THEORETICAL ANALYSIS OF MICROPHONE ARRAYS WITH POSTFILTERING FOR COHERENT AND INCOHERENT NOISE SUPPRESSION IN NOISY ENVIRONMENTS

Junfeng Li and *Masato Akagi*

School of Information Science
Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa, 923-1292, Japan
{junfeng, akagi}@jaist.ac.jp

ABSTRACT

Microphone arrays have shown substantial capability in reducing noise due to their spatial selectivity. In this paper, we first introduce two noise reduction algorithms using microphone arrays to deal with both coherent and incoherent noise components. Two algorithms have the same components, a *generalized sidelobe canceller* (GSC) and a Wiener filter, but with different structures. Their theoretical performance are then analyzed in different theoretically defined noise fields. The performance is further evaluated by using the real-world multi-channel noise recordings. With regard to the theoretical and experimental results, some discussions on two algorithms are also presented.

1. INTRODUCTION

Combating undesired noise signal is still a challenging research topic, although it has been studied for several decades. The difficulty is mainly caused by the complexity and variation of the environments. In practical environments, noise signals generally consist of coherent and incoherent components and noise characteristics also vary with time and place. To deal with this problem, in recent years, microphone arrays have been widely researched and shown the substantial superiority due to its spatial filtering capability.

A variety of microphone arrays have been reported so far. The conventional beamformer, *delay-and-sum beamformer*, enhances the desired speech signal by summing the in-phase microphone signals. The linearly constrained adaptive beamformer, first presented by Frost, keeps the signals arriving from the desired look-direction distortionless while suppressing the signals from other directions by minimizing the power of the beamformer output [1]. A *generalized sidelobe canceller* (GSC) beamformer, first presented by Griffiths and Jim as an alternative implementation structure of the Frost beamformer, has also been widely researched [2]. Bitzer proved that the theoretical noise reduction performance of the GSC beamformer reaches infinite in a coherent noise field, while limited in a diffuse noise field and an incoherent noise field [3]. To further deal with incoherent noise component, Fischer *et al.* suggested to apply a Wiener post-filter in the upper path of the GSC beamformer [4]. While, other systems exploited a Wiener post-filter at the beamformer output [5][6]. However, no comparison of their noise reduction performance was reported.

In this paper, we first introduce two noise reduction algorithms using microphone arrays with post-filtering which are designed to suppress coherent and incoherent noises in noisy environments. Two algorithms consist of the same components, a GSC beamformer and a Wiener filter, but with different structures. Then we mainly focus on discussing their theoretical performance in well defined noise fields. The theoretical discussion results are further confirmed by the experimental results using multi-channel noise recordings. Finally, some discussions on two studied algorithms are described.

2. NOISE REDUCTION ALGORITHMS USING MICROPHONE ARRAYS

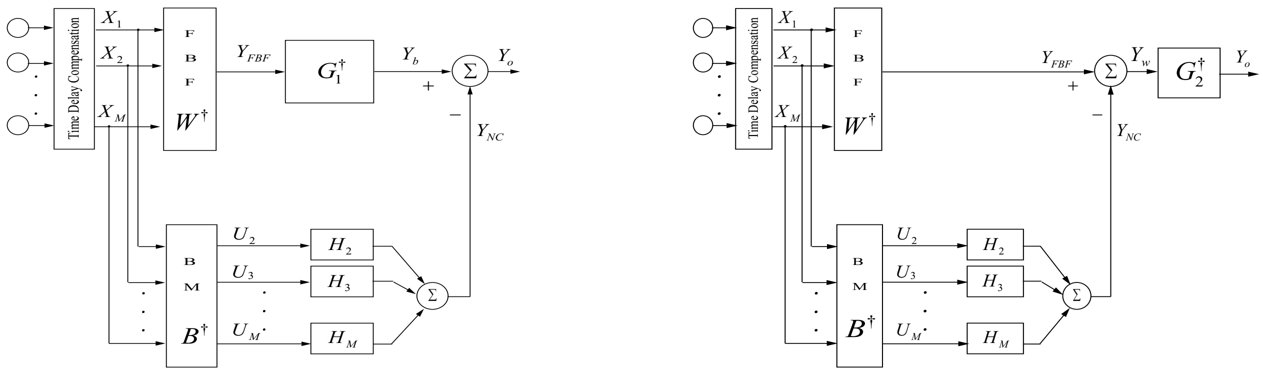
In this section, we describe two noise reduction algorithms to be investigated, which have the same components and different structures.

The block diagrams of the studied algorithms are shown in Fig. 1. As Fig. 1 shows, two algorithms have the same components: a GSC beamformer and a Wiener post-filter $G^\dagger$. The GSC beamformer consists of three components: a *Fixed Beamformer* (FBF) $\mathbf{W}^\dagger$, a *Blocking Matrix* (BM) $\mathbf{B}^\dagger$ and multi-channel noise canceller $\mathbf{H}^\dagger$, respectively, defined by:

$$\mathbf{W}^\dagger = \frac{1}{M}[1, 1, \cdots, 1], \tag{1}$$

$$\mathbf{H}^\dagger(\omega) = [H_2^*(\omega), H_3^*(\omega), \cdots, H_M^*(\omega)], \tag{2}$$

where $H_i, (i = 2, 3, \cdots, M)$ should be determined accord-



(a) (b)

Fig. 1. Block diagrams of the studied noise reduction algorithm 1 (a) and algorithm 2 (b).

ing to the input signals, and

$$\mathbf{B}^\dagger = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{bmatrix} \tag{3}$$

where superscript † and * are the conjugative transpose and conjugate, and ω denotes the frequency index.

In the following derivation, for simplicity, we omit the frequency index ω and assume: speech and noise are independent, and noise *power spectral density* (PSD) is identical on each microphone. As Fig. 1 shows, the difference between two studied algorithms can be summarized as:

- *Algorithm 1.* In this algorithm, to suppress incoherent noise, the Wiener filter is applied on the FBF output in the upper path of the GSC beamformer. And coherent noise is then suppressed by subtracting the noise estimate from the speech reference signal (the Wiener filter output).

  To avoid speech distortion, the noise cancellers are determined when no speech is present, given by:

$$\mathbf{H}_1 = \left(\mathbf{B}^\dagger \boldsymbol{\Gamma} \mathbf{B}\right)^{-1} \mathbf{B}^\dagger \boldsymbol{\Gamma} \mathbf{W} G_1, \tag{4}$$

  where $\boldsymbol{\Gamma}$ is the noise coherence matrix and the Wiener filter $G_1^\dagger$ is implemented by the improved Zelinski post-filter, given by [7]:

$$\begin{aligned} G_1^\dagger &= \frac{\frac{2}{M(M-1)} \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \phi_{n_i n_j}}{\phi_{Y_{FBF} Y_{FBF}}} \\ &= \frac{\frac{2}{M(M-1)} \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \Gamma_{n_i n_j}}{\mathbf{W}^\dagger \boldsymbol{\Gamma} \mathbf{W}}. \end{aligned} \tag{5}$$

  And the system output is:

$$Y_1 = G_1^\dagger \mathbf{W}^\dagger \mathbf{X} - \mathbf{H}_1^\dagger \mathbf{B}^\dagger \mathbf{X}. \tag{6}$$

- *Algorithm 2.* In this algorithm, the GSC beamformer is first applied to the multi-channel inputs to suppress coherent noise components. And the GSC output is then processed by the Wiener filter to suppress the residual incoherent noise components.

  The noise cancellers are still determined during speech-free periods as:

$$\mathbf{H}_2 = \left(\mathbf{B}^\dagger \Gamma \mathbf{B}\right)^{-1} \mathbf{B}^\dagger \Gamma \mathbf{W}, \tag{7}$$

  and the Wiener filter $G_2^\dagger$ is implemented as:

$$\begin{aligned} G_2^\dagger &= \frac{\frac{2}{M(M-1)} \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \Gamma_{n_i n_j}}{\phi_{Y_w Y_w}} \\ &= \frac{\frac{2}{M(M-1)} \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \Gamma_{n_i n_j}}{\mathbf{W}^\dagger \boldsymbol{\Gamma} \mathbf{W} - \mathbf{W}^\dagger \boldsymbol{\Gamma} \mathbf{B} \left(\mathbf{B}^\dagger \boldsymbol{\Gamma} \mathbf{B}\right)^{-1} \mathbf{B}^\dagger \boldsymbol{\Gamma} \mathbf{W}}. \end{aligned} \tag{8}$$

  Thus, the system output is:

$$Y_2 = G_2^\dagger \mathbf{W}^\dagger \mathbf{X} - G_2^\dagger \mathbf{H}_2^\dagger \mathbf{B}^\dagger \mathbf{X}. \tag{9}$$

3. THEORETICAL ANALYSIS OF TWO ALGORITHMS

In this section, we first give a measure used to show the noise reduction performance of two algorithms. Then their performance is examined in theoretically defined noise fields.

3.1. Performance evaluation measure

To examine the performance of the noise reduction algorithms, a measure, *noise reduction performance* (NR), is defined as the ratio of PSD of system input ϕ_{xx}^n and that of output $\phi_{Y_o Y_o}^n$ when desired speech is absent, given by:

$$\mathrm{NR} = \frac{\phi_{xx}^n}{\phi_{Y_o Y_o}^n}, \tag{10}$$

Note, that NR will achieve the infinite value if noise components are reduced completely.

3.1.1. *Noise reduction of Algorithm 1*

For algorithm 1, with Eqs. (1) - (6), the PSD of the system output when speech is absent is described as:

$$\phi^n_{Y_1Y_1} = G_1^\dagger \mathbf{W}^\dagger \Phi_{\mathbf{NN}} \mathbf{W} G_1 - G_1^\dagger \mathbf{W}^\dagger \Phi_{\mathbf{NN}} \mathbf{B} \left(\mathbf{B}^\dagger \Phi_{\mathbf{NN}} \mathbf{B}\right)^{-1} \mathbf{B}^\dagger \Phi_{\mathbf{NN}} \mathbf{W} G_1, \tag{11}$$

where, under the assumption of identical noise PSD ϕ_{nn} on each microphone, $\Phi_{\mathbf{NN}} = \phi_{nn}\mathbf{\Gamma}$, and the PSD of system input for speech free periods will be $\phi^n_{xx} = \phi_{nn}$. Therefore, NR of the algorithm 1 can be calculated as:

$$\mathrm{NR}_1 = [G_1^\dagger \mathbf{W}^\dagger \mathbf{\Gamma} \mathbf{W} G_1 - G_1^\dagger \mathbf{W}^\dagger \mathbf{\Gamma} \mathbf{B} \left(\mathbf{B}^\dagger \mathbf{\Gamma} \mathbf{B}\right)^{-1} \mathbf{B}^\dagger \mathbf{\Gamma} \mathbf{W} G_1]^{-1}. \tag{12}$$

3.1.2. *Noise reduction of Algorithm 2*

For algorithm 2, with Eqs. (1)-(3) and Eqs. (7)-(9), the PSD of the system output when speech is absent is given by:

$$\phi^n_{Y_2Y_2} = G_2^\dagger \mathbf{W}^\dagger \Phi_{\mathbf{NN}} \mathbf{W} G_2 - G_2^\dagger \mathbf{W}^\dagger \Phi_{\mathbf{NN}} \mathbf{B} \left(\mathbf{B}^\dagger \Phi_{\mathbf{NN}} \mathbf{B}\right)^{-1} \mathbf{B}^\dagger \Phi_{\mathbf{NN}} \mathbf{W} G_2. \tag{13}$$

With the assumption of identical noise PSD on each microphone, NR of the algorithm 2 is given by:

$$\mathrm{NR}_2 = [G_2^\dagger \mathbf{W}^\dagger \mathbf{\Gamma} \mathbf{W} G_2 - G_2^\dagger \mathbf{W}^\dagger \mathbf{\Gamma} \mathbf{B} \left(\mathbf{B}^\dagger \mathbf{\Gamma} \mathbf{B}\right)^{-1} \mathbf{B}^\dagger \mathbf{\Gamma} \mathbf{W} G_2]^{-1}. \tag{14}$$

3.2. Theoretical performance analysis

In the following, we examine the performance of the studied algorithms in theoretically defined noise fields.

1. Coherence noise field. In a coherent noise field, the coherence function is given by:

$$\Gamma_{n_in_j}(\omega) = e^{-j\omega\delta} \tag{15}$$

In this case, both algorithms reduce to the GSC beamformer since the Wiener filters are to be all pass filters. And the GSC beamformer show the infinite performance in a coherent noise field [3], Therefore, both studied algorithms also achieve the infinite noise reduction performance in this condition.

2. Incoherent noise field. In an incoherent noise field, the coherence function is zero for all frequencies, $\Gamma_{n_in_j}(\omega) = 0, \forall\omega$. In this noise field, since the performance of the adaptive path of the GSC beamformer is zeros [3], two algorithms will reduce to the delay-and-sum beamformer followed by the Wiener filter. The infinite noise reduction performance is achieved due to the contribution of the Wiener filter [8].

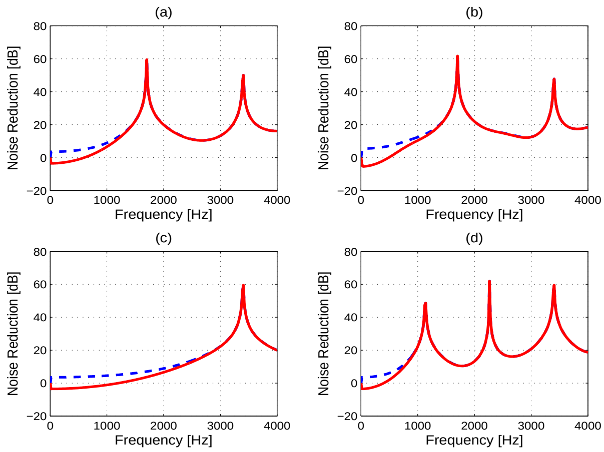


Fig. 2. Noise Reduction performance of the algorithm 1 (dashed) and of the algorithm 2 (solid) in a diffuse noise field for different array apertures. (a) $M = 3, d = 10cm$; (b) $M = 5, d = 10cm$; (c) $M = 3, d = 5cm$; (d) $M = 3, d = 15cm$.

3. Diffuse noise field. A diffuse noise field is characterized by the following coherent function:

$$\Gamma(\omega) = \frac{\sin(\omega d/c)}{\omega d/c}, \tag{16}$$

where d and c represent the inter-element spacing and the velocity of sound. The noise reduction performance depends on the inter-element spacing d and the number of microphones M. Fig. 2 plots the noise reduction performance of the studied algorithms for different number of microphones and different inter-element spacings. As Figs. 2 (a)(b) show, the performance of two systems demonstrates slight difference with the increasing number of microphones and the algorithm 1 outperforms the algorithm 2 in the low frequency region, while the same performance in the relatively high frequency region. Figs. 2 (c)(d) show that both algorithm provide higher noise reduction performance for all frequencies with the larger inter-element spacing.

4. EXPERIMENTS AND RESULTS

The comparison of the two noise reduction algorithms was further performed by experiments using multi-channel noise recordings. For this purpose, an 3-sensor equally-spaced linear array with inter-element spacing of 10 cm, was mounted in a car. The noise recordings were performed across all channels when car was running at the speed of 100km/h. The clean speech signals, consisting of 20 Japanese sentences from an ATR database. Both speech and noise signals were first re-sampled to 8kHz at 16 bit accuracy. We generated the multi-channel noisy signals by artificially mixing clean speech signals and real-world multi-channel car noise signals at different global SNR levels [-5, 15] dB.

Table 1. Segmental SNR [dB].

	Input global SNR				
	-5dB	0dB	5dB	10dB	15dB
Noisy	-29.14	-24.14	-19.14	-14.14	-9.14
Alg. 1	-16.59	-12.56	-8.39	-4.64	-1.46
Alg. 2	-16.67	-12.63	-8.42	-4.70	-1.51

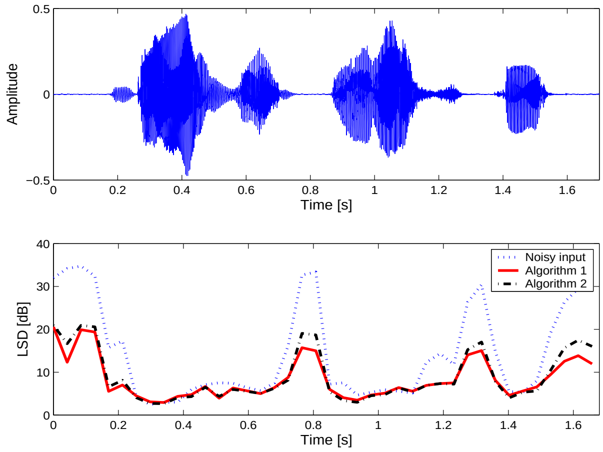


Fig. 3. Typical waveform of the clean speech /dozo yoroshiku/ (upper) and the LSDs (lower) of the noisy signal (SNR=10dB) and enhanced signals.

The experimental results of the average *Segmental SNR* (SEGSNR) are listed in Table 1. Both algorithms achieved the high SEGSNRs in all SNR levels. And compared to the algorithm 2, the algorithm 1 shows the slightly higher SEGSNR improvements in all conditions. These results confirmed the theoretical analysis results obtained in the previous section. A typical example is shown in Fig. 3 in terms of *Log-Spectral Distance* (LSD). As Fig. 3 shows, with regard to the algorithm 2, the algorithm 1 provides more "clean" enhanced signal, corresponding to a lower LSDs, in all periods, especially in the speech free periods. This improvement indicates a higher noise reduction performance was obtained by the algorithm 1. Whereas, listening tests showed that enhanced signals processed by the algorithm 1 involve some artifacts (eg. "musical noises"). Comparatively, musical noises are partly dealt with by the post-filter in the algorithm 2.

5. DISCUSSIONS

Based on the theoretical and experimental results, we present some discussions. In two algorithms, the noise PSDs are estimated using the cross-spectral densities of multi-channel inputs at beamformer output. For the algorithm 2, this approach overestimates the noise PSD since noise has partly been suppressed by the GSC beamformer, especially in low frequencies. Thus, the overestimated noise PSDs result in a larger gain function of the post-filter, further enlarging the noise spectra in the enhanced signal. Whereas, a practical problem associated with the algorithm 1 is the "musical noises" introduced by the spectral subtraction, as indicated by listening tests. While the artificial noises can be dealt with by the post-filter in the algorithm 2. Therefore, for both algorithms, there still have a large room to improve their performance in practical environments.

6. CONCLUSION

In this paper, we first described two noise reduction systems which have the same components, a GSC beamformer and a Wiener filter, with different structures. Then,we analyzed their theoretical noise reduction performance in well defined noise fields. And the theoretical analysis results were further confirmed by the experimental results using real-world car noises. Two algorithms were also discussed based on the theoretical and experimental results.

7. REFERENCES

[1] O. L. Frost, "An algorithm for linearly constrained adaptive array processing," in Proc. IEEE, vol. 60, pp. 926-935, 1972.

[2] L.J. Griffths and C.W. Jim. " alternative approach to linearly constrained adaptive beamforming". IEEE Trans. on Antennas Propagat., vol. AP-30, pp. 27-34, 1982.

[3] J. Bitzer, K.U. Simmer and K.-D. Kammeyer. Theoretical noise reduction limits of the generalized sidelobe canceller (GSC) for speech enhancement". In Proc. ICASSP'99, pp. 2965-2968, 1999.

[4] S. Fischer and K. U. Simmer, Beamforming microphone array for speech acquisition in noisy envi- ronments", Speech Communication, vol. 20, pp.215-227, 1996.

[5] R. Zelinski, Microphone Array With Adaptive Post-Filtering for Noise Reduction in Reverberant Rooms," in Proc. ICASSP'88, pp. 2578-2581, 1988.

[6] I. A. McCowan and H. Bourlard, Microphone array post-filter based on noise eld coherence," IEEE Trans. on Speech and Audio Proc., vol. 11, no. 6, pp. 709-716, 2003.

[7] K. U. Simmer and A. Wasilje. "Adaptive microphone arrays for noise suppression in the frequency domain," Second cost 229 Workshop on Adaptive Algorithm in communication, pp. 185-194, 1992.

[8] C. Marro, Y. Mahieux and K. U. Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postltering," IEEE Trans. on Speech and Audio Processing, vol. 6, no. 3, May, 1998.