Performance Improvement of Adaptive Noise Cancellation Using a Speech Detector

Jang Sik Park*, Kyung Sik Son*

ABSTRACT

The performance of two-channel adaptive noise canceller is often degraded by the weights perturbation due to the speech signal. In this paper, an adaptive noise canceller employing a speech detector and two adaptation algorithms which are switched according to the speech detector is proposed. When highly correlated speech signal is detected, the tap weights of the adaptive filter are adapted by the sign algorithm. On the other hand, the weights are adapted by the NLMS algorithm when silence is detected or when the characteristics of the noise propagation channel is changed. The employed speech detector utilizes the power ratio of the input and the output of an adaptive linear prediction-error filter. According to the computer simulation, the proposed method yields better performance than conventional ones.

I. Introduction

Separating a desired signal from background noise is an important and common problem in signal processing [1]. In particular, noise reduction is required in many speech processing systems. The performance of background noise causes the quality or intelligibility of speech to degrade. Reducing background noise in aircraft cockpit has received much attention in the literature[1]. Recently, growth of mobile communication networks has led to noise reduction techniques being applied to hands-free telephone speech in automobiles[2, 3]. Front-end speech enhancement can also be useful for automatic speech recognition[4].

Two channel adaptive noise cancellation(ANC) is known to be an effective technique for estimating a signal corrupted by additive noise or interference, when at least two inputs are available. The conventional ANC scheme proposed by Widrow et al. consists of an adaptive filter that operates on the reference input noise to produce an estimate of the noise in the primary input, which is then substracted from the primary input noise at the output of the system[5, 6]. The tap weights of the filter is adapted by the well-known Widrow-Hoff LMS algorithm[7]. The LMS algorithm is used for many applications because of its simplicity and good performance. The ANC method using the LMS algorithm was shown to reduce ambient noise power by at least 20 dB with minimal speech distortion[8].

Unfortunately, the performance of the ANC using LMS algorithms is degraded by the processed speech signal. The adaptive filter of the ANC acts as a system identifier which is modeling the noise propagation channel from the noise source to the primary microphone. It produces significant weight-misalignment due to the speech signal which acts as the measurement noise at the adaptive system identification. When strong speech signal is present at the primary microphone, the SNR of the ANC is severely decreased.

In this paper, to alleviate the above-mentioned drawback of the conventional ANC scheme and to achieve better SNR ratio at the system output, an ANC which employs speech detector and two adaptation algorithms is proposed. When the speech detector indicates the absense of speech or low energy speech present, the proposed ANC acts as the conventional ANC using the LMS (more precisely, normalized LMS (NLMS)) algorithm. But, on the other hand, when highly correlated speech signal is detected at primary input, the adaptive filter weights of the proposed ANC are adapted by another adaptation algorithm which yields less perturbation in the tap weights than the conventional one. As a results, the weights are not perturbated by input speech signal. The speech detector used in the proposed ANC utilizes the power ratio of input and output of an adaptive linear prediction-error filter followed by the conventional ANC. When the strong speech signal which is known to be highly correlated presents at the input of the adaptive linear prediction-error filter, the output power is much lower than in-

*Pusan National University
Manuscript Received April 23, 1996.

put power. But, if uncorrelated signal is applied into the adaptive linear prediction-error filter, the output power is not decreased at all. So, we can detect the strong speech signal at the primary input. The computer simulation shows that the proposed ANC has better SNR at the system output than the conventional one.

Ⅱ. Two channel ANC

The ANC method employs two sensors, commonly called the reference and the primary input. The primary input collects the noisy speech signal, whereas the reference input gains additional information about the background noise. The conventional two-channel ANC scheme is illustrated in Fig. 1.

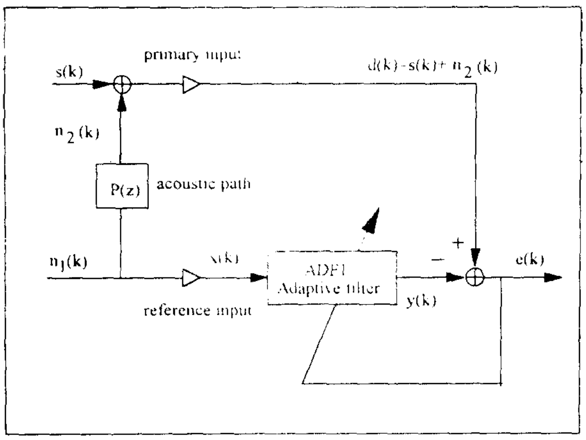

Fig 1. The structure of adaptive noise cancellation with two input sensor

In this figure, $x(k)$, $d(k)$, and $y(k)$ are the reference input, primary input, and output signal of the adaptive filter, respectively. Then reference input, $x(k)$ is placed near the noise source, and the primary input of the ANC collects input speech signal, $s(k)$, and the additive noise, $n_2(k)$, through the acoustic path, $P(z)$, from noise source to the primary input. The output of the ANC system, $e(k)$, is the difference between primary input and adaptive filter output. The LMS algorithm updates the tap weight vector of the adaptive filter, $W(k)$, or size L as

$$W(k+1)=W(k)+\alpha e(k)X(k) \tag{1}$$

where

$$e(k)=d(k)-W^T(k)X(k) \tag{2}$$

$$d(k)=s(k)+n_2(k) \tag{3}$$

$$n_2(k)=W_o^T X(k) \tag{4}$$

W_o denotes optimal weight vector which implys the tap weights of the filter modeling the acoustic path from the noise source to the primary input. $X(k)$ denotes the reference input vector to the adaptive filter defined as

$$X(k)=[x(k), x(k-1), \cdots, x(k-L+1)]^T \tag{5}$$

$[*]^T$ denotes the matrix transpose of $*$, and α denotes the size of the adaptive filter. To analyse the effect of speech signal on the perturbation of the tap weight vector, misalignment vector $V(k)$ is introduced as

$$V(k)=W(k)-W_o \tag{6}$$

Substituting (3), (4) and (6) in (2), the estimation error is given by

$$e(k)=s(k)-W^T(k)X(k)+W_o^T X(k)=s(k)-V^T(k)X(k) \tag{7}$$

It can be seen from (7) that $e(k)$ approximates $s(k)$ when the adaptive filter is well adjusted. Substituting (7) in (1), and substracting both sides with W_o, updating equation for the misalignment vector is taken by

$$V(k+1)=(1-\alpha X(k)X^T(k))V(k)+\alpha s(k)X(k) \tag{8}$$

Squaring and taking expectations of both sides of (8) gives

$$\begin{aligned} Q(k+1) &= E[(1-\alpha X(k)X^T(k))^2 V(k)V^T(k)+\alpha^2 s^2(k)X(k)X^T(k) \\ &\quad +2\alpha(1-\alpha X(k)X^T(k))V(k)s(k)X(k)] \\ &= (1-\alpha R)^2 Q(k)+\alpha^2 E[s^2(k)]R \end{aligned} \tag{9}$$

where

$$Q(k)=E[V(k)V^T(k)],$$

and

$$R=E[X(k)X^T(k)]. \tag{10}$$

Here we assumed that

$$E[s(k)X(k)]=0. \tag{11}$$

From (9), one can see that the second moment behavior of the misalignment vector is subject to the second moment behavior of speech signal. The stronger the energy of speech signal is, the larger the second moment of misalignment vector is. Squaring and taking expections of (2), it follows that

$$E[e^2(k)] = E[s^2(k)] + Tr[Q(k)R] \tag{12}$$

Also, we assumed that $s(k)$ is independent of $X(k)$ and $V(k)$.

From (12), it is clear that mean-square estimation error or the mean-square of the ANC output is affected by the second moment of misalignment vector. Since the SNR of the ANC is degraded due to the term $Tr[Q(k)R]$, a new ANC method is required to compensate this degradation. From the view of system identification, the speech signal acts as the measurement noise for the adaptive system identification. If there is no measurement noise, the excess minimum mean-squared estimation error is zero. But when the measurement noise is added, that can not be converged to zero[5]. Double-talk problem in acoustic echo cancellation(AEC) faces the same situation.

Ⅲ. The proposed ANC with speech detector

To reduce above problem, the proposed ANC method is illustrated in Fig. 2. The proposed ANC method consists of a speech detector and two adaptive algorithms which are switched by the speech detector. One of two algorithms is the sign algorithm(SA) which updates the adaptive filter weights when speech signal is included at the primary input. The other is the NLMS algorithm, which is updates the adaptive filter weights when the acoustic path is changed or the silence of the speech signal is present at the primary input.

Generally, the adaptive filter of the ANC can be converged in a short duration from the beginning of adaptation. After converged, the processed speech signal of the ANC is nearly clear, that is, $e(k) \doteqdot s(k)$, but the adaptive filter is perturbated by the processed speech signal because the processed speech acts as the estimation error of the adaptive filter in the conventional ANC. From (9), we knew that voiced speech severely perturbates the weights due to its large energy.

The adaptation should be halted to alleviate weights variation due to the processed speech when speech presents at the primary input. Speech segments can be detected by measurng the energy of the primary input[1]. However, in case that the primary input has low SNR, it was very difficult to detect the speech segments. As a second choice, the energy of the estimation error can be used to distinguish speech from silence bacause the estimation error is close to input speech after converged. But, if only the energy of the estimation error is used to detect speech signal, the change of the path will be considered that speech signal is present. In spite of fast adapting for tracking the change, the adaptation is stopped. This

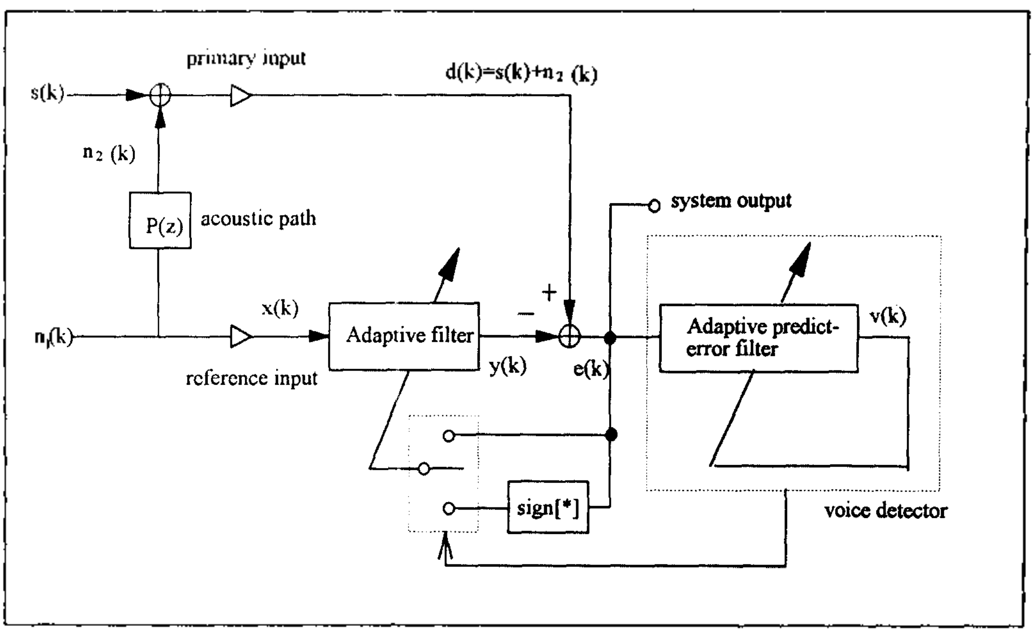

Fig 2. The proposed ANC structure with speech detector

causes the output of the ANC to be noisy.

To avoid this problem, we use adaptive prediction-error filter as the speech detector. A prediction-error filter can reduce the correlation component of the input, provided that the order of the filter is high enough[10]. It is well known that voice segments of the speech with large energy are highly correlated. Applying speech for the input of prediction-error filter, the power-ratio of input and output signal becomes lower than specific threshold. When uncorrelated signal is applied at the input, the output power of prediction-error filter is not decreased. As a result, speech signal can be distinguished from silence or acoustic path changing, provided that the noise signal is uncorrelated white signal.

In the proposed ANC method, the adaptive filter weights are updated by the sign algorithm(SA)[11] during intervals when there is speech present. The reason for updating with the SA is to adjust the weights which is misadjusted due to the misalignment of the speech detector. If the adaptation is stopped with misadjusted weights, the performance of the ANC will be degraded at the silence segments as well as at the speech. The SA updates the weights as

$$W(k+1)=W(k)+\beta\, sign[e(k)]\, X(k) \tag{13}$$

Table 1. The proposed ANC method

$$e(k)=d(k)-W^T(k)X(k)$$

$$v(k)=e(k)-C^T E(k-1)$$

$$P_{th}(k)=\frac{P_v(k)}{P_e(k)}$$

$$P_e(k)=\frac{1}{N}\sum_{i=0}^{N} e^2(k-i)$$

$$P_v(k)=\frac{1}{N}\sum_{i=0}^{N} v^2(k-i)$$

$$a(k)=\begin{cases}\gamma\, sign[e(k)], & P_{th}(k)\ge T\\ \dfrac{\alpha e(k)}{X^T(k)X(k)}, & P_{th}(k)<T\end{cases}$$

α; the NLMS step size of the adaptive filter
γ; the SA step size of the adaptive filter
T; threshold of the power-ratio

$$W(k+1)=W(k)+a(k)X(k)$$

$$C(k+1)=C(k)+\beta v(k)E(k-1)$$

β; step size of the APE

where

$$sign[x]=\begin{cases}1 & x\ge 0\\ -1 & x<0\end{cases}$$

The proposed ANC method is summarized in table 1. $E(k-1)$ denotes estimation error vector defined as

$$E(k-1)=[e(k-1),\, e(k-2),\, \cdots,\, e(k-M+1)]^T \tag{14}$$

$P_{th}(k)$, $P_e(k)$, and $P_v(k)$ are the power-ratio, input and output power of the adaptive linear prediction-error filter, respectively. The SA normalizes the magnitude of the estimation error while preserving its direction. By the SA, the weights are updated as

$$W(k+1)=W(k)+\gamma\, sign[s(k)-V^T(k)X(k)]\, X(k). \tag{15}$$

If tap weights of the adaptive filter are well adjusted during the silence or background noise period by the NLMS algorithm, $W(k)$ is close to W_o, and $V^T(k)X(k)$ is nearly zero. As results, (15) is approximated as

$$V(k+1)=\gamma\, sign[s(k)]\, X(k) \tag{16}$$

The second moment of misalignment of the SA is approximated as

$$Q(k+1)=\gamma^2 E[(sign[s(k)])^2]R. \tag{17}$$

At the same situation, the second moment of misalignment of the LMS algorithm is taken as

$$Q(k+1)=\alpha^2 E[s^2(k)]R \tag{18}$$

Comparing (17) and (18), we can see that they are significantly less perturbated than the LMS algorithm. The performance of the proposed ANC is not degraded by the processed speech and the adaptive filter showly adapts to compensate the misalignment due to the mal-detection of the speech segments.

Ⅳ. Simulation results

In order to evaluate the performance of the proposed ANC method, computer simulation was conducted on the task of cancelling noise in the speech signals. As the performance measure, the SNR and the segmental SNR with the segment size of 20 msec were used. It is known that the segmental SNR has significantly higher correlation with subjective speech quality than the SNR[6]. In this simulation, speech signal was sampled at 8 kHz, and zero-mean white Gaussian noise was used as the reference input noise. The acoustic transmission path from the reference input to the primary input was modelled as a 32 order FIR system which approximates the IIR system function of

$$P(z) = \frac{0.5}{1 - 1.272792z^{-1} + 0.81z^{-2}} \tag{19}$$

Step size of the adaptive filter is set to 0.1 and 0.7E-07 for the NLMS algorithm and the SA algorithm, respectively. Step size of the adaptive linear prediction-error filter is 0.01, and the length for estimating the power is 1 frame(20 msec), that is 160 samples. The threshold of power-ratio puts 0.35, it is power-ratio value, when the adaptive filter converged during the silence segment.

A speech signal and the noisy signal at the primary input used in this simulation are plotted in Fig. 3(a) and (b). In Fig. 3(c), the power-ratio of the adaptive linear prediction-error filter is shown, and the output of the speech detector used in this paper is plotted in Fig. 3(d). The detected speech segments are denoted '0' level. It shows that the speech detector is well operating.

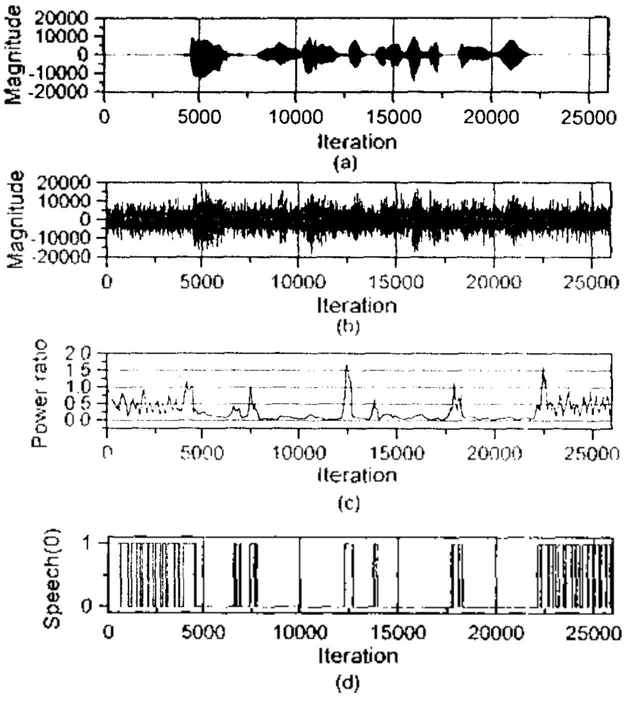

Fig 3. Operation of the speech detector used in the proposed ANC method
(a) clean speech
(b) noisy speech at primary input(− 3 dB SNR case)
(c) power ratio
(d) output of speech detector

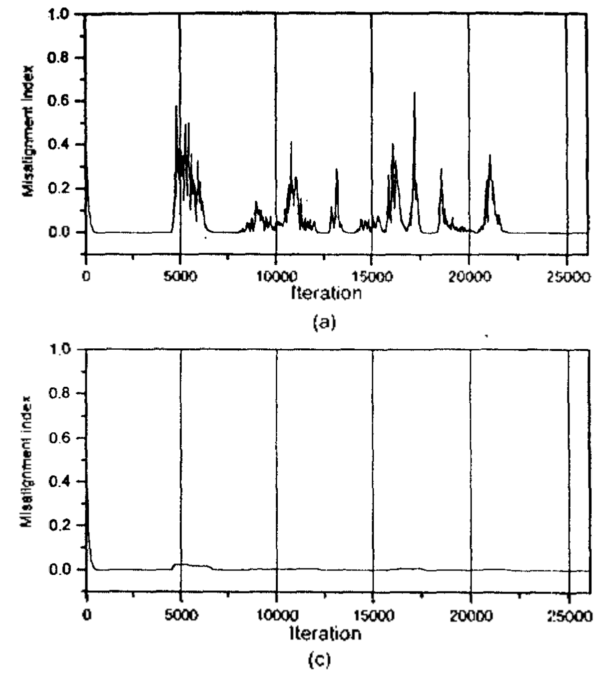

Fig 4. Weight-misalignment index of the ANCs
(a) Conventional ANC (b) Proposed ANC

As a measure to compare the performance of the ANC schemes, we used the misalignment index, which is defined as

$$W_{var}(k) = \frac{\sum_{i=0}^{L-1} (W_k(i) - W_o(i))^2}{\sum_{i=0}^{L-1} W_o^2(i)} \tag{20}$$

Fig. 4 shows that the simulation results on the performance comparison between the conventional scheme and the proposed one. In Fig. 4(a), the misalignment index of the conventional ANC is shown, and in Fig. 4(b), that of the proposed ANC is plotted. In this simulation,

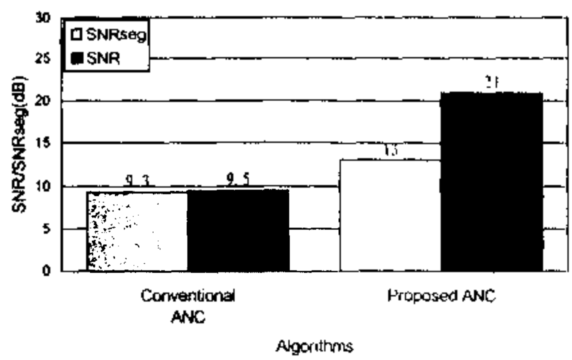

Fig 5. The performance comparison by the segmental SNR and the SNR

the relative gain of the speech signal over the noise in the primary input was set so that the SNR at the primary input becomes −3 dB. The misalignment index for the conventional ANC is significantly increased at the speech segments, while the misalignment index for the proposed method is smaller than the conventional for most of the range.

In Fig. 5, the SNR and the segmental SNR for the two ANC schemes are compared. The proposed ANC method improves about 3.7 dB of the segmental SNR, and about 11.7 dB of the SNR.

V. Conclusions

An adaptive noise cancellation employing speech detector and two adaptation algorithms which are switched by the speech detector have been presented. In the proposed method, the speech detector discriminates speech segment using the power-ratio of input and output of the adaptive linear prediction-error filter. The performance of the proposed scheme has been investigated by simulations. Although the reference noise is restricted to white, the speech detector well detects speech signal. The simulation results also shows that the proposed method yields better noise cancelling capability with compared to the conventional one.

References

1. W. A. Harrison, J. S. Lim, and E. Singer, "A New Application of Adaptive Noise Cancellation," IEEE Trans., ASSP., Vol. ASSP-34, No. 1, pp. 21-27, Feb. 1986.
2. I. Lecomt, M. Lever, J. Boudy and A. Tassy, "Car Noise Processing for Speech," in proc. IEEE ICASSP 89, Glasgow, Scotland, pp. 512-515, May 1989.
3. R. B. Wallace and R. A. Goubran, "Improved Tracking Adaptive Noise Canceler for Nonstationary Environments," IEEE Trans. SP, Vol. 40, No. 3, pp. 700-703, Mar. 1992.
4. L. Rabiner and B. H. Juang, *Fundametals of Speech Recognition*, Prentice-Hall, Englewood Cliffs, N.J., 1993.
5. B. Widrow and S. D. Stern, *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, N.J., 1985, p. 303.
6. J. R. Jr., J. G. Proakis, John H. L. Hansen, *Discrete-Time Processing of Speech Signals*, Macmillan, N.J., 1993, p. 528.
7. B. Widrow, et. al, "Adaptive Noise Cancelling: Principles and Applications," Proc. IEEE, Vol. 63, No. 12, pp. 1692-1716, Dec. 1975.
8. Steven F. Boll and Dennis C. Pulsipher, "Suppression of Acoustic Noise in Speech Using Two Microphone Adaptive Noise Cancelleation," IEEE Trans., ASSP, Vol. ASSP-28, No., 6, pp. 752-753, Dec. 1980.
9. L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, Englewood Cliffs, N.J., 1978, p. 122.
10. S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Englewood Cliffs, N.J., 1991, p. 216.
11. S. H. Cho and V. J. Mathew, "Tracking Analysis of the Sign Algorithm in Nonstationary Environments," IEEE Trans ASSP, Vol. 38, No. 12, pp. 2046-2057, Dec. 1990.

▲Kyung Sik Son

Kyung Sik Son received his BS and MS degrees in electronics engineering in 1973 and 1977, respectively, from Pusan National University, Korea, and the PhD in electronics engineering in 1991 from Kyung Puk National University in Korea. Since 1979, he has been a member of the faculty of the Department of Electronics Engineering at Pusan National University, where he is currently an associate professor. His research interests include digital signal processing, adaptive signal processing, and neural networks.

▲Jang Sik Park

Jang Sik Park received his BS and MS degrees in electronics engineering in 1992 and 1994, respectively, from Pusan National University, Korea. He is currently working towards PhD degree at Pusan National Unversity. His current research interests include adaptive signal processing and digital signal processing.