Diffusion Noise Suppression by Crystal-Shape Subtraction Array

Akira Tanaka* and Ryo Takahashi*
* Division of Computer Science, Hokkaido University,
N14W9, Kita-ku, Sapporo, 060-0814 Japan
E-mail: {takira,ryo-t}@main.ist.hokudai.ac.jp Tel: +81-11-706-6809

***Abstract*—Noise suppression of diffusion noise by microphone arrays is discussed in this paper. In our previous work, we proposed a method for jointly estimating signal and noise correlation matrices from observations with diffusion noise by using so-called crystal shape microphone arrays; and discussed the performance of the Wiener filter based on those correlation matrices. In this paper, we propose a novel method for noise suppression of diffusion noise based on the newly adopted spectral subtraction scheme with the estimated correlation matrices by our previous work. We also verify the efficacy of the proposed method by some computer simulations and show that the proposed method outperforms our previous method by the Wiener filter.**

I. Introduction

Noise suppression for audio signals is one of important topics in the field of speech and acoustic signal processing. A microphone-array-based noise suppression scheme is known as one of effective approaches for this problem. The minimum variance distortionless response filter (MVDRF) [1] and the Wiener filter (WF) (see [2], [3] for instance) are representative linear methods in the scheme. As one of array-based non-linear noise suppressors, Takahashi et al., proposed a novel method in [4], named 'spatial subtraction array with independent component analysis (ICA-SSA)', in which the amplitude spectrum of noise signal is estimated by the ICA and the noise suppression is conducted by the spectral subtraction (SS) [5] with the estimate. It is reported in [4] that an additive noise can be effectively suppressed by this method. However, there exist some drawbacks in this method such as 1) the number of target signals is assumed to be unity, 2) the sum of the number of target signals and the number of noise signals must be less than or equal to the number of microphones, and 3) noise signals are basically assumed to be directional ones.

Recently, Ono et al. clarified that eigenvectors of noise correlation matrices for diffusion noise are invariant with specific crystal-shape microphone arrays [6], [7], [2]. On the basis of the knowledge, they also succeeded in widening the application area of the WF [2]. Moreover, we introduced a method for jointly estimating signal and correlation matrices on the basis of their idea; and improved the performance of the WF [3].

In this paper, we construct a novel non-linear noise suppressor for diffusion noise, named 'crystal-shape subtraction array (CSSA)' which incorporates the SS and the estimation of the amplitude spectrum of noise by the method proposed in [3]. This method can be applied to the cases where the number of microphones is equal to the number of target signals in which the ICA-SSA can not be used. We also verify the efficacy of the proposed method by some computer simulations and show that the proposed method outperforms our previous method by the Wiener filter [3].

This work was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (C), 24500001.

II. Problem Formulation and Some Preliminaries

Let n, m $(m \leq n)$, t, ω be the number of observations (microphones), the number of target signals, and the time and frequency bin indices in the short time Fourier domain, respectively. Note that $m = n$ is permitted. Let $\boldsymbol{s}(t,\omega) \in \mathbf{C}^m$, $\boldsymbol{n}(t,\omega) \in \mathbf{C}^n$, and $A(\omega) \in \mathbf{C}^{n\times m}$ be an unknown target signal vector, an observation noise vector, and an observation matrix consisting of steering vectors of $\boldsymbol{s}(t,\omega)$ (or corresponding to a mixing matrix related with impulse responses between the microphones and the target signal sources) with $\mathrm{rank}(A) = m$, where $\mathbf{C}^n$ and $\mathbf{C}^m$ are n-dimensional and m-dimensional unitary spaces. We assume that an observation vector $\boldsymbol{x}(t,\omega) \in \mathbf{C}^n$ is given by the following model:

$$\boldsymbol{x}(t,\omega) = A(\omega)\boldsymbol{s}(t,\omega) + \boldsymbol{n}(t,\omega). \tag{1}$$

The aim of the noise suppression is to estimate the unknown target signal vectors $\boldsymbol{s}(t,\omega)$ or corresponding waveforms. Note that $A(\omega)$ can be estimated by DOA estimation methods in case of $n > m$, or noisy-BSS methods such as [8] for the case of $m = n$. In this paper, we assume that $A(\omega)$ is given. We also assume that the target signal vector $\boldsymbol{s}(t,\omega)$ and the observation noise vector $\boldsymbol{n}(t,\omega)$ is uncorrelated each other, that is

$$E_t[\boldsymbol{s}(t,\omega)\boldsymbol{n}^*(t,\omega)] = O_{m,n} \tag{2}$$

holds for each ω, where the superscript * denotes the adjoint (conjugate and transposition) operator; E_t denotes the expectation operator over t; and $O_{m,n}$ denotes the m by n zero matrix. On the basis of the assumption Eq.(2), the correlation matrix of the observation vector $\boldsymbol{x}(t,\omega)$ is reduced to

$$\begin{aligned} X(\omega) &= E_t[\boldsymbol{x}(t,\omega)\boldsymbol{x}^*(t,\omega)] \\ &= A(\omega)R(\omega)A^*(\omega) + Q(\omega), \end{aligned} \tag{3}$$

where $R(\omega)$ and $Q(\omega)$ denote the correlation matrices of the unknown target signal vector and the noise vector, defined as

$$\begin{aligned} R(\omega) &= E_t[\boldsymbol{s}(t,\omega)\boldsymbol{s}^*(t,\omega)], \\ Q(\omega) &= E_t[\boldsymbol{n}(t,\omega)\boldsymbol{n}^*(t,\omega)], \end{aligned}$$

respectively. In this paper, we assume that $R(\omega)$ is diagonal for each ω, which implies that the target signals are uncorrelated each other. We also assume that the noise vector $\boldsymbol{n}(t,\omega)$ is obtained by observation of diffusion noise. In [7], the diffusion noise is defined by 1) the power spectrum of each microphone is identical, and 2) the cross-power spectrum only depends on the distance between corresponding two microphones. Let

$$Q(\omega) = P(\omega)\Lambda(\omega)P^*(\omega) \tag{4}$$

be the eigenvalue decomposition of $Q(\omega)$. When we observe the diffusion noise by a so-called crystal-shape array, the unitary matrix $P(\omega)$ is reduced to a constant matrix as shown in [7]. In this paper, we use a crystal-shape microphone array, which implies that we can assume that $P(\omega)$ is given.

III. Crystal-Shape-Array-Based Wiener Filtering

In this section, we give a overview of the method proposed in [3]. Firstly, we give some definitions and a proposition as preliminaries.

Definition 1: [9] Let $A = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m]$, $\boldsymbol{a}_i \in \mathbf{C}^n$, then the vectored version of A is defined as

$$\mathrm{vec}(A) = [\boldsymbol{a}_1', \ldots, \boldsymbol{a}_m']' \in \mathbf{C}^{mn}, \tag{5}$$

where the superscript $'$ denotes the transposition operator.

Definition 2: [9] Let $A \in \mathbf{C}^{p\times q}$ and $B \in \mathbf{C}^{m\times n}$ be arbitrary matrices and $B = (b_{ij})$, then the Kronecker product of B and A is defined as

$$B \otimes A = \begin{bmatrix} b_{11}A & \cdots & b_{1n}A \\ \vdots & \ddots & \vdots \\ b_{m1}A & \cdots & b_{mn}A \end{bmatrix} \in \mathbf{C}^{mp\times nq}. \tag{6}$$

Proposition 1: [9] Let M_1, M_2, and M_3 be arbitrary matrices such that the product $M_1M_2M_3$ is defined. Then,

$$\mathrm{vec}(M_1M_2M_3) = (M_3' \otimes M_1)\mathrm{vec}(M_2) \tag{7}$$

holds.

Definition 3: Let $Z_n \in \mathbf{C}^{(n^2-n)\times n^2}$ be the matrix that extracts non-diagonal elements of 'vec'-ed version of a square matrix of degree n.

For example, Z_2 is given by

$$Z_2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{8}$$

In fact, for

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

we have

$$Z_2\mathrm{vec}(M) = \begin{bmatrix} c \\ b \end{bmatrix}.$$

From Eqs.(3) and (4), and Proposition 1, we obtain

$$\begin{aligned} \mathrm{vec}(X(\omega)) &= (\overline{A(\omega)} \otimes A(\omega))\mathrm{vec}(R(\omega)) \\ &\quad +(\overline{P(\omega)} \otimes P(\omega))\mathrm{vec}(\Lambda(\omega)), \end{aligned} \tag{9}$$

where the overline denotes the conjugate operator. Also from the assumption that $R(\omega)$ and $\Lambda(\omega)$ are diagonal, we obtain

$$Z_m\mathrm{vec}(R(\omega)) = \mathbf{0}_{m^2-m}, \tag{10}$$

$$Z_n\mathrm{vec}(\Lambda(\omega)) = \mathbf{0}_{n^2-n}, \tag{11}$$

where $\mathbf{0}_n$ denotes the zero vector in $\mathbf{C}^n$. From Eqs.(9), (10), and (11), we have the linear equation

$$G(\omega)\begin{bmatrix} \mathrm{vec}(R(\omega)) \\ \mathrm{vec}(\Lambda(\omega)) \end{bmatrix} = \begin{bmatrix} \mathrm{vec}(X(\omega)) \\ \mathbf{0}_{m^2-m} \\ \mathbf{0}_{n^2-n} \end{bmatrix}, \tag{12}$$

where

$$G(\omega) = \begin{bmatrix} (\overline{A(\omega)} \otimes A(\omega)) & (\overline{P(\omega)} \otimes P(\omega)) \\ Z_m & O_{m^2-m,n^2} \\ O_{n^2-n,m^2} & Z_n \end{bmatrix}. \tag{13}$$

Thus, we have the estimates $\hat{R}(\omega)$ and $\hat{\Lambda}(\omega)$ of $R(\omega)$ and $\Lambda(\omega)$ by solving Eq.(12); and substituting $\hat{\Lambda}(\omega)$ to Eq.(4) yields the estimate $\hat{Q}(\omega)$ of $Q(\omega)$.

Note that the matrix $G(\omega)$ must be full column rank so that $\mathcal{R}(G^*(\omega))$ (the linear subspace spanned by the column vectors of $G^*(\omega)$) is identical to $\mathbf{C}^{m^2+n^2}$, or the estimates $\hat{R}(\omega)$ and $\hat{\Lambda(\omega)}$ have some biases from their true matrices.

Based on these estimates, we can construct the WF as

$$B_{WF}(\omega) = R(\omega)A^*(\omega)(A(\omega)R(\omega)A^*(\omega) + Q(\omega))^+, \tag{14}$$

where the superscript $^+$ denotes the Moore-Penrose generalized inverse [10]; and the final estimate of $\boldsymbol{s}(t,\omega)$ is given as

$$\hat{\boldsymbol{s}}(t,\omega) = B_{WF}(\omega)\boldsymbol{x}(t,\omega). \tag{15}$$

Hereafter, we call this method 'Crystal-shape-array-based Wiener Filter' which is abbreviated by CWF.

IV. The Proposed Method

The key idea of the proposed method is adopting the SS for noise suppression in which the amplitude spectrum of the observation noise is estimated by the method shown in the previous section.

Firstly, we obtain the estimated target signals by

$$\begin{aligned} \boldsymbol{y}(t,\omega) &= A^+(\omega)\boldsymbol{x}(t,\omega) \\ &= \boldsymbol{s}(t,\omega) + A^+(\omega)\boldsymbol{n}(t,\omega). \end{aligned} \tag{16}$$

Note that $A^+(\omega)$ is a left inverse matrix of $A(\omega)$, that is,

$$A^+(\omega)A(\omega) = I_m, \tag{17}$$

where I_m denotes the identity matrix of degree m since $A(\omega)$ is full column rank matrix. Let $y_i(t,\omega)$ be the i-th element of $\boldsymbol{y}(t,\omega)$ and let $\tilde{n}_i(t,\omega)$ be the i-th element of $\tilde{\boldsymbol{n}}(t,\omega) = A^+(\omega)\boldsymbol{n}(t,\omega)$. The correlation matrix of $\tilde{\boldsymbol{n}}(t,\omega)$ is trivially given as

$$\tilde{Q}(\omega) = A^+\hat{Q}(\omega)(A^+)^*, \tag{18}$$

where $\hat{Q}(\omega)$ is the estimated correlation matrix of $\boldsymbol{n}(t,\omega)$ given by the method shown in the previous section. Thus, the power spectrum of the noise included in $y_i(t,\omega)$ is given as the i-th diagonal element of $\tilde{Q}(\omega)$, written as $Q_{ii}(\omega)$. Accordingly, the amplitude spectrum of the noise included in $y_i(t,\omega)$ is given as

$$|\tilde{n}_i(t,\omega)| = \sqrt{\tilde{Q}_{ii}(\omega)}. \tag{19}$$

We conduct the SS for noise suppression by using this estimate $|\tilde{n}_i(t,\omega)|$. Final estimated i-the target signal $\hat{s}_i(t,\omega)$ is given by

$$\hat{s}_i(t,\omega) = \max(|y_i(t,\omega)| - |\tilde{n}_i(t,\omega)|, 0)\frac{y_i(t,\omega)}{|y_i(t,\omega)|}. \tag{20}$$

We call this method 'Crystal-Shape Subtraction Array with INVerse' abbreviated by CSSA-INV.

When the power of the noise in $\boldsymbol{y}(t,\omega)$ is comparatively large, which may be caused by small singular values in $A(\omega)$, the performance of the CSSA-INV may be deteriorated. In such cases, we can adopt $B_{WF}(\omega)$ in Eqs.(16) and (18) instead of $A^{+}(\omega)$. We abbreviate the proposed crystal-shape subtraction array using Wiener filter $B_{WF}(\omega)$ as CSSA-WF.

Note that we can adopt more sophisticated SS in Eq.(20). Also note that the fidelity of waveforms by the proposed method may deteriorated since we use the phase spectrum of $y_i(t,\omega)$ for $\hat{s}_i(t,\omega)$ as the same with the general SS scheme.

V. Computer Simulations

In this section, we numerically investigate the performance of the proposed methods (CSSA-INV and CSSA-WF) by comparing with CWF and the simple inverse filtering (INV), that is, $A^{+}(\omega)$ is used for B_{WF} in Eq.(15), in computer simulations.

Let $m = n = 3$ and we adopt a regular-triangle-shape array for a crystal-shape microphone array in which the distance between two microphones is set to 9.086 cm. As target signals we use three music samples of 3.0 s with $f_s = 44.1$ kHz shown in Fig.1. The layout of the microphone array and the directions of the target signals is shown in Fig.2. Note that ICA-SSA can not deal with this setting since $m = n$. As the noise signals, we use temporary white Gaussian noise whose spatial correlation matrix in the time domain is given by

$$Q = \sigma^2 \begin{bmatrix} 1.0 & 0.8 & 0.8 \\ 0.8 & 1.0 & 0.8 \\ 0.8 & 0.8 & 1.0 \end{bmatrix}, \tag{21}$$

where σ^2 is the variance of the noise. The short-time Fourier transforms are conducted with the frame size of 512 samples; the half-shift; and the hamming window.

As the evaluation measures, we use the waveform-based SNR defined as

$$S_w = \max_{\alpha} 10\log \sum_{t=1}^{T} \frac{||\boldsymbol{s}(t)||^2}{||\boldsymbol{s}(t) - \alpha\hat{\boldsymbol{s}}(t)||^2} \tag{22}$$

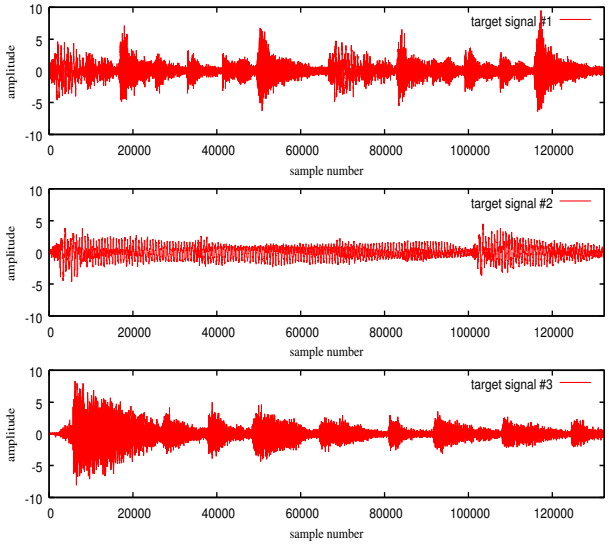

Fig. 1. The target signals.

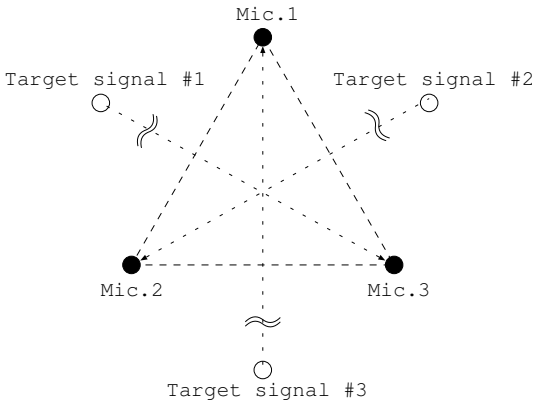

Fig. 2. The layout of the microphone array and the directions of the target signals.

and the amplitude-spectrum-based SNR defined as

$$S_s = \max_{\alpha} 10\log \sum_{\omega=1}^{T} \frac{|(F\boldsymbol{s})(\omega)|^2}{(|(F\boldsymbol{s})(\omega)| - \alpha|(F\hat{\boldsymbol{s}})(\omega)|)^2}, \tag{23}$$

where T denotes the number of samples; $\boldsymbol{s}(t)$ and $\hat{\boldsymbol{s}}(t)$ denotes the waveforms corresponding to $\boldsymbol{s}(t,\omega)$ and $\hat{\boldsymbol{s}}(t,\omega)$; and $F\boldsymbol{s}$ and $F\hat{\boldsymbol{s}}$ denotes the Fourier transforms of $\boldsymbol{s}(t)$ and $\hat{\boldsymbol{s}}(t)$, respectively. Note that α is the regression coefficient to absorb the scale differences.

Figure 3 shows the transition of S_w by the INV, CWF, CSSA-INV, and CSSA-WF with respect to the SNR of the observations defined as

$$S_o = 10\log \sum_{t=1}^{T} \frac{||\boldsymbol{x}(t) - \boldsymbol{n}(t)||^2}{||\boldsymbol{n}(t)||^2}, \tag{24}$$

where $\boldsymbol{x}(t)$ and $\boldsymbol{n}$ denotes the waveforms corresponding to $\boldsymbol{x}(t,\omega)$ and $\boldsymbol{n}(t,\omega)$; and Figure 4 show those of S_s. Figure 5 shows the three observations with $So = 9.93$ dB and Figure

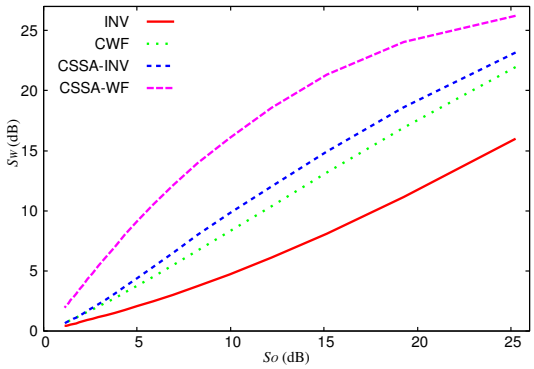

Fig. 3. Transition of S_w by each method with respect to S_o.

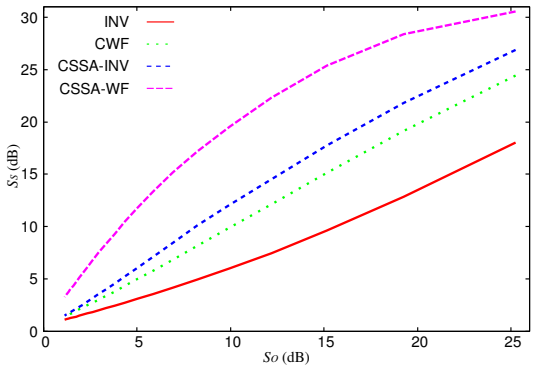

Fig. 4. Transition of S_s by each method with respect to S_o.

6 shows the estimated target signal #1 corresponding to the upper graph in Fig.1.

From these results, it is confirmed that the proposed CSSA outperforms the CWF (and INV) in this setting and the CSSA-WF gives the best performance among adopted competitors.

VI. Conclusion

In this paper, we proposed a novel noise suppression method, named crystal-shape subtraction array, that is based on the noise estimation scheme by crystal-shape microphone arrays and the spectral subtraction. We also investigated the performance of the proposed method by computer simulations and confirmed that the proposed method outperforms the conventional ones.

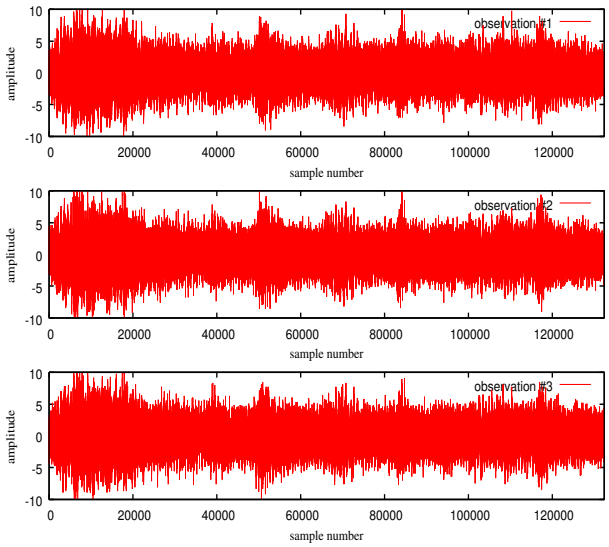

Fig. 5. The observed signals.

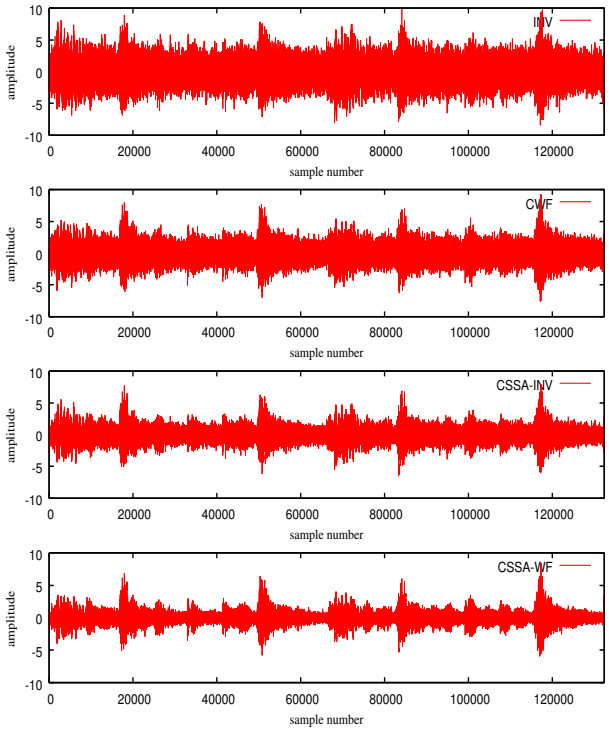

Fig. 6. The estimated signals #1 (top:INV / 2nd.:CWF / 3rd.:CSSA-INV / bottom:CSSA-WF).

References

[1] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[2] N. Ito, N. Ono, and S. Sagayama, "A blind noise decorrelation approach with crystal arrays on designing post-filters for diffuse noise suppression," in *Proc. ICASSP*, 2008, pp. 317–320.

[3] A. Tanaka and M. Miyakoshi, "Joint estimation of signal and noise correlation matrices and its application to inverse filtering," in *Proc. ICASSP*, 2009, pp. 2181–2184.

[4] Y. Takahashi, T. Takatani, H. Saruwatari, and K. Shikano, "Robust spatial subtraction array with independent component analysis for speech enhancement," in *Proc. ISSPA*, 2007, pp. 1–4.

[5] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 27, no. 2, pp. 113–120, 1979.

[6] H. Shimizu, N. Ono, K. Matsumoto, and S. Sagayama, "Isotropic noise suppression on power spectrum domain by symmetric microphone array," in *Proc. WASPAA*, 2007, pp. 54–57.

[7] N. Ono, N. Ito, and S. Sagayama, "Five classes of crystal arrays for blind decorrelation of diffuse noise," in *Proc. SAM*, 2008, pp. 151–154.

[8] A. Tanaka, H. Imai, and M. Miyakoshi, "Noisy bss based on joint diagonalization of differences of correlation matrices," *Proc. IASTED SIP*, pp. 368–373, 2008.

[9] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, 1988.

[10] C. R. Rao and S. K. Mitra, *Generalized Inverse of Matrices and its Applications*, John Wiley & Sons, 1971.