Adaptive Noise Canceling for Speech Signals

MARVIN R. SAMBUR, MEMBER, IEEE

Abstract—A least mean-square (LMS) adaptive filtering approach has been formulated for removing the deleterious effects of additive noise on the speech signal. Unlike the classical LMS adaptive filtering scheme, the proposed method is designed to cancel out the clean speech signal. This method takes advantage of the quasi-periodic nature of the speech signal to form an estimate of the clean speech signal at time t from the value of the signal at time t minus the estimated pitch period. For additive white noise distortion, preliminary tests indicate that the method improves the perceived speech quality and increases the signal-to-noise ratio (SNR) by 7 dB in a 0 dB environment. The method has also been shown to partially remove the perceived granularity of CVSD coded speech signals and to lead to an improvement in the linear prediction analysis/synthesis of noisy speech.

I. INTRODUCTION

IN ALL PRACTICAL situations, the received speech waveform contains some form of noise component. The noise may be a result of the finite precision involved in coding the transmitted waveform (quantization noise), or due to the addition of acoustically coupled background noise. Depending on the amount and type of noise, the quality of the received waveform can range from being slightly degraded to being annoying to listen to, and finally to being totally unintelligible.

The problem of removing the unwanted noise component from a received signal has been the subject of numerous investigations. The pioneering work of Wiener and others give an optimum approach for deriving a filter that tends to suppress the noise while leaving the desired signal relatively unchanged [1]–[3]. The design of these filters requires that the signal and the noise be stationary and that the statistics of both signals be known *a priori*. In practice, these conditions are rarely met.

In this paper, we present a technique for optimum noise filtering for speech signals based upon the principles of least mean-square (LMS) adaptive filtering [4]. This technique has the advantage of requiring no *a priori* knowledge of the detailed properties of the noise signal. The technique takes advantage of the quasi-periodic nature of the speech, and preliminary results indicate that the technique improves the perceived speech quality of a signal corrupted by additive white noise. In addition, the method also improves the perceived quality of CVSD coded speech and the linear prediction analysis/synthesis of noisy speech.

The organization of this paper is as follows. In the next section we shall review the concept of adaptive filtering for noise canceling. We shall then indicate how this concept can be applied to filtering noise from a speech waveform. In Section V, we discuss the experimental evaluation of the proposed

technique for removing additive white noise distortions, and for filtering the granularity in CVSD coded speech. In Section VI, we discuss the effectiveness of the algorithm in improving the performance of the LPC analysis/synthesis of noisy speech.

II. CONCEPT OF ADAPTIVE NOISE CANCELING

Fig. 1 illustrates the basic principles of adaptive noise canceling.[1] The input to the adaptive filter is a noise signal $w_1(n)$ that is highly correlated with the additive disturbance, $w(n)$, but is uncorrelated with the clean signal $s(n)$. (One can think of $w_1(n)$ as being derived from a sensor located at a point in the noise field where the signal is undetectable.) The reference signal $w_1(n)$ is filtered to produce the output $\hat{w}(n)$ that is an estimate of the additive noise $w(n)$. This output is then subtracted from the noisy signal $x(n)$ to produce the system output $z(n)$. The system output is used to control the adaptive filter and is an estimate of $s(n)$.

Provided $s(n)$ is uncorrelated with both $w_1(n)$ and $w(n)$, and the adaptive filter is adjusted to give a system output $z(n)$ that has the least possible energy, then $z(n)$ is a best least-squares fit to the clean signal $s(n)$. To prove this, we note that the power in $z(n)$ is given by

$$E(z^2(n)) = E(s^2(n) + (w(n) - \hat{w}(n))^2 + 2s(n)(w(n) - \hat{w}(n)))$$

where $E(.)$ denotes expected value. Now since the noise terms and the signal $s(n)$ are assumed uncorrelated,

$$E(z^2(n)) = E(s^2(n)) + E((w(n) - \hat{w}(n))^2).$$

Since the signal energy is a fixed quantity for the frame of interest, minimizing the output energy yields

$$\min E(z^2(n)) = E(s^2(n)) + \min E((w(n) - \hat{w}(n))^2).$$

Thus, when the noise canceling filter is adjusted so that $E(z(n))$ is minimized, $E((w(n) - \hat{w}(n)))^2$ is also minimized. The filter output $\hat{w}(n)$ is then a best least-squares estimate of the primary noise $w(n)$. Moreover, when $E((w(n) - \hat{w}(n))^2)$ is minimized, $E((z(n) - s(n))^2)$ is also minimized since

$$z(n) - s(n) = w(n) - \hat{w}(n).$$

Thus, $z(n)$ is a best least-squares estimate of the clean signal $s(n)$.

III. NOISE CANCELING FOR SPEECH INPUTS

The success of the adaptive noise canceling filter is dependent on obtaining an external reference noise input that is uncorrelated with the signal and highly correlated with the additive noise corruption. In most applications, a reference noise

---

[1] Much of the material in this section is abstracted from the excellent paper cited in [4].
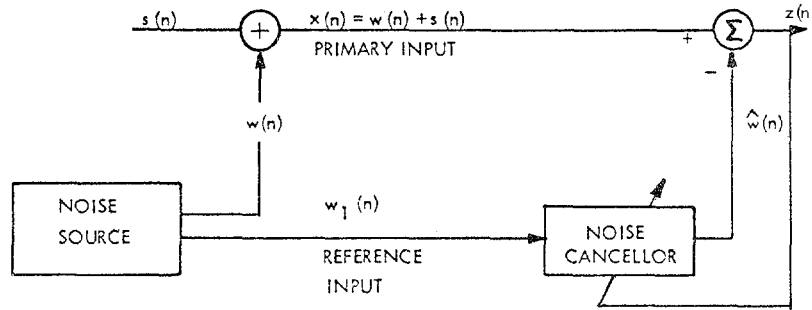
Fig. 1. Noise canceling model.

signal is not available from a second sensor. To avoid this problem, previous investigators have tried to form a reference noise signal by assuming that the noise is stationary, and that the average signal determined during periods classified (or known to be signal free) as "silence" are representative of the noise [5], [6]. Unfortunately, this approach suffers because the noise is rarely stationary, a finite sample may be insufficient to estimate the noise signal, the silence decision is not error free, and finally, the technique cannot be applied for quantization noise.

Although it may be difficult to form a reference noise input, it is quite easy to form a reference input of the original speech signal. Since speech is quasi-periodic, a section of speech delayed by a small amount (one or two pitch periods) will be highly correlated with the true speech signal $s(n)$ and will be uncorrelated with the additive noise $w(n)$ (provided that $w(n)$ is sufficiently broad band).

Fig. 2 illustrates how we can take advantage of the quasi-periodic nature of speech for effective noise removal. In this system, we take advantage of the fact that $s(n)$ and $s(n - T)$ are highly correlated, and $w(n)$ and $w(n - T)$ are not correlated with themselves and the speech signal. Using the arguments of Section II, it can be seen that minimizing the energy in the system output $\hat{w}(n)$ will result in a signal $\hat{s}(n)$ that is a best least-squares fit to $s(n)$.

IV. Noise Canceling Algorithm

The structure of the proposed adaptive filter is illustrated in Fig. 3 and represents a filter with a finite impulse response with output

$$\hat{s}(n) = \sum_{i=0}^{L} b_i \, x(n - i - T)$$

where x is the noisy speech signal and T is the analyzed pitch period for the frame of interest. The pitch period is automatically calculated using a modified AMDF approach with nonlinear smoothing [7]-[9]. This algorithm works well even in the presence of a high degree of noise. During analysis frame classified as unvoiced, the output is given by[2]

$$\hat{s}(n) = x(n).$$

The coefficients of the filter, the b_i's, are updated on a sample

---

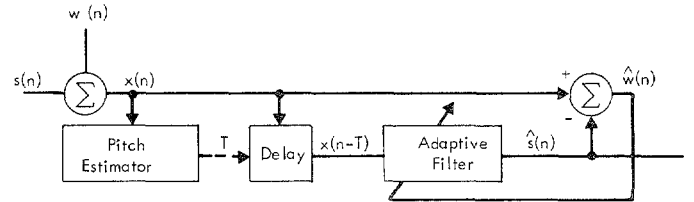[2]Another effective procedure is to keep the LMS filter constant (no updating) during unvoiced frames.

Fig. 2. Adaptive filtering approach for removing noise from speech.

by sample basis according to an LMS algorithm proposed by Widrow and Hoff [4], [10], [11]. Letting B_n denote the coefficient vector $(b_0, b_1, \cdots, b_L)$ at time n, then the coefficients to be used at time $n + 1$ are

$$B_{n+1} = B_n + 2u\hat{w}(n) X_{n-T}$$

where

$$\hat{w}(n) = x(n) - \hat{s}(n),$$
$$X_n = (x(n), x(n - 1), \cdots, x(n - L)),$$

and u is a factor that controls stability and rate of convergence. It can be shown [12], [13] that starting with an arbitrary coefficient vector, the algorithm will converge in the mean and will remain stable as long as the parameter u is greater than zero but less than the reciprocal of the largest eigenvalue of the matrix R

$$R = E\left[X_n X_n^T\right].$$

In the next section we shall indicate the results of our preliminary evaluation of the proposed adaptive filter for a variety of noise distortions.

V. Effectiveness in Noise Canceling

A. Additive White Noise

Pseudo-random additive white noise was added to a high quality speech signal to evaluate the effectiveness of the LMS adaptive filtering algorithm (Figs. 2 and 3) for background noise removal. The implemented LMS adaptive filter used pitch estimates automatically computed every 22.5 ms frame. The order of the filter was set to 10 ($L = 10$) and u was approximated as

$$u = \frac{0.01}{h_{max}}$$

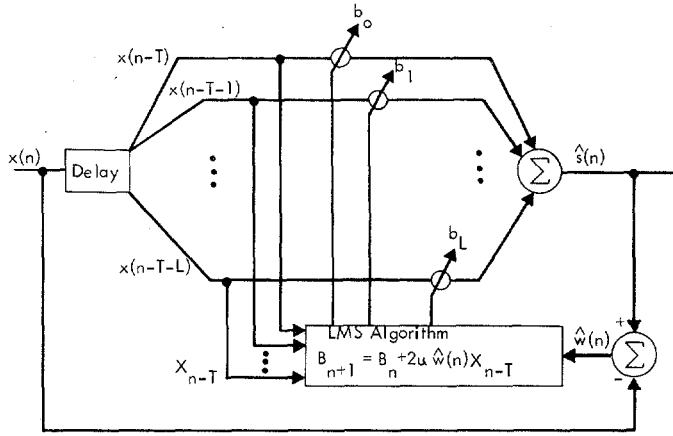where h_{max} is the largest eigenvalue of the correlation matrix R_0 of the first voiced frame

Fig. 3. LMS adaptive algorithm for removing noise from speech.

$$R_0 = \begin{bmatrix} V_{00} & V_{01} & & \\ V_{10} & V_{11} & \cdot & \\ V_{20} & \cdot & \cdot & \cdot \\ \cdot & & & \\ \cdot & & & \end{bmatrix}$$

where

$$V_{ij} = \frac{1}{N} \sum_{n=1}^{180} x(n-i) x(n-j)$$

and where N is the number of samples in the frame (in this case $N = 180$ as the sampling rate was 8 kHz). The adaption parameter, u, was deliberately set at a conservative value in order to avoid any problems in ensuring convergence and stability of the adaptive filter.

In the experimental evaluation, additive white noise with signal-to-noise ratio (SNR) values of 0 dB, 5 dB, and 10 dB were analyzed. In all cases, the quality of the processed speech was improved over that of the noisy input.[3] The more severe the noise, the more dramatic the improvement in quality of the filtered speech. The actual SNR[4] values computed for the processed speech as a function of L are illustrated in the following table.

	Input SNR		
	0 dB	5 dB	10 dB
$L = 6$	6.5	8.3	10.8
$L = 10$	6.9	8.5	10.9
$L = 14$	7.1	8.7	10.95

The table shows that at 0 dB, the filtered speech demonstrates about a 7 dB improvement in SNR. At 10 dB, there is a very small, barely perceptible improvement.

Using the adaption rate $u = 0.01/h_{max}$, the filter would con-

---

[3] By improvement, we mean that the speech was more pleasant to listen to and "appeared" to have more intelligibility. More extensive testing is needed to confirm the intelligibility improvement.

[4] $SNR = \log 10 \dfrac{s^2(n)}{(s(n) - \hat{s}(n))^2}$.

verge and remain stable after about $\frac{3}{4}$ s of processed speech. As long as u was sufficiently small, the quality of the filtered speech was insensitive to the exact value of u. A good compromise value for u is 10^{-7} when the speech is digitized with 12 bits.

B. Quantization Noise

The LMS adaptive noise filter was evaluated to determine its effectiveness in removing quantization noise. In the evaluation, the original speech sample was coded by continuously variable slope delta (CVSD) modulation at 16 kbits/s [14].[5] The CVSD coded speech was then processed by the LMS adaptive filter implementation and the resulting speech was perceptually evaluated by informal listening experiments.

The informal perceptual experiment indicated that the LMS adaptive filter removed a portion of the "granular" quality of the CVSD input.[6] Fig. 4 illustrates a section of the CVSD coded speech and the corresponding section of the filtered speech. The CVSD coded speech demonstrates the familiar granular quantization noise which is missing in the filtered speech.

This perceptual result is consistent with the theoretical behavior of the adaptive filter and the types of noise introduced by the CVSD coder. The CVSD quantization noise is one of two types: slope overload noise which is signal dependent and occurs when the step size is too small, and granular noise which is signal independent and broad band and occurs when the step size is too large [14]. The slope overload noise is the most prevalent type of distortion in the CVSD signal, but the granular noise is perceptually the most annoying [14]. As discussed in Section II, the adaptive filter will tend to remove the granular noise since this noise is signal independent and broad band and will tend to leave the slope overload noise unaffected since this noise is signal dependent.

VI. LINEAR PREDICTION ANALYSIS/SYNTHESIS IMPROVEMENT

The quality and intelligibility of linear prediction coded (LPC) speech is quite markedly degraded when performed on noisy speech [15]. The additive noise alters the underlying all-pole structure of the speech and causes the linear prediction analysis to incorrectly estimate the actual formant frequencies and bandwidths of the signal. Fig. 5 illustrates the poor performance of the LPC method when used to analyze a section of speech corrupted by additive white noise (SNR = 0 dB). In the same figure, the LPC analysis of the same speech section after LMS adaptive filtering is shown. The noise canceled speech has an LPC computed spectrum that is more nearly equal to the original. The improvements in the LPC analysis of noisy speech after being processed by the LMS adaptive filtering algorithm ranged from slight to the magnitude indicated in Fig. 5.

The LPC synthesis of noisy speech is quite poor compared to the quality of the LPC synthesis of the clean speech even when

---

[5] A standard Harris semiconductor HC 55532 all Digital CVSD package was used to generate the CVSD signal.

[6] Again we should point out that this improvement may not mean that the filtered speech is more intelligible than the original.
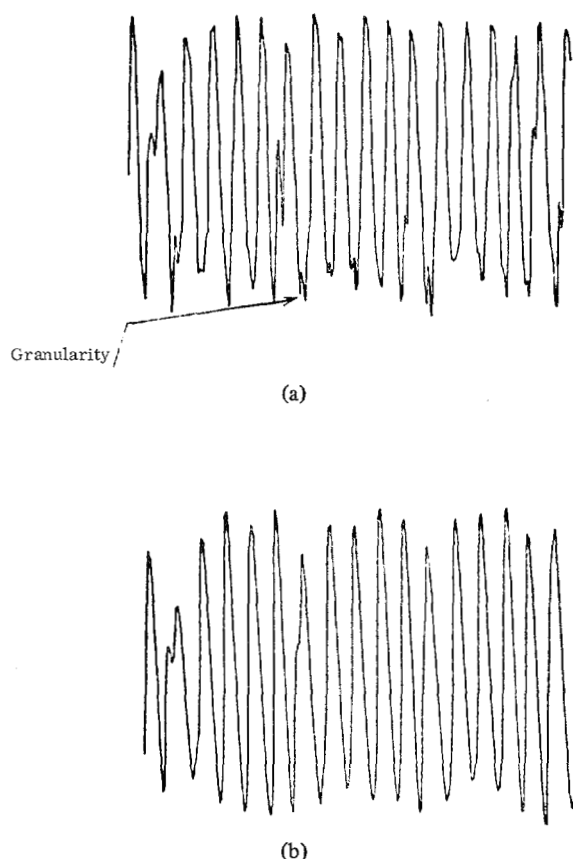
Granularity

(a)

(b)

Fig. 4. LMS algorithm performance for CVSD coded speech. (a) Original CVSD coded speech. (b) Filtered speech.

the noise is not severe (15 dB). However, when the noisy speech is processed by the LMS adaptive filter, preliminary experiments indicate that the LPC synthesis is more pleasant to listen to and appears to have more intelligibility than the LPC version of the noisy speech.

An objective indication of the improvement afforded by the adaptive noise canceling scheme is illustrated in Fig. 6. In this figure, we have plotted the dissimilarity between the LPC coefficients derived from the clean speech signal and the LPC coefficients derived first from the noisy speech (SNR = 0 dB) and then from the noise canceled speech. The dissimilarity is measured by an LPC distance function discussed in [15]. The higher the LPC distance, the more perceptually dissimilar are the compared waveforms. Fig. 6 indicates that the adaptive noise canceling method results in a set of LPC coefficients that are more nearly equal to the coefficients obtained from the clean signal than are the coefficients derived from the noisy signal. The cumulative distance between the LPC coefficients from the clean signal and the LPC coefficients calculated from the noise canceled speech is 377. By contrast, the cumulative distance between the coefficients derived from the noisy speech and those from the clean speech is 901.

VII. DISCUSSION

A technique for filtering the noise from a speech waveform has been proposed. The technique is based upon the concept

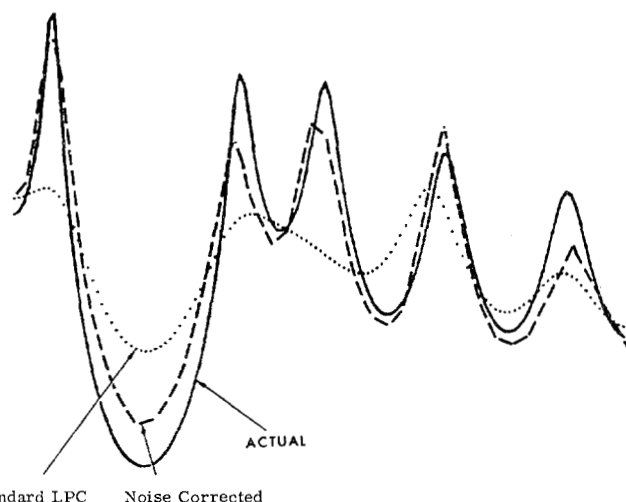ACTUAL

Standard LPC   Noise Corrected

Fig. 5. Illustration of improvement in spectral representation of noisy speech. White noise with SNR equals 0 dB.

of adaptive filtering and takes advantage of the quasi-periodic nature of the speech waveform to supply a reference signal to the adaptive filter. Preliminary tests indicate that the technique appears to improve the quality of noise speech and slightly reduce granular quantization noise. The technique also appears to improve the performance of the linear prediction analysis/synthesis of noisy speech. Further experimentation is planned to confirm these preliminary conclusions and to formally measure diagnostic rhyme test (DRT) scores to determine the improvement in intelligibility this scheme affords.

In addition, future plans call for the use of multiple reference signals to improve the effectiveness of the adaptive noise canceling algorithm. For example, $s(n)$ is correlated with both $s(n - T)$ and $s(n + T)$ and both signals can be used as reference inputs (Fig. 2).

REFERENCES

[1] N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series, with Engineering Applications.* New York: Wiley, 1949.

[2] R. Kalman, "On the general theory of control," in *Proc. 1st IFAC Congress.* London: Butterworths, 1960.

[3] R. Kalman and R. Bucy, "New results in linear filtering and prediction theory," *Trans. ASME (ser. D, J. Basic Eng.),* vol. 83, pp. 95–107, Dec. 1961.

[4] B. Widrow *et al.,* "Adaptive noise canceling: Principles and applications," *Proc. IEEE,* vol. 63, pp. 1692–1716, Dec. 1975.

[5] S. Boll, "Improving linear prediction analysis of noisy speech by predictive noise cancellation," in *1977 IEEE Proc. Int. Conf. Acoust., Speech, Signal Processing,* pp. 10–12, May 9, 1977.

[6] R. McAulay, "Optimum classification of voiced speech, unvoiced speech and silence in the presence of noise and interference," Lincoln Lab. Tech. Note, 1976-7, June 1976.

[7] M. J. Ross, H. L. Shaffer, A. Cohen, R. Freudberg, and H. J. Manley, "Average magnitude difference function pitch extractor," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-22, pp. 353–362, Oct. 1974.

[8] "Technical description operation software package for the narrowband terminal processor assembly development," vol. 1, pt. 2, Contr. MDA 904-76-C-0313.

[9] L. Rabiner, M. Sambur, and C. Schmidt, "Applications of a nonlinear smoothing algorithm to speech processing," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-23, pp. 552–557, Dec. 1975.

CUMULATIVE DISTANCE OF NOISY SP = 901.
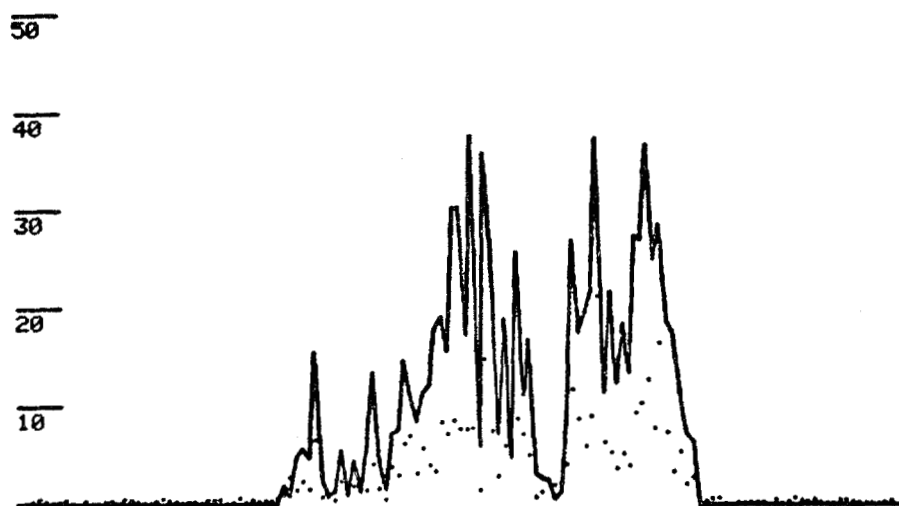CUMULATIVE DISTANCE OF NOISE CANCELLED SP = 377.

Fig. 6. LPC distance analysis for the sentence "Happy hour is here." The solid line represents the distance between the LPC coefficients for the noisy speech (SNR = 0 dB) and the coefficients obtained from the clean speech. The dotted line represents the distance between the coefficients calculated from the adaptive noise canceled signal and those obtained from the clean signal.

[10] B. Widrow and M. Hoff, Jr., "Adaptive switching circuits," in *IRE WESCON Conv. Rec.*, pt. 4, pp. 96-104, 1960.

[11] J. S. Koford and G. F. Groner, "The use of an adaptive threshold element to design a linear optimal pattern classifier," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 42-50, Jan. 1966.

[12] B. Widrow, P. E. Mantey, L. J. Griffiths, and B. B. Goode, "Adaptive antenna systems," *Proc. IEEE*, vol. 55, pp. 2143-2159, Dec. 1967.

[13] R. L. Riegler and R. T. Compton, Jr., "An adaptive array for in-terference rejection," *Proc. IEEE*, vol. 61, pp. 748-758, June 1973.

[14] N. S. Jayant, "Digital coding of speech waveforms: PCM, DPCM, and DM quantizers," *Proc. IEEE*, vol. 62, pp. 611-632, May 1974.

[15] M. R. Sambur and N. S. Jayant, "LPC analysis/synthesis from speech inputs containing quantizing noise or additive white noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 488-494, Dec. 1976.