A BLIND NOISE DECORRELATION APPROACH WITH CRYSTAL ARRAYS ON DESIGNING POST-FILTERS FOR DIFFUSE NOISE SUPPRESSION

Nobutaka Ito, Nobutaka Ono, and Shigeki Sagayama

Graduate School of Information Science and Technology, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan
{ito, onono, sagayama}@hil.t.u-tokyo.ac.jp

ABSTRACT

This paper describes a new framework for extracting the target signal in diffuse noise environments. We utilize *crystal arrays*, a certain class of symmetrical microphone arrays with crystal-like geometries, which enable interchannel decorrelation of isotropic noise without knowing the value of its covariance matrix. We refer to this decorrelation as *blind noise decorrelation*. Using an improved estimation of the signal power spectrum obtained by the blind noise decorrelation, the multichannel Wiener filter is properly implemented, which is the optimal estimator of the target signal in the minimum mean square error sense. Simulated experiments have shown the effectiveness of the proposed method.

Index Terms— Array signal processing, covariance matrix, diffuse noise, post-filter, power spectrum estimation.

1. INTRODUCTION

In the field of array signal processing, considerable research has been conducted on extracting the target signal from the observed noisy signals in various situations [1]. A fundamental delay-and-sum beamformer with a huge number of microphones arrayed in a sufficiently large aperture could achieve sharp directivity, but it is infeasible. The Minimum Variance Distortionless Response (MVDR) beamformer works well even with small arrays, especially for localized noise sources by steering its nulls in the direction of them. In diffuse noise situations such as cocktail parties, stations, or reverberant rooms, combining the MVDR beamformer with post-filtering techniques [2] is more effective. It is shown that the multichannel Wiener filter, which is the optimal estimator of the target signal in the Minimum Mean Square Error (MMSE) sense, can be decomposed into the MVDR beamformer and a subsequent Wiener post-filter, where the most important issue is the implementation of the Wiener post-filter.

So far, several methods have been proposed for implementing the Wiener post-filter. The most common approach is to utilize the power spectra and cross-spectra of the observed signals. Zelinski's method [3] is based on the assumption that the noise components in the observed signals are mutually uncorrelated. However, this assumption is inaccurate, especially for small arrays or at low frequencies. For ideal noise fields such as spherically isotropic noise fields, the theoretical coherence function is available. The methods proposed by McCowan *et al.* [5] and Lefkimmiatis *et al.* [6] are based on the assumption of a known noise field coherence function, improving the Zelinski's method. However, these methods may still yield an inaccurate result when the assumed coherence function is far from the actual one.

Instead of assuming a known noise field coherence function, our method is based on interchannel decorrelation of isotropic noise without knowing the value of its covariance matrix, utilizing symmetrical microphone arrays with crystal-like geometries. This decorrelation enables the accurate estimation of the signal power spectrum, so that the Wiener post-filter is implemented properly. This paper is organized as follows. In section 2, the multichannel Wiener filter is reviewed briefly. In section 3, the novel method for implementing the Wiener post-filter is described. In section 4, experimental results are presented to verify the effectiveness of the proposed method.

2. MULTICHANNEL WIENER FILTER

2.1. Observation Model

We assume that each of the M microphones in the array receives a delayed and attenuated version of the target signal corrupted by diffuse noise. This observation model can be written in the time-frequency domain as

$$\boldsymbol{X}(t,\omega) = \boldsymbol{d}(\omega)S(t,\omega) + \boldsymbol{N}(t,\omega), \tag{1}$$

where $\boldsymbol{X}(t,\omega)$ denotes the observation vector, $S(t,\omega)$ the target signal, $\boldsymbol{d}(\omega)$ the known steering vector, and $\boldsymbol{N}(t,\omega)$ the diffuse noise component.

2.2. Wiener Post-filter

We consider estimating $S(t,\omega)$ from $\boldsymbol{X}(t,\omega)$ by

$$\hat{S}(t,\omega) \triangleq \boldsymbol{w}^{\mathsf{H}}(t,\omega)\boldsymbol{X}(t,\omega), \tag{2}$$

where $\boldsymbol{w}(t,\omega)$ is a deterministic weight vector. For brevity, we will omit the arguments t and ω hereafter. We assume that S and $\boldsymbol{N}$ are zero-mean and mutually uncorrelated. The optimal weight vector in the MMSE sense is the multichannel Wiener filter

$$\boldsymbol{w}_{\text{opt}} \triangleq \boldsymbol{\Phi}_{\boldsymbol{XX}}^{-1} \phi_{SS} \boldsymbol{d}, \tag{3}$$

where $\boldsymbol{\Phi}_{\boldsymbol{XX}} \triangleq E[\boldsymbol{X}\boldsymbol{X}^{\mathsf{H}}]$ and $\phi_{SS} \triangleq E[|S|^2]$, where similar notations will be used throughout this paper. Noting that

$$\boldsymbol{\Phi}_{\boldsymbol{XX}} = \boldsymbol{d}\boldsymbol{d}^{\mathsf{H}} \phi_{SS} + \boldsymbol{\Phi}_{\boldsymbol{NN}}, \tag{4}$$

we can factorize $\boldsymbol{w}_{\text{opt}}$ into the MVDR beamformer ($\boldsymbol{w}_{\text{MVDR}}$) and the Wiener post-filter (H_{post}) [2]:

$$\boldsymbol{w}_{\text{opt}} = \underbrace{\frac{\phi_{SS}}{\phi_{SS} + (\boldsymbol{d}^{\mathsf{H}} \boldsymbol{\Phi}_{\boldsymbol{NN}}^{-1} \boldsymbol{d})^{-1}}}_{H_{\text{post}}} \cdot \underbrace{\frac{\boldsymbol{\Phi}_{\boldsymbol{XX}}^{-1} \boldsymbol{d}}{\boldsymbol{d}^{\mathsf{H}} \boldsymbol{\Phi}_{\boldsymbol{XX}}^{-1} \boldsymbol{d}}}_{\boldsymbol{w}_{\text{MVDR}}}. \tag{5}$$

$\boldsymbol{w}_{\text{MVDR}}$ can be easily calculated from the observed signals. Therefore, the key for implementing the multichannel Wiener filter is how to estimate ϕ_{SS} accurately from the observed noisy signals.

2.3. Zelinski's Method

If the noise components in the observed signals are uncorrelated, ϕ_{SS} can be obtained from each nondiagonal element in (4), which is noise-free according to the above assumption, by

$$\phi_{SS} = \frac{\phi_{X_m X_n}}{d_m d_n^*} \quad (m \neq n), \tag{6}$$

where * denotes complex conjugation. Zelinski [3] estimates ϕ_{SS} by averaging (6) for all m and n such that $m < n$ and taking the real part:

$$\phi_{SS} = \Re\left[\frac{2}{M(M-1)} \sum_{m<n} \frac{\phi_{X_m X_n}}{d_m d_n^*}\right]. \tag{7}$$

However, this estimation may be inaccurate for small arrays or at low frequencies, because the noise components in the observed signals are highly correlated in such cases.

3. PROPOSED METHOD

3.1. Blind Noise Decorrelation by Crystal Arrays

Let $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_M$ be orthonormal eigenvectors of $\boldsymbol{\Phi}_{\boldsymbol{NN}}$ and $\boldsymbol{E} \triangleq [\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_M]$, then the elements of the vector

$$\tilde{\boldsymbol{N}} \triangleq \boldsymbol{E}^{\mathsf{H}} \boldsymbol{N} \tag{8}$$

are uncorrelated. The transformed version of (4) is

$$\boldsymbol{\Phi}_{\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}} = \tilde{\boldsymbol{d}}\tilde{\boldsymbol{d}}^{\mathsf{H}} \phi_{SS} + \boldsymbol{\Phi}_{\tilde{\boldsymbol{N}}\tilde{\boldsymbol{N}}}, \tag{9}$$

where $\tilde{\boldsymbol{X}}$ and $\tilde{\boldsymbol{d}}$ are defined in the same way as $\tilde{\boldsymbol{N}}$. Because $\boldsymbol{\Phi}_{\tilde{\boldsymbol{N}}\tilde{\boldsymbol{N}}}$ is diagonal, the nondiagonal elements in (9) are noise-free, from which ϕ_{SS} can be obtained.

Generally, we need to know the value of $\boldsymbol{\Phi}_{\boldsymbol{NN}}$ to obtain $\boldsymbol{E}$. Although possible based on Voice Activity Detection (VAD) when the target signal is speech and the noise is non-speech, this is difficult in general. However, we have shown that $\boldsymbol{E}$ can be obtained without knowing the value of $\boldsymbol{\Phi}_{\boldsymbol{NN}}$, if we use *crystal arrays*, a certain class of symmetrical microphone arrays with crystal-like geometries (see Fig. 1) and the noise field satisfies the following isotropy conditions [7].

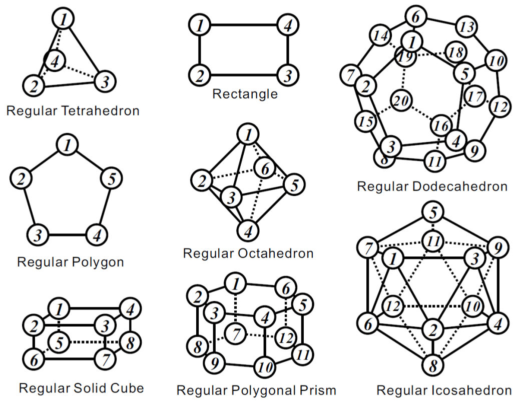

Fig. 1. Examples of crystal arrays.

- The power spectrum is independent of the observation position.
- The cross-spectrum is independent of the orientation between the two observation positions.

Therefore, if the noise field is isotropic, we can decorrelate $\boldsymbol{N}$ without knowing the value of $\boldsymbol{\Phi}_{\boldsymbol{NN}}$ using crystal arrays. We refer to this technique as *Blind Noise Decorrelation* (BND). For example, consider an array with microphones arrayed at the vertices of a square and numbered cyclically, then the noise covariance matrix is a circulant matrix

$$\boldsymbol{\Phi}_{\boldsymbol{NN}} = \begin{bmatrix} \alpha & \beta & \gamma & \beta \\ \beta & \alpha & \beta & \gamma \\ \gamma & \beta & \alpha & \beta \\ \beta & \gamma & \beta & \alpha \end{bmatrix}. \tag{10}$$

A corresponding matrix $\boldsymbol{E}$ is the fourth-order DFT matrix [8]

$$\boldsymbol{E} = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & j & j^2 & j^3 \\ 1 & j^2 & j^4 & j^6 \\ 1 & j^3 & j^6 & j^9 \end{bmatrix} \quad (j \triangleq \sqrt{-1}). \tag{11}$$

3.2. Power Spectrum Estimation

Assume that BND has been performed, then it follows from each nondiagonal element in (9) that

$$\phi_{\tilde{X}_m \tilde{X}_n} = \tilde{d}_m \tilde{d}_n^* \phi_{SS} \quad (m \neq n). \tag{12}$$

We estimate ϕ_{SS} from (12) for m and n such that $m < n$, based on the least square method, as

$$\phi_{SS} = \Re\left[\frac{\sum_{m<n} \tilde{d}_m^* \tilde{d}_n \phi_{\tilde{X}_m \tilde{X}_n}}{\sum_{m<n} |\tilde{d}_m|^2 |\tilde{d}_n|^2}\right]. \tag{13}$$

If estimated ϕ_{SS} is negative, then we reset it to zero.

3.3. Wiener Post-filter Design

Combining the estimation of the signal power spectrum in (13) with the estimation of the denominator proposed by Zelinski [3], we design the Wiener post-filter as

$$H_{\text{post}} = \frac{\Re\left[\dfrac{\sum_{m<n} \tilde{d}_m^* \tilde{d}_n \phi_{\tilde{X}_m \tilde{X}_n}}{\sum_{m<n} |\tilde{d}_m|^2 |\tilde{d}_n|^2}\right]}{\dfrac{1}{M}\displaystyle\sum_{m=1}^{M} \dfrac{\phi_{X_m X_m}}{|d_m|^2}}. \tag{14}$$

Since H_{post} satisfies $0 \le H_{\text{post}} \le 1$ by definition, we modify its value in (14) by

$$H_{\text{post}} \leftarrow \begin{cases} 0 & (H_{\text{post}} < 0) \\ 1 & (H_{\text{post}} > 1) \end{cases}. \tag{15}$$

4. EXPERIMENTAL EVALUATION

In this section, we present results of simulated experiments to verify the effectiveness of the proposed method. The experimental conditions were as follows (see Fig. 2).

- array geometry: square
- target: speech (DOA: 60°)
- noise: white noise or speech (DOA: $\frac{360^\circ}{64}i$, $0 \le i \le 63$)
- propagation: plane wave, single-path
- sampling rate: 16kHz
- frame length: 2^8 samples
- shift: 2^4 samples
- window: Hamming

The speech data used were continuous speech data in the ATR Japanese speech database (set B) [9]. The target speech was extracted by the multichannel Wiener filter with the proposed post-filter (called below *the proposed method*) and that with Zelinski's post-filter (called below *the Zelinski's method*). $\boldsymbol{\Phi}_{\boldsymbol{XX}}$ in $\boldsymbol{w}_{\text{MVDR}}$ was estimated by averaging in time $\boldsymbol{X}\boldsymbol{X}^{\mathsf{H}}$ over all the frames. As the matrix $\boldsymbol{E}$ for BND, the fourth-order DFT matrix (11) was used. $\phi_{\tilde{X}_m \tilde{X}_n}$ and $\phi_{X_m X_m}$ in (14) were estimated every 2^4 frames by averaging in time $\tilde{X}_m \tilde{X}_n^*$ and $|X_m|^2$. The range restriction (15) was applied to the Zelinski's method as well.

In Fig. 3, the noise covariance matrices before and after BND (density plots) are shown. They were calculated by averaging $\boldsymbol{X}\boldsymbol{X}^{\mathsf{H}}$ and $\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^{\mathsf{H}}$ over all the time-frequency slots. The interferences were speech signals. It is seen that the noise covariance matrix is almost circulant before BND and almost diagonal after that, which shows the effectiveness of BND. The ratio of the absolute sum of the diagonal elements to that of all the elements increased from 37.1% to 96.3 % through BND.

Next, we compare the accuracy of the power spectrum estimation for the two methods mentioned above. In Fig. 4, the scatter graph of the *true* value (horizontal axis) and estimated

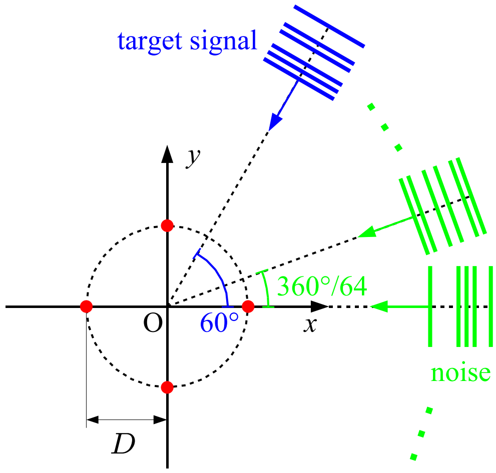

Fig. 2. The arrangement of the sources and the microphones.

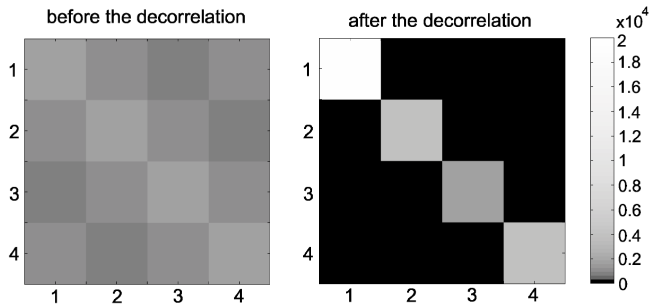

Fig. 3. The density plots (absolute values) of the noise covariance matrices before and after BND.

value (vertical axis) of ϕ_{SS} is shown. The array radius was set at 0.015m, and the input SNR at -5.0dB. The *true* signal power spectrum was calculated every 2^4 frames by averaging in time $|S|^2$. For the proposed method, the correlation coefficient is 0.919, and the points lie near the line $y = x$. This shows the accuracy of the estimation. On the other hand, for the Zelinski's method, the correlation coefficient is 0.777, and many points lie far from the line.

Finally, we compare the noise reduction performance of the proposed method with that of the Zelinski's method. Shown in Fig. 5 is the graph of the SNR enhancement (SNRE), which is defined as the difference of the output and input SNR, as a function of the input SNR. To show the SNR gain achieved by the post-filtering, the result for the MVDR beamformer without post-filtering is also shown. The array radius was set at 0.015m to avoid the spatial aliasing. It is seen that the proposed method gives superior SNR than the Zelinski's method. In Fig. 6, the output SNR is plotted as a function of the array radius. The input SNR was set at 0.0dB. For small arrays, the proposed method gives superior SNR than the Zelinski's method, and little SNR gain is achieved by post-filtering in the Zelinski's method. For larger arrays, the difference in the SNR between the two methods is small. This result is well explained by the fact that the assumption of the uncorrelated noise, upon which the Zelinski's method is based, is valid only for large arrays.

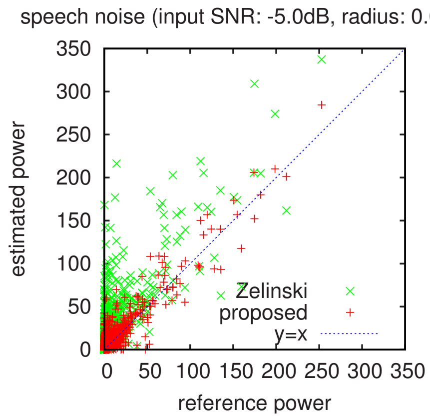

Fig. 4. The scatter graph of the true value (horizontal axis) and estimated value (vertical axis) of ϕ_{SS}.

5. CONCLUSION

We presented a novel method for implementing the multichannel Wiener filter for diffuse noise suppression, based on BND using crystal arrays. Simulated experiments have shown that 1) diffuse noise is decorrelated effectively by BND, that 2) the estimation of the signal power spectrum by the proposed method is more accurate than that by the Zelinski's method, and that 3) the proposed method gives higher SNR than the Zelinski's method especially for small arrays.

6. REFERENCES

[1] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing*, Prentice Hall, New York, 1993.

[2] K. U. Simmer, J. Bitzer, and C. Marro, "Post-filtering techniques," in *Microphone Arrays*, M. Brandstein and D. Ward Eds., Springer-Verlag, Berlin, ch. 3, pp. 39–60, 2001.

[3] R. Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms," *Proc. ICASSP*, pp. 2578–2581, 1988.

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

[5] 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, 2003.

[6] 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, 2007.

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

[8] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, 1983.

[9] 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," *Speech Commun.*, vol. 9, no. 4, pp. 357–363, 1990.

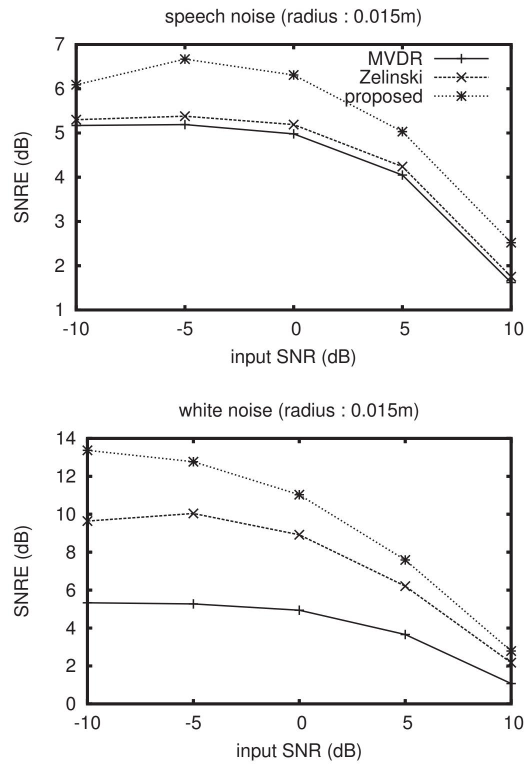

Fig. 5. The graph of the SNR enhancement (SNRE) as a function of the input SNR.

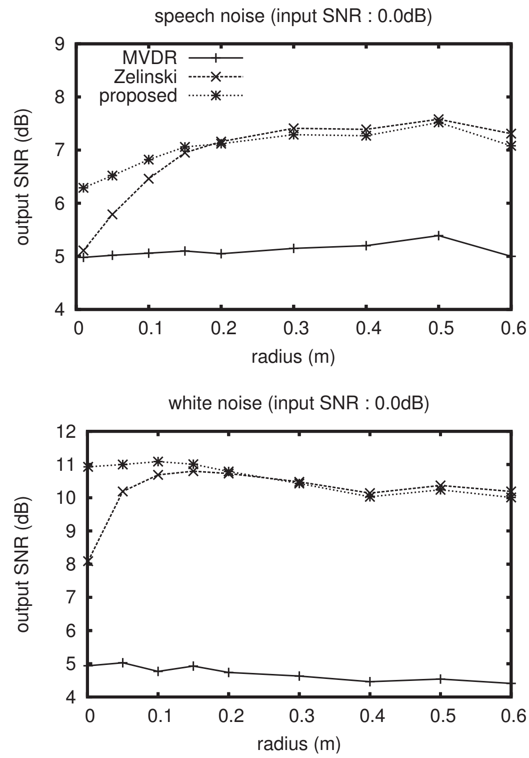

Fig. 6. The graph of the output SNR as a function of the array radius.