Diffuse Noise Suppression Using Crystal-shaped Microphone Arrays

Nobutaka Ito, *Student Member, IEEE*, Hikaru Shimizu, *Non-Member, IEEE*, Nobutaka Ono, *Member, IEEE*, and Shigeki Sagayama, *Member, IEEE*

***Abstract*— This paper describes novel methods for diffuse noise suppression using crystal-shaped microphone arrays. The two-stage processing of the observed signals by the Minimum Variance Distortionless Response (MVDR) beamformer and the subsequent Wiener post-filter is effective for diffuse noise suppression and gives the Linear Minimum Mean Square Error (LMMSE) estimator of the target signal. It is essential in this framework to accurately estimate the power spectrogram and the steering vectors of the target signal from the noisy observations. Our methods diagonalize the spatial noise covariance matrix and utilizes the denoised off-diagonal entries of the spatial covariance matrix to accurately estimate the power spectrogram and the steering vectors of the target signal. We employ *crystal arrays*, certain classes of crystal-shaped array geometries, which make it possible to diagonalize the unknown noise covariance matrix by a constant unitary matrix regardless of its value as long as noise meets an isotropy condition. It is shown through experiments with simulated and real environmental noise that the proposed methods outperform previous methods substantially for real world noise and in the presence of reverberation.**

***Index Terms*— Diffuse noise, microphone arrays, noise suppression, post-filtering, reverberation.**

I. Introduction

MUCH research has been devoted to microphone array signal processing for noise suppression. Beamforming techniques suppress noise mainly based on spatial information by forming directivity. The fundamental delay-and-sum beamformer requires a aperture large enough compared to the wavelength in order to form sharp directivity. Adaptive beamformers such as the MVDR beamformer [1] effectively suppress noise arriving only from a few point sources regardless of the aperture, by forming zeros in the directivity pattern in the direction of the noise sources. However, they do not sufficiently suppress diffuse noise arriving from many directions, encountered, *e.g.* at cocktail parties or in vehicles.

Recently an approach of post-filtering, *i.e.* time-frequency masking at the output of a beamformer, has been studied [2]–[11]. This framework is suitable for suppression of diffuse noise thanks to the time-frequency masking. Among all, it has been shown by Simmer *et al.* [5] and Van Trees [12] that the LMMSE estimate of the target signal is obtained by the MVDR beamformer followed by a time-frequency mask called the Wiener post-filter [5], [7], [11]. In the design of the Wiener post-filter, it is essential to accurately estimate the power spectrogram or equivalently the short-time autocorrelation function of the target signal from the observed noisy signals at the microphones.

Zelinski [2] proposed estimating the target autocorrelation function from the inter-channel observation cross-correlation functions, which are noise-free under the assumption that noise components in different channels are uncorrelated. Since the assumption is valid only when the distances between microphones are large enough compared to the wavelength, the method does not work well with a small-aperture array or at low frequencies. Aiming at effective noise suppression with a small-aperture array, McCowan *et al.* [7] proposed estimating the target power spectrogram assuming that the inter-channel noise coherences are known. This is the case for some idealistic noise models such as spherically isotropic noise model where uncorrelated noise waves with an equal power spectrogram propagate in all directions and the coherence function is given by a sinc function [13]. However, noise environments in the real world do not always obey such idealistic models because of the distribution of the noise sources, the room shape, the diffraction by a rigid mount, *etc.* Consequently, the assumption of explicit values of noise coherences may cause significant errors in the estimation.

In this paper, we present a novel design of the Wiener post-filter [14], aiming at effective diffuse noise suppression in the real world. Our design is based on diagonalization of the spatial noise covariance matrix, *i.e.* noise decorrelation, and estimation of the target power spectrogram using the denoised off-diagonal entries of the spatial observation covariance matrix. Certain classes of symmetrical array geometries allow us to diagonalize it in a blind manner by a constant unitary matrix independent of its value, under an assumption related to noise isotropy. Taking into account inter-channel noise correlation and performing noise decorrelation, the proposed method suppresses diffuse noise well even with small-aperture arrays unlike Zelinski's method. Besides, not assuming explicit noise coherences but only isotropy, our method based on the blind diagonalization has the potential of being robust against deviation of noise from the sinc coherence model encountered, *e.g.* when microphones are mounted on a symmetical object.

Although these methods for diffuse noise suppression as-

N. Ito is with INRIA Rennes - Bretagne Atlantique, Campus de Beaulieu, 35042 Rennes Cedex, France. He is also with the Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan. (Tel: +33-2-9984-7524. e-mail: nito@irisa.fr; ito@hil.t.u-tokyo.ac.jp.)

H. Shimizu was with the Graduate School of Information Science and Technology, The University of Tokyo. He is now with Mitsubishi Corporation, 3-1, Marunouchi 2-chome, Chiyodaku, Tokyo, Japan (e-mail: hikaru/shimizu@mitsubishicorp.com).

N. Ono and S. Sagayama are with the Graduate School of Information Science and Technology, The University of Tokyo (e-mail: {onono,sagayama}@hil.t.u-tokyo.ac.jp).

sume that the steering vectors are known, their accurate estimation from the observed signals is important for the effective noise suppression in the real world, especially when there is a high reverberation. In a general transfer function generalized sidelobe canceller proposed by Gannot *et al.* [15], normalized transfer functions are estimated on the assumption that noise is stationary for a longer period compared to the target signal. Methods proposed by Benesty *et al.* [16] and Doclo *et al.* [17] manage to avoid the problem by calculating a multichannel noise suppression filter using solely the spatial covariance matrices of the observed signals and noise in order to estimate the spatial image(s) of the target signal. In this paper, we also present a method for the joint estimation of the steering vectors and the short-time power spectra of the target signal from the denoised off-diagonal entries of the spatial observation covariance matrix. This is the main novelty of the paper compared to our previous conference papers [14], [18], [19].

We use the following notation throughout the paper. The superscripts *, $^\mathsf{T}$, and $^\mathsf{H}$ denote complex conjugation, transposition, and Hermitian transposition, respectively. Signals are represented in the time-frequency domain with τ and ω representing the frame index and the angular frequency, respectively. The cross-spectrum of scalar signals $\alpha(\tau,\omega)$ and $\beta(\tau,\omega)$ is denoted by

$$\phi_{\alpha\beta}(\tau,\omega) \triangleq \mathcal{E}[\alpha(\tau,\omega)\beta^*(\tau,\omega)], \tag{1}$$

and the covariance matrix of a zero-mean random vector $\boldsymbol{\gamma}(\tau,\omega)$ by

$$\boldsymbol{\Phi}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}(\tau,\omega) \triangleq \mathcal{E}[\boldsymbol{\gamma}(\tau,\omega)\boldsymbol{\gamma}^\mathsf{H}(\tau,\omega)], \tag{2}$$

where $\mathcal{E}[\cdot]$ denotes expectation. We denote

$$\mathrm{circ}(\alpha_1,\alpha_2,\dots,\alpha_k) \triangleq \begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_k \\ \alpha_k & \alpha_1 & \dots & \alpha_{k-1} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_2 & \alpha_3 & \dots & \alpha_1 \end{bmatrix}, \tag{3}$$

and

$$\boldsymbol{F}_k \triangleq \frac{1}{\sqrt{k}} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & \zeta_k & \dots & \zeta_k^{k-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \zeta_k^{k-1} & \dots & \zeta_k^{(k-1)(k-1)} \end{bmatrix}, \tag{4}$$

where $\zeta_k \triangleq \mathrm{e}^{\mathrm{j}2\pi/k}$.

The rest of the paper is organized as follows. Section II reviews the Wiener post-filter and its previous designs. In Section III, we describe the proposed method for designing the Wiener post-filter and also the joint estimation of the short-time power spectra and the steering vectors of the target signal. In Section IV, five classes of array geometries which enable the diagonalization of the noise covariance matrix are presented. We show some results of experiments with simulated and real environmental noise in Section V and conclude in Section VI.

II. Review of Wiener Post-filtering Techniques

A. Observation Model

We assume that an array of M microphones receives a target signal emitted from a point source in the presence of diffuse noise and reverberation. Let $s(\tau,\omega)$ be the target signal at a reference microphone (let it be the first microphone without loss of generality in the following), and $x_m(\tau,\omega)$ and $v_m(\tau,\omega)$ be the observed signal and the diffuse noise component at the m-th microphone.

The target source is assumed to be static and the transfer function from $s(\tau,\omega)$ to $x_m(\tau,\omega)$ is denoted by $h_m(\omega)$. Therefore, our observation model is given by

$$x_m(\tau,\omega) = s(\tau,\omega)h_m(\omega) + v_m(\tau,\omega). \tag{5}$$

This equation can be rewritten in a vector form as

$$\boldsymbol{x}(\tau,\omega) = s(\tau,\omega)\boldsymbol{h}(\omega) + \boldsymbol{v}(\tau,\omega), \tag{6}$$

where $\boldsymbol{x}(\tau,\omega)$, $\boldsymbol{h}(\omega)$, and $\boldsymbol{v}(\tau,\omega)$ are defined as the column vectors with the m-th entry equal to $x_m(\tau,\omega)$, $h_m(\omega)$, and $v_m(\tau,\omega)$, respectively.

We assume $s(\tau,\omega)$ and $\boldsymbol{v}(\tau,\omega)$ to be uncorrelated zero-mean random variables. Therefore, $\boldsymbol{x}(\tau,\omega)$ is regarded as a zero-mean random vector whose covariance matrix is

$$\boldsymbol{\Phi}_{\boldsymbol{x}\boldsymbol{x}}(\tau,\omega) = \phi_{ss}(\tau,\omega)\boldsymbol{h}(\omega)\boldsymbol{h}^\mathsf{H}(\omega) + \boldsymbol{\Phi}_{\boldsymbol{v}\boldsymbol{v}}(\tau,\omega). \tag{7}$$

B. Wiener Post-filter

Of the linear estimators of $s(\tau,\omega)$ of the form

$$\hat{s}(\tau,\omega) \triangleq \boldsymbol{w}^\mathsf{H}(\tau,\omega)\boldsymbol{x}(\tau,\omega), \tag{8}$$

the one minimizing the mean square error is given by [5], [12]:

$$\hat{s}_{\text{LMMSE}}(\tau,\omega) \triangleq \phi_{ss}(\tau,\omega)\boldsymbol{h}^\mathsf{H}(\omega)\boldsymbol{\Phi}_{\boldsymbol{x}\boldsymbol{x}}^{-1}(\tau,\omega)\boldsymbol{x}(\tau,\omega). \tag{9}$$

We can show that the LMMSE estimator (9) is closely related to the output of the MVDR beamformer given by

$$y(\tau,\omega) \triangleq \frac{\boldsymbol{h}^\mathsf{H}(\omega)\boldsymbol{\Phi}_{\boldsymbol{x}\boldsymbol{x}}^{-1}(\tau,\omega)\boldsymbol{x}(\tau,\omega)}{\boldsymbol{h}^\mathsf{H}(\omega)\boldsymbol{\Phi}_{\boldsymbol{x}\boldsymbol{x}}^{-1}(\tau,\omega)\boldsymbol{h}(\omega)}. \tag{10}$$

From this equation, the short-time power spectrum of $y(\tau,\omega)$ is given by

$$\phi_{yy}(\tau,\omega) = \frac{1}{\boldsymbol{h}^\mathsf{H}(\omega)\boldsymbol{\Phi}_{\boldsymbol{x}\boldsymbol{x}}^{-1}(\tau,\omega)\boldsymbol{h}(\omega)}. \tag{11}$$

Therefore, (9) is rewritten as follows [5], [12]:

$$\hat{s}_{\text{LMMSE}}(\tau,\omega) = \underbrace{\frac{\phi_{ss}(\tau,\omega)}{\phi_{yy}(\tau,\omega)}}_{\triangleq\, p(\tau,\omega)} \cdot \underbrace{\frac{\boldsymbol{h}^\mathsf{H}(\omega)\boldsymbol{\Phi}_{\boldsymbol{x}\boldsymbol{x}}^{-1}(\tau,\omega)\boldsymbol{x}(\tau,\omega)}{\boldsymbol{h}^\mathsf{H}(\omega)\boldsymbol{\Phi}_{\boldsymbol{x}\boldsymbol{x}}^{-1}(\tau,\omega)\boldsymbol{h}(\omega)}}_{=\, y(\tau,\omega)}. \tag{12}$$

This means that $\hat{s}_{\text{LMMSE}}(\tau,\omega)$ is obtained by post-processing the MVDR beamformer's output $y(\tau,\omega)$ with the time-frequency mask $p(\tau,\omega)$ called the Wiener post-filter. In the design of $p(\tau,\omega)$, it is crucial to accurately estimate the numerator $\phi_{ss}(\tau,\omega)$ from the noisy observed signals, whereas the denominator $\phi_{yy}(\tau,\omega)$ can be easily estimated using the beamformer output $y(\tau,\omega)$.

C. Zelinski's Design of the Wiener Post-filter

Zelinski's estimator of $\phi_{ss}(\tau,\omega)$ is based on the assumption of uncorrelated noise. Although Zelinski's method [2] was originally presented in the time domain, we describe here its equivalence in the time-frequency domain for simplicity. Under the assumption, the interchannel cross-spectra of observed signals are noise-free:

$$\phi_{x_m x_n}(\tau,\omega) = \phi_{ss}(\tau,\omega) h_m(\omega) h_n^*(\omega) \ (m \neq n). \tag{13}$$

Solving Eq. (13) for $\phi_{ss}(\tau,\omega)$ and averaging the result over the microphone pairs, we obtain the estimator

$$\hat{\phi}_{ss}^{\mathrm{Z}}(\tau,\omega) = \frac{2}{M(M-1)} \sum_{m<n} \Re\Big[\frac{\phi_{x_m x_n}(\tau,\omega)}{h_m(\omega)h_n^*(\omega)}\Big]. \tag{14}$$

Here, $h_m(\omega)$ in the denominator is assumed to have been calculated based on *e.g.* planewave assumption.

On the other hand, the denominator $\phi_{yy}(\tau,\omega)$ of $p(\tau,\omega)$ is estimated by the following equation:

$$\hat{\phi}_{yy}^{\mathrm{Z}}(\tau,\omega) \triangleq \frac{1}{M} \sum_{m=1}^{M} \phi_{x_m x_m}(\tau,\omega). \tag{15}$$

The power/cross spectra $\phi_{x_m x_n}(\tau,\omega)$ in Eqs. (14) and (15) can be estimated by, for example, averaging $x_m(\tau,\omega)x_n^*(\tau,\omega)$ temporally over several adjacent frames.

D. McCowan's Design of the Wiener Post-filter

Instead of neglecting inter-channel noise correlation like Zelinski's method, McCowan's estimator of $\phi_{ss}(\tau,\omega)$ [7] is based on the assumption that the inter-channel noise coherences are given. Here we present a slightly modified version of McCowan's method, which is more theoretically sound as explained later. The assumption is based on the fact that they are known for some ideal noise fields. For example, the noise coherence between the m-th and n-th microphones in spherically isotropic noise fields is

$$\begin{aligned} \gamma_{v_m v_n}(\tau,\omega) &\triangleq \frac{\phi_{v_m v_n}(\tau,\omega)}{\sqrt{\phi_{v_m v_m}(\tau,\omega)}\sqrt{\phi_{v_n v_n}(\tau,\omega)}} && (16)\\ &= \mathrm{sinc}\left(\frac{r_{mn}\omega}{c}\right), && (17) \end{aligned}$$

where r_{mn} is the distance between the microphones, and c the velocity of sound. It is also assumed that the short-time power spectrum of noise is identical at all microphones, which is true for spherically/cylindrically isotropic noise:

$$\phi_{v_1 v_1}(\tau,\omega) = \cdots = \phi_{v_M v_M}(\tau,\omega) =: \phi_{vv}(\tau,\omega). \tag{18}$$

In this case, we have

$$\phi_{v_m v_n}(\tau,\omega) = \phi_{vv}(\tau,\omega)\gamma_{v_m v_n}(\tau,\omega), \tag{19}$$

and therefore

$$\phi_{x_m x_n} = \phi_{ss} h_m h_n^* + \phi_{vv}\gamma_{v_m v_n}. \tag{20}$$

Here we abbreviated the arguments τ and ω because of space limitations. McCowan's estimator of ϕ_{ss} is obtained based on solving the system of equations (20) for ϕ_{ss} and averaging the result over all microphone pairs, and is given by Eq. (21) on the top of the next page. $h_m(\omega)$ is assumed to have been calculated as in Zelinski's method. Zelinski's estimator $\hat{\phi}_{yy}^{\mathrm{Z}}(\tau,\omega)$ is used for the denominator of the post-filter.

Note that Eq. (21) differs slightly from the original version proposed by McCowan *et al.*, where the difference lies in the term $\frac{|h_m(\omega)|^2+|h_n(\omega)|^2}{2}$. In this method, it is assumed that the noise signals were aligned with respect to the target signal, namely $v_m(\tau,\omega)/h_m(\omega)$, is spherically/cylindrically isotropic [13], [20], which resulted in

$$\frac{\phi_{v_m v_n}(\tau,\omega)}{h_m(\omega)h_n^*(\omega)} = \begin{cases} \phi_{vv}(\tau,\omega)\mathrm{sinc}\left(\dfrac{r_{mn}\omega}{c}\right) \text{ (spherical)},\\ \phi_{vv}(\tau,\omega)J_0\left(\dfrac{l_{mn}\omega}{c}\right) \text{ (cylindrical)}. \end{cases} \tag{22}$$

Here $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind and l_{mn} is the distance between the orthogonal projections onto the xy-plane of the m-th and n-th microphones. However, the factor $\frac{1}{h_m(\omega)h_n^*(\omega)}$ caused by the alignment changes the phase for the planewave case and both the phase and the magnitude in general. The slightly modified version presented here models the original noise signals $v_m(\tau,\omega)$ as spherically/cylindrically isotropic, which results in Eq. (19). Therefore, we have

$$\phi_{v_m v_n}(\tau,\omega) = \begin{cases} \phi_{vv}(\tau,\omega)\mathrm{sinc}\left(\dfrac{r_{mn}\omega}{c}\right) \text{ (spherical)},\\ \phi_{vv}(\tau,\omega)J_0\left(\dfrac{l_{mn}\omega}{c}\right) \text{ (cylindrical)}. \end{cases} \tag{23}$$

This is theoretically sounder and resulted in much better results in our experiments.

III. Proposed Method

A. Our Model of Diffuse Noise

Focusing on the isotropic characteristics of diffuse noise, we put the following two assumptions [14], [18], [19]:

1) The short-time power spectrum of noise at all microphones as in Eq. (18).
2) The inter-channel noise cross-spectrogram is identical for all microphone pairs with an equal distance:

$$r_{mn} = r_{kl} \ \Rightarrow \ \phi_{v_m v_n}(\tau,\omega) = \phi_{v_k v_l}(\tau,\omega). \tag{24}$$

Note that McCowan's method makes a stronger assumption of 1) and Eq. (23).

B. Proposed Design of the Wiener Post-filter

Our design of the Wiener post-filter is based on the diagonalization of the noise covariance matrix. In the following, the arguments τ and ω are often abbreviated. With $\boldsymbol{U}$ denoting a unitary diagonalization matrix of $\boldsymbol{\Phi_{vv}}$, Eq. (7) can be transformed into

$$\boldsymbol{U}^{\mathsf{H}}\boldsymbol{\Phi_{xx}}\boldsymbol{U} = \phi_{ss}\boldsymbol{U}^{\mathsf{H}}\boldsymbol{h}\boldsymbol{h}^{\mathsf{H}}\boldsymbol{U} + \boldsymbol{U}^{\mathsf{H}}\boldsymbol{\Phi_{vv}}\boldsymbol{U}. \tag{25}$$

Since $\boldsymbol{U}^{\mathsf{H}}\boldsymbol{\Phi_{vv}}\boldsymbol{U}$ is diagonal, the off-diagonal entries of $\boldsymbol{U}^{\mathsf{H}}\boldsymbol{\Phi_{xx}}\boldsymbol{U}$ are noise-free as follows:

$$\boldsymbol{u}_m^{\mathsf{H}}\boldsymbol{\Phi_{xx}}\boldsymbol{u}_n = \phi_{ss}\boldsymbol{u}_m^{\mathsf{H}}\boldsymbol{h}\boldsymbol{h}^{\mathsf{H}}\boldsymbol{u}_n \ (m \neq n), \tag{26}$$

$$\hat{\phi}^{\mathsf{M}}_{ss}(\tau,\omega) = \frac{2}{M(M-1)} \sum_{m<n} \frac{\Re\left[\dfrac{\phi_{x_m x_n}(\tau,\omega)}{h_m(\omega)h_n^*(\omega)}\right] - \dfrac{\phi_{x_m x_m}(\tau,\omega)+\phi_{x_n x_n}(\tau,\omega)}{2}\Re\left[\dfrac{\gamma_{v_m v_n}(\tau,\omega)}{h_m(\omega)h_n^*(\omega)}\right]}{1-\dfrac{|h_m(\omega)|^2+|h_n(\omega)|^2}{2}\Re\left[\dfrac{\gamma_{v_m v_n}(\tau,\omega)}{h_m(\omega)h_n^*(\omega)}\right]} \tag{21}$$

where $\boldsymbol{u}_m$ denotes the m-th column of $\boldsymbol{U}$. Defining $\tilde{x}_m(\tau,\omega) \triangleq \boldsymbol{u}_m^{\mathsf{H}}\boldsymbol{x}(\tau,\omega)$ and $\tilde{h}_m(\omega) \triangleq \boldsymbol{u}_m^{\mathsf{H}}\boldsymbol{h}(\omega)$, this is simplified to

$$\phi_{\tilde{x}_m\tilde{x}_n}(\tau,\omega) = \phi_{ss}(\tau,\omega)\tilde{h}_m(\omega)\tilde{h}_n^*(\omega). \tag{27}$$

Since Eq. (27) is an overdetermined set of equations on $\phi_{ss}(\tau,\omega)$, we estimate it by minimizing the square error

$$J \triangleq \sum_{\tau,\omega}\sum_{m,n,m\neq n} |\phi_{\tilde{x}_m\tilde{x}_n}(\tau,\omega) - \phi_{ss}(\tau,\omega)\tilde{h}_m(\omega)\tilde{h}_n^*(\omega)|^2, \tag{28}$$

with respect to $\{\phi_{ss}(\tau,\omega)\}_{\tau,\omega}$. Therefore the estimator is given by

$$\hat{\phi}^{\mathsf{P}}_{ss}(\tau,\omega) = \frac{\sum_{m,n,m\neq n}\Re[\phi_{\tilde{x}_m\tilde{x}_n}(\tau,\omega)\tilde{h}_m^*(\omega)\tilde{h}_n(\omega)]}{\sum_{m,n,m\neq n}|\tilde{h}_m(\omega)|^2|\tilde{h}_n(\omega)|^2}. \tag{29}$$

On the other hand, the estimate $\hat{\phi}^{\mathsf{P}}_{yy}(\tau,\omega)$ of $\phi_{yy}(\tau,\omega)$ is calculated by averaging $|y(\tau,\omega)|^2$ temporally over several adjacent frames. Additionally, based on the knowledge that $p(\tau,\omega)$ lies in the range 0 to 1, we restrict its estimate $\hat{p}^{\mathsf{P}}(\tau,\omega)$ in the range by post-processing it in the following way:

$$\begin{cases}\hat{p}^{\mathsf{P}}(\tau,\omega) \leftarrow 0, \text{ if } \hat{p}^{\mathsf{P}}(\tau,\omega) < 0,\\ \hat{p}^{\mathsf{P}}(\tau,\omega) \leftarrow 1, \text{ if } \hat{p}^{\mathsf{P}}(\tau,\omega) > 1.\end{cases} \tag{30}$$

C. Blind Noise Decorrelation

There still remains a problem of obtaining a diagonalization matrix $\boldsymbol{U}$ because $\boldsymbol{\Phi}_{\boldsymbol{vv}}$ is normally unknown. Spherically isotropic noise, which results from the superposition of planewaves of an equal power from all directions in free field, can be decorrelated via spherical harmonic decomposition [21]. However noise is not always spherically isotropic, because of the distribution of the noise sources, the room shape, the diffraction by a rigid mount, etc.

We take another approach which can decorrelate larger class of noise satisfying the isotropy assumptions 1) and 2) in Section III-A. We found out that, under these assumptions, there exist some classes of arrays such that $\boldsymbol{\Phi}_{\boldsymbol{vv}}$ is diagonalized by a constant unitary matrix independent of its value [18], [19]. We refer to such diagonalization as *Blind Noise Decorrelation* (BND) because it does not require the knowledge of the explicit value of $\boldsymbol{\Phi}_{\boldsymbol{vv}}$ and the diagonalization of $\boldsymbol{\Phi}_{\boldsymbol{vv}}$ means the decorrelation of noise. Not assuming explicit noise coherences but only isotropy, our method has the potential of being robust against deviation of noise from the sinc coherence model encountered, *e.g.* when microphones are mounted on a symmetical object.

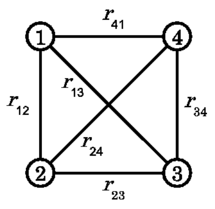


Fig. 1. Square array. Since $r_{12} = r_{23} = r_{34} = r_{41}$ and $r_{13} = r_{24}$, the spatial covariance matrix of isotropic noise becomes circulant for this array geometry.

As an example of BND, consider a 4-element array with its microphones at the vertices of a square (Fig. 1). From assumption 1), we have

$$\phi_{v_1v_1} = \phi_{v_2v_2} = \phi_{v_3v_3} = \phi_{v_4v_4} =: \alpha. \tag{31}$$

Furthermore, we have, from assumption 2),

$$\phi_{v_1v_2} = \phi_{v_2v_1} = \phi_{v_2v_3} = \phi_{v_3v_2} = \cdots = \phi_{v_1v_4} =: \beta, \tag{32}$$

$$\phi_{v_1v_3} = \phi_{v_3v_1} = \phi_{v_2v_4} = \phi_{v_4v_2} =: \gamma, \tag{33}$$

because of $r_{12} = r_{23} = r_{34} = r_{41}$ and $r_{13} = r_{24}$. Consequently, $\boldsymbol{\Phi}_{\boldsymbol{vv}}$ has the following structure: $\boldsymbol{\Phi}_{\boldsymbol{vv}} = \mathrm{circ}(\alpha,\beta,\gamma,\beta)$. Being a circulant matrix, it is diagonalized by the 4×4 DFT matrix $\boldsymbol{F}_4$ for any values of α, β, and γ.

D. Joint Estimation of the Short-time Power Spectra and the Steering Vectors

Aiming at robust diffuse noise suppression in reverberant environments, we present here a method for joint estimation of $\boldsymbol{h}(\omega)$ and $\phi_{ss}(\tau,\omega)$ from the denoised off-diagonal entries of the observation covariance matrix.

Our idea consists in minimizing J in Eq. (28) with respect to $\{\tilde{h}_m(\omega)\}_m$ and $\{\phi_{ss}(\tau,\omega)\}_{\tau,\omega}$. Equating the complex partial derivatives of J with respect to $\phi_{ss}(\tau,\omega)$ and $h_k(\omega)\,(k = 1,2,\ldots,M)$ to zero we have Eq. (29) and

$$\tilde{h}_k(\omega) = \frac{\sum_\tau \phi_{ss}(\tau,\omega)\sum_{m,m\neq k}\phi_{\tilde{x}_k\tilde{x}_m}(\tau,\omega)\tilde{h}_m(\omega)}{\left[\sum_\tau \phi_{ss}^2(\tau,\omega)\right]\left[\sum_{m,m\neq k}|\tilde{h}_m(\omega)|^2\right]}. \tag{34}$$

Note here that, without any constraint on $\phi_{ss}(\tau,\omega)$ or $\tilde{h}_m(\omega)$, there is scale indeterminacy between them, because replacing $\{\phi_{ss}(\tau,\omega)\}_\tau$ with $\{K\phi_{ss}(\tau,\omega)\}_\tau$ and $\{\tilde{h}_m(\omega)\}_m$ with $\{\frac{1}{\sqrt{K}}\tilde{h}_m(\omega)\}_m$ (K: arbitrary positive number) does not change the value of J. Because the signal $s(\tau,\omega)$ that we are trying to estimate is the target signal observed at the first microphone, the transfer function $h_1(\omega)$ from $s(\tau,\omega)$ to $x_1(\tau,\omega)$ is unity:

$$h_1(\omega) = 1. \tag{35}$$

By definition, $h_1(\omega)$ is written as

$$h_1(\omega) = \sum_{m=1}^{M} u_{1m}\tilde{h}_m(\omega), \tag{36}$$

where u_{1m} denotes the $(1,m)$-entry of $\boldsymbol{U}$. From these equations, we have

$$\sum_{m=1}^{M} u_{1m}\tilde{h}_m(\omega) = 1. \tag{37}$$

We use this equation to solve the indeterminacy.

Consequently, the joint estimation consists in iterating the following updates alternately:

$$\hat{\phi}_{ss}(\tau) \leftarrow \frac{\sum_{m,n,m\neq n} \Re[\phi_{\tilde{x}_m\tilde{x}_n}(\tau)\hat{\tilde{h}}_m^*\hat{\tilde{h}}_n]}{\sum_{m,n,m\neq n} |\hat{\tilde{h}}_m|^2|\hat{\tilde{h}}_n|^2}, \tag{38}$$

$$\hat{\tilde{h}}_k \leftarrow \frac{\sum_\tau \hat{\phi}_{ss}(\tau)\sum_{m,m\neq k}\phi_{\tilde{x}_k\tilde{x}_m}(\tau)\hat{\tilde{h}}_m}{\left[\sum_\tau \hat{\phi}_{ss}^2(\tau)\right]\left[\sum_{m,m\neq k}|\hat{\tilde{h}}_m|^2\right]}, \; k = 1, 2, \ldots, M, \tag{39}$$

$$\hat{\tilde{\boldsymbol{h}}} \leftarrow \frac{\hat{\tilde{\boldsymbol{h}}}}{\sum_{m=1}^{M} u_{1m}\hat{\tilde{h}}_m}, \tag{40}$$

where the last updating rule scales $\hat{\tilde{\boldsymbol{h}}}$ so that it satisfies the constraint (37). $\hat{\tilde{\boldsymbol{h}}}$ can be initialized with *e.g.* the planewave model. Here, we omitted the variable ω to save space because the above update is performed separately in each frequency bin. The estimate of $\boldsymbol{h}(\omega)$ is obtained by $\hat{\boldsymbol{h}}(\omega) = \boldsymbol{U}\hat{\tilde{\boldsymbol{h}}}(\omega)$. $\hat{\boldsymbol{h}}(\omega)$ is used to design the MVDR beamformer and $\hat{\phi}_{ss}(\tau,\omega)$ is used to design the Wiener post-filter. $\phi_{yy}(\tau,\omega)$ is estimated as explained in Section III-B using the output of the MVDR beamformer designed using $\hat{\boldsymbol{h}}(\omega)$.

IV. Five Classes of Arrays for BND

In this section, we show that BND is enabled by five classes of crystal-shaped arrays comprising 1) regular polygonal arrays, 2) (twisted) rectangular arrays, 3) (twisted) regular polygonal prism arrays, 4) rectangular solid arrays, and 5) regular polyhedral arrays (see Fig. 2). We refer to these arrays as *crystal arrays*.

We use the property of circulant matrices that $\mathrm{circ}(\alpha_1, \alpha_2, \ldots, \alpha_k)$ is diagonalized by $\boldsymbol{F}_k$, regardless of the values of $\alpha_1, \alpha_2, \ldots, \alpha_k \in \mathbb{C}$ [22].

A. Regular Polygonal Arrays

A regular M-gonal array is an array with the microphones at the vertices of a regular M-gon. If we number the microphones as in Fig. 2, $\boldsymbol{\Phi}_{\boldsymbol{vv}}$ has the following structure:

$$\boldsymbol{\Phi}_{\boldsymbol{vv}} = \mathrm{circ}(\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_2, \alpha_1). \tag{41}$$

Therefore, it is diagonalized by $\boldsymbol{F}_M$.

B. (Twisted) Rectangular Arrays

A rectangular array is an array with the microphones at the vertices of a rectangle. If we number the microphones as in Fig. 2, $\boldsymbol{\Phi}_{\boldsymbol{vv}}$ has the following structure:

$$\boldsymbol{\Phi}_{\boldsymbol{vv}} = \begin{bmatrix} \boldsymbol{C}_1 & \boldsymbol{C}_2 \\ \boldsymbol{C}_2 & \boldsymbol{C}_1 \end{bmatrix}, \tag{42}$$

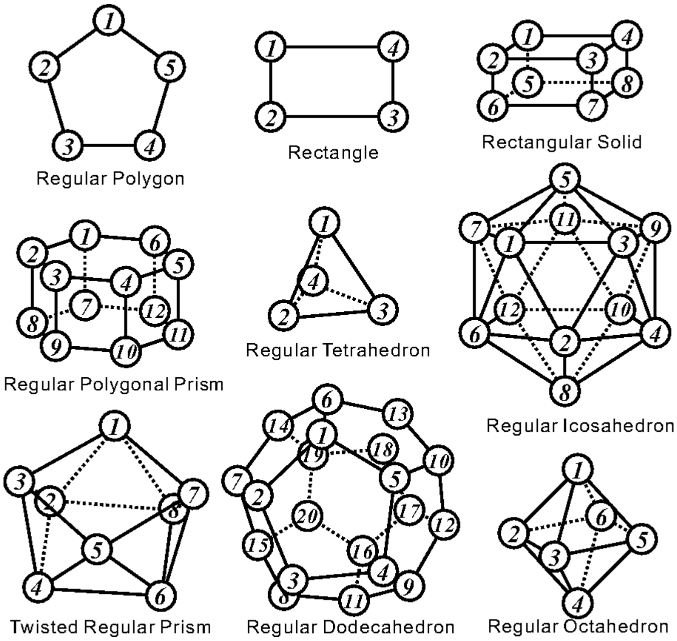


Fig. 2. Examples of crystal arrays for blind noise decorrelation: (from the top-left) a regular pentagonal array (belonging to the regular polygonal arrays), a regular hexagonal prism array (belonging to the regular polygonal prism arrays), a twisted square prism array (belonging to the twisted regular polygonal arrays), a rectangular array, a regular tetrahedral array (belonging to the regular polyhedral arrays), a regular dodecahedral array (belonging to the regular polyhedral arrays), a rectangular solid array, a regular icosahedral array (belonging to the regular polyhedral arrays), and a regular octahedral array (belonging to the regular polyhedral arrays).

where $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ are 2×2 circulant matrices. It is diagonalized by

$$\boldsymbol{F}_2 \otimes \boldsymbol{F}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} \boldsymbol{F}_2 & \boldsymbol{F}_2 \\ \boldsymbol{F}_2 & -\boldsymbol{F}_2 \end{bmatrix}, \tag{43}$$

where $\otimes$ denotes the Kronecker product. This can be easily shown using the following equation:

$$\boldsymbol{\Phi}_{\boldsymbol{vv}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \boldsymbol{C}_1 + \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \otimes \boldsymbol{C}_2, \tag{44}$$

and the following properties of the Kronecker product [22]:

$$(\boldsymbol{A}\otimes\boldsymbol{B})^{\mathsf{H}} = \boldsymbol{A}^{\mathsf{H}}\otimes\boldsymbol{B}^{\mathsf{H}}, \tag{45}$$

$$(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{C}\otimes\boldsymbol{D}) = (\boldsymbol{AC})\otimes(\boldsymbol{BD}). \tag{46}$$

The above discussion also holds true for a rectangular array twisted about one of its mirror symmetry axes by any angle because this deformation does not change the structure of $\boldsymbol{\Phi}_{\boldsymbol{vv}}$ in Eq. (42).

C. (Twisted) Regular Polygonal Prism Arrays

A regular $M/2$-gonal prism array (M: even positive integer) is an array with the microphones at the vertices of a regular $M/2$-gonal prism. If we number the microphones as in Fig 2, $\boldsymbol{\Phi}_{\boldsymbol{vv}}$ has the same structure as in Eq. (42) with $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ being in turn $M/2 \times M/2$ circulant matrices of the following forms:

$$\boldsymbol{C}_1 \triangleq \mathrm{circ}(\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_2, \alpha_1), \tag{47}$$

$$\boldsymbol{C}_2 \triangleq \mathrm{circ}(\beta_0, \beta_1, \beta_2, \ldots, \beta_2, \beta_1). \tag{48}$$

Hence, $\boldsymbol{\Phi}_{\boldsymbol{v}\boldsymbol{v}}$ is diagonalized by

$$\boldsymbol{F}_2 \otimes \boldsymbol{F}_{M/2} = \frac{1}{\sqrt{2}} \begin{bmatrix} \boldsymbol{F}_{M/2} & \boldsymbol{F}_{M/2} \\ \boldsymbol{F}_{M/2} & -\boldsymbol{F}_{M/2} \end{bmatrix}, \tag{49}$$

just as in the case of the rectangular array.

A regular $M/2$-gonal prism twisted about its rotation symmetry axis by $2\pi/M$ (*i.e.* regular $M/2$-gonal antiprism) is also allowable. In this case, if we number the microphones as in Fig 2, $\boldsymbol{\Phi}_{\boldsymbol{v}\boldsymbol{v}}$ has the structure of Eq. (41), and therefore it is diagonalized by $\boldsymbol{F}_M$.

D. Rectangular Solid Arrays

A rectangular solid array is an array with the microphones at the vertices of a rectangular solid. If we number the microphones as in Fig 2, $\boldsymbol{\Phi}_{\boldsymbol{v}\boldsymbol{v}}$ has the following structure:

$$\boldsymbol{\Phi}_{\boldsymbol{v}\boldsymbol{v}} = \begin{bmatrix} \boldsymbol{C}_1 & \boldsymbol{C}_2 & \boldsymbol{C}_3 & \boldsymbol{C}_4 \\ \boldsymbol{C}_2 & \boldsymbol{C}_1 & \boldsymbol{C}_4 & \boldsymbol{C}_3 \\ \boldsymbol{C}_3 & \boldsymbol{C}_4 & \boldsymbol{C}_1 & \boldsymbol{C}_2 \\ \boldsymbol{C}_4 & \boldsymbol{C}_3 & \boldsymbol{C}_2 & \boldsymbol{C}_1 \end{bmatrix}, \tag{50}$$

with $\boldsymbol{C}_k$ being 2×2 circulant matrices. Hence, $\boldsymbol{\Phi}_{\boldsymbol{v}\boldsymbol{v}}$ is diagonalized by

$$\boldsymbol{F}_2 \otimes \boldsymbol{F}_2 \otimes \boldsymbol{F}_2 = \frac{1}{2} \begin{bmatrix} \boldsymbol{F}_2 & \boldsymbol{F}_2 & \boldsymbol{F}_2 & \boldsymbol{F}_2 \\ \boldsymbol{F}_2 & -\boldsymbol{F}_2 & \boldsymbol{F}_2 & -\boldsymbol{F}_2 \\ \boldsymbol{F}_2 & \boldsymbol{F}_2 & -\boldsymbol{F}_2 & -\boldsymbol{F}_2 \\ \boldsymbol{F}_2 & -\boldsymbol{F}_2 & -\boldsymbol{F}_2 & \boldsymbol{F}_2 \end{bmatrix}, \tag{51}$$

just as in the case of the rectangular array.

E. Regular Polyhedral Arrays

A regular polyhedral array is an array with the microphones at the vertices of a regular polyhedron. A regular tetrahedral array is a twisted rectangular array, a regular octahedral array a twisted equilateral triangular prism array, and a regular hexahedral (cubic) array a rectangular solid array. Therefore, only regular icosahedral and dodecahedral arrays are left to be discussed in the following.

For an icosahedral array, if we number the microphones as in Fig. 2, $\boldsymbol{\Phi}_{\boldsymbol{v}\boldsymbol{v}}$ has the following structure:

$$\boldsymbol{\Phi}_{\boldsymbol{v}\boldsymbol{v}} = \begin{bmatrix} \boldsymbol{C}_1 & \boldsymbol{C}_2 & \boldsymbol{C}_3 & \boldsymbol{C}_4 \\ \boldsymbol{C}_2 & \boldsymbol{C}_5 & \boldsymbol{C}_6 & \boldsymbol{C}_3 \\ \boldsymbol{C}_3 & \boldsymbol{C}_6 & \boldsymbol{C}_5 & \boldsymbol{C}_2 \\ \boldsymbol{C}_4 & \boldsymbol{C}_3 & \boldsymbol{C}_2 & \boldsymbol{C}_1 \end{bmatrix}, \tag{52}$$

with $\boldsymbol{C}_k$ being 3×3 circular matrices of the following forms:

$$\boldsymbol{C}_1 = \mathrm{circ}(a, b, b), \qquad \boldsymbol{C}_2 = \mathrm{circ}(c, b, b), \tag{53}$$

$$\boldsymbol{C}_3 = \mathrm{circ}(b, c, c), \qquad \boldsymbol{C}_4 = \mathrm{circ}(d, c, c), \tag{54}$$

$$\boldsymbol{C}_5 = \mathrm{circ}(a, c, c), \qquad \boldsymbol{C}_6 = \mathrm{circ}(d, b, b). \tag{55}$$

We found out that $\boldsymbol{\Phi}_{\boldsymbol{v}\boldsymbol{v}}$ is diagonalized by

$$\frac{1}{2} \begin{bmatrix} \boldsymbol{F}_3 & \boldsymbol{F}_3 & \boldsymbol{F}_3\boldsymbol{P}_{3+} & \boldsymbol{F}_3\boldsymbol{P}_{3-} \\ \boldsymbol{F}_3 & -\boldsymbol{F}_3 & -\boldsymbol{F}_3\boldsymbol{Q}_{3+} & -\boldsymbol{F}_3\boldsymbol{Q}_{3-} \\ \boldsymbol{F}_3 & -\boldsymbol{F}_3 & \boldsymbol{F}_3\boldsymbol{Q}_{3+} & \boldsymbol{F}_3\boldsymbol{Q}_{3-} \\ \boldsymbol{F}_3 & \boldsymbol{F}_3 & -\boldsymbol{F}_3\boldsymbol{P}_{3+} & -\boldsymbol{F}_3\boldsymbol{P}_{3-} \end{bmatrix}, \tag{56}$$

where

$$\boldsymbol{P}_{3\pm} \triangleq \mathrm{diag}\left(\frac{1}{\sqrt{4\mathrm{g}_\pm + 3}}, \frac{1}{\sqrt{\mathrm{g}_\pm/2 + 1}}, \frac{1}{\sqrt{\mathrm{g}_\pm/2 + 1}} \right), \tag{57}$$

$$\boldsymbol{Q}_{3\pm} \triangleq \mathrm{diag}\left(\frac{2\mathrm{g}_\pm + 1}{\sqrt{4\mathrm{g}_\pm + 3}}, \frac{\mathrm{g}_\pm}{\sqrt{\mathrm{g}_\pm/2 + 1}}, \frac{\mathrm{g}_\pm}{\sqrt{\mathrm{g}_\pm/2 + 1}} \right), \tag{58}$$

$$\mathrm{g}_\pm \triangleq \frac{1 \pm \sqrt{5}}{2}. \tag{59}$$

On the other hand, for a dodecahedral array, if we number the microphones as in Fig. 2, $\boldsymbol{\Phi}_{\boldsymbol{v}\boldsymbol{v}}$ has the same structure as in Eq. (52) with $\boldsymbol{C}_k$ being 5×5 circular matrices of the following forms:

$$\boldsymbol{C}_1 = \mathrm{circ}(a, b, c, c, b), \qquad \boldsymbol{C}_2 = \mathrm{circ}(b, c, d, d, c), \tag{60}$$

$$\boldsymbol{C}_3 = \mathrm{circ}(e, d, c, c, d), \qquad \boldsymbol{C}_4 = \mathrm{circ}(f, e, d, d, e), \tag{61}$$

$$\boldsymbol{C}_5 = \mathrm{circ}(a, c, e, e, c), \qquad \boldsymbol{C}_6 = \mathrm{circ}(f, d, b, b, d). \tag{62}$$

We found out that $\boldsymbol{\Phi}_{\boldsymbol{v}\boldsymbol{v}}$ is diagonalized by

$$\frac{1}{2} \begin{bmatrix} \boldsymbol{F}_5\boldsymbol{P}_{5+} & \boldsymbol{F}_5\boldsymbol{P}_{5-} & \boldsymbol{F}_5\boldsymbol{Q}_{5+} & \boldsymbol{F}_5\boldsymbol{Q}_{5-} \\ \boldsymbol{F}_5\boldsymbol{R}_{5+} & -\boldsymbol{F}_5\boldsymbol{R}_{5-} & -\boldsymbol{F}_5\boldsymbol{S}_{5+} & -\boldsymbol{F}_5\boldsymbol{S}_{5-} \\ \boldsymbol{F}_5\boldsymbol{R}_{5+} & -\boldsymbol{F}_5\boldsymbol{R}_{5-} & \boldsymbol{F}_5\boldsymbol{S}_{5+} & \boldsymbol{F}_5\boldsymbol{S}_{5-} \\ \boldsymbol{F}_5\boldsymbol{P}_{5+} & \boldsymbol{F}_5\boldsymbol{P}_{5-} & -\boldsymbol{F}_5\boldsymbol{Q}_{5+} & -\boldsymbol{F}_5\boldsymbol{Q}_{5-} \end{bmatrix}, \tag{63}$$

where the matrices $\boldsymbol{P}_{5\pm}$, $\boldsymbol{Q}_{5\pm}$, $\boldsymbol{R}_{5\pm}$, and $\boldsymbol{S}_{5\pm}$ are defined in Eqs. (64)–(67) at the top of the next page.

We have neither proved that the five classes presented here are the only geometries enabling BND nor found any other one. A method for judging whether a given array geometry enables BND was described in [23].

V. Experiments and Results

We conducted some experiments with simulated and real-world noise in order to demonstrate the effectiveness of the proposed methods. We examined the performance of the MVDR beamformer and the proposed, Zelinski's, and McCowan's methods. Each post-filter was preceded by the MVDR beamformer. The observed signals were analyzed by STFT, where the frame length and the frame shift were 512 and 32, respectively, unless otherwise noted, and the Hamming window was used. We calculated $\boldsymbol{\Phi}_{\boldsymbol{x}\boldsymbol{x}}$ for the beamformer (10) by averaging $\boldsymbol{x}(\tau, \omega)\boldsymbol{x}^{\mathsf{H}}(\tau, \omega)$ temporally over all frames. On the other hand, $\boldsymbol{\Phi}_{\boldsymbol{x}\boldsymbol{x}}(\tau, \omega)$ and $\phi_{x_m x_n}(\tau, \omega)$ for the post-filters were calculated by averaging $\boldsymbol{x}(\tau, \omega)\boldsymbol{x}^{\mathsf{H}}(\tau, \omega)$ and $x_m(\tau, \omega)x_n^*(\tau, \omega)$ temporally over 16 consecutive frames unless otherwise noted, where we can reasonably assume signal stationarity.

We used an output Signal-to-Noise Ratio (SNR), *Signal Distortion (SD)*, and *Noise Reduction (NR)*, to evaluate the

$$\boldsymbol{P}_{5\pm} \triangleq \operatorname{diag}\left(1, \frac{2}{\sqrt{6\mathrm{g}_{\mp}+6}}, \frac{2}{\sqrt{6\mathrm{g}_{\pm}+6}}, \frac{2}{\sqrt{6\mathrm{g}_{\pm}+6}}, \frac{2}{\sqrt{6\mathrm{g}_{\mp}+6}}\right) \tag{64}$$

$$\boldsymbol{Q}_{5\pm} \triangleq \operatorname{diag}\left(\frac{1}{\sqrt{4\mathrm{g}_{\pm}+3}}, \frac{2}{\sqrt{2\mathrm{g}_{\pm}+4}}, \frac{2}{\sqrt{2\mathrm{g}_{\pm}+4}}, \frac{2}{\sqrt{2\mathrm{g}_{\pm}+4}}, \frac{2}{\sqrt{2\mathrm{g}_{\pm}+4}}\right) \tag{65}$$

$$\boldsymbol{R}_{5\pm} \triangleq \operatorname{diag}\left(1, \frac{2\mathrm{g}_{\mp}^2}{\sqrt{6\mathrm{g}_{\mp}+6}}, \frac{2\mathrm{g}_{\pm}^2}{\sqrt{6\mathrm{g}_{\pm}+6}}, \frac{2\mathrm{g}_{\pm}^2}{\sqrt{6\mathrm{g}_{\pm}+6}}, \frac{2\mathrm{g}_{\mp}^2}{\sqrt{6\mathrm{g}_{\mp}+6}}\right) \tag{66}$$

$$\boldsymbol{S}_{5\pm} \triangleq \operatorname{diag}\left(\frac{2\mathrm{g}_{\pm}+1}{\sqrt{4\mathrm{g}_{\pm}+3}}, \frac{2\mathrm{g}_{\pm}}{\sqrt{2\mathrm{g}_{\pm}+4}}, \frac{2\mathrm{g}_{\pm}}{\sqrt{2\mathrm{g}_{\pm}+4}}, \frac{2\mathrm{g}_{\pm}}{\sqrt{2\mathrm{g}_{\pm}+4}}, \frac{2\mathrm{g}_{\pm}}{\sqrt{2\mathrm{g}_{\pm}+4}}\right) \tag{67}$$

overall performance, target distortion, and noise reduction, respectively. Let

$$\boldsymbol{s} \triangleq \begin{bmatrix} s(0) & s(1) & \cdots & s(N-1) \end{bmatrix}^{\mathsf{T}}, \tag{68}$$

$$\hat{\boldsymbol{s}} \triangleq \begin{bmatrix} \hat{s}(0) & \hat{s}(1) & \cdots & \hat{s}(N-1) \end{bmatrix}^{\mathsf{T}} \tag{69}$$

be the vectors comprising the samples in the target signal $s(t)$ and its estimate (the output of a noise suppression algorithm) $\hat{s}(t)$, respectively. The output SNR is

$$20 \log_{10} \frac{\|\hat{\boldsymbol{s}}_{\|}\|_2}{\|\hat{\boldsymbol{s}}_{\perp}\|_2}, \tag{70}$$

where the vectors $\hat{\boldsymbol{s}}_{\|}$ and $\hat{\boldsymbol{s}}_{\perp}$ are the components of $\hat{\boldsymbol{s}}$ parallel and perpendicular to $\boldsymbol{s}$, respectively:

$$\hat{\boldsymbol{s}}_{\|} \triangleq \frac{\hat{\boldsymbol{s}}^{\mathsf{T}}\boldsymbol{s}}{\|\boldsymbol{s}\|_2^2}\boldsymbol{s}, \tag{71}$$

$$\hat{\boldsymbol{s}}_{\perp} \triangleq \hat{\boldsymbol{s}} - \hat{\boldsymbol{s}}_{\|}. \tag{72}$$

The use of SD and NR is aimed to separately evaluate target cancellation and noise reduction. The output of a noise suppression multichannel filter $\boldsymbol{w}(\tau,\omega)$ is decomposed into the signal-originated component $s'(\tau,\omega)$ and the noise-originated component $v'(\tau,\omega)$ as follows:

$$\boldsymbol{w}^{\mathsf{H}}(\tau,\omega)\boldsymbol{x}(\tau,\omega) = \underbrace{\boldsymbol{w}^{\mathsf{H}}(\tau,\omega)\boldsymbol{c}(\tau,\omega)}_{s'(\tau,\omega)} + \underbrace{\boldsymbol{w}^{\mathsf{H}}(\tau,\omega)\boldsymbol{v}(\tau,\omega)}_{v'(\tau,\omega)}, \tag{73}$$

where $\boldsymbol{c}(\tau,\omega)$ is the column vector whose m-th entry is the target signal observed at the m-th microphone. Using these signals s' and v', SD and NR are defined as follows:

$$\text{SD} \triangleq 10 \log_{10} \frac{\sum_{t=0}^{N-1} \{s'(t) - s(t)\}^2}{\sum_{t=0}^{N-1} s^2(t)}, \tag{74}$$

$$\text{NR} \triangleq 10 \log_{10} \frac{\sum_{t=0}^{N-1} v_1^2(t)}{\sum_{t=0}^{N-1} v'^2(t)}, \tag{75}$$

which are basically equivalent to criteria proposed in ref. [24] except that these are defined for broadband signals in the time domain. For the output SNR and NR, a higher value means better performance, whereas a lower value of SD means better performance.

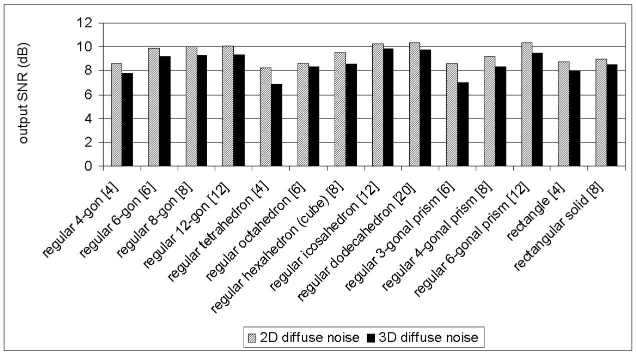


Fig. 3. The output SNR of the proposed method presented in Section III-B with various array geometries in simulated diffuse noise fields. (Input SNR: 0dB.)

A. Performance of Different Crystal Array Geometries

First we examine the performance of the post-filter proposed in Section III-B for different crystal array geometries. The array diameter was assumed to be 5 cm. We simulated the observed signals at the microphones assuming anechoic planewave propagation. The target speech arrived from a known direction (azimuth $\phi = 60°$; zenith angle $\theta = 90°$), and distinct speech interferences arrived from 64 randomly chosen directions. The speech files were taken from the ATR Japanese speech database [25]. We considered two types of diffuse noise:

- 2D diffuse noise, where the noise directions were chosen randomly from the region $(\theta,\phi) \in [70°, 110°] \times [0°, 360°]$.
- 3D diffuse noise, where the noise directions were chosen randomly from the region $(\theta,\phi) \in [0°, 180°] \times [0°, 360°]$.

The input SNR at the first microphone was adjusted to 0 dB. The duration of the observed signals was 4 s, and the sampling frequency was 16 kHz. We assumed the steering vectors to be given.

Fig. 3 shows the output SNR averaged over two (male and female) speakers for different crystal array geometries and different noise types. We can see that all geometries worked. The output SNR did not increase so much with an increasing number of microphones. However, the use of more microphones may be beneficial when, for example, there are strong directional noise sources because more nulls can be formed.

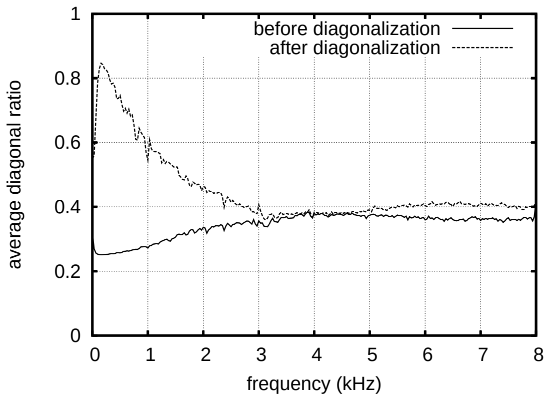
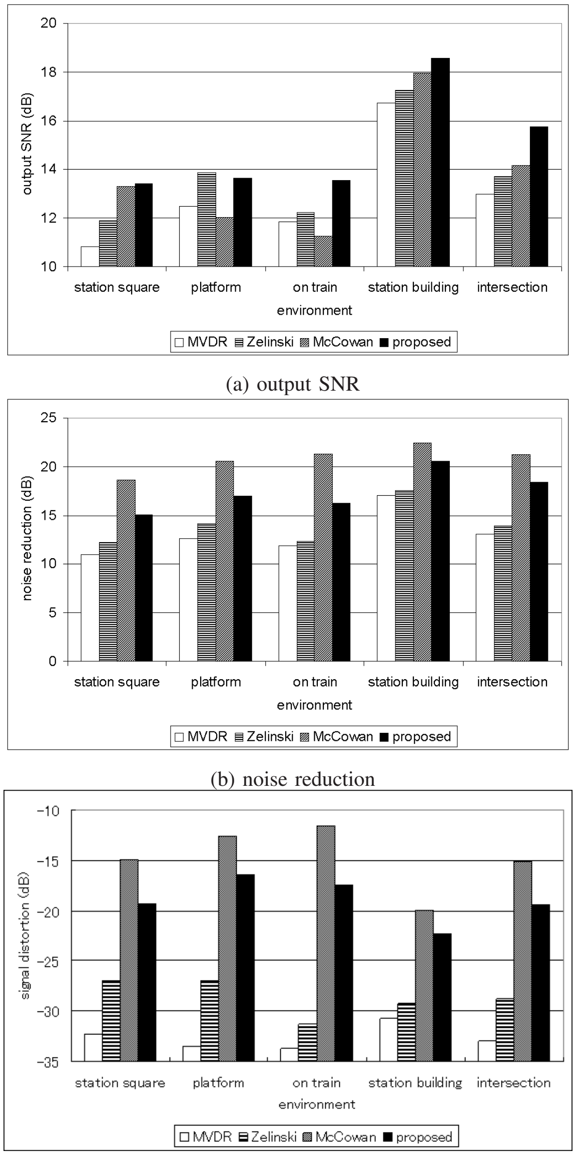


Fig. 4. Average diagonal ratios before and after BND as a function of the frequency (environment: station square).



(a) output SNR

(b) noise reduction

(c) signal distortion

Fig. 5. Performance comparison of the algorithms using objective measures for several environments.

B. Comparison between Proposed and Conventional Methods (Anechoic Case)

Next we compare the performance of the proposed method presented in Section III-B and the previous methods for real-world noise. We fabricated a square array with a diameter of 5 cm and recorded noise in Tokyo in a station square, in a platform, on a train, in a station building, and at an intersection. We used electret-type microphones (SONY ECM-C10) and a multi-channel input board with microphone amplifiers (Tokyo Electron Device TD-BD-8CSUSB). The noise was recorded at the sampling frequency of 44.1 kHz, and then down-sampled to 16 kHz. The target components observed at the microphones were simulated by a spherical wave assumption. The observed signals were obtained by adding the noise recording and the simulated target components. The duration of the observed signals was 10 s. As the model of inter-channel noise coherences for McCowan's method, the cylindrically isotropic model was used, which always resulted in a better result than the spherically isotropic model in our experiments. The other conditions were the same as those in Section V-A.

To demonstrate the effectiveness of BND for real-world noise, we define a *diagonal ratio* of $\boldsymbol{\Phi}_{\boldsymbol{vv}}(\tau,\omega)$ as

$$\frac{\sum_{m=1}^{M}|\phi_{v_m v_m}(\tau,\omega)|}{\sum_{m=1}^{M}\sum_{n=1}^{M}|\phi_{v_m v_n}(\tau,\omega)|}. \tag{76}$$

The diagonal ratio of the noise covariance matrix after BND $\boldsymbol{U}^{\mathsf{H}}\boldsymbol{\Phi}_{\boldsymbol{vv}}(\tau,\omega)\boldsymbol{U}$ is defined in the same manner. The diagonal ratio lies in the range 0 to 1 and equals 1 if and only if the matrix is perfectly diagonal. The *average diagonal ratio* is then defined by the temporal average of the diagonal ratio. Fig. 4 plots the average diagonal ratio before and after BND as a function of the frequency for noise in the station square. The results for the other environments were similar. We can see that the average diagonal ratio increased through BND, which shows its effectiveness for real-world noise.

Then we examine the overall performance of the methods. The output SNRs, NRs, and SDs of the methods for each environment are shown in Fig. V-B. We see that Zelinski's method causes little target distortion but does not suppress noise well. In contrast, McCowan's method reduced noise significantly but at the cost of much target distortion, likely because of mismatch between the assumed and the actual coherences. On the other hand, the proposed method suppressed noise well with less target distortion, and consequently gave a highest overall SNR value among all methods in most cases. The only exception was that it was outperformed by Zelinski's method slightly in terms of the output SNR for the platform noise, which was not very isotropic because there was a loud announcement during recording. These results show that the proposed method suppresses diffuse noise without causing much distortion.

C. Comparison between Proposed and Conventional Methods (Echoic Case)

Finally we consider the echoic case. We examined

- The MVDR beamformer, Zelinski's and McCowan's post-filters, and the proposed post-filter presented in

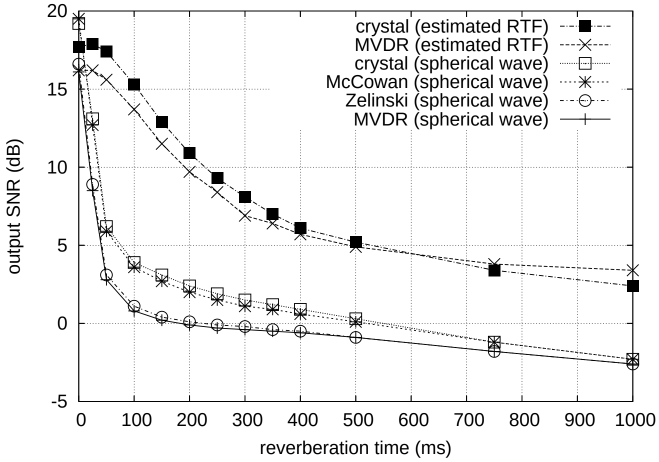


Fig. 6. Output SNR vs. reverberation time (environment: station building).

Section III-B. The steering vectors were calculated on the assumption of spherical wave propagation.

- The MVDR beamformer and post-filtering method using the short-time power spectra and the steering vectors estimated jointly by the method in Section III-D.

The target signals at the microphones were simulated via the image method [26]. (We used a matlab code ”roomsim_single.m” written by Vincent [27].) The dimension of the room was assumed to be $7 \times 5 \times 3\,\mathrm{m}$, the array to be situated at the center of the room, and the distance between the target source to the array centroid to be $1\,\mathrm{m}$ (azimuth $\phi = 60°$; zenith angle $\theta = 90°$). The noise recorded in the station building was used. The frame length was 1024, frame shift was 32, and the interchannel cross-spectra of the observed signals were calculated using consecutive 32 frames. The joint estimation method was initialized with the steering vectors based on the spherical wave model. The number of iterations was 1000. Fig. 6 plots the output SNR as a function of the reverberation time RT_{60}. We altered the reverberation time to desired values by adjusting the absorption coefficients of the walls. In the presence of reverberation, the methods using the steering vectors estimated by the method worked significantly better than those with spherical wave steering vectors. For extremely high reverberation times, the proposed post-filter was inferior to the MVDR beamformer. This is likely because the proposed post-filter reduces more the trans-frame reverberation compared to the beamformer. This result clearly shows the effectiveness of the joint estimation method for diffuse noise suppression in reverberant environments.

VI. Conclusion

This paper described a new design of the Wiener post-filter for diffuse noise suppression using the crystal arrays. Our design is based on diagonalization of the spatial noise covariance matrix and estimation of the short-time power spectra of the target signal using the denoised off-diagonal entries of the spatial observation covariance matrix. The crystal arrays make it possible to diagonalize any covariance matrix of isotropic noise by a constant unitary matrix. The crystal arrays comprise 1) regular polygonal arrays, 2) (twisted) rectangular arrays, 3) (twisted) regular polygonal prism arrays, 4) rectangular solid arrays, and 5) regular polyhedral arrays. We presented a constant unitary diagonalization matrix for BND for each class. Through experiments using real-world noise, it was shown that the proposed method works better than Zelinski's and McCowan's post-filter.

Furthermore, we presented a method for the joint estimation of the steering vectors and the short-time power spectra of the target signal from the denoised off-diagonal entries of the spatial observation covariance matrix aiming at effective diffuse noise suppression in reverberant environments. The experimental result showed that the proposed post-filtering method and the MVDR beamformer using the estimated steering vectors worked substantially better than methods using the spherical wave steering vectors in the presence of reverberation.

Since the proposed approach assumes that there is only one directional sound, the performance may degrade when there are more than two strong directional sounds. Therefore the future work includes the generalization of the approach to the case of multiple directional sounds.

Acknowledgments

This work was partly supported by Grant-in-Aid for Young Scientists (B) 21760309 from MEXT, Japan.

References

[1] M. Brandstein and D. Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*. Berlin: Springer-Verlag, 2001.

[2] R. Zelinski, “A microphone array with adaptive post-filtering for noise reduction in reverberant rooms,” in *Proc. ICASSP ’88*, New York, Apr. 1988, pp. 2578–2581.

[3] K. U. Simmer and A. Wasiljeff, “Adaptive microphone arrays for noise suppression in the frequency domain,” in *Second Cost 229 Workshop on Adaptive Algorithms in Communications*, Bordeaux, Oct. 1992, pp. 185–194.

[4] S. Fischer and K. U. Simmer, “Beamforming microphone arrays for speech acquisition in noisy environments,” *Speech Commun.*, vol. 20, no. 3–4, pp. 215–227, Dec. 1996.

[5] K. U. Simmer, J. Bitzer, and C. Marro, “Post-filtering techniques,” in *Microphone Arrays: Signal Processing Techniques and Applications*, M. Brandstein and D. Ward, Eds. Berlin: Springer-Verlag, 2001, ch. 3, pp. 39–60.

[6] J. Bitzer, K. U. Simmer, and K.-D. Kammeyer, “Multi-microphone noise reduction techniques as front-end devices for speech recognition,” *Speech Commun.*, vol. 34, no. 1–2, pp. 3–12, Apr. 2001.

[7] I. A. McCowan and H. Bourlard, “Microphone array post-filter based on noise field coherence,” *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 709–716, Nov. 2003.

[8] I. Cohen, “Multichannel post-filtering in nonstationary noise environments,” *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1149–1160, May 2004.

[9] S. Gannot and I. Cohen, “Speech enhancement based on the general transfer function GSC and postfiltering,” *IEEE Trans. Speech Audio Process.*, vol. 12, no. 6, pp. 561–571, Nov. 2004.

[10] J. Li and M. Akagi, “A noise reduction system based on hybrid noise estimation technique and post-filtering in arbitrary noise environments,” *Speech Commun.*, vol. 48, no. 2, pp. 111–126, Feb. 2006.

[11] S. Lefkimmiatis and P. Maragos, “A generalized estimation approach for linear and nonlinear microphone array post-filters,” *Speech Commun.*, vol. 49, no. 7–8, pp. 657–666, July–Aug. 2007.

[12] H. L. V. Trees, *Optimum Array Processing*. New York: John Wiley & Sons, 2002.

[13] R. K. Cook, R. V. Waterhouse, R. D. Berendt, S. Edelman, and J. M. C. Thompson, “Measurement of correlation coefficients in reverberant sound fields,” *J. Acoust. Soc. Am.*, vol. 27, no. 6, pp. 1072–1077, Nov. 1955.

[14] 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*, Las Vegas, USA, Apr. 2008, pp. 317–320.

[15] S. Gannot, D. Burshtein, and E. Weinstein, “Signal enhancement using beamforming and nonstationarity with applications to speech,” *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.

[16] J. Benesty, J. Chen, and Y. A. Huang, “A minimum speech distortion multichannel algorithm for noise reduction,” in *Proc. ICASSP 2008*, Las Vegas, USA, 2008, pp. 321–324.

[17] S. Doclo and M. Moonen, “GSVD-based optimal filtering for single and multimicrophone speech enhancement,” *IEEE Trans. Signal Process.*, vol. 50, no. 9, pp. 2230–2244, Sept. 2002.

[18] H. Shimizu, N. Ono, K. Matsumoto, and S. Sagayama, “Isotropic noise suppression in the power spectrum domain by symmetric microphone arrays,” in *Proc. WASPAA*, New Paltz, NY, Oct. 2007, pp. 54–57.

[19] N. Ono, N. Ito, and S. Sagayama, “Five classes of crystal arrays for blind decorrelation of diffuse noise,” in *Proc. SAM*, Darmstadt, Germany, July 2008, pp. 151–154.

[20] G. W. Elko, “Spatial coherence functions for differential microphones in isotropic noise fields,” in *Microphone Arrays: Signal Processing Techniques and Applications*, M. Brandstein and D. Ward, Eds. Berlin: Springer-Verlag, 2001, ch. 4, pp. 61–85.

[21] H. Sun, S. Yan, and U. P. Svensson, “Robust spherical microphone array beamforming with multi-beam-multi-null steering and sidelobe control,” in *Proc. WASPAA*, New Paltz, NY, Oct. 2009, pp. 113–116.

[22] G. A. F. Seber, *A Matrix Handbook for Statisticians*. New Jersey: John Wiley & Sons, Inc., 2008.

[23] A. Tanaka, M. Miyakoshi, and N. Ono, “Analysis on blind decorrelation of isotropic noise correlation matrices based on symmetric decomposition,” in *Proc. SSP*, Cardiff, UK, Sept. 2009, pp. 421–424.

[24] M. Souden, J. Benesty, and S. Affes, “On optimal frequency-domain multichannel linear filtering for noise reduction,” *IEEE Trans. Acoust., Speech, Lang. Process.*, vol. 18, no. 2, pp. 260–276, 2010.

[25] A. Kurematsu, K. Takeda, Y. Sagisaka, S. Katagiri, H. Kuwabara, and K. Shikano, “ATR Japanese speech database as a tool of speech recognition and synthesis,” vol. 9, no. 4, pp. 357–363, Aug. 1990.

[26] J. B. Allen and D. A. Berkley, “Image method for efficiently simulating small-room acoustics,” *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[27] E. Vincent and D. R. Campbell. (2010, Nov. 29). [Online]. Available: http://www.irisa.fr/metiss/members/evincent/Roomsimove.zip.