrees from Nagoya University in 2006 and 2008. Since joining Nippon Telegraph and Telephone Corporation (NTT) in 2008, he has been engaged in research on microphone arrays. Currently, he is a researcher at NTT Cyber Space Laboratories. He was awarded the Awaya Prize by the Acoustical Society of Japan (ASJ) in 2010. He is also a member of the IEICE and ASJ.

Yoichi Haneda (A'92-M'97-SM'06) received the B.S., M.S., and Ph.D. degrees from Tohoku University, Sendai, in 1987, 1989, and 1999. From 1989 to 2012, he was with the Nippon Telegraph and Telephone Corporation (NTT), Japan. In 2012, he joined the University of Electro-Communications, where he is a Professor. His research interests include modeling of acoustic transfer functions, microphone arrays, loudspeaker arrays, and acoustic echo cancellers. He received paper awards from the ASJ and from the IEICE of Japan in 2002. Dr. Haneda is a senior member of the IEICE and the ASJ.