

Adaptive Filtering for Speech Enhancement

M. B. Yeary[†], P. C. Loizou^{††}

[†]Dept. of Elec. Engr., Texas A&M Univ.,
College Station, TX 77843-3128, mbyear@ee.tamu.edu

^{††}Dept. of Elec. Engr., Univ. of Texas at Dallas,
Richardson, TX 75083-0688, loizou@utdallas.edu

Abstract— This paper introduces an adaptive filter that employs a single input sensor and oversampling for speech enhancement. The architecture employs a hardware efficient IIR filter that utilizes the recursive least squares algorithm. To demonstrate the effectiveness of the new filter, it is compared to the spectral subtraction technique using the Itakura-Saito distance measure. The new filtering technique produces results that are slightly more favorable than spectral subtraction. This hardware efficient filter would be particularly suitable for cellular telephone applications where space is a premium.

I. INTRODUCTION

ADaptive noise canceling (ANC) techniques have been successfully applied to speech, electrocardiography, etc. since the 1960s [1]. These techniques employ the LMS algorithm to minimize the mean square error between a primary channel composed of speech plus uncorrelated noise, and a second reference signal consisting of noise. In general, ANC can be employed only when a second channel is available. But suppose that we can somehow generate a reference signal from the primary signal. Such an approach was proposed by Sambur [2] who used as a reference signal a delayed version of the primary signal. He delayed the speech signal by one pitch period, so that the delayed signal will be correlated with the original speech signal. Sambur's approach was shown to improve speech quality for additive white noise in the SNR range 0-10 dB [2].

One of the main limitations of Sambur's approach is the requirement of accurate and robust pitch estimation in noise. Other methods [3] removed the pitch estimator using forward and backward-adaptive filters, but required a speech/silence discriminator. Pitch estimation and speech/silence detectors, however, are prone to errors, particularly in noise.

In this paper, we propose a single-channel ANC technique which does not rely on pitch estimation or speech/silence detector to derive the reference signal. Instead, it produces a reference signal by oversampling the input (primary) signal. The proposed method is compared against the spectral subtraction algorithm.

This paper is organized as follows. Section II provides the theoretical formulation of the proposed algorithm, and section III provides a few comments about its implementation. Section IV gives a brief description of the implementation of spectral subtraction, and section V presents the

results.

II. THEORETICAL DEVELOPMENT

The block diagram in Figure 1 depicts the proposed filter architecture in which a single sensor captures both the desired signal and the noise [4]. This architecture has the following characteristics: (1) it does not require any statistics about the desired signal, (2) the desired signal need not be stationary, and (3) it only requires one input source which is sampled above the Nyquist frequency (oversampled).

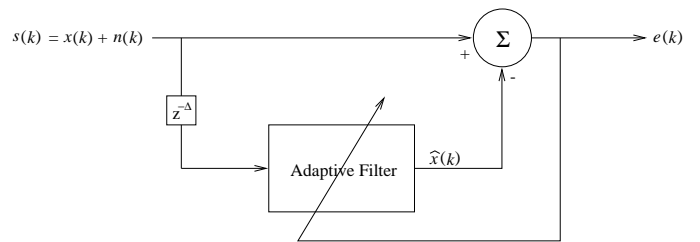


Fig. 1. Adaptive Filter Architecture

A reference signal is produced by delaying the input signal. Thus if two signals, $x(k)$ and $x(k - \Delta)$, are correlated, then $x(k)$ may be estimated by $\hat{x}(k)$ from $x(k - \Delta)$. The signal path with the delay element is important to provide a reference signal, and Δ ensures that $n(k)$ and $n(k - \Delta)$ are not correlated. It is also assumed that $n(k)$ is not correlated with $x(k)$. By minimizing the mean squared error of $e(k)$, $\hat{x}(k)$ will be the best estimate of $x(k)$ as shown below.

$$\begin{aligned} E[e^2(k)] &= E[(x(k) - \hat{x}(k) + n(k))^2] \\ &= E[(x(k) - \hat{x}(k))^2] + 2E[(x(k) - \hat{x}(k))n(k)] \\ &\quad + E[n^2(k)] \end{aligned} \quad (1)$$

Assuming $E[x(k)n(k)] = 0$ and $E[n(k)n(k - \Delta)] = 0$, then $2E[(x(k) - \hat{x}(k))n(k)] = 0$, and the mean squared error is

$$J = E[e^2(k)] = E[(x(k) - \hat{x}(k))^2] + E[n^2(k)] \quad (2)$$

Minimizing J is equivalent to minimizing $E[(x(k) - \hat{x}(k))^2]$. Therefore, minimizing J will cause $\hat{x}(k)$ to be the minimum mean-square estimate of $x(k)$ [4]. Estimating $\hat{x}(k)$ depends

on several factors including the type of filter, either IIR or FIR, and the strategy of how the cost function is to be minimized, be it either least mean squares or recursive least squares [4]. For this paper, the type of filter is IIR, and it minimizes J based on the recursive least squares algorithm.

It should be pointed out that Romdhane and Madiseti [5] have designed a filter based on the architecture in Figure 1, but it was FIR in nature.

In reference to Figure 1, the least squares algorithm is designed to minimize the sum of squared errors as defined by

$$\varepsilon(k) = \sum_{i=0}^k e^2(i) \quad (3)$$

where

$$\begin{aligned} e(i) &= s(i) - \hat{x}(i) \\ &= s(i) - \mathbf{w}^T \mathbf{v}(i) \quad , \end{aligned} \quad (4)$$

and

$$\mathbf{w} = [a_1, a_2, \dots, a_N, b_0, b_1, \dots, b_M]^T \quad , \quad (5)$$

and

$$\begin{aligned} \mathbf{v}(i) &= [\hat{x}(i-1), \hat{x}(i-2), \dots, \hat{x}(i-N), \\ & \quad s(i-\Delta), s(i-1-\Delta), \dots, s(i-M-\Delta)]^T \end{aligned} \quad (6)$$

The vector $\mathbf{v}(i)$ has $N + M + 1$ elements, indexed from $v(i)$ to $v(i - (N + M))$. The vector \mathbf{w} contains a_1, a_2, \dots, a_N which are known as the feedback coefficients, and b_0, b_1, \dots, b_M which are known as the feedforward coefficients. This vector has $N + M + 1$ elements. To minimize the sum of the squared errors, the partial derivative of $\varepsilon(k)$ is evaluated as

$$\begin{aligned} \frac{\partial \varepsilon(k)}{\partial w(j)} &= \frac{\partial}{\partial w(j)} \left[\sum_{i=0}^k e^2(i) \right] \\ &= 2 \sum_{i=0}^k e(i) \frac{\partial e(i)}{\partial w(j)} \quad , \quad j = 0, 1, \dots, N + M \end{aligned} \quad (7)$$

$$= -2 \sum_{i=0}^k e(i) v(i-j) \quad (8)$$

To minimize this error, the partial derivative is set to zero

$$\frac{\partial \varepsilon(k)}{\partial w(j)} = \sum_{i=0}^k e(i) v(i-j) = 0 \quad , \quad j = 0, 1, \dots, N + M \quad . \quad (9)$$

By combining (4) and (9), it follows

$$\sum_{i=0}^k \left[s(i) - \sum_{l=0}^{N+M} w(l) v(i-l) \right] v(i-j) = 0 \quad , \quad j = 0, 1, \dots, N+M \quad (10)$$

$$\sum_{l=0}^{N+M} w(l) \sum_{i=0}^k v(i-j) v(i-l) = \sum_{i=0}^k s(i) v(i-j) \quad j = 0, 1, \dots, N + M \quad (11)$$

Recognizing outer products, yields $\mathbf{r}_{vv}(k) \mathbf{w} = \mathbf{r}_{sv}(k)$. Therefore, the optimum coefficients are

$$\mathbf{w} = \mathbf{r}_{vv}^{-1}(k) \mathbf{r}_{sv}(k) \quad , \quad (12)$$

where

$$\mathbf{r}_{vv}(k) = \sum_{i=0}^k \mathbf{v}(i) \mathbf{v}^T(i) \quad (13)$$

and

$$\mathbf{r}_{sv}(k) = \sum_{i=0}^k s(i) \mathbf{v}(i) \quad . \quad (14)$$

Rather than solving equation (12) by computing the inverse of $\mathbf{r}_{vv}(k)$, the inverse will be recursively computed by making use of the matrix inversion lemma. The weight vector \mathbf{w} will become a function of discrete time, and will assume the notation $\mathbf{w}(k)$. The following recursive equation is the first step towards determining a recursive formula that will allow the weight vector to be updated at each k .

$$\begin{aligned} \mathbf{r}_{vv}(k) &= \sum_{i=0}^k \mathbf{v}(i) \mathbf{v}^T(i) \\ &= \sum_{i=0}^{k-1} \mathbf{v}(i) \mathbf{v}^T(i) + \mathbf{v}(k) \mathbf{v}^T(k) \\ &= \mathbf{r}_{vv}(k-1) + \mathbf{v}(k) \mathbf{v}^T(k) \end{aligned} \quad (15)$$

Similarly, $\mathbf{r}_{sv} = \mathbf{r}_{sv}(k-1) + s(k) \mathbf{v}(k)$. The inverse of $\mathbf{r}_{vv}(k)$ can be recursively computed using the matrix inversion lemma:

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1} \quad . \quad (16)$$

by allowing $A = \mathbf{r}_{vv}(k-1)$, $B = \mathbf{v}(k)$, $C = 1$, and $D = \mathbf{v}^T(k)$, i.e.:

$$\mathbf{r}_{vv}^{-1}(k) = \mathbf{r}_{vv}^{-1}(k-1) - \frac{\mathbf{r}_{vv}^{-1}(k-1) \mathbf{v}(k) \mathbf{v}^T(k) \mathbf{r}_{vv}^{-1}(k-1)}{1 + \mathbf{v}^T(k) \mathbf{r}_{vv}^{-1}(k-1) \mathbf{v}(k)} \quad . \quad (17)$$

At any instant in time, $\mathbf{r}_{vv}^{-1}(k)$ may also be determined as

$$\mathbf{r}_{vv}^{-1}(k) = \frac{\mathbf{r}_{vv}^{-1}(k-1)}{1 + \mathbf{v}^T(k) \mathbf{r}_{vv}^{-1}(k-1) \mathbf{v}(k)} \quad , \quad (18)$$

which will prove useful in the next several steps. To recursively update $\mathbf{w}(k)$, the following equations were used

$$\mathbf{w}(k) = \mathbf{r}_{vv}^{-1}(k) \mathbf{r}_{sv}(k) \quad (19)$$

and

$$\mathbf{r}_{sv}(k) = \mathbf{r}_{sv}(k-1) + s(k)\mathbf{v}(k) \quad (20)$$

hence

$$\begin{aligned} \mathbf{w}(k) &= \mathbf{r}_{vv}^{-1}(k)[\mathbf{r}_{sv}(k-1) + s(k)\mathbf{v}(k)] \\ &= \mathbf{r}_{vv}^{-1}(k)\mathbf{r}_{sv}(k-1) + s(k)\mathbf{r}_{vv}^{-1}(k)\mathbf{v}(k) . \end{aligned} \quad (21)$$

Then it follows,

$$\begin{aligned} \mathbf{w}(k) &= \left[\mathbf{r}_{vv}^{-1}(k-1) - \frac{\mathbf{r}_{vv}^{-1}(k-1)\mathbf{v}(k)\mathbf{v}^T(k)\mathbf{r}_{vv}^{-1}(k-1)}{1 + \mathbf{v}^T(k)\mathbf{r}_{vv}^{-1}(k-1)\mathbf{v}(k)} \right] \\ &\quad \cdot \mathbf{r}_{sv}(k-1) + s(k)\mathbf{r}_{vv}^{-1}(k)\mathbf{v}(k) \\ &= \mathbf{w}(k-1) \\ &\quad - \left[\frac{\mathbf{r}_{vv}^{-1}(k-1)\mathbf{v}(k)}{1 + \mathbf{v}^T(k)\mathbf{r}_{vv}^{-1}(k-1)\mathbf{v}(k)} \right] \mathbf{v}^T(k)\mathbf{w}(k-1) \\ &\quad + s(k)\mathbf{r}_{vv}^{-1}(k)\mathbf{v}(k) . \end{aligned} \quad (22)$$

By replacing $\mathbf{r}_{vv}^{-1}(k-1)$ in the third term on the right hand side of the previous equation with equation (18), the following expression is found:

$$\begin{aligned} \mathbf{w}(k) &= \mathbf{w}(k-1) \\ &\quad - \left[\frac{\mathbf{r}_{vv}^{-1}(k-1)\mathbf{v}(k)}{1 + \mathbf{v}^T(k)\mathbf{r}_{vv}^{-1}(k-1)\mathbf{v}(k)} \right] \mathbf{v}^T(k)\mathbf{w}(k-1) \\ &\quad + s(k) \left[\frac{\mathbf{r}_{vv}^{-1}(k-1)\mathbf{v}(k)}{1 + \mathbf{v}^T(k)\mathbf{r}_{vv}^{-1}(k-1)\mathbf{v}(k)} \right] . \end{aligned} \quad (23)$$

Then by recognizing the bracketed term in the above equation as a time varying gain term that modulates how much the error influences the magnitude of the update at each iteration, the following update equation is realized:

$$\mathbf{w}(k) = \mathbf{w}(k-1) + \mathbf{G}(k)[s(k) - \mathbf{v}^T(k)\mathbf{w}(k-1)] , \quad (24)$$

where

$$\mathbf{G}(k) = \frac{\mathbf{r}_{vv}^{-1}(k-1)\mathbf{v}(k)}{1 + \mathbf{v}^T(k)\mathbf{r}_{vv}^{-1}(k-1)\mathbf{v}(k)} . \quad (25)$$

This single sensor, IIR recursive least squares filter with one input signal is referred to as the IIR-RLS filter from this point forward.

III. IIR-RLS FILTER AND NOISE REMOVAL

The previous section provided the mathematical development of the IIR-RLS filter. Since the IIR-RLS filter is primarily intended to be used for non-stationary signals, the error term, $e(k)$, that influences the weights may be modified so that only the relatively recent values of $e(k)$ will be significant. Therefore equation (3) was modified to reflect this change:

$$E(k) = \sum_{i=0}^k \lambda^{k-i} e^2(i) \quad (26)$$

This consequently influences the time varying filter gain defined by equation (25), which becomes

$$\mathbf{G}(k) = \frac{\mathbf{r}_{vv}^{-1}(k-1)\mathbf{v}(k)}{\lambda + \mathbf{v}^T(k)\mathbf{r}_{vv}^{-1}(k-1)\mathbf{v}(k)} . \quad (27)$$

In general, the Nyquist sampling rate must be obeyed during sampling; furthermore, the sampling rate should be significantly high enough to justify correlation between $x(k)$ and $x(k-\Delta)$. Thus, by delaying an oversampled signal even by a small amount, say Δ in this case, $x(k)$ may be estimated from $x(k-\Delta)$. On the other hand, Δ should be large enough to decorrelate $n(k)$ and $n(k-\Delta)$. The net result of this will be an adaptive filter whose transfer function favors the spectrum of the narrowband signal; hence the additive noise will be attenuated.

It is also important to mention that $\det[\mathbf{r}_{vv}(k)] = 0$ must not occur during filtering. This was accomplished by following a technique proposed in [6].

IV. SPECTRAL SUBTRACTION

The spectral subtraction method is one of the most popular methods used for removing noise from the speech signal [7][8][9]. Unlike the proposed IIR-RLS algorithm, the processing is done on a frame-by-frame basis (typically 20-30 ms) in the frequency domain. Consider a speech signal $x(k)$ corrupted by additive stationary background noise $n(k)$. The noisy speech signal can be written as:

$$s(k) = x(k) + n(k) \quad (28)$$

The enhanced speech short-time magnitude $|\hat{X}(\omega)|$ is computed by subtracting from the noisy speech magnitude $|S(\omega)|$ the noise magnitude estimate $|\hat{N}(\omega)|$ estimated during speech pauses. The subtraction of noise is expressed as

$$|\hat{X}(\omega)|^2 = \begin{cases} |S(\omega)|^2 - \alpha|\hat{N}(\omega)|^2 & \text{if } \frac{|S(\omega)|^2}{|\hat{N}(\omega)|^2} > \alpha + \beta \\ \beta|\hat{N}(\omega)|^2 & \text{otherwise} \end{cases} \quad (29)$$

where $|\hat{N}(\omega)|^2$ represents the noise power spectrum estimate, α is a subtraction factor ($\alpha \geq 1$), and β ($0 < \beta \ll 1$) is the spectral floor parameter [8]. The parameter α varied between 1 and 5 depending on the segmental SNR [8], and the spectral floor parameter β was set to $\beta = 0.02$. The enhanced speech signal $\hat{x}(k)$ was obtained by taking the IDFT of the enhanced magnitude spectrum $|\hat{X}(\omega)|$, using the phase of the noisy speech signal. For the experiments described below, speech was segmented into 20-ms frames using a Tukey window with 10% overlap. The overlap-and-add method was used to reconstruct the signal.

V. EXPERIMENTAL PROCEDURE AND RESULTS

For comparative purposes, we processed the word “head”, sampled at a rate of 16 KHz, by both spectral subtraction and IIR-RLS algorithms. Prior to applying the IIR-RLS filter, the speech signal was upsampled to 200 KHz, and white Gaussian noise was added to obtain an SNR of 5 dB. After processing the signal through the IIR-RLS filter, the output signal was downsampled to 16 KHz for a fair comparison with the spectral subtraction technique. The IIR-RLS filter had 9 taps (7 feedforward and 2 feedback), and the delay variable, Δ , was assigned a value of 1, since this was enough to decorrelate the additive noise, $n(k)$. The parameter λ was set to be 0.998.

The performance of the two algorithms was evaluated using the Itakura-Saito measure [10][11][12][13]. The Itakura-Saito distance compares the all-pole models of the signal before and after filtering. It will be used as the figure of merit for comparing performance of the IIR-RLS filter against spectral subtraction. For the j^{th} stationary speech segment, the Itakura-Saito distance is defined to be

$$d(\vec{\rho}_{\hat{x}}, \vec{\rho}_x) = \log \left[\frac{\vec{\rho}_{\hat{x}}^T \mathbf{R}_x \vec{\rho}_{\hat{x}}}{\vec{\rho}_x^T \mathbf{R}_x \vec{\rho}_x} \right] \quad (30)$$

where $\vec{\rho}_{\hat{x}} = [1, -\rho_{\hat{x}}(1), \dots, -\rho_{\hat{x}}(N_r)]^T$ is the vector containing the LPC coefficients of the filtered segment. The vector $\vec{\rho}_x = [1, -\rho_x(1), \dots, -\rho_x(N_r)]^T$ contains the LPC coefficients of the original noise free segment. N_r was set equal to 16. \mathbf{R}_x is the corresponding autocorrelation matrix of the j^{th} noise-free segment [14].

Figure 2 depicts Itakura-Saito distance for both filtering techniques. During the duration of the digitized signal, sixteen 10-ms segments containing the primary utterance were examined to compute the mean distance. The mean value of the Itakura-Saito distance for the IIR-RLS filter was 0.5900, with a standard deviation of 0.2396. On the other hand, the mean value of the Itakura-Saito distance for the spectral subtraction technique was 0.9050, with a standard deviation of 0.6955. In general, as the distance is reduced, the autoregressive model of the filtered signal more closely resembles the model of the clean signal. Hence the IIR-RLS filter is slightly more competitive than the spectral subtraction technique.

VI. SUMMARY

An adaptive IIR-RLS filter was developed which uses a single input sensor and oversampling to derive a reference signal. A recursive least squares approach was used to derive the filter coefficients on a sample by sample basis. This filter was used to remove noise from the speech signal, and evaluated against the spectral subtraction algorithm. The proposed filtering technique produced slightly more favorable results than the spectral subtraction technique. One of the main advantages of the IIR-RLS filter is that it processes data on sample per sample basis, which lends itself to a more efficient real-time implementation.

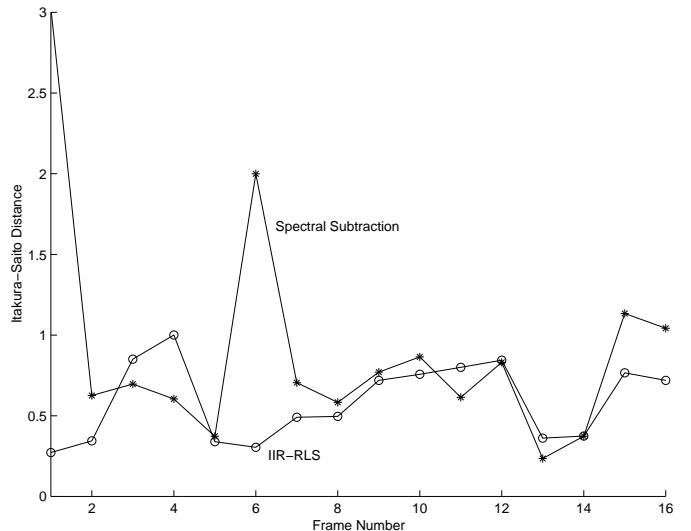


Fig. 2. Comparison Plot of the Itakura-Saito Distances

REFERENCES

- [1] B. Widrow, S. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, N.J. : Prentice-Hall, 1985.
- [2] M. Sambur, “Adaptive noise cancelling for speech signals,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-26, No. 5, pp. 419-423, 1978.
- [3] J. Kim, and C. Un “Enhancement of noisy speech by forward/backward adaptive digital filtering,” *Proc. ICASSP*, Tokyo, Japan, vol. 1, pp. 89-92, 1986.
- [4] M. Hayes, *Statistical Digital Signal Processing and Modeling*. New York: John Wiley and Sons, 1996.
- [5] M. Romdhane and V. Madisetti, “All-digital oversampled front-end sensors,” *IEEE Signal Processing Letters*, vol. 3, No. 2, pp 38-39, 1996.
- [6] M. Tham and S. Mansoori, “Covariance resetting in recursive least squares estimation,” *1988 International Conference on Control*, pp. 128-133. Conference Publication Number 285. April, 1988.
- [7] S. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-27, No. 2, pp. 113-120, 1979.
- [8] M. Berouti, J. Makhoul, and R. Schwartz, “Enhancement of speech corrupted by acoustic noise,” *Proc. of ICASSP*, Washington, DC., pp. 208-211, 1979
- [9] J. Hansen, M. Clements, “Use of objective speech quality measures in selecting effective spectral estimation techniques for speech enhancement,” *Proceedings of the 32nd Midwest Symposium on Circuits and Systems*, vol. 1, pp. 105-108, 1989.
- [10] J. Deller, J. Proakis, and J. Hansen, *Discrete Time Processing of Speech Signals*. New York: Prentice Hall, 1993.
- [11] S. Gannot, D. Burshtein, E. Weinstein, “Iterative and sequential Kalman filter-based speech enhancement algorithms,” *IEEE Transactions on Speech and Audio Processing*, vol. 6, No. 4, pp. 373-385, 1998.
- [12] J. Gibson, B. Koo, and S. Gray, “Filtering of Colored Noise for Speech Enhancement and Coding,” *IEEE Transactions on Signal Processing*, vol. 39, No. 8, pp. 1732-1742, 1991.
- [13] D. Popescu and I. Zeljkovic, “Kalman filtering of colored noise for speech enhancement,” *Proc. of ICASSP*, vol. 1, pp. 17-20, May, 1998.
- [14] S. Quackenbush, T. Barnwell III, and M. Clements, *Objective Measures of Speech Quality*. Englewood Cliffs, New Jersey: Prentice Hall, 1988.