A SOFT-DECISION ADAPTATION MODE CONTROLLER FOR AN EFFICIENT FREQUENCY-DOMAIN GENERALIZED SIDELOBE CANCELLER

Min-Seok Choi[1], *Chang-Hyun Baik*[1], *Young-Cheol Park*[2] *and Hong-Goo Kang*[1]

Yonsei Univ. Dept. of Electrical & Electronic Eng., Seoul, 120-749, Republic of Korea[1]
Yonsei Univ. Computer and Telecommunications Eng. Division, Wonju, 220-710, Republic of Korea[2]

ABSTRACT

In this paper, we propose a new soft-decision adaptation mode controller (SD-AMC) for frequency domain generalized sidelobe canceller (GSC) as a speech enhancement system. Contrarily to conventional systems that update filter coefficients in a hard-decision manner using voice activity detection (VAD), the proposed method flexibly controls the step-sizes of adaptive filters depending on the probability of speech presence in each frequency bin. Therefore, it further improves the system performance for various environments without much consideration on noise type and signal to noise ratio (SNR) of input signal. It also improves the robustness of GSC system by avoiding the miss-classification error by the hard-decision logic. Experimental results with speech recognition systems verify that the SD-AMC shows higher performance than ideally designed hard-decision approaches.

Index Terms— Speech enhancement, Microphone array, Generalized sidelobe canceller, Adaptation mode controller, Adaptive filter.

1. INTRODUCTION

The GSC is one of the most commonly used structures for microphone array systems. It contains three functional blocks as depicted in Fig. 1. The adaptive blocking matrix (ABM) derives a noise reference of the multiple input canceller (MIC) by eliminating the desired speech component from the noisy input signal. The MIC removes the noise signal from the fixed beamformer (FBF) output by using the ABM output signal. Both the ABM and MIC block adopt an adaptive filtering (ADF) structure that takes a criterion of minimizing output power [1][2][3].

In speech absence region, the target signal of the ABM filter does not exist, thus it is recommended not to update filter coefficients during speech pause to avoid wrong adaptation. On the contrary, filter coefficients of the MIC block should be updated during speech absence region only to avoid speech removal caused by correlated noise or leaked speech in the ABM output [1][3][4]. To control the adaptation interval of ABM and MIC filter, GSC generally has a voice activity decision module called adaptation mode controller (AMC). Overall performance of the GSC system is highly related to the accuracy of AMC module because a miss-classification caused by adopting a hard-decision approach results in wrong adaptation of filter coefficients which leads to severe quality degradation [1][4].

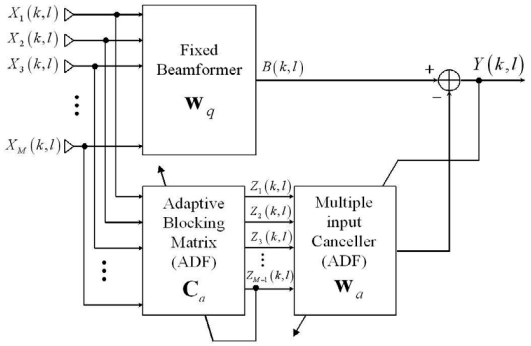

Fig. 1. Block diagram of GSC

The adaptation speed and steady state error of the ADF are highly related to the step-size constant, but it is very hard to find optimal step-size which guarantees the good performance in a general environment. Moreover, since the two performance metrics have trade-off relationship, fixing the step-size constant can not achieve two goals simultaneously [5].

The objective of this paper is to minimize the problems mentioned above in a frequency-domain GSC as a preprocessor of speech recognition system. In SD-AMC, the step-sizes of ABM and MIC filters are softly and automatically decided by utilizing the speech presence probability in each frequency band thus it minimizes miss-detection problem in a hard-decision method. The SD-AMC also helps achieving optimal performance without any load on step-size decision. According to the performance evaluation with speech recognition tests in various environments, GSC system with the SD-AMC shows higher recognition rates than with hard-decision AMC using perfect VAD for most cases.

2. CONVENTIONAL ADAPTATION MODE CONTROLLER (AMC) ALGORITHM

The VAD-based AMC uses the power ratio of FBF and ABM output [1]. Let $P\{\bullet\}$ define the power of the signal, the AMC

with the power ratio method is represented as

$$VAD\,(l) = \begin{cases} H_0\,, & P_{ratio}\,(l) < \eta \\ H_1\,, & P_{ratio}\,(l) \geq \eta \end{cases} \\ where,\ \ P_{ratio} = P\{B(k,l)\}/P\{Z_i(k,l)\}. \tag{1}$$

$B(k,l)$ and $Z_i(k,l)$ is the frequency-domain representation of FBF output and the ABM output, respectively. η in Eq. (1) denotes the threshold to make a binary decision. H_0 denotes the state of speech absence and H_1 the state of speech presence. In general, the $P\{B(k,l)\}$ and $P\{Z_i(k,l)\}$ are estimated using a first-order recursive averaging for real-time implementation. After it decides the voice activity, the filter of ABM is updated in the speech presence region and viceversa for the filter of MIC.

The AMC using the power ratio is simple to implement and shows reasonable performance if the FBF and ABM work well. However, the performance of the VAD significantly degrades if there exists of target leakage at ABM output or speech attenuation at FBF output. The other problem is that it is hard to set the threshold, η,suitable for various environments such as SNR or noise type variation [1][4].

In the next section, we propose a new SD-AMC that controls the step-size of ABM and MIC filters depending on the speech presence probability. It not only reduces artifacts caused by detection errors in the hard-decision logic, but also improves accuracy of the ADFs by adaptively controlling step-sizes to various type of input signals.

3. PROPOSED SOFT-DECISION AMC (SD-AMC) ALGORITHM

3.1. Step-size decision by utilizing the speech presence probability

Let the ADFs of ABM and MIC block in frequency-domain GSC are defined as $C_{a,i}(k,l)$ and $W_{a,i}(k,l)$, respectively, the weight vectors can be updated with a normalized least mean square (NLMS) ADF [5][6],

$$\begin{aligned} Z_i(k,l) &= X_i(k,l) - C_{a,i}(k,l)X_1(k,l), \\ C_{a,i}(k,l+1) &= C_{a,i}(k,l) + \frac{\mu_z}{\lambda_{x_1}(k,l)} Z_i(k,l)X_1(k,l), \\ Y(k,l) &= B(k,l) - \sum_{i=1}^{M} W_{a,i}(k,l)Z_i(k,l), \\ W_{a,i}(k,l+1) &= W_{a,i}(k,l) + \frac{\mu_y}{\lambda_z(k,l)} Y(k,l)Z_i(k,l), \end{aligned} \tag{2}$$

where, $\lambda_{\{\bullet\}}(k,l)$ denotes estimated signal power at each frequency bin of current frame. In general, it can be calculated by recursively averaging the signal. $\mu_{\{\bullet\}}$ represents step-sizes of each ADF which defines adaptation speed. In conventional algorithm, the filter coefficients of ADFs in GSC are updated alternately with a fixed value of step-size according to a voice activity.

The proposed method in this paper variably and automatically controls the step-size, $\mu_z(k,l)$ and $\mu_y(k,l)$, rather than uses fixed values, μ_z and μ_y in Eq. (2). Specifically, we use the speech presence probability of the FBF output, $P(H_1|B(k,l))$, to determine the step-size. Based on numerous simulation results, we concluded that the linear relation between the optimal step-size and the speech presence probability was a reasonable choice. Therefore, a decision rule of the step-sizes is obtained by linearly mapping it to the speech presence probability. If we define the maximum bounds of the step-size of the ABM and the MIC as $\mu_{z,M}$ and $\mu_{y,M}$, the function to calculate step-sizes of the adaptive filters is represented follows.

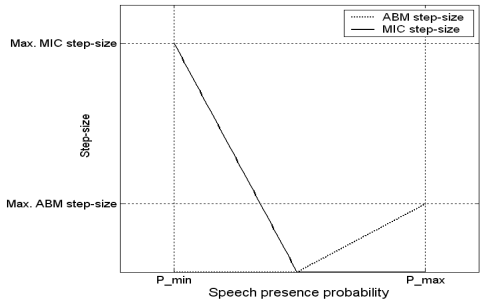

Fig. 2. Decision rule of step-size of ABM (solid line) and MIC (dotted line)

$$\begin{aligned} \mu_z(k,l) &= \begin{cases} \tilde{\mu}_z(k,l), & \tilde{\mu}_z(k,l) \geq 0 \\ 0, & \tilde{\mu}_z(k,l) < 0, \end{cases} \\ \tilde{\mu}_z(k,l) &= \left[P\left(H_1|B\left(k,l\right)\right) - \frac{P_M + P_m}{2} \right] \frac{2\,\mu_{z,M}}{P_M - P_m}, \\ \mu_y(k,l) &= \begin{cases} \tilde{\mu}_y(k,l), & \tilde{\mu}_y(k,l) \geq 0 \\ 0, & \tilde{\mu}_y(k,l) < 0, \end{cases} \\ \tilde{\mu}_y(k,l) &= \left[\frac{P_M + P_m}{2} - P\left(H_1|B\left(k,l\right)\right) \right] \frac{2\,\mu_{y,M}}{P_M - P_m}, \end{aligned} \tag{3}$$

where P_m and P_M represent minimum and maximum bounds of the speech presence probability. Fig. 2 shows an example of step-size decision curves for the ABM and the MIC. In this paper, we set the maximum bound of step-size for the ABM and the MIC 0.2 and 0.7, respectively. The last step of the proposed SD-AMC is defining the speech presence probability.

3.2. Speech presence probability of the FBF output

The speech presence probability of the FBF output can be represented as follows [7].

$$\begin{aligned} P(H_1|B(k,l)) &= \frac{\Lambda(k,l)}{1+\Lambda(k,l)} \\ \Lambda(k,l) &= \frac{(1-q(k,l))}{q(k,l)} \frac{\exp(n(k,l))}{1+\xi(k,l)}, \end{aligned} \tag{4}$$

where $n(k,l)$ is the function of *a priori* SNR, $\xi(k,l)$, and *a posteriori* SNR, $\gamma_B(k,l)$.

$$\begin{aligned} n(k,l) &= \frac{\xi(k,l)}{1+\xi(k,l)}\,\gamma(k,l) \\ \xi(k,l) &= \frac{\lambda_s(k,l)}{\lambda_{v_B}(k,l)}\,,\quad \gamma_B(k,l) = \frac{|B(k,l)|^2}{\lambda_{v_B}(k,l)}. \end{aligned} \tag{5}$$

$q(k,l)$ in Eq. (4) is defined as *a priori* speech absence probability. $\lambda_s(k,l)$ and $\lambda_{v_B}(k,l)$ denote the variance of desired speech and residual noise in FBF output, respectively. To obtain the speech presence probability in Eq. (4), we need an SNR of FBF output and the *a priori* speech absence probability (SAP) [7][8].

3.2.1. SNR of FBF output estimation using the noise reduction ratio (NRR)

To estimate speech presence probability from noisy input signal, we should first estimate the SNR of FBF output. The first step to estimate the SNR is to obtain the noise power of the input. In general, the noise power spectral density (PSD) can be computed by averaging the signal power during the period of speech pause. However, they cannot track noise variation in speech presence region.

The main idea of the proposed noise PSD estimator is to treat the output of ABM as a biased value of the noise component in FBF output. In general, the noise signals at each sensors are originated from the same sources. Therefore, if we assume the lossless propagation environment, the input noises to multiple sensors have same magnitude.

$$\begin{aligned} |V_i(k,l)| &= |V_j(k,l)| \\ V_i(k,l) &= V_j(k,l)e^{j\tau_{v_{i,j}}(k,l)}, \end{aligned} \tag{6}$$

where V_i represents the noise signal of i-th sensor and $\tau_{v_{i,j}}(k,l)$ represents the phase difference of noise between the i-th sensor and the j-th sensor. Therefore, we can rewrite the residual noise of the FBF output and the output of the ABM as Eq. (7) and Eq. (8), respectively. We assume that the filter of the ABM is optimally adjusted to eliminate the speech components [2].

$$V_B(k,l) = \frac{1}{M}\left[\sum_{i=1}^{M} e^{j\tau_{v_{i,1}}(k,l)}e^{j\omega_k\tau(i-1)}\right]V_1(k,l), \tag{7}$$

$$Z_i(k,l) = \left[e^{j\tau_{v_{i,1}}(k,l)} - C_{a,i}(k)\right]V_1(k,l). \tag{8}$$

As a result, the noise component of the FBF output and the ABM output is represented as a multiplication of certain filter and $V_1(k,l)$. Therefore, for simplification, the power of $V_B(k,l)$ and $Z_i(k,l)$ can be defined as the multiplication of compensation factor and the power of $V_1(k,l)$.

$$\begin{aligned} \lambda_{v_B}(k,l) &= R_b(k,l)\lambda_{v_1}(k,l) \\ \lambda_z(k,l) &= R_z(k,l)\lambda_{v_1}(k,l), \end{aligned} \tag{9}$$

where λ_{v_B}, λ_{v_1} and $\lambda_z(k,l)$ represents the power of noise of the FBF output, noise of the first sensor input and the ABM output, respectively. In summary, we can obtain the noise PSD of the FBF output by compensating the bias of ABM output. Lets define the noise reduction ratio (NRR), $R(k,l)$, as the ratio between $R_b(k,l)$ and $R_z(k,l)$, the noise power of FBF output can be represented as Eq. (10) by Eq. (9).

$$\begin{aligned} \lambda_{v_B}(k,l) &= R(k,l)\,\lambda_z(k,l), \\ R(k,l) &= R_b(k,l)/R_z(k,l). \end{aligned} \tag{10}$$

The power of ABM output, $\lambda_z(k,l)$, can be obtained by using the recursive averaging of the ABM output. The NRR is a function which depends on the location of signal sources because the filter of FBF and ABM in Eq. (7) and Eq. (8) is a function of propagation delays. Since the proposed algorithm uses the power of the ABM output, the noise PSD estimator operates always. After the noise PSD estimation, we can calculate the SNR in Eq. (5)

3.2.2. Estimation of noise reduction ratio (NRR)

The proposed algorithm estimates the noise PSD of FBF output by compensating the ABM output by using the NRR. However, NRR is an unknown value in real environment because we do not know the location of signal sources. Therefore, we estimate NRR by dividing the noise power of the FBF output with the power of the ABM output.

$$R(k,l) = \frac{R_b(k,l)}{R_z(k,l)} = \frac{R_b(k,l)\lambda_{v_1}(k,l)}{R_z(k,l)\lambda_{v_1}(k,l)} = \frac{\lambda_{v_B}(k,l)}{\lambda_z(k,l)}. \tag{11}$$

In general, the NRR is more stationary than the noise itself because it relates to the source location which changes very slowly. Therefore, the NRR can be updated only in speech absence region as Eq. (12) and it shows an efficient performance to calculate the SNR.

$$\begin{aligned} &R(k,l+1) \\ &\quad = \alpha_R R(k,l) + (1-\alpha_R)\frac{\lambda_{v_B}(k,l)}{\lambda_z(k,l)}\,,\ \ if\ H_0\ state. \end{aligned} \tag{12}$$

α_R is set to 0.85 for the simulation.

To estimate the variance of the noise PSD of FBF output and the variance of ABM output in Eq. (10), (11), (12), we introduce the minima controlled recursive averaging (MCRA) method which considers the speech absence probability to calculate the PSD of the noise [9]. Though it is possible to calculate the power of ABM output using conventional recursive averaging, the MCRA method further improves the performance and robustness of the SNR estimator if there exists speech leakage in ABM output. Since the MCRA reduces the effect of current frame in speech presence region, it is helpful to avoid the effect of speech leakage in the ABM output. Speech absence probability of the FBF output is used to implement the MCRA method and the minimum bound of recursive averaging factor in the MCRA method is set to 0.95 and 0.8 for the λ_{v_B} and λ_z, respectively [9].

Table 1. The correction rate of speech recognition test

	step size		AMC type	noise type					
				babble			pink		
	BM	MIC		5dB	10dB	20dB	5dB	10dB	20dB
enhanced data	0.1	0.1	hand marked	73.00	90.50	99.00	28.50	66.00	99.00
			hard-decision	71.50	89.50	99.00	29.50	65.00	98.50
	0.1	0.5	hand marked	72.00	88.50	99.00	37.50	71.00	99.00
			hard-decision	65.50	86.00	98.50	38.50	70.50	98.50
	0.5	0.1	hand marked	75.00	87.50	93.50	33.00	64.00	94.50
			hard-decision	76.00	88.50	96.00	32.00	65.00	93.50
	0.5	0.5	hand marked	69.00	84.00	90.50	42.50	67.00	92.00
			hard-decision	72.50	82.50	91.50	44.50	67.00	87.50
	SD-AMC			74.50	91.00	99.50	45.50	75.50	99.00
noisy data				55.00	84.00	98.50	15.50	32.00	88.00

4. SIMULATION RESULTS

In order to evaluate the performance of the proposed AMC under realistic environments, noisy speech for tests is recorded in a multi-media room. Speech and noise signals are recorded by using a uniformly spaced, with a spacing of 4.5 cm, line array with 6 microphones. Two different noise types, babble and pink noise, are located about 30 degrees from the center, each with the SNR of 5dB, 10dB, and 20dB. 20 speakers are uttered 40 kinds of words at a distance of 2.5 meters from the array.

We evaluate the recognition performance of enhanced outputs with three different AMC methods : proposed SD-AMC, hand-marked AMC, and conventional hard-decision AMC with the power ratio method. In the hand-marked AMC we manually discriminate speech region, thus the VAD is nearly perfect. Simulation results are summarized in table 1. It shows the correction rate of the recognition tests to the three different AMC methods by varying ABM and MIC step-size.

Comparing with the hand-marked and the hard-decision AMC, the proposed SD-AMC shows best performance for most cases. Other two methods show reasonable performance in a couple of special cases. However, to obtain the good performance using these methods, we must decide the adequate step-size of ABM and MIC mainly depending on noise types and SNRs. It indicate that the conventional methods are not suitable for various environments. On the contrary, the proposed SD-AMC method shows best or near the best performance independent of the noise characteristics and the SNR.

5. CONCLUSION

The SD-AMC method using speech presence probability was proposed. Since the proposed algorithm automatically controls the step-size of ADF rather than hardly decides the adaptation mode, it enhances the accuracy and robustness of the system, without regarding to the input noise types and SNRs.

In terms of recognition rates, the GSC system with SD-AMC technique shows higher performance than with the hand-marked and the conventional hard-decision method. The proposed method can be easily applied to many other speech signal processing systems utilizing ADFs.

6. REFERENCES

[1] M. Brandstein and D. Ward, *Microphone Array*, Springer, 2001.

[2] L. J. Griffiths and C. W. Jim, "An alternative approach to linear constrained adaptive beamforming," *IEEE Trans. Antennas Propagat.*, vol. AP-30, no. 1, pp. 27-34, Jan. 1982.

[3] O. Hoshuyama, A. Sugiyama and A. Hirano, "A robust adaptive beamformer for microphone array with blocking matrix using constrained adaptive filter," *IEEE Trans. Signal Processing*, vol. 47, no. 10, pp. 2677-2684, Oct. 1999.

[4] O. Hoshuyama, B. Begasse, A. Sugiyama and A. Hirano, "A real-time robust adaptive microphone array controlled by an SNR estimate," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 3605-3678, 1998.

[5] S. Haykin, *Adaptive filter theory*, Prentice Hall, 1991.

[6] Y. H. Chen and H. D. Fang, "Frequency-domain implementation of Griffiths-Jim adaptive beamformer," *Jounal of Acoustic Soc. of America*, vol. 91, no. 6, Jum. 1992.

[7] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal processing*, vol. ASSP-32, pp. 1109-1121, Dec. 1984.

[8] M. S. Choi and H. G. Kang, "An improved estimation of a priori speech absence probability for speech enhancement : in perspective of speech perception," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 1117-1120, 2005.

[9] I. Cohen and B. Berdugo "Speech enhancement for non-statinary noise environments," *Signal Process.*, vol. 81, pp. 2403-2418, Nov. 2001.