An Improved Affine Projection Algorithm Based Crosstalk Resistant Adaptive Noise Canceller

Guo Chen[1], Vijay Parsa[1,2]

[1]Department of Electrical & Computer Engineering and [2]National Centre for Audiology
University of Western Ontario, London, Ontario, Canada
E-mails: guo.chen@nca.uwo.ca, parsa@nca.uwo.ca

Abstract

This paper presents an improved affine projection algorithm based crosstalk resistant adaptive noise canceller (CRANC). The proposed CRANC consists of three adaptive filters, one of which is used for detecting speech activity and the others are employed to estimate the interference and crosstalk transfer functions, respectively. An efficient sequential speech detector is also proposed in this paper, which is completely built into the adaptive strategy and thus causes no delay in the input and output of the noise canceller. The proposed CRANC was tested and compared with the conventional CRANC in the white Gaussian noise environment. Experimental results show that the proposed CRANC achieved higher signal-to-noise ratio (SNR) improvements than the conventional CRANC.

Index Terms: adaptive noise cancellation, crosstalk resistance, affine projection algorithm

1. Introduction

Recovering a desired speech signal in noisy backgrounds is an important issue in speech communication applications, such as hand-free telephone in vehicle & mobile systems [1] and hearing aids [2]. In such applications, adaptive noise cancellation (ANC) is an effective technique for enhancing noisy speech [3]. The block diagram of the basic adaptive noise canceller is shown in Fig. 1. Normally the adaptive noise canceller requires two inputs from two microphones. One is called the primary input $y(k)$ consisting of the desired signal $s(k)$ corrupted by additive noise $\tilde{n}(k)$, where k is time index. The other is called the reference input $x(k)$ that will only include the noise signal $n(k)$ if there is no leakage of the signal, i.e. $G(\omega) = 0$. The reference input is processed by an adaptive filter whose weights are adjusted to minimize the power in the output signal. This minimization is achieved by filtering the reference input to approximate the noise signal in the primary input and then subtracting it from the primary input. The adaptive weights are typically adjusted using the least mean squares (LMS) algorithm [3] because of its simplicity, ease of implementation, and low computational complexity. Although the ANC in its basic form is an appealing technique, there are other factors which seriously influence the performance of the ANC and degrade the quality of the recovered speech signal. One such factor is the presence of a partial speech component in the reference input, i.e. the leakage of the signal called the crosstalk signal $\tilde{s}(k)$. The presence of the crosstalk will not only make the adaptive filter partially cancel the speech components which directly leads to a speech distortion and a lower output SNR, but also cause the adaptive algorithm biased away from the optimal weights. Another
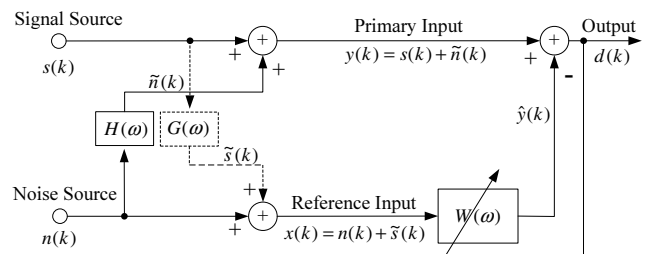

Figure 1: The block diagram of the basic adaptive noise canceller (ANC)

important factor to upset the noise cancelling performance is the excess mean-squared error (MSE) of adaptive algorithms due to the power fluctuations of the desired speech signal [4]. The excess MSE increases linearly with the desired signal power and greatly degrades the performance of the ANC when the power of the desired signal is strong and fluctuates frequently. This is a serious problem in many practical applications [4, 5] but hitherto ignored in the development of a crosstalk resistant adaptive noise canceller (CRANC) [6].

During the past years, there have been a number of reports published addressing the crosstalk problem [6, 7, 8, 9, 10]. Basically, there are two types of the CRANC. One is based on a cascade structure as reported in [8, 9, 10], while the other employs a dual joint process estimator as presented in [7, 6]. In the first type, one adaptive filter is utilized to estimate the interference transfer function while a second adaptive filter is used to alleviate the influence of the crosstalk transfer function. The main characteristic of this type CRANC is that the first adaptive filter is driven to the optimal weights of the interference transfer function during the crosstalk is absent. This CRANC structure has successfully been applied in processing Brainstem Auditory Evoked Potentials (BAEPs) [8] and Somatosensory Evoked Potential (SEP) [10]. The second type of CRANC is consisted of a dual adaptive filters connected in a feedback structure [7, 6]. The main characteristic of this type of CRANC is its joint process estimation which can estimate the interference and crosstalk transfer functions recursively. Due to this passive adaptive strategy, the noise cancelling ability is limited and the recovered signal is always accompanied with a large signal distortion and poor speech quality.

In this paper, we propose a new crosstalk resistant adaptive noise canceller (CRANC). The proposed CRANC consists of three
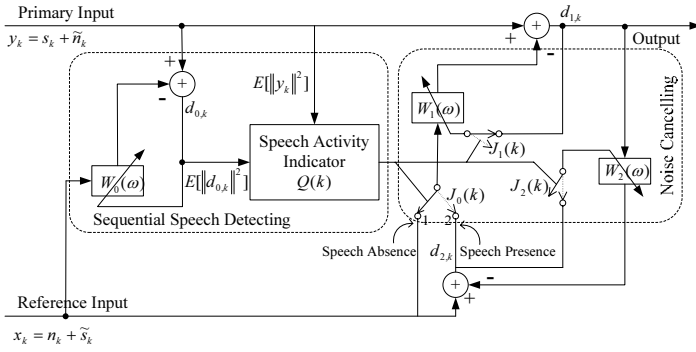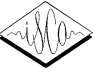
Figure 2: The block diagram of the proposed CRANC

adaptive filters, one of which is used to detect speech activity sequentially and the others are employed to estimate the interference and crosstalk transfer functions, respectively. The affine projection algorithm (APA), which provides less computational complexity than RLS but much faster convergence than NLMS, is used in the proposed CRANC to drive the adaptive filters. An efficient sequential speech detector is also proposed in this paper. The proposed detector is completely built into the adaptive strategy and thus causes no delay in the input and output of the proposed CRANC.

The remainder of the paper is organized as follows. In the next section, we outline the proposed CRANC. Section III presents the proposed sequential speech detector. In Section IV, we describe the filter adaptive algorithm and noise cancelling of the proposed CRANC. Section V presents the experimental results. Conclusions are drawn in Section VI.

2. The structure of the proposed CRANC

The block diagram of the proposed CRANC is shown in Fig. 2. It is assumed that the primary and reference inputs are short-time wide-sense stationary (WSS) processes and that the speech signal is statistically independent of the noise signal. Let the primary and reference input vectors be denoted by $\mathbf{y_k}$ and $\mathbf{x_k}$, respectively, i.e.

$$\mathbf{y_k} = \mathbf{s_k} + \tilde{\mathbf{n}}_\mathbf{k}, \quad \mathbf{x_k} = \mathbf{n_k} + \tilde{\mathbf{s}}_\mathbf{k}, \tag{1}$$

where $\mathbf{s_k} = [s(k), s(k-1), \cdots, s(k-L+1)]^T$, $\mathbf{n_k} = [n(k), n(k-1), \cdots, n(k-L+1)]^T$, $\tilde{\mathbf{s}}_\mathbf{k} = [\tilde{s}(k), \tilde{s}(k-1), \cdots, \tilde{s}(k-L+1)]^T$, $\tilde{\mathbf{n}}_\mathbf{k} = [\tilde{n}(k), \tilde{n}(k-1), \cdots, \tilde{n}(k-L+1)]^T$. The parameters L and k are the projection order of the APA and time index, respectively. Here, $\tilde{s}(k)$ stands for the crosstalk signal produced by $s(k)$ passing through the crosstalk transfer function $G(\omega)$, and $\tilde{n}(k)$ stands for the interference signal presenting in the primary input caused by $n(k)$ passing through interference transfer function $H(\omega)$.

As shown in Fig. 2, the proposed CRANC structure consists of two parts: sequential speech detection and noise cancellation. The speech detection part is composed of the adaptive filter $W_0(\omega)$ and speech activity indicator $Q(k)$. The noise cancellation part is made up of two joint adaptive filters $W_1(\omega)$ and $W_2(\omega)$, as well as three switches ($J_0(k)$, $J_1(k)$ and $J_2(k)$). Compared to the passive adaptive strategy used in [7, 6], the proposed CRANC is based on an active adaptive strategy with the following properties. First, the interference transfer function, $H(\omega)$, will be estimated using the adaptive filter $W_1(\omega)$ driven by the APA during the intervals

when speech is absent. This is motivated by the fact that, if we know that speech is absent at time k, it is unnecessary to subtract the output of the adaptive filter $W_2(\omega)$ from the reference input as does the conventional CRANC [7, 6]. When speech is detected active at time k, the switch $J_0(k)$ is shifted to 2 and the adaptive filter $W_1(\omega)$ is frozen (i.e., set $J_1(k) = 0$) at its previous weights. The frozen weights of the $W_1(\omega)$ are used to compute the output of the $W_1(\omega)$ throughout the detected period of active speech. However, the weights of the adaptive filter $W_2(\omega)$ are not updated immediately (i.e., $J_2(k)$ is still at 0) unless the power of the speech is below a certain threshold. This can mitigate the influence of a large excess MSE of the adaptive algorithm on the noise cancellation. As a result, the adaptive weights of the $W_2(\omega)$ converge during the intervals when the speech is active but at a relatively low power, and are not updated during the intervals when the speech is detected as having a strong power.

3. The sequential speech detector

The proposed speech detector is based on the Euclidean norms of the primary input and the output of the adaptive filter $W_0(\omega)$. In the detector, the output signal, $\mathbf{d_{0,k}}$, is calculated by

$$\mathbf{d_{0,k}} = \mathbf{y_k} - \mathbf{X_k} \cdot \mathbf{w_{0,k}}. \tag{2}$$

The APA computes the weights of the adaptive filter $W_0(\omega)$ by

$$\mathbf{w_{0,k+1}} = \mathbf{w_{0,k}} + \mu_0 \mathbf{X_k}^T \left(\mathbf{X_k} \mathbf{X_k}^T \right)^{-1} \mathbf{d_{0,k}} \tag{3}$$

where

$$\mathbf{w_{0,k}} = [w_0(k), w_0(k-1), \cdots, w_0(k-M+1)]^T,$$

$$\mathbf{d_{0,k}} = [d_0(k), d_0(k-1), \cdots, d_0(k-L+1)]^T,$$

$$\mathbf{X_k} = \begin{pmatrix} x(k) & x(k-1) & \ldots & x(k-M+1) \\ x(k-1) & x(k-2) & \ldots & x(k-M) \\ \vdots & \vdots & \ddots & \vdots \\ x(k-L+1) & x(k-L) & \ldots & x(k-M-L) \end{pmatrix}.$$
$$\tag{4}$$

The parameter μ_0 is the step size. A speech activity indicator with respect to time k is defined by

$$Q(k) = \frac{\beta \cdot U_k}{V_k + L\epsilon} \tag{5}$$

where $U_k = E[\|\mathbf{d_{0,k}}\|^2]$, $V_k = E[\|\mathbf{y_k}\|^2]$. The notation $\|\cdot\|$ is the Euclidean norm and E is the expectation operator. The parameters β and ϵ are the amplification and regularization factors, respectively. It can be observed that the indicator $Q(k)$ approaches zero when speech is absent, and that $Q(k)$ is kept as a positive value when speech is present. This can be clarified as follows. According to the definition of the Euclidean norm, we have

$$U_k = E\left(\sum_{i=0}^{L-1} [d_0(k-i)]^2 \right) = \sum_{i=0}^{L-1} E\left([d_0(k-i)]^2 \right) \tag{6}$$

Since $d_0(k)$ is a wide sense stationary (W.S.S.) stochastic process in the short-time period $[0, L-1]$, The Eq.(6) implies

$$U_k = LR_{\mathbf{d_{0,k}}}(0). \tag{7}$$
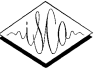
Meanwhile, in light of the Einstein-Wiener-Khinchin theorem [11], we can express the average short-time power of $\mathbf{d_{0,k}}$ in frequency domain as $U_k = \frac{L}{2\pi} \int_{-\pi}^{\pi} P_{\mathbf{d_{0,k}}}(\omega) d\omega$. Similarly, we also have $V_k = LR_{\mathbf{y_k}}(0) = \frac{L}{2\pi} \int_{-\pi}^{\pi} P_{\mathbf{y_k}}(\omega) d\omega$. Thus, the speech activity indicator defined by Eq.(5) can be expressed as

$$Q(k) = \frac{\beta \int_{-\pi}^{\pi} P_{\mathbf{d_{0,k}}}(\omega) d\omega}{\int_{-\pi}^{\pi} P_{\mathbf{y_k}}(\omega) d\omega + L\epsilon} \qquad (8)$$

On the other hand, from the basic ANC as shown in Fig. 1, we readily have

$$P_{\mathbf{y_k}}(\omega) = P_{\mathbf{s_k}}(\omega) + |H(\omega)|^2 P_{\mathbf{n_k}}(\omega)$$
$$P_{\mathbf{d_{0,k}}}(\omega) = P_{\mathbf{s_k}}(\omega) \left(1 - |W_0(\omega)G(\omega)|^2\right) +$$
$$P_{\mathbf{n_k}} \left(|H(\omega)|^2 - |W_0(\omega)|^2\right) \qquad (9)$$

We can therefore rewrite Eq.(8) as

$$Q(k) = \frac{\beta \int_{-\pi}^{\pi} P_{\mathbf{s_k}}(\omega) \left(1 - |W_0(\omega)G(\omega)|^2\right) d\omega}{\int_{-\pi}^{\pi} P_{\mathbf{s_k}}(\omega) d\omega + \int_{-\pi}^{\pi} P_{\mathbf{n_k}}(\omega) |H(\omega)|^2 d\omega + L\epsilon} +$$
$$\frac{\beta \int_{-\pi}^{\pi} P_{\mathbf{n_k}}(\omega) \left(|H(\omega)|^2 - |W_0(\omega)|^2\right) d\omega}{\int_{-\pi}^{\pi} P_{\mathbf{s_k}}(\omega) d\omega + \int_{-\pi}^{\pi} P_{\mathbf{n_k}}(\omega) |H(\omega)|^2 d\omega + L\epsilon}. \qquad (10)$$

Obviously, in the case of the speech absence, i.e. $P_{\mathbf{s_k}}(\omega) = 0$ for any ω, we can get

$$Q(k) = \frac{\beta \int_{-\pi}^{\pi} P_{\mathbf{n_k}}(\omega) \left(|H(\omega)|^2 - |W_0(\omega)|^2\right) d\omega}{\int_{-\pi}^{\pi} P_{\mathbf{n_k}}(\omega) |H(\omega)|^2 d\omega + L\epsilon}. \qquad (11)$$

During the intervals when speech is absent, the APA drives the weights of the adaptive filter $W_0(\omega)$ to converge to the transfer function $H(\omega)$, i.e., $|H(\omega)|^2$ is equal to $|W_0(\omega)|^2$. Correspondingly, $Q(k)$ will go to zero. On the other hand, in the case of the speech presence, because $Q(k)$ is a function not less than zero, we will hold $Q(k) > 0$ if the following equation is satisfied,

$$\int_{-\pi}^{\pi} P_{\mathbf{s_k}}(\omega) \left(1 - |W_0(\omega)G(\omega)|^2\right) d\omega \neq 0. \qquad (12)$$

Motivated by these facts, we propose to estimate U_k and V_k by time-averaging as follows:

$$\hat{U}_k = \alpha \hat{U}_{k-1} + (1-\alpha)\|\mathbf{d_{0,k}}\|^2,$$
$$\hat{V}_k = \alpha \hat{V}_{k-1} + (1-\alpha)\|\mathbf{y_k}\|^2 \qquad (13)$$

with a smoothing factor α $(0 \leq \alpha < 1)$.

Using $\hat{U}_k$ and $\hat{V}_k$ instead of U_k and V_k in Eq.(5), the proposed speech activity indicator becomes

$$Q(k) = \frac{\beta \cdot \hat{U}_k}{\hat{V}_k + L\epsilon} \qquad (14)$$

In actual calculations, we can use a threshold Q'_T of a small positive value to indicate the speech absence. At the same time, the value of $Q(k)$ can also be utilized to indicate the the power fluctuation of the input speech signal. Correspondingly, we set another threshold Q''_T to indicate whether or not the active speech has a strong power (i.e., $Q(k) \geq Q''_T$).

4. The Noise Cancellation

The noise cancellation of the proposed CRANC is performed by the two jointed adaptive filters ($W_1(\omega)$ and $W_2(\omega)$) and three switches ($J_0(k)$, $J_1(k)$ and $J_2(k)$). The status of the switches are dependent on the output values of the speech detector, $Q(k)$. The adaptive algorithm of the proposed CRANC is summarized as follows:

● Filtering:

$$\mathbf{d_{0,k}} = \mathbf{y_k} - \mathbf{X_k} \cdot \mathbf{w_{0,k}}$$
$$\mathbf{d_{1,k}} = \mathbf{y_k} - \mathbf{D_{2,k}} \cdot \mathbf{w_{1,k}}$$
$$\mathbf{d_{2,k}} = \mathbf{x_k} - \mathbf{D_{1,k}} \cdot \mathbf{w_{2,k}}$$

● Speech activity detecting:

$$Q(k) = \frac{\beta \cdot E[\|\mathbf{d_{0,k}}\|^2]}{E[\|\mathbf{y_k}\|^2] + L\epsilon}$$

● Tap-weight update:

$\mathbf{w_{0,k+1}} = \mathbf{w_{0,k}} + \mu_0 \mathbf{X_k}^T (\mathbf{X_k}\mathbf{X_k}^T)^{-1} \mathbf{d_{0,k}}$

if $Q(k) < Q'_T$ (i.e.,speech is absent)

    $\mathbf{D_{2,k}} = \mathbf{X_k}$ (i.e.,$J_0(k) = 1$)

    $\mathbf{w_{1,k+1}} = \mathbf{w_{1,k}} + \mu_1 \mathbf{D_{2,k}}^T (\mathbf{D_{2,k}}\mathbf{D_{2,k}}^T)^{-1} \mathbf{d_{1,k}}$

    (i.e.,$J_1(k)$=1)

    $\mathbf{w_{2,k+1}} = \mathbf{w_{2,k}}$ (i.e.,$J_2(k) = 0$)

else   (i.e.,speech is present)

    $\mathbf{w_{1,k+1}} = \mathbf{w_{1,k}}$ (i.e.,$J_1(k)$= 0 and $J_0(k)$=2)

    if $Q(k) > Q''_T$ (i.e.,speech with a strong power)

        $\mathbf{w_{2,k+1}} = \mathbf{w_{2,k}}$ (i.e.,$J_2(k) = 0$)

    else

        $\mathbf{w_{2,k+1}} = \mathbf{w_{2,k}} + \mu_2 \mathbf{D_{1,k}}^T (\mathbf{D_{1,k}}\mathbf{D_{1,k}}^T)^{-1} \mathbf{d_{2,k}}$

        (i.e.,$J_2(k)$=1)

    end

end

where,

$$\mathbf{w_{i,k}} = [w_i(k), w_i(k-1), \cdots, w_i(k-M+1)]^T,$$
$$\mathbf{d_{i,k}} = [d_i(k), d_i(k-1), \cdots, d_i(k-L+1)]^T,$$
$$\mathbf{D_{i,k}} = \begin{pmatrix} d_i(k) & d_i(k-1) & \ldots & d_i(k-M+1) \\ d_i(k-1) & d_i(k-2) & \ldots & d_i(k-M) \\ \vdots & \vdots & \ddots & \vdots \\ d_i(k-L+1) & d_i(k-L) & \ldots & d_i(k-M-L) \end{pmatrix},$$
$$i = 0, 1, 2.$$

5. Experimental Results

Performance of the proposed CRANC was evaluated by computer simulations in comparison with that of the conventional CRANC [6]. White Gaussian noise with different input levels was added to the original clean speech signal to construct the primary and reference signals with different SNR levels. The SNR of primary signal varies from 0 dB to 6 dB with a step size of 3 dB while the crosstalk SNR, defined by $10log\left(\|\tilde{\mathbf{s}}\|^2/\|\mathbf{s}\|^2\right)$, varies from -9

Table 1: The parameters used in the proposed CRANC

Parameter	Value	Description
M	16	Filter length
L	10	Projection order
α	0.98	Smoothing factor
μ_0	0.1	Step size for the filter $W_0(\omega)$
μ_1	1	Step size for the filter $W_1(\omega)$
μ_2	0.01	Step size for the filter $W_2(\omega)$
β	2.5	Amplification factor
ϵ	10^{-9}	Regularization factor
Q_T'	0.01	Threshold for detecting speech absent
Q_T''	1.2	Threshold for detecting active speech with a strong power

Table 2: Performance evaluation by the SNR measurement

Primary input y(dB)	Crosstalk signal $\tilde{s}(dB)$	Reference input x(dB)	Conventional CRANC (dB)	Proposed CRANC (dB)
6	-9	-3	12.37	20.02
	-6	0	9.83	17.64
	-3	3	7.23	15.06
	0	6	5.09	12.29
3	-9	-6	13.71	18.45
	-6	-3	11.75	16.28
	-3	0	8.88	13.80
	0	3	6.48	11.17
0	-9	-9	14.27	16.58
	-6	-6	12.80	14.71
	-3	-3	10.68	12.45
	0	0	8.03	9.87

dB to 0 dB with a step size 3 dB. All signals are with a sampling rate of 8 kHz. The values of the parameters used in our tests are shown in Table. 1. For the primary signal, the additive noise $\tilde{n}$ was obtained as the output of a 16-th order FIR filter which was set up as the transfer function $H(\omega)$. For the reference signal, the crosstalk signal $\tilde{s}$ was obtained as the output of another 16-th order FIR filter which was set up as the transfer function $G(\omega)$. The weights of the filters of $H(\omega)$ and $G(\omega)$ are the values between 0 and 1 generated randomly.

The performance of the proposed CRANC was evaluated by using the signal-to-noise ratio (SNR) measure. The results are shown in Table 2. It can be seen from Table 2 that noticeable higher SNR improvements have been achieved by the proposed CRANC for all input levels. The average higher SNR improvements by the proposed CRANC against the conventional one attained around 4.77 dB. Meanwhile, it can also be observed from Table. 2 that the higher the input SNR of the primary input, the higher improvements the proposed CRANC achieves.

6. Conclusion

In this paper a new crosstalk resistant adaptive noise canceller was proposed. The proposed noise canceller, based on the affine projection algorithm, employed a sequential speech detector to manipulate the behavior of the noise cancelling filters. In contrast to the passive adaptive scheme used by the conventional CRANC, the proposed CRANC was based on an active adaptive strategy and therefore the performance of noise cancelling was substantially improved. Simulation results showed that the proposed CRANC achieved around 4.77 dB higher SNR than the conventional CRANC.

7. Acknowledgements

8. References

[1] G.Chen, S.Koh, and I.Soon, "Speech enhancement based on f-norm constrained truncated svd algorithm," in *DSP for In-Vehicle and Mobile Systems*, H.Abut, J.H.L.Hansen, and K.Takeda, Eds., chapter 11, pp. 169–178. Springer-Verlag, 2005.

[2] J.B.Maj, L.Royackers, M.Moonen, and J.Wouters, "A real time implementation and an evaluation of an optimal filtering technique for noise reduction in dual microphone hearing aids," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Montreal, Canada, May 2004, vol. 4, pp. 9–12.

[3] B.Widrow, J.R.Glover, J.M.McCool, J.Kaunitz, C.S.Williams, R.H.Hearn, J.R.Zeidler, E.Dong, and R.C.Goodlin, "Adaptive noise cancelling: Principles and applications," *Proceedings of The IEEE*, vol. 63, no. 12, pp. 1692–1716, Dec. 1975.

[4] J.E.Greenberg, "Modified lms algorithms for speech processing with an adaptive noise canceller," *IEEE Transactions on speech and audio processing*, vol. 6, no. 4, pp. 338–351, Jul. 1998.

[5] S.Ikeda and A.Sugiyama, "An adaptive noise canceller with low signal distortion for speech codecs," *IEEE Trans. on signal processing*, vol. 47, no. 3, pp. 665–674, Mar. 1999.

[6] M.Gabrea, "Two microphones speech enhancement system based on a double affine projection algorithm," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Hongkong, China, Apr. 2003, vol. 2, pp. 544–547.

[7] G.Mirchandani, R.L.Zinser, and J.B.Evans, "A new adaptive noise cancellation scheme in the presence of crosstalk," *IEEE Trans. Circuits and Systems II: Analog and Digital Signal Processing*, vol. 39, no. 10, pp. 681–694, Oct. 1992.

[8] G.Madhavan and H. de Bruin, "Crosstalk resistant adaptive noise cancellation," *Annals of Biomedical Engineering*, vol. 18, pp. 57–67, 1990.

[9] V.Parsa and P.A.Parker, "Constrained crosstalk resistant adaptive noise canceller," *IEE Electronics Letters*, vol. 30, no. 16, pp. 1276–1277, Aug. 1994.

[10] V.Parsa, P.A.Parker, and R.N.Scott, "Performance analysis of a crosstalk resistant adaptive noise canceller," *IEEE Trans. Circuits and Systems II: Analog and Digital Signal Processing*, vol. 43, no. 10, pp. 473–482, Jul. 1996.

[11] Simon Haykin, *Adaptive filter theory*, Prentice-Hall, Englewood Cliffs, NJ, third edition, 1996.