ADAPTIVE SIGNAL BLOCKING FOR GENERALIZED SIDELOBE CANCELLER USING MATCHED FILTER ARRAY

Jwu-Sheng Hu and Chia-Hsing Yang
Department of Electrical and Control Engineering, National Chiao-Tung University.
Hsinchu, Taiwan, R.O.C.
jshu@cn.nctu.edu.tw, chyang.ece92g@nctu.edu.tw

ABSTRACT

This work proposes an adaptive beamformer based on generalized sidelobe canceller (GSC) structure with novel blocking matrix design. The classical GSC presented by Griffiths and Jim suffers from desired signal cancellation problems due to the complicated acoustic environment. This work utilizes the pseudo-inverse property of matched filters and subarray structure to design a new blocking matrix of the GSC. For practical implementation, matched filter ratios between microphone pairs are estimated, instead of estimating matched filters from the sound source to each microphone. Simulation results are presented to show the effectiveness of the proposed method.

***Index Terms*—**GSC, matched filters, subarray

1. INTRODUCTION

The quality of captured speech signal is degraded when the environment is contaminated by interference signals such as room reverberation and ambient noise. Adaptive beamformer techniques, such as generalized sidelobe canceller, have been developed to improve the quality of acquired sound signal for a long time. Among various adaptive beamformers, the Griffiths and Jim beamformer (GJBF) [1] is the most widely known method which assumes that the sound wave propagation behavior is simple time-delay. However, in GJBF, the steering vector errors which cause the desired signal cancellation may occur due to the reverberant environment or microphone mismatch problems.

Jan and Flanagan proposed the matched filter array processing of microphone arrays [2]. It is shown in [3] that the distinct advantage of the matched filter beamformer over delay-and-sum beamformer is its ability to suppress the reverberation effect. They have also discussed the relation between the truncation length of the room transfer function and SNR quality. Affes and Grenier proposed the GSC-based near-field enhancement algorithms [4] and they utilized the matched filter with signal subspace tracking to design the fixed beamformer of the GSC.

Since matched filters show a superior property in enhancing source signal under reverberation, they can be used to block undesired source and have a minimal effect on desired signal cancellation. In this paper, an adaptive beamformer based on GSC architecture is presented. The proposed beamformer exploits the matched filter array property which can mitigate the effects of enclosure reverberation and subarray structure to design the blocking matrix of the GSC. The room impulse response measurements are required to construct the matched filter. Usually the most direct method to estimate room impulse response is the Maximal Length (ML) sequence method [5]. This is not practical if the environment changes frequently and the measurement has to be taken repeatedly. Hence, in this paper, matched filter measurement is replaced by estimating the matched filter ratio. It is not necessary to estimate the impulse response from the source to each microphone. The proposed algorithm is implemented in the frequency domain and the performance is evaluated by the simulations.

2. PRINCIPLES OF THE PROPOSED BEAMFORMER

Matched-filtering techniques can be applied to microphone arrays to improve signal to noise ratio (SNR) in reverberant environment. Consider M microphones in a reverberant and noisy environment, the received signal of the m-th microphone can be model as

$$x_m(t) = s(t) * a_m(t) + n_m(t) \tag{1}$$

where $*$ denotes the convolution; $a_m(t)$ is the transfer function from sound source to the m-th microphone and $n_m(t)$ is the interference noise. In the matched filter beamformer processing, the output of each microphone is convolved with the matched filter which is the time reverse of the transfer function from the source to that microphone and the beamformer output is the summation of the individual matched filter output. Therefore, the matched filter beamformer output of M microphones is expressed as:

$$\begin{aligned} y_{MFBF} &= \sum_{m=1}^{M} x_m(t) * a_m(-t) \\ &= s(t) * \sum_{m=1}^{M} a_m(t) * a_m(-t) + \sum_{m=1}^{M} n_m(t) * a_m(-t) \end{aligned} \tag{2}$$

where the term $\sum_{m=1}^{M} a_m(t) * a_m(-t)$ is defined as the matched

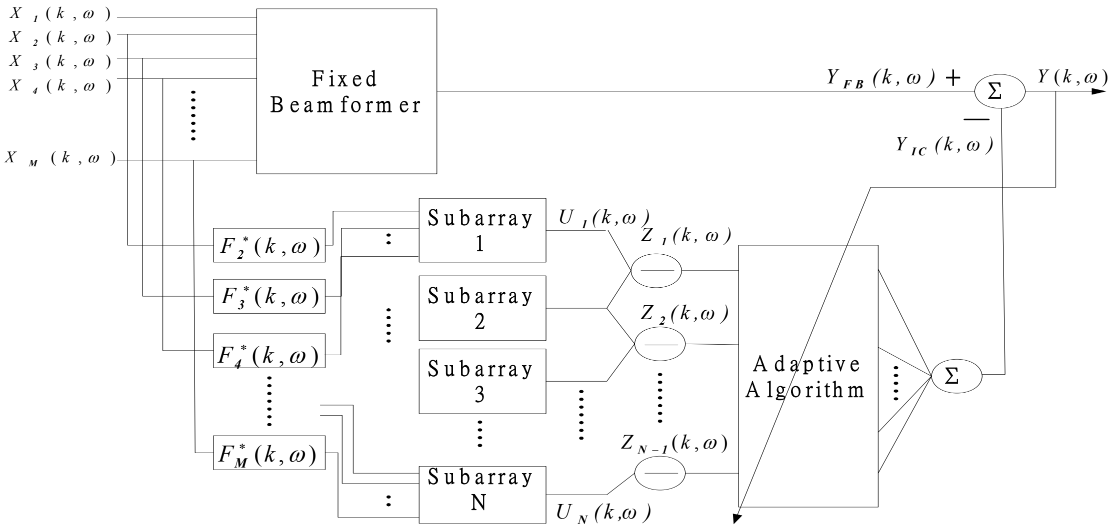


Figure 1. Overall system architecture

filter array response. The response of matched filter array approximates the large amplitude impulse if M is sufficiently large and the source signal is enhanced. It means the effect of reverberation can be suppressed and the approximated matched filter array response is represented as:

$$\sum_{m=1}^{M} a_m(t) * a_m(-t) \approx \alpha \cdot \delta(t) \quad (3)$$

where α is a scalar and $\delta(t)$ is the unit impulse and rewrite equation (3) using fast Fourier transform：

$$\sum_{m=1}^{M} A_m^*(\omega) A_m(\omega) \approx \alpha \quad (4)$$
$$\omega = 0,1,\cdots NFFT-1$$

where $*$ denotes complex conjugation and ω is the frequency band. This work adopted the delta-like property of the matched filter array response to design the blocking matrix of GSC. The proposed beamformer architecture is shown in Figure 1 and this architecture is comprised of three components including fixed beamformer, blocking matrix used to block desired signal and adaptive interference canceller.

2.1. Blocking matrix using matched filtering and subarrays techniques

In practice, it might not be easy to obtain the transfer function between the source and microphone to construct the matched filter. Instead, the matched filter ratios among microphones and subarray structure are proposed to design the blocking matrix. As shown in Figure 1, $X_m(l,\omega)$ is the Fourier transform of $x_m(t)$ at frame l. The inputs of blocking matrix are separated into N subarrays $\{G_1, G_2, \cdots, G_N\}$, and each subarray comprises P microphones. Notably, subarray G_n and G_{n+1} have no microphones in common. To be practical, considering only microphone received signals can be used to estimate the matched filter. Therefore, the cross-spectral method that exploits nonstationary features [6] is employed here to estimate the matched filter ratio described as:

$$F_m^*(\omega) = \frac{A_m^*(\omega)}{A_1^*(\omega)} \quad (5)$$
$$= \left(\frac{\left\langle \hat{\Phi}_{x_1x_1}(\omega)\hat{\Phi}_{x_mx_1}(\omega) \right\rangle_{avg} - \left\langle \hat{\Phi}_{x_1x_1}(\omega) \right\rangle_{avg} \left\langle \hat{\Phi}_{x_mx_1}(\omega) \right\rangle_{avg}}{\left\langle \hat{\Phi}^2_{x_1x_1}(\omega) \right\rangle_{avg} - \left\langle \hat{\Phi}_{x_1x_1}(\omega) \right\rangle^2_{avg}} \right)^*$$

where $\hat{\Phi}_{x_mx_1}$ is the estimated cross power spectral density between x_m and x_1 and $\langle \cdot \rangle_{avg}$ denotes the average operation. Hence, the output of the n-th subarray can be described as:

$$U_n(k,\omega) = \sum_{m \in G_n} F_m^* X_m(k,\omega)$$
$$= \sum_{m \in G_n} \left(S(k,\omega) A_m(\omega) \frac{A_m^*(\omega)}{A_1^*(\omega)} + N_m(k,\omega) \frac{A_m^*(\omega)}{A_1^*(\omega)} \right) \quad (6)$$
$$n = 1, \cdots, N$$

As a result, the n-th output $U_n(k,\omega)$ of the blocking matrix (see Figure 1) can be shown as:

$$U_n(k,\omega)-U_{n+1}(k,\omega)=\sum_{m\in\boldsymbol{G}_n}\left(\frac{S(k,\omega)}{A_1^*(\omega)}A_m(\omega)A_m^*(\omega)+N_m(k,\omega)\frac{A_m^*(\omega)}{A_1^*(\omega)}\right)$$
$$-\sum_{m\in\boldsymbol{G}_{n+1}}\left(\frac{S(k,\omega)}{A_1^*(\omega)}A_m(\omega)A_m^*(\omega)+N_m(k,\omega)\frac{A_m^*(\omega)}{A_1^*(\omega)}\right) \tag{7}$$

Since the matched filter array frequency response can approximate a scalar as explained in (4), the blocking matrix outputs with normalization ($\boldsymbol{Z_1}(k,\omega)\sim\boldsymbol{Z_{N-1}}(k,\omega)$ in Figure 1) can be approximated as follows:

$$\begin{bmatrix} Z_1(k,\omega) \\ Z_2(k,\omega) \\ \vdots \\ Z_{N-1}(k,\omega) \end{bmatrix}=\begin{bmatrix} \langle U_1(k,\omega)\rangle_{nor}-\langle U_2(k,\omega)\rangle_{nor} \\ \langle U_2(k,\omega)\rangle_{nor}-\langle U_3(k,\omega)\rangle_{nor} \\ \vdots \\ \langle U_N(k,\omega)\rangle_{nor}-\langle U_{N-1}(k,\omega)\rangle_{nor} \end{bmatrix}$$
$$\approx\begin{bmatrix} \frac{1}{U_1^{\max}(k)}\left(\sum_{m\in\boldsymbol{G_1}}N_m(k,\omega)\frac{A_m^*(\omega)}{A_1^*(\omega)}\right)-\frac{1}{U_2^{\max}(k)}\left(\sum_{m\in\boldsymbol{G}_2}N_m(k,\omega)\frac{A_m^*(\omega)}{A_1^*(\omega)}\right) \\ \frac{1}{U_2^{\max}(k)}\left(\sum_{m\in\boldsymbol{G}_2}N_m(k,\omega)\frac{A_m^*(\omega)}{A_1^*(\omega)}\right)-\frac{1}{U_3^{\max}(k)}\left(\sum_{m\in\boldsymbol{G}_3}N_m(k,\omega)\frac{A_m^*(\omega)}{A_1^*(\omega)}\right) \\ \vdots \\ \frac{1}{U_{N-1}^{\max}(k)}\left(\sum_{m\in\boldsymbol{G}_{N-1}}N_m(k,\omega)\frac{A_m^*(\omega)}{A_1^*(\omega)}\right)-\frac{1}{U_N^{\max}(k)}\left(\sum_{m\in\boldsymbol{G}_N}N_m(k,\omega)\frac{A_m^*(\omega)}{A_1^*(\omega)}\right) \end{bmatrix} \tag{8}$$

where $\langle\cdot\rangle_{nor}$ is the normalization operation which is defined as:

$$\langle U_n(k,\omega)\rangle_{nor}\equiv\frac{U_n(k,\omega)}{U_n^{\max}(k)}=\frac{U_n(k,\omega)}{\max([U_n(k,0),U_n(k,1),\cdots,U_n(k,NFFT-1)])} \tag{9}$$

The approximated blocking matrix outputs contain only interference signal which is suitable for adaptation algorithm.

2.2. Signal processing in the proposed beamformer

The fixed beamformer in Figure 1 is realized using delay and sum beamformer, and the fixed beamformer output can be written as:

$$Y_{FB}(k,\omega)=\frac{\boldsymbol{W}^{\mathrm{H}}(\omega)\boldsymbol{X}(k,\omega)}{M} \tag{10}$$

where H denotes conjugation transpose and

$$\boldsymbol{W}(\omega)=\left[e^{j2\pi\tau_1(\omega/NFFT)}\quad e^{j2\pi\tau_2(\omega/NFFT)}\quad\cdots\quad e^{j2\pi\tau_M(\omega/NFFT)}\right]^{\mathrm{T}}$$
$$\boldsymbol{X}(k,\omega)=\left[X_1(k,\omega)\quad X_2(k,\omega)\quad\cdots\quad X_M(k,\omega)\right]^{\mathrm{T}}$$

$\tau_1,\tau_2,\cdots,\tau_M$ are time delay steering elements which are used to point the array to the direction of sound source. In multi-channel adaptive filter, the overall output can be obtained by subtracting the components correlated to $Z_n(k,\omega)$ from $Y_{FB}(k,\omega)$. Let $\boldsymbol{Q}(k,\omega)$ and $\boldsymbol{Z}(k,\omega)$ be the filter coefficients vector and signal vector of the subarray output. Then the output of the proposed speech enhancement algorithm can be written as:

$$Y(k,\omega)=Y_{FB}(k,\omega)-\boldsymbol{Q}^{\mathrm{H}}(k,\omega)\boldsymbol{Z}(k,\omega) \tag{11}$$

where

$$\boldsymbol{Q}(k,\omega)=\left[Q_1(k,\omega)\quad Q_2(k,\omega)\quad\cdots\quad Q_{N-1}(k,\omega)\right]^{\mathrm{T}}$$
$$\boldsymbol{Z}(k,\omega)=\left[Z_1(k,\omega)\quad Z_2(k,\omega)\quad\cdots\quad Z_{N-1}(k,\omega)\right]^{\mathrm{T}}$$

The adaptive filters are adjusted to minimize the output $Y(k,\omega)$. The well-known normalized least mean square is utilized to adapt the filter coefficients and the iterative equation is shown as:

$$\boldsymbol{Q}(k+1,\omega)=\boldsymbol{Q}(k,\omega)+\mu\frac{\beta\boldsymbol{Z}(k,\omega)Y_{FB}^*(k,\omega)}{\gamma+\boldsymbol{Z}^{\mathrm{H}}(k,\omega)\boldsymbol{Z}(k,\omega)} \tag{12}$$

where μ is the step size; β is a scalar for adjusting the normalized values and γ is a small constant included to ensure that the update term does not become excessively large when $\boldsymbol{Z}^{\mathrm{H}}(k,\omega)\boldsymbol{Z}(k,\omega)$ temporarily become small.

3. SIMULATION RESULTS

This section provides simulation results to evaluate the capability of the proposed speech enhancement algorithm. A uniform linear array with 10 microphones and 7 cm spaces is constructed for this simulation. The arrangement of microphone array and sound source is illustrated in Figure 2. The speech source is male speech in English which is also used to obtain the matched filter ratios and the noise is white noise with various gain levels. To model the room acoustics, the image method model [7] is adopted to simulate the room impulse response. In this simulation, the enclosure room size is $10\mathrm{m}\times10\mathrm{m}\times3\mathrm{m}$ with reflection coefficient $\lambda=0.9$. The parameters N and P in Figure 1 are set to 9 and 4 in this simulation. The performance index, signal to noise ratio (SNR), is defined in equation (13) to evaluate the performance of the proposed algorithm.

$$SNR=\frac{\sum_{t\in T_S}x^2(t)-\sum_{t\in T_N}x^2(t)}{\sum_{t\in T_N}x^2(t)} \tag{13}$$

where T_S and T_N denote the time that the speech signal is active and inactive respectively. The proposed approach is demonstrated by six conditions which are shown in Table I and the simulation results are compared with those of

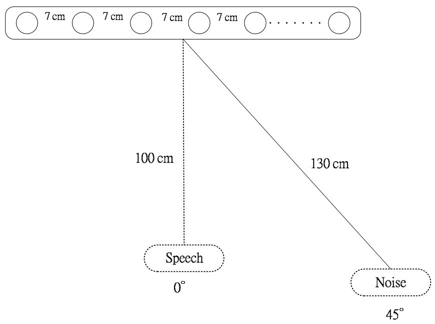


Figure 2. Arrangements of microphone array, noise and speech source

Table I
The experimental conditions and average SNRs

Condition	Input SNR(dB)	Condition	Input SNR(dB)
C1	2.98	C4	8.55
C2	4.9	C5	9.85
C3	6.96	C6	11.19

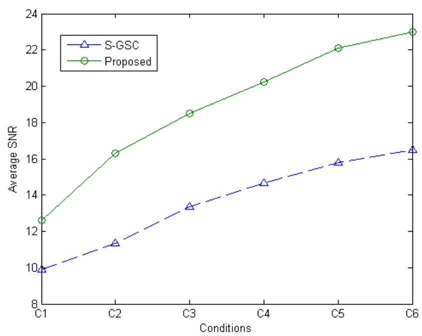
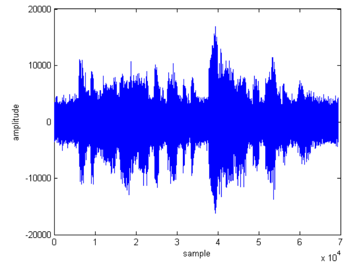
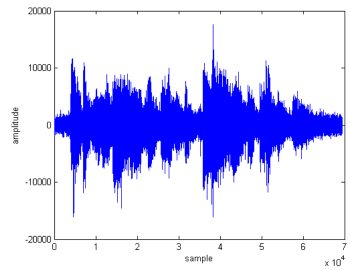


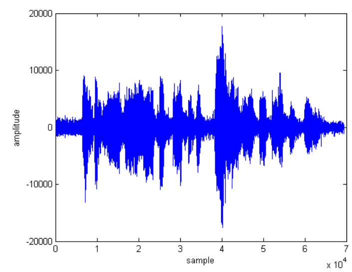
Figure 3. SNR of conditions C1 to C6



(a) Contaminated speech of Mic.#1



(b) Enhanced speech (S-GSC)



(c) Enhanced speech (proposed)

Figure 4. Speech waveforms

standard GSC (S-GSC) algorithm [1]. As can be seen from Figure 3, the average SNRs of the proposed approach are higher than those of the standard GSC algorithm. This is because the standard GSC with only time delay assumption about the propagation behavior degrades severely in the highly reflective environment ($\lambda = 0.9$). However, the blocking matrix using matched filter processing with subarrays can suppress the effect of the reverberation and produce purer interference signals than the standard GSC. Figure 4 shows one of the simulation results at different stage including contaminated signal of the first microphone and enhanced signal of the S-GSC and the proposed algorithm.

4. CONCLUSION

It was shown that the summation of matched filters approaches a delta impulse response. When applied to an enclosure, this implies that the reverberation effect can be suppressed. The property is utilized in this paper to design a novel blocking matrix which is robust in a reverberant environment. The use of matched filter ratios rather than matched filter makes the proposed algorithm practical and efficient. Simulation results show that the proposed adaptive beamformer can achieve the better SNR values than conventional GJBF using delay elements.

REFERENCES

[1] L. J. Griffiths and C. W. Jim, “An Alternative Approach to Linearly Constrained Adaptive Beamforming,” IEEE Trans. Antennas Propagat., vol. AP-30, pp. 27-34, Jan. 1982.

[2] E. E. Jan and J. Flanagan, “Sound capture from spatial volumes: Matched-filter processing of microphone arrays having randomly-distributed sensors,” in Proc. Int. Conf. Acoust., Speech, Signal Process., Atlanta, GA, vol. 2, pp. 917–920, May 1996.

[3] D. V. Rabinkin, “Optimum Sensor Placement for Microphone Arrays,” Ph.D Dissertation, Rutgers University, May 1998.

[4] S. Affes and Y. Grenier, "A signal subspace tracking algorithm for microphone array processing of speech," IEEE Trans. Speech Audio Process., vol.5, no.5, pp.425–437, Sept. 1997.

[5] F. J. Macwilliams, N. J. A. Sloane, Pseudo-random sequences and arrays, Proc. IEEE, vol. 68, pp.593-619, May 1980.

[6] O. Shalvi and E. Weinstein, “System identification using non-stationary signals,” IEEE Trans. Signal Processing, vol. 44, no. 8, pp. 2055-2063, Aug. 1996.

[7] J.B. Allen and D.A. Berkley, “Image method for efficiently simulating small-room acoustics.” J . Acoust. Soc. Am., vol. 65, pp. 943-950, Apr. 1978.