tuations in the atmosphere" (in Russian), *Izv. Vysch. Ucheb. Zaved., Radiofiz.*, vol. 17, pp. 105-112, 1974.

Section VII

[181] D. Fried, "Differential angle of arrival; theory, evaluation, and measurement feasibility," *Radio Sci.*, vol. 10, pp. 71-76, Jan. 1975.
[182] J. Lowry, J. Wolf, and J. Carter, "Acquisition and tracking assembly," McDonnell Douglas Tech. Rep. for Air Force Avionics Lab., Tech. Rep. AFAL-TR-73-380, Feb. 15, 1974.
[183] V. Lukin, V. Pokasov, and S. Khmelevstov, "Investigation of the time characteristics of fluctuations of the phases of optical waves propagating in the bottom layer of the atmosphere," *Radiophys. Quantum Electron.*, vol. 15, pp. 1426-1430, Dec. 1972.
[184] R. Lutomirski and R. Buser, "Phase difference and angle-of-arrival fluctuations in tracking a moving point source," *Appl. Opt.*, vol. 13, pp. 2869-2873, Dec. 1974.
[185] R. Lutomirski and R. Warren, "Atmospheric distortions in a retroreflected laser signal," *Appl. Opt.*, vol. 14, pp. 840-846, Apr. 1975.
[186] R. Lutomirski and H. Yura, "Imaging of extended objects through a turbulent atmosphere," *Appl. Opt.*, vol. 13, pp. 431-437, Feb. 1974.
[187] J. Pearson *et al.*, "Atmospheric turbulence compensation using coherent optical adaptive techniques," presented at OSA Topical Meeting on Propagation Through Turbulence, Boulder, Colo., paper ThB5-1, July 1974.
[188] D. Slepian, "Linear least-squares filtering of distorted images," *J. Opt. Soc. Amer.*, vol. 57, pp. 918-922, July 1967.
[189] H. Yura, Holography in a random spatially inhomogeneous medium," *Appl. Opt.*, vol. 12, pp. 1188-1192, June 1973.

Section VIII

[190] C. Gardner and M. Plonus, "The effects of atmospheric turbulence on the propagation of pulsed laser beams," *Radio Sci.*, vol. 10, pp. 129-137, Jan. 1975.

Section IX

[191] L. Apresyan, "The radiative-transfer equation with allowance for longitudinal waves," *Radiophys. Quantum Electron.*, vol. 16, pp. 348-356, Mar. 1973.
[192] Y. Barabanenkov, A. Vinogradov, Y. Kravtsov, and V. Tatarski, "Application of the theory of multiple scattering of waves to the derivation of the radiation transfer equation for a statistically inhomogeneous medium," *Radiophys. Quantum Electron.*, vol. 15, pp. 1420-1425, Dec. 1972.
[193] C. Martens and N. Jen, "Electromagnetic wave scattering from a turbulent plasma," *Radio Sci.*, vol. 10, pp. 221-228, Feb. 1975.
[194] I. Besieris, "Long-range electromagnetic random wave propagation using the parabolic equation method," *Dig. 1975 URSI Meet.*, p. 15, June 1975.

Adaptive Noise Cancelling: Principles and Applications

BERNARD WIDROW, SENIOR MEMBER, IEEE, JOHN R. GLOVER, JR., MEMBER, IEEE, JOHN M. McCOOL, SENIOR MEMBER, IEEE, JOHN KAUNITZ, MEMBER, IEEE, CHARLES S. WILLIAMS, STUDENT MEMBER, IEEE, ROBERT H. HEARN, JAMES R. ZEIDLER, EUGENE DONG, JR., AND ROBERT C. GOODLIN

Abstract—This paper describes the concept of adaptive noise cancelling, an alternative method of estimating signals corrupted by additive noise or interference. The method uses a "primary" input containing the corrupted signal and a "reference" input containing noise correlated in some unknown way with the primary noise. The reference input is adaptively filtered and subtracted from the primary input to obtain the signal estimate. Adaptive filtering before subtraction allows the treatment of inputs that are deterministic or stochastic, stationary or time variable. Wiener solutions are developed to describe asymptotic adaptive performance and output signal-to-noise ratio for stationary stochastic inputs, including single and multiple reference inputs. These solutions show that when the reference input is free of signal and certain other conditions are met noise in the primary input can be essentially eliminated without signal distortion. It is further shown that in treating periodic interference the adaptive noise canceller acts as a notch filter with narrow bandwidth, infinite null, and the capability of tracking the exact frequency of the interference; in this case the canceller behaves as a linear, time-invariant system, with the adaptive filter converging on a dynamic rather than a static solution. Experimental results are presented that illustrate the usefulness of the adaptive noise cancelling technique in a variety of practical applications. These applications include the cancelling of various forms of periodic interference in electrocardiography, the cancelling of periodic interference in speech signals, and the cancelling of broad-band interference in the sidelobes of an antenna array. In further experiments it is shown that a sine wave and Gaussian noise can be separated by using a reference input that is a delayed version of the primary input. Suggested applications include the elimination of tape hum or turntable rumble during the playback of recorded broad-band signals and the automatic detection of very-low-level periodic signals masked by broad-band noise.

Manuscript received March 24, 1975; August 7, 1975.
This work was supported in part by the National Science Foundation under Grant ENGR 74-21752, the National Institutes of Health under Grant 1R01HL18307-01CVB, and the Naval Ship Systems Command under Task Assignment SF 11-121-102.

B. Widrow and C. S. Williams are with the Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, Calif. 94305.
J. R. Glover, Jr., was with the Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, Calif. He is now with the Department of Electrical Engineering, University of Houston, Houston, Tex.
J. M. McCool, R. H. Hearn, and J. R. Zeidler are with the Fleet Engineering Department, Naval Undersea Center, San Diego, Calif. 92132.
J. Kaunitz was with the Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, Calif. He is now with Computer Sciences of Australia, St. Leonards, N. S. W., Australia, 2065.
E. Dong, Jr., and R. C. Goodlin are with the School of Medicine, Stanford University, Stanford, Cailf. 94305.

I. INTRODUCTION

THE USUAL method of estimating a signal corrupted by additive noise[1] is to pass it through a filter that tends to suppress the noise while leaving the signal relatively unchanged. The design of such filters is the domain of optimal filtering, which originated with the pioneering work of Wiener

[1] For simplicity the term "noise" is used in this paper to signify all forms of interference, deterministic as well as stochastic.

Amazon v. SoundClear
US Patent 9,031,259
Amazon EX-1019

and was extended and enhanced by the work of Kalman, Bucy, and others [1]-[5].

Filters used for the above purpose can be fixed or adaptive. The design of fixed filters is based on prior knowledge of both the signal and the noise. Adaptive filters, on the other hand, have the ability to adjust their own parameters automatically, and their design requires little or no *a priori* knowledge of signal or noise characteristics.

Noise cancelling is a variation of optimal filtering that is highly advantageous in many applications. It makes use of an auxiliary or reference input derived from one or more sensors located at points in the noise field where the signal is weak or undetectable. This input is filtered and subtracted from a primary input containing both signal and noise. As a result the primary noise is attenuated or eliminated by cancellation.

At first glance, subtracting noise from a received signal would seem to be a dangerous procedure. If done improperly it could result in an increase in output noise power. If, however, filtering and subtraction are controlled by an appropriate adaptive process, noise reduction can be accomplished with little risk of distorting the signal or increasing the output noise level. In circumstances where adaptive noise cancelling is applicable, levels of noise rejection are often attainable that would be difficult or impossible to achieve by direct filtering.

The purpose of this paper is to describe the concept of adaptive noise cancelling, to provide a theoretical treatment of its advantages and limitations, and to describe some of the applications where it is most useful.

II. Early Work in Adaptive Noise Cancelling

The earliest work in adaptive noise cancelling known to the authors was performed by Howells and Applebaum and their colleagues at the General Electric Company between 1957 and 1960. They designed and built a system for antenna sidelobe cancelling that used a reference input derived from an auxiliary antenna and a simple two-weight adaptive filter [6].

At the time of this work, only a handful of people were interested in adaptive systems, and development of the multiweight adaptive filter was just beginning. In 1959, Widrow and Hoff at Stanford University were devising the least-mean-square (LMS) adaptive algorithm and the pattern recognition scheme known as Adaline (for "adaptive linear threshold logic element") [7], [8]. Rosenblatt had recently built his Perceptron at the Cornell Aeronautical Laboratory [9]-[11].[2] Aizermann and his colleagues at the Institute of Automatics and Telemechanics in Moscow, U.S.S.R., were constructing an automatic gradient searching machine. In Great Britain, D. Gabor and his associates were developing adaptive filters [12]. Each of these efforts was proceeding independently.

In the early and middle 1960's, work on adaptive systems intensified. Hundreds of papers on adaptation, adaptive controls, adaptive filtering, and adaptive signal processing appeared in the literature. The best known commercial application of adaptive filtering grew from the work during this period of Lucky at the Bell Laboratories [13], [14]. His high-speed MODEM's for digital communication are now widely used in connecting remote terminals to computers as well as one computer to another, allowing an increase in the rate and accuracy of data transmission by a reduction of intersymbol interference.

[2] This pioneering equipment now resides at the Smithsonian Institution in Washington, D.C.

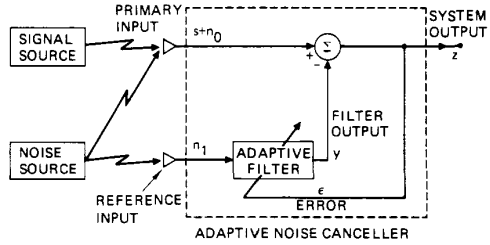

Fig. 1. The adaptive noise cancelling concept.

The first adaptive noise cancelling system at Stanford University was designed and built in 1965 by two students. Their work was undertaken as part of a term paper project for a course in adaptive systems given by the Electrical Engineering Department. The purpose was to cancel 60-Hz interference at the output of an electrocardiographic amplifier and recorder. A description of the system, which made use of a two-weight analog adaptive filter, together with results recently obtained by computer implementation, is presented in Section VIII.

Since 1965, adaptive noise cancelling has been successfully applied to a number of additional problems, including other aspects of electrocardiography, also described in Section VIII, to the elimination of periodic interference in general [15], and to the elimination of echoes on long-distance telephone transmission lines [16], [17]. A recent paper on adaptive antennas by Riegler and Compton [18] generalizes the work originally performed by Howells and Applebaum. Riegler and Compton's approach is based on the LMS algorithm and is an application of the adaptive antenna concepts of Widrow *et al.* [19], [20].

III. The Concept of Adaptive Noise Cancelling

Fig. 1 shows the basic problem and the adaptive noise cancelling solution to it. A signal s is transmitted over a channel to a sensor that also receives a noise n_0 uncorrelated with the signal. The combined signal and noise $s + n_0$ form the primary input to the canceller. A second sensor receives a noise n_1 uncorrelated with the signal but correlated in some unknown way with the noise n_0. This sensor provides the reference input to the canceller. The noise n_1 is filtered to produce an output y that is as close a replica as possible of n_0. This output is subtracted from the primary input $s + n_0$ to produce the system output $z = s + n_0 - y$.

If one knew the characteristics of the channels over which the noise was transmitted to the primary and reference sensors, it would theoretically be possible to design a fixed filter capable of changing n_1 into n_0. The filter output could then be subtracted from the primary input, and the system output would be signal alone. Since, however, the characteristics of the transmission paths are as a rule unknown or known only approximately and are seldom of a fixed nature, the use of a fixed filter is not feasible. Moreover, even if a fixed filter were feasible, its characteristics would have to be adjusted with a precision difficult to attain, and the slightest error could result in an increase in output noise power.

In the system shown in Fig. 1 the reference input is processed by an adaptive filter. An adaptive filter differs from a fixed filter in that it automatically adjusts its own impulse response. Adjustment is accomplished through an algorithm that responds to an error signal dependent, among other things, on the filter's output. Thus with the proper algorithm, the filter can operate under changing conditions and can readjust itself continuously to minimize the error signal.

The error signal used in an adaptive process depends on the nature of the application. In noise cancelling systems the practical objective is to produce a system output $z = s + n_0 - y$ that is a best fit in the least squares sense to the signal s. This objective is accomplished by feeding the system output back to the adaptive filter and adjusting the filter through an LMS adaptive algorithm to minimize total system output power.[3] In an adaptive noise cancelling system, in other words, the system output serves as the error signal for the adaptive process.

It might seem that some prior knowledge of the signal s or of the noises n_0 and n_1 would be necessary before the filter could be designed, or before it could adapt, to produce the noise cancelling signal y. A simple argument will show, however, that little or no prior knowledge of s, n_0, or n_1, or of their interrelationships, either statistical or deterministic, is required.

Assume that s, n_0, n_1, and y are statistically stationary and have zero means. Assume that s is uncorrelated with n_0 and n_1, and suppose that n_1 is correlated with n_0. The output z is

$$z = s + n_0 - y. \tag{1}$$

Squaring, one obtains

$$z^2 = s^2 + (n_0 - y)^2 + 2s(n_0 - y). \tag{2}$$

Taking expectations of both sides of (2), and realizing that s is uncorrelated with n_0 and with y, yields

$$\begin{aligned} E[z^2] &= E[s^2] + E[(n_0 - y)^2] + 2E[s(n_0 - y)] \\ &= E[s^2] + E[(n_0 - y)^2]. \end{aligned} \tag{3}$$

The signal power $E[s^2]$ will be unaffected as the filter is adjusted to minimize $E[z^2]$. Accordingly, the minimum output power is

$$\min E[z^2] = E[s^2] + \min E[(n_0 - y)^2]. \tag{4}$$

When the filter is adjusted so that $E[z^2]$ is minimized, $E[(n_0 - y)^2]$ is, therefore, also minimized. The filter output y is then a best least squares estimate of the primary noise n_0. Moreover, when $E[(n_0 - y)^2]$ is minimized, $E[(z - s)^2]$ is also minimized, since, from (1),

$$(z - s) = (n_0 - y). \tag{5}$$

Adjusting or adapting the filter to minimize the total output power is thus tantamount to causing the output z to be a best least squares estimate of the signal s for the given structure and adjustability of the adaptive filter and for the given reference input.

The output z will contain the signal s plus noise. From (1), the output noise is given by $(n_0 - y)$. Since minimizing $E[z^2]$ minimizes $E[(n_0 - y)^2]$, *minimizing the total output power minimizes the output noise power.* Since the signal in the output remains constant, *minimizing the total output power maximizes the output signal-to-noise ratio.*

It is seen from (3) that the smallest possible output power is $E[z^2] = E[s^2]$. When this is achievable, $E[(n_0 - y)^2] = 0$. Therefore, $y = n_0$, and $z = s$. In this case, minimizing output power causes the output signal to be perfectly noise free.[4]

These arguments can readily be extended to the case where the primary and reference inputs contain, in addition to n_0 and n_1, additive random noises uncorrelated with each other and with s, n_0, and n_1. They can also readily be extended to the case where n_0 and n_1 are deterministic rather than stochastic.

IV. Wiener Solutions to Statistical Noise Cancelling Problems

In this section, optimal unconstrained Wiener solutions to certain statistical noise cancelling problems are derived. The purpose is to demonstrate analytically the increase in signal-to-noise ratio and other advantages of the noise cancelling technique. Though the idealized solutions presented do not take into account the issues of finite filter length or causality, which are important in practical applications, means of approximating optimal unconstrained Wiener performance with physically realizable adaptive transversal filters are readily available and are described in Appendix B.

As previously noted, fixed filters are for the most part inapplicable in noise cancelling because the correlation and cross correlation functions of the primary and reference inputs are generally unknown and often variable with time. Adaptive filters are required to "learn" the statistics initially and to track them if they vary slowly. For stationary stochastic inputs, however, the steady-state performance of adaptive filters closely approximates that of fixed Wiener filters, and Wiener filter theory thus provides a convenient method of mathematically analyzing statistical noise cancelling problems.

Fig. 2 shows a classic single-input single-output Wiener filter. The input signal is x_j, the output signal y_j, and the desired response d_j. The input and output signals are assumed to be discrete in time, and the input signal and desired response are assumed to be statistically stationary. The error signal is $\epsilon_j = d_j - y_j$. The filter is linear, discrete, and designed to be optimal in the minimum mean-square-error sense. It is composed of an infinitely long, two-sided tapped delay line.

The optimal impulse response of this filter may be described in the following manner. The discrete autocorrelation function of the input signal x_j is defined as

$$\phi_{xx}(k) \triangleq E[x_j x_{j+k}]. \tag{6}$$

The cross-correlation function between x_j and the desired response d_j is similarly defined as

$$\phi_{xd}(k) \triangleq E[x_j d_{j+k}]. \tag{7}$$

The optimal impulse response $w^*(k)$ can then be obtained from the discrete Wiener–Hopf equation:

$$\sum_{l=-\infty}^{\infty} w^*(l)\, \phi_{xx}(k - l) = \phi_{xd}(k). \tag{8}$$

[3] The characteristics and terminology of the LMS adaptive filter used in the noise cancelling systems described in this paper are presented in Appendix A.

[4] Note that, on the other hand, when the reference input is completely uncorrelated with the primary input, the filter will "turn itself off" and will not increase output noise. In this case the filter output y will be uncorrelated with the primary input. The output power will be $E[z^2] = E[(s + n_0)^2] + 2E[-y(s + n_0)] + E[y^2] = E[(s + n_0)^2] + E[y^2]$. Minimizing output power requires that $E[y^2]$ be minimized, which is accomplished by making all weights zero, bringing $E[y^2]$ to zero.

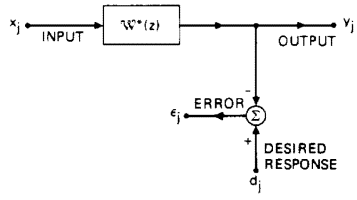

Fig. 2. Single-channel Wiener filter.

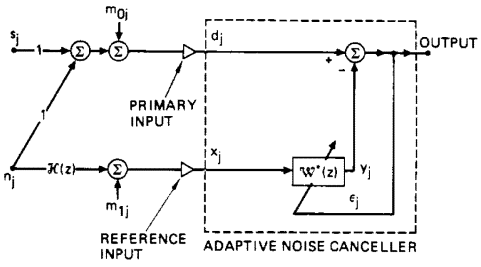

Fig. 3. Single-channel adaptive noise canceller with correlated and uncorrelated noises in the primary and reference inputs.

The convolution can be more simply written as

$$w^*(k) * \phi_{xx}(k) = \phi_{xd}(k). \tag{9}$$

This form of the Wiener solution is unconstrained in that the impulse response $w^*(k)$ may be causal or noncausal and of finite or infinite extent to the left or right of the time origin.[5]

The transfer function of the Wiener filter may now be derived as follows. The power-density spectrum of the input signal is the Z transform of $\phi_{xx}(k)$:

$$\mathcal{S}_{xx}(z) \triangleq \sum_{k=-\infty}^{\infty} \phi_{xx}(k)\, z^{-k}. \tag{10}$$

The cross power spectrum between the input signal and desired response is

$$\mathcal{S}_{xd}(z) \triangleq \sum_{k=-\infty}^{\infty} \phi_{xd}(k)\, z^{-k}. \tag{11}$$

The transfer function of the Wiener filter is

$$\mathfrak{W}^*(z) \triangleq \sum w^*(k)\, z^{-k}. \tag{12}$$

Transforming (8) then yields the optimal unconstrained Wiener transfer function:

$$\mathfrak{W}^*(z) = \frac{\mathcal{S}_{xd}(z)}{\mathcal{S}_{xx}(z)}. \tag{13}$$

The application of Wiener filter theory to adaptive noise cancelling may now be considered. Fig. 3 shows a single-channel adaptive noise canceller with a typical set of inputs. The primary input consists of a signal s_j plus a sum of two noises m_{0j} and n_j. The reference input consists of a sum of two other noises m_{1j} and $n_j * h(j)$, where $h(j)$ is the impulse response of the channel whose transfer function is $\mathcal{H}(z)$.[6] The noises n_j and $n_j * h(j)$ have a common origin, are correlated with each other, and are uncorrelated with s_j. They further are assumed to have a finite power spectrum at all frequencies. The noises m_{0j} and m_{1j} are uncorrelated with each other, with s_j, and with n_j and $n_j * h(j)$. For the purposes of analysis all noise propagation paths are assumed to be equivalent to linear, time-invariant filters.

The noise canceller of Fig. 3 includes an adaptive filter whose input x_j, the reference input to the canceller, is $m_{1j} + n_j * h(j)$ and whose desired response d_j, the primary input to the canceller, is $s_j + m_{0j} + n_j$. The error signal ϵ_j is the noise canceller's output. If one assumes that the adaptive process has converged and the minimum mean-square-error solution has been found, then the adaptive filter is equivalent to a Wiener filter. The optimal unconstrained transfer function of the adaptive filter is thus given by (13) and may be written as follows.

The spectrum of the filter's input $\mathcal{S}_{xx}(z)$ can be expressed in terms of the spectra of its two mutually uncorrelated additive components. The spectrum of the noise m_1 is $\mathcal{S}_{m_1m_1}(z)$, and that of the noise n arriving via $\mathcal{H}(z)$ is $\mathcal{S}_{nn}(z)|\mathcal{H}(z)|^2$. The filter's input spectrum is thus

$$\mathcal{S}_{xx}(z) = \mathcal{S}_{m_1m_1}(z) + \mathcal{S}_{nn}(z)|\mathcal{H}(z)|^2. \tag{14}$$

The cross power spectrum between the filter's input and the desired response depends only on the mutually correlated primary and reference components and is given by

$$\mathcal{S}_{xd}(z) = \mathcal{S}_{nn}(z)\,\mathcal{H}(z^{-1}). \tag{15}$$

The Wiener transfer function is thus

$$\mathfrak{W}^*(z) = \frac{\mathcal{S}_{nn}(z)\,\mathcal{H}(z^{-1})}{\mathcal{S}_{m_1m_1}(z) + \mathcal{S}_{nn}(z)|\mathcal{H}(z)|^2}. \tag{16}$$

Note that $\mathfrak{W}^*(z)$ is independent of the primary signal spectrum $\mathcal{S}_{ss}(z)$ and of the primary uncorrelated noise spectrum $\mathcal{S}_{m_0m_0}(z)$.

An interesting special case occurs when the additive noise m_1 in the reference input is zero. Then $\mathcal{S}_{m_1m_1}(z)$ is zero and the optimal transfer function (16) becomes

$$\mathfrak{W}^*(z) = 1/\mathcal{H}(z). \tag{17}$$

This result is intuitively appealing. The adaptive filter, as in the balancing of a bridge, causes the noise n_j to be perfectly nulled at the noise canceller output. The primary uncorrelated noise m_{0j} remains uncancelled.

The performance of the single-channel noise canceller can be evaluated more generally in terms of the ratio of the signal-to-noise density ratio at the output, $\rho_{\text{out}}(z)$ to the signal-to-noise density ratio at the primary input $\rho_{\text{pri}}(z)$.[7] Assuming that the signal spectrum is greater than zero at all frequencies and

[5] The Shannon–Bode realization of the Wiener solution, by contrast, is constrained to a causal response. This constraint generally leads to a loss of performance and, as shown in Appendix B, can normally be avoided in adaptive noise cancelling applications.

[6] To simplify the notation the transfer function of the noise path from n_j to the primary input has been set at unity. This procedure does not restrict the analysis, since by a suitable choice of $\mathcal{H}(z)$ and of statistics for n_j any combination of mutually correlated noises can be made to appear at the primary and reference inputs. Though $\mathcal{H}(z)$ may consequently be required to have poles inside and outside the unit circle in the Z-plane, a stable two-sided impulse response $h(j)$ will always exist.

[7] Signal-to-noise density ratio is here defined as the ratio of signal power density to noise power density and is thus a function of frequency.

factoring out the signal power spectrum yields

$$\frac{\rho_{\text{out}}(z)}{\rho_{\text{pri}}(z)} = \frac{\text{primary noise power spectrum}}{\text{output noise power spectrum}} = \frac{\mathcal{S}_{nn}(z) + \mathcal{S}_{m_0 m_0}(z)}{\mathcal{S}_{\text{output noise}}(z)}. \tag{18}$$

The canceller's output noise power spectrum, as may be seen from Fig. 3, is a sum of three components, one due to the propagation of m_{0j} directly to the output, another due to the propagation of m_{1j} to the output via the transfer function $-\mathfrak{W}^*(z)$, and another due to the propagation of n_j to the output via the transfer function $1 - \mathcal{H}(z)\,\mathfrak{W}^*(z)$. The output noise power spectrum is thus

$$\mathcal{S}_{\text{output noise}}(z) = \mathcal{S}_{m_0 m_0}(z) + \mathcal{S}_{m_1 m_1}(z)\,|\mathfrak{W}^*(z)|^2 + \mathcal{S}_{nn}(z)\,|[1 - \mathcal{H}(z)\,\mathfrak{W}^*(z)]|^2. \tag{19}$$

If one lets the ratios of the spectra of the uncorrelated to the spectra of the correlated noises ("noise-to-noise density ratios") at the primary and reference inputs now be defined as

$$A(z) \triangleq \frac{\mathcal{S}_{m_0 m_0}(z)}{\mathcal{S}_{nn}(z)} \tag{20}$$

and

$$B(z) \triangleq \frac{\mathcal{S}_{m_1 m_1}(z)}{\mathcal{S}_{nn}(z)\,|\mathcal{H}(z)|^2} \tag{21}$$

then the transfer function (17) can be written as

$$\mathfrak{W}^*(z) = \frac{1}{\mathcal{H}(z)[B(z) + 1]}. \tag{22}$$

The output noise power spectrum (19) can accordingly be rewritten as

$$\begin{aligned}\mathcal{S}_{\text{output noise}}(z) &= \mathcal{S}_{m_0 m_0}(z) + \frac{\mathcal{S}_{m_1 m_1}(z)}{|\mathcal{H}(z)|^2\,|B(z) + 1|^2} + \mathcal{S}_{nn}(z)\left|1 - \frac{1}{B(z) + 1}\right|^2 \\ &= \mathcal{S}_{nn}(z)\,A(z) + \mathcal{S}_{nn}(z)\,\frac{B(z)}{B(z) + 1}.\end{aligned} \tag{23}$$

The ratio of the output to the primary input noise power spectra is

$$\begin{aligned}\frac{\rho_{\text{out}}(z)}{\rho_{\text{pri}}(z)} &= \frac{\mathcal{S}_{nn}(z)[1 + A(z)]}{\mathcal{S}_{\text{output noise}}(z)} \\ &= \frac{1 + A(z)}{A(z) + B(z)/(B(z) + 1)} \\ &= \frac{[A(z) + 1][B(z) + 1]}{A(z) + A(z)\,B(z) + B(z)}.\end{aligned} \tag{24}$$

This expression is a general representation of ideal noise canceller performance with single primary and reference inputs and stationary signals and noises. It allows one to estimate the level of noise reduction to be expected with an ideal noise cancelling system. In such a system the signal propagates to the output in an undistorted fashion (with a transfer function of unity).[8] Classical configurations of Wiener, Kalman, and adaptive filters, in contrast, generally introduce some signal distortion in the process of noise reduction.

It is apparent from (24) that the ability of a noise cancelling system to reduce noise is limited by the uncorrelated-to-correlated noise density ratios at the primary and reference inputs. The smaller are $A(z)$ and $B(z)$, the greater will be $\rho_{\text{out}}(z)/\rho_{\text{pri}}(z)$ and the more effective the action of the canceller. The desirability of low levels of uncorrelated noise in both inputs is made still more evident by considering the following special cases.

1) Small $A(z)$:

$$\frac{\rho_{\text{out}}(z)}{\rho_{\text{pri}}(z)} \cong \frac{1 + B(z)}{B(z)}. \tag{25}$$

2) Small $B(z)$:

$$\frac{\rho_{\text{out}}(z)}{\rho_{\text{pri}}(z)} \cong \frac{1 + A(z)}{A(z)}. \tag{26}$$

3) Small $A(z)$ and $B(z)$:

$$\frac{\rho_{\text{out}}(z)}{\rho_{\text{pri}}(z)} \cong \frac{1}{A(z) + B(z)}. \tag{27}$$

Infinite improvement is implied by these relations when both $A(z)$ and $B(z)$ are zero. In this case there is complete removal of noise at the system output, resulting in perfect signal reproduction. When $A(z)$ and $B(z)$ are small, however, other factors become important in limiting system performance. These factors include the finite length of the adaptive filter in practical systems, discussed in Appendix B, and "misadjustment" caused by gradient estimation noise in the adaptive process, discussed in [19] and [20]. A third factor, signal components sometimes present in the reference input, is discussed in the following section.

V. Effect of Signal Components in the Reference Input

In certain instances the available reference input to an adaptive noise canceller may contain low-level signal components in addition to the usual correlated and uncorrelated noise components. There is no doubt that these signal components will cause some cancellation of the primary input signal. The question is whether they will cause sufficient cancellation to render the application of noise cancelling useless. An answer is provided in the present section through a quantitative analysis based, like that of the previous section, on unconstrained Wiener filter theory. In this analysis expressions are derived for signal-to-noise density ratio, signal distortion, and noise spectrum at the canceller output.

Fig. 4 shows an adaptive noise canceller whose reference input contains signal components and whose primary and reference inputs contain additive correlated noises. Additive uncorrelated noises have been omitted to simplify the analysis. The signal components in the reference input are assumed to be propagated through a channel with the transfer function $\mathcal{J}(z)$. The other terminology is the same as that of Fig. 3.

[8] Some signal cancellation is possible when adaptation is rapid (that is, when the value of the adaptation constant μ, defined in Appendix A, is large) because of the dynamic response of the weight vector, which approaches but does not equal the Wiener solution. In most cases this effect is negligible; a particular case where it is not negligible is described in Section VI.

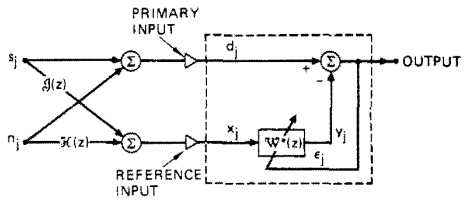

Fig. 4. Adaptive noise canceller with signal components in the reference input.

The spectrum of the signal in Fig. 4 is $\mathcal{S}_{ss}(z)$ and that of the noise $\mathcal{S}_{nn}(z)$. The spectrum of the reference input, which is identical to the spectrum of the input x_j to the adaptive filter, is thus

$$\mathcal{S}_{xx}(z) = \mathcal{S}_{ss}(z)|\mathcal{G}(z)|^2 + \mathcal{S}_{nn}(z)|\mathcal{H}(z)|^2. \tag{28}$$

The cross spectrum between the reference and primary inputs, identical to the cross spectrum between the filter's input x_j and desired response d_j, is similarly

$$\mathcal{S}_{xd}(z) = \mathcal{S}_{ss}(z)\,\mathcal{G}(z^{-1}) + \mathcal{S}_{nn}(z)\,\mathcal{H}(z^{-1}). \tag{29}$$

When the adaptive process has converged, the unconstrained Wiener transfer function of the adaptive filter, given by (13), is thus

$$\mathcal{W}^*(z) = \frac{\mathcal{S}_{ss}(z)\,\mathcal{G}(z^{-1}) + \mathcal{S}_{nn}(z)\,\mathcal{H}(z^{-1})}{\mathcal{S}_{ss}(z)|\mathcal{G}(z)|^2 + \mathcal{S}_{nn}(z)|\mathcal{H}(z)|^2}. \tag{30}$$

The first objective of the analysis is to find the signal-to-noise density ratio $\rho_{\text{out}}(z)$ at the noise canceller output. The transfer function of the propagation path from the signal input to the noise canceller output is $1 - \mathcal{G}(z)\,\mathcal{W}^*(z)$ and that of the path from the noise input to the canceller output is $1 - \mathcal{H}(z)\cdot\mathcal{W}^*(z)$. The spectrum of the signal component in the output is thus

$$\begin{aligned}\mathcal{S}_{ss_{\text{out}}}(z) &= \mathcal{S}_{ss}(z)|1 - \mathcal{G}(z)\,\mathcal{W}^*(z)|^2 \\ &= \mathcal{S}_{ss}(z)\left|\frac{[\mathcal{H}(z) - \mathcal{G}(z)]\,\mathcal{S}_{nn}(z)\,\mathcal{H}(z^{-1})}{\mathcal{S}_{ss}(z)|\mathcal{G}(z)|^2 + \mathcal{S}_{nn}(z)|\mathcal{H}(z)|^2}\right|^2\end{aligned} \tag{31}$$

and that of the noise component is similarly

$$\begin{aligned}\mathcal{S}_{nn_{\text{out}}}(z) &= \mathcal{S}_{nn}(z)|1 - \mathcal{H}(z)\,\mathcal{W}^*(z)|^2 \\ &= \mathcal{S}_{nn}(z)\left|\frac{[\mathcal{G}(z) - \mathcal{H}(z)]\,\mathcal{S}_{ss}(z)\,\mathcal{G}(z^{-1})}{\mathcal{S}_{ss}(z)|\mathcal{G}(z)|^2 + \mathcal{S}_{nn}(z)|\mathcal{H}(z)|^2}\right|^2.\end{aligned} \tag{32}$$

The output signal-to-noise density ratio is thus

$$\begin{aligned}\rho_{\text{out}}(z) &= \frac{\mathcal{S}_{ss}(z)}{\mathcal{S}_{nn}(z)}\left|\frac{\mathcal{S}_{nn}(z)\,\mathcal{H}(z^{-1})}{\mathcal{S}_{ss}(z)\,\mathcal{G}(z^{-1})}\right|^2 \\ &= \frac{\mathcal{S}_{nn}(z)|\mathcal{H}(z)|^2}{\mathcal{S}_{ss}(z)|\mathcal{G}(z)|^2}.\end{aligned} \tag{33}$$

The output signal-to-noise density ratio can be conveniently expressed in terms of the signal-to-noise density ratio at the reference input $\rho_{\text{ref}}(z)$ as follows. The spectrum of the signal component in the reference input is

$$\mathcal{S}_{ss_{\text{ref}}}(z) = \mathcal{S}_{ss}(z)|\mathcal{G}(z)|^2 \tag{34}$$

and that of the noise component is similarly

$$\mathcal{S}_{nn_{\text{ref}}}(z) = \mathcal{S}_{nn}(z)|\mathcal{H}(z)|^2. \tag{35}$$

The signal-to-noise density ratio at the reference input is thus

$$\rho_{\text{ref}}(z) = \frac{\mathcal{S}_{ss}(z)|\mathcal{G}(z)|^2}{\mathcal{S}_{nn}(z)|\mathcal{H}(z)|^2}. \tag{36}$$

The output signal-to-noise density ratio (33) is, therefore,

$$\rho_{\text{out}}(z) = \frac{1}{\rho_{\text{ref}}(z)}. \tag{37}$$

This result is exact and somewhat surprising. It shows that, assuming the adaptive solution to be unconstrained and the noises in the primary and reference inputs to be mutually correlated, the signal-to-noise density ratio at the noise canceller output is simply the reciprocal at all frequencies of the signal-to-noise density ratio at the reference input.

The next objective of the analysis is to derive an expression for signal distortion at the noise canceller output. The most useful reference input is one composed almost entirely of noise correlated with the noise in the primary input. When signal components are present some signal distortion will generally occur. The amount will depend on the amount of signal propagated through the adaptive filter, which may be determined as follows. The transfer function of the propagation path through the filter is

$$-\mathcal{G}(z)\,\mathcal{W}^*(z) = -\mathcal{G}(z)\frac{\mathcal{S}_{ss}(z)\,\mathcal{G}(z^{-1}) + \mathcal{S}_{nn}(z)\,\mathcal{H}(z^{-1})}{\mathcal{S}_{ss}(z)|\mathcal{G}(z)|^2 + \mathcal{S}_{nn}(z)|\mathcal{H}(z)|^2}. \tag{38}$$

When $|\mathcal{G}(z)|$ is small, this function can be approximated as

$$-\mathcal{G}(z)\,\mathcal{W}^*(z) \cong -\mathcal{G}(z)/\mathcal{H}(z). \tag{39}$$

The spectrum of the signal component propagated to the noise canceller output through the adaptive filter is thus approximately

$$\mathcal{S}_{ss}(z)|\mathcal{G}(z)/\mathcal{H}(z)|^2. \tag{40}$$

The combining of this component with the signal component in the primary input involves complex addition and is the process that results in signal distortion. The worst case, bounding the distortion to be expected in practice, occurs when the two signal components are of opposite phase.

Let "signal distortion" $\mathcal{D}(z)$ be defined[9] as a dimensionless ratio of the spectrum of the output signal component propagated through the adaptive filter to the spectrum of the signal component at the primary input:

$$\begin{aligned}\mathcal{D}(z) &\triangleq \frac{\mathcal{S}_{ss}(z)|\mathcal{G}(z)\,\mathcal{W}^*(z)|^2}{\mathcal{S}_{ss}(z)} \\ &= |\mathcal{G}(z)\,\mathcal{W}^*(z)|^2.\end{aligned} \tag{41}$$

From (39) it can be seen that, when $\mathcal{G}(z)$ is small, (41) reduces to

$$\mathcal{D}(z) \cong |\mathcal{G}(z)/\mathcal{H}(z)|^2. \tag{42}$$

This expression may be rewritten in a more useful form by combining the expressions for the signal-to-noise density ratio at the primary input:

$$\rho_{\text{pri}}(z) \triangleq \mathcal{S}_{ss}(z)/\mathcal{S}_{nn}(z) \tag{43}$$

[9] Note that signal distortion as defined here is a linear phenomenon related to alteration of the signal waveform as it appears at the noise canceller output and is not to be confused with nonlinear harmonic distortion.

and the signal-to-noise density ratio at the reference input (36):

$$\mathfrak{D}(z) \cong \rho_{\text{ref}}(z)/\rho_{\text{pri}}(z). \tag{44}$$

Equation (44) shows that, with an unconstrained adaptive solution and mutually correlated noises at the primary and reference inputs, low signal distortion results from a high signal-to-noise density ratio at the primary input and a low signal-to-noise density ratio at the reference input. This conclusion is intuitively reasonable.

The final objective of the analysis is to derive an expression for the spectrum of the output noise. The noise n_j propagates to the output with a transfer function

$$1 - \mathcal{H}(z)\,\mathfrak{W}^*(z) = 1 - \mathcal{H}(z)\left[\frac{\mathcal{S}_{ss}(z)\,\mathcal{J}(z^{-1}) + \mathcal{S}_{nn}(z)\,\mathcal{H}(z^{-1})}{\mathcal{S}_{ss}(z)\,|\mathcal{J}(z)|^2 + \mathcal{S}_{nn}(z)\,|\mathcal{H}(z)|^2}\right]$$

$$= \frac{\mathcal{S}_{ss}(z)\,\mathcal{J}(z^{-1})[\mathcal{J}(z) - \mathcal{H}(z)]}{\mathcal{S}_{ss}(z)\,|\mathcal{J}(z)|^2 + \mathcal{S}_{nn}(z)\,|\mathcal{H}(z)|^2}. \tag{45}$$

When $|\mathcal{J}(z)|$ is small, (45) reduces to

$$1 - \mathcal{H}(z)\,\mathfrak{W}^*(z) \cong \frac{-\mathcal{S}_{ss}(z)\,\mathcal{J}(z^{-1})}{\mathcal{S}_{nn}(z)\,\mathcal{H}(z^{-1})}. \tag{46}$$

The output noise spectrum is

$$\mathcal{S}_{\text{output noise}} = \mathcal{S}_{nn}(z)\,|1 - \mathcal{H}(z)\,\mathfrak{W}^*(z)|^2. \tag{47}$$

When $|\mathcal{J}(z)|$ is small, (47) similarly reduces to

$$\mathcal{S}_{\text{output noise}}(z) \cong \mathcal{S}_{nn}(z)\left|\frac{\mathcal{S}_{ss}(z)\,\mathcal{J}(z^{-1})}{\mathcal{S}_{nn}(z)\,\mathcal{H}(z^{-1})}\right|^2. \tag{48}$$

This equation can be more conveniently expressed in terms of the signal-to-noise density ratios at the reference input (36) and primary input (43):

$$\mathcal{S}_{\text{output noise}}(z) \cong \mathcal{S}_{nn}(z)\,|\rho_{\text{ref}}(z)|\,|\rho_{\text{pri}}(z)|. \tag{49}$$

This result, which may appear strange at first glance, can be understood intuitively as follows. The first factor implies that the output noise spectrum depends on the input noise spectrum and is readily accepted. The second factor implies that, if the signal-to-noise density ratio at the reference input is low, the output noise will be low; that is, the smaller the signal component in the reference input, the more perfectly the noise will be cancelled. The third factor implies that, if the signal-to-noise density ratio in the primary input (the desired response of the adaptive filter) is low, the filter will be trained most effectively to cancel the noise rather than the signal and consequently output noise will be low.

The above analysis shows that signal components of low signal-to-noise ratio in the reference input, though undesirable, do not render the application of adaptive noise cancelling useless.[10] For an illustration of the level of performance attainable in practical circumstances consider the following example. Fig. 5 shows an adaptive noise cancelling system designed to pass a plane-wave signal received in the main beam of an antenna array and to discriminate against strong interference in the near field or in a minor lobe of the array. If one assumes that the signal and interference have overlapping and similar power spectra and that the interference power density is twenty times greater than the signal power density at the individual array element, then the signal-to-noise ratio at the reference input ρ_{ref} is 1/20. If one further assumes that, because of array gain, the signal power equals the interference power at the array output, then the signal-to-noise ratio at the primary input ρ_{pri} is 1. After convergence of the adaptive filter the signal-to-noise ratio at the system output will thus be

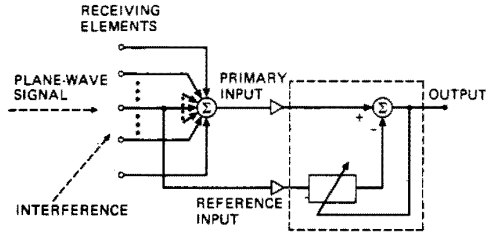

Fig. 5. Adaptive noise cancelling applied to a receiving array.

$$\rho_{\text{out}} = 1/\rho_{\text{ref}} = 20.$$

The maximum signal distortion will similarly be

$$\mathfrak{D} = \rho_{\text{ref}}/\rho_{\text{pri}} = (1/20)/1 = 5 \text{ percent}.$$

In this case, therefore, adaptive noise cancelling improves signal-to-noise ratio twentyfold and introduces only a small amount of signal distortion.

VI. The Adaptive Noise Canceller as a Notch Filter

In certain situations a primary input is available consisting of a signal component with an additive undesired sinusoidal interference. The conventional method of eliminating such interference is through the use of a notch filter. In this section an unusual form of notch filter, realized by an adaptive noise canceller, is described. The advantages of this form of notch filter are that it offers easy control of bandwidth, an infinite null, and the capability of adaptively tracking the exact frequency of the interference. The analysis presented deals with the formation of a notch at a single frequency. Analytical and experimental results show, however, that if more than one frequency is present in the reference input a notch for each will be formed [21].

Fig. 6 shows a single-frequency noise canceller with two adaptive weights. The primary input is assumed to be any kind of signal—stochastic, deterministic, periodic, transient, etc.—or any combination of signals. The reference input is assumed to be a pure cosine wave $C\cos(\omega_0 t + \phi)$. The primary and reference inputs are sampled at the frequency $\Omega = 2\pi/T$ rad/s. The reference input is sampled directly, giving x_{1j}, and after undergoing a 90° phase shift, giving x_{2j}^{Δ}. The samplers are synchronous and strobe at $t = 0, \pm T, \pm 2T$, etc.

A transfer function for the noise canceller of Fig. 6 may be obtained by analyzing signal propagation from the primary input to the system output.[11] For this purpose the flow diagram of Fig. 7, showing the operation of the LMS algorithm in detail, is constructed. Note that the procedure for updating

[10] It should be noted that if the reference input contained signal components but no noise components, correlated or uncorrelated, then the signal would be completely cancelled. When the reference input is properly derived, however, this condition cannot occur.

[11] It is not obvious, from inspection of Fig. 6, that a transfer function for this propagation path in fact exists. Its existence is shown, however, by the subsequent analysis.

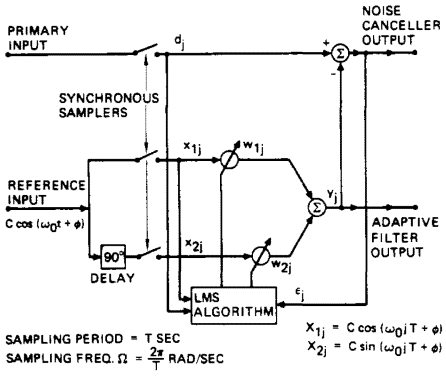

Fig. 6. Single-frequency adaptive noise canceller.

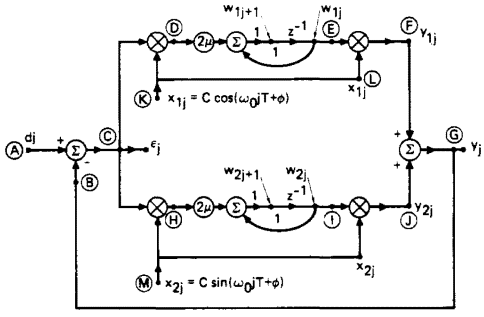

Fig. 7. Flow diagram showing signal propagation in single-frequency adaptive noise canceller.

the weights, as indicated in the diagram, is given by

$$w_{1j+1} = w_{1j} + 2\mu\epsilon_j x_{1j}$$
$$w_{2j+1} = w_{2j} + 2\mu\epsilon_j x_{2j}. \tag{50}$$

The sampled reference inputs are

$$x_{1j} = C\cos(\omega_0 jT + \phi) \tag{51}$$

and

$$x_{2j} = C\sin(\omega_0 jT + \phi). \tag{52}$$

The first step in the analysis is to obtain the isolated impulse response from the error ϵ_j, point C, to the filter output, point G, with the feedback loop from point G to point B broken. Let an impulse of amplitude α be applied at point C at discrete time $j = k$; that is,

$$\epsilon_j = \alpha\delta(j - k) \tag{53}$$

where

$$\delta(j - k) = \begin{cases} 1, & \text{for } j = k \\ 0, & \text{for } j \neq k. \end{cases} \tag{54}$$

The response at point D is then

$$\epsilon_j x_{1j} = \begin{cases} \alpha C\cos(\omega_0 kT + \phi), & \text{for } j = k \\ 0, & \text{for } j \neq k \end{cases} \tag{55}$$

which is the input impulse scaled in amplitude by the instantaneous value of x_{1j} at $j = k$. The signal flow path from point D to point E is that of a digital integrator with transfer function $2\mu/(z - 1)$ and impulse response $2\mu u(j - 1)$, where $u(j)$ is the discrete unit step function

$$u(j) = \begin{cases} 0, & \text{for } j < 0 \\ 1, & \text{for } j \geqslant 0. \end{cases} \tag{56}$$

Convolving $2\mu u(j - 1)$ with $\epsilon_j x_{1j}$ yields the response at point E:

$$w_{1j} = 2\mu\alpha C\cos(\omega_0 kT + \phi) \tag{57}$$

where $j \geqslant k + 1$. When the scaled and delayed step function is multiplied by x_{1j}, the response at point F is obtained:

$$y_{1j} = 2\mu\alpha C^2\cos(\omega_0 jT + \phi)\cos(\omega_0 kT + \phi) \tag{58}$$

where $j \geqslant k + 1$. The corresponding response at point J, obtained in a similar manner, is

$$y_{2j} = 2\mu\alpha C^2\sin(\omega_0 jT + \phi)\sin(\omega_0 kT + \phi) \tag{59}$$

where $j \geqslant k + 1$. Combining (58) and (59) yields the response at the filter output, point G:

$$\begin{aligned} y_j &= 2\mu\alpha C^2\cos\omega_0 T(j - k) \\ &= 2\mu\alpha C^2 u(j - k - 1)\cos\omega_0 T(j - k). \end{aligned} \tag{60}$$

Note that (60) is a function only of $(j - k)$ and is thus a time-invariant impulse response, proportional to the input impulse.

A linear transfer function for the noise canceller may now be derived in the following manner. If the time k is set equal to zero, the unit impulse response of the linear time-invariant signal-flow path from point C to point G is

$$y_j = 2\mu C^2 u(j - 1)\cos(\omega_0 jT) \tag{61}$$

and the transfer function of this path is

$$\begin{aligned} G(z) &= 2\mu C^2\left[\frac{z(z - \cos\omega_0 T)}{z^2 - 2z\cos\omega_0 T + 1} - 1\right] \\ &= \frac{2\mu C^2(z\cos\omega_0 T - 1)}{z^2 - 2z\cos\omega_0 T + 1}. \end{aligned} \tag{62}$$

This function can be expressed in terms of a radian sampling frequency $\Omega = 2\pi/T$ as

$$G(z) = \frac{2\mu C^2[z\cos(2\pi\omega_0\Omega^{-1}) - 1]}{z^2 - 2z\cos(2\pi\omega_0\Omega^{-1}) + 1}. \tag{63}$$

If the feedback loop from point G to point B is now closed, the transfer function $H(z)$ from the primary input, point A, to the noise canceller output, point C, can be obtained from the feedback formula:

$$H(z) = \frac{z^2 - 2z\cos(2\pi\omega_0\Omega^{-1}) + 1}{z^2 - 2(1 - \mu C^2)z\cos(2\pi\omega_0\Omega^{-1}) + 1 - 2\mu C^2}. \tag{64}$$

Equation (64) shows that the single-frequency noise canceller has the properties of a notch filter at the reference frequency ω_0. The zeros of the transfer function are located in the Z plane at

$$z = \exp(\pm i 2\pi\omega_0\Omega^{-1}) \tag{65}$$

and are precisely on the unit circle at angles of $\pm 2\pi\omega_0\Omega^{-1}$ rad. The poles are located at

$$\begin{aligned} z = (1 - \mu C^2)\cos(2\pi\omega_0\Omega^{-1}) \pm i[(1 - 2\mu C^2) \\ - (1 - \mu C^2)\cos^2(2\pi\omega_0\Omega^{-1})]^{1/2}. \end{aligned} \tag{66}$$

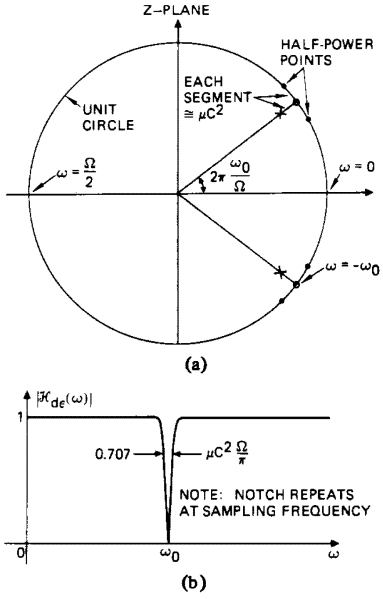

Fig. 8. Properties of transfer function of single-frequency adaptive noise canceller. (a) Location of poles and zeros. (b) Magnitude of transfer function.

The poles are inside the unit circle at a radial distance $(1 - 2\mu C^2)^{1/2}$, approximately equal to $1 - \mu C^2$, from the origin and at angles of

$$\pm \arccos\left[(1 - \mu C^2)(1 - 2\mu C^2)^{-1/2} \cos(2\pi\omega_0\Omega^{-1})\right].$$

For slow adaptation (that is, small values of μC^2) these angles depend on the factor

$$\frac{1 - \mu C^2}{(1 - 2\mu C^2)^{1/2}} = \left(\frac{1 - 2\mu C^2 + \mu^2 C^4}{1 - 2\mu C^2}\right)^{1/2}$$
$$\cong (1 - \mu^2 C^4 + \cdots)^{1/2}$$
$$\cong 1 - \tfrac{1}{2}\mu^2 C^4 + \cdots \tag{67}$$

which differs only slightly from a value of one. The result is that, in practical instances, the angles of the poles are almost identical to those of the zeros.

The location of the poles and zeros and the magnitude of the transfer function in terms of frequency are shown in Fig. 8. Since the zeros lie on the unit circle, the depth of the notch in the transfer function is infinite at the frequency $\omega = \omega_0$. The sharpness of the notch is determined by the closeness of the poles to the zeros. Corresponding poles and zeros are separated by a distance approximately equal to μC^2. The arc length along the unit circle (centered at the position of a zero) spanning the distance between half-power points is approximately $2\mu C^2$. This length corresponds to a notch bandwidth of

$$\mathrm{BW} = \mu C^2 \Omega/\pi. \tag{68}$$

The Q of the notch is determined by the ratio of the center frequency to the bandwidth:

$$Q \cong \frac{\omega_0 \pi}{\mu C^2 \Omega}. \tag{69}$$

The single-frequency noise canceller is, therefore, equivalent to a stable notch filter when the reference input is a pure cosine wave. The depth of the null achievable is generally superior to that of a fixed digital or analog filter because the adaptive process maintains the null exactly at the reference frequency.

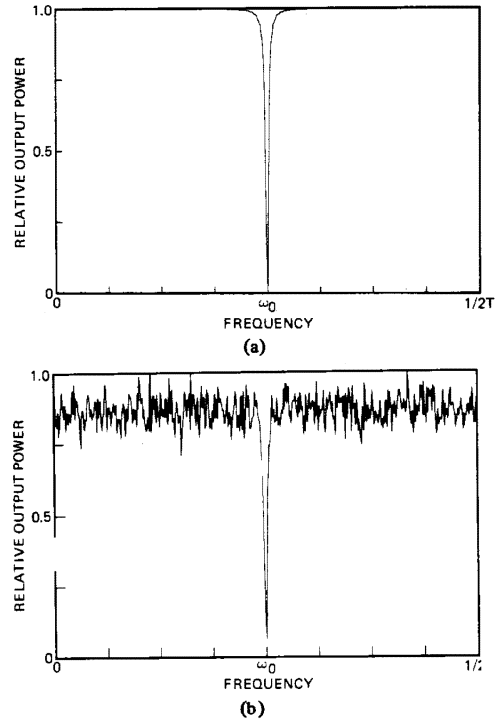

Fig. 9. Results of single-frequency adaptive noise cancelling experiments. (a) Primary input composed of cosine wave at 512 discrete frequencies. (b) Primary input composed of uncorrelated samples of white noise.

Fig. 9 shows the results of two experiments performed to demonstrate the characteristics of the adaptive notch filter. In the first the primary input was a cosine wave of unit power stepped at 512 discrete frequencies. The reference input was a cosine wave with a frequency ω_0 of $\pi/2T$ rad/s. The value of C was 1, and the value of μ was 1.25×10^{-2}. The frequency resolution of the fast Fourier transform was 512 bins. The output power at each frequency is shown in Fig. 9(a). As the primary frequency approaches the reference frequency, significant cancellation occurs. The weights do not converge to stable values but "tumble" at the difference frequency,[12] and the adaptive filter behaves like a modulator, converting the reference frequency into the primary frequency. The theoretical notch width between half-power points, $1.59 \times 10^{-2}\ \omega_0$, compares closely with the measured notch width of $1.62 \times 10^{-2}\ \omega_0$.

In the second experiment, the primary input was composed of uncorrelated samples of white noise of unit power. The reference input and the processing parameters were the same as in the first experiment. An ensemble average of 4096 power spectra at the noise canceller output is shown in Fig. 9(b). An infinite null was not obtained in this experiment because of the finite frequency resolution of the spectral analysis algorithm.

[12] When the primary and reference frequencies are held at a constant difference, the weights develop a sinusoidal steady state at the difference frequency. In other words, they converge on a dynamic rather than a static solution. This is an unusual form of adaptive behavior.

In these experiments the filtering of a reference cosine wave of a given frequency caused cancellation of primary input components at adjacent frequencies. This result indicates that, under some circumstances, primary input components may be partially cancelled and distorted even though the reference input is uncorrelated with them. In practice this kind of cancellation is of concern only when the adaptive process is rapid; that is, when it is effected with large values of μ. When the adaptive process is slow, the weights converge to values that are nearly stable, and though signal cancellation as described in this section occurs it is generally not significant.

Additional experiments have recently been conducted with reference inputs containing more than one sinusoid. The formation of multiple notches has been achieved by using an adaptive filter with multiple weights (typically an adaptive transversal filter). Two weights are required for each sinusoid to achieve the necessary filter gain and phase. Uncorrelated broad-band noise superposed on the reference input creates a need for additional weights. A full analysis of the multiple notch problem can be found in [21].

VII. The Adaptive Noise Canceller as a High-Pass Filter

The use of a bias weight in an adaptive filter to cancel low-frequency drift in the primary input is a special case of notch filtering with the notch at zero frequency. The method of incorporating the bias weight is shown in Appendix A. Because there is no need to match the phase of the signal, only one weight is needed. The reference input is set to a constant value of one.

The transfer function from the primary input to the noise canceller output is derived as follows. Applying equations (A.3) and (A.15) of Appendix A yields

$$y_j = w_j \cdot 1 = w_j \tag{70}$$

$$w_{j+1} = w_j + 2\mu(\epsilon_j x_j) \tag{71}$$

or

$$\begin{aligned} y_{j+1} &= y_j + 2\mu(d_j - y_j) \\ &= (1 - 2\mu)\, y_j + 2\mu d_j. \end{aligned} \tag{72}$$

Taking the Z transform of (72) yields the steady-state solution:

$$Y(z) = \frac{2\mu}{z - (1 - 2\mu)} D(z). \tag{73}$$

The transfer function is then obtained by substituting $E(z) = D(z) - Y(z)$ in (73):

$$D(z) - E(z) = \frac{2\mu}{z - (1 - 2\mu)} D(z) \tag{74}$$

which reduces to

$$H(z) = \frac{E(z)}{D(z)} = \frac{z - 1}{z - (1 - 2\mu)}. \tag{75}$$

Equation (75) shows that the bias-weight filter is a high-pass filter with a zero on the unit circle at zero frequency and a pole on the real axis at a distance 2μ to the left of the zero. Note that this corresponds to a single-frequency notch filter, described by (64), for the case where $\omega_0 = 0$ and $C = 1$. The half-power frequency of the notch is at $\mu\Omega/\pi$ rad/s.

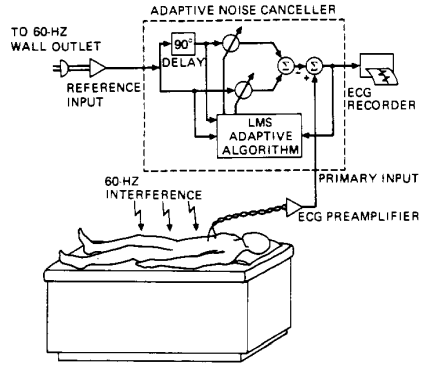

Fig. 10. Cancelling 60-Hz interference in electrocardiography.

The single-weight noise canceller acting as a high-pass filter is capable of removing not only a constant bias but also slowly varying drift in the primary input. Moreover, though it is not demonstrated in this paper, experience has shown that bias or drift removal can be accomplished simultaneously with cancellation of periodic or stochastic interference.

VIII. Applications

The principles of adaptive noise cancelling, including a description of the concept and theoretical analyses of performance with various kinds of signal and noise, have been presented in the preceding pages. This section describes a variety of practical applications of the technique. These applications include the cancelling of several kinds of interference in electrocardiography, of noise in speech signals, of antenna sidelobe interference, and of periodic or broad-band interference for which there is no external reference source. Experimental results are presented that demonstrate the performance of adaptive noise cancelling in these applications and that show its potential value whenever suitable inputs are available.

A. Cancelling 60-Hz Interference in Electrocardiography

In a recent paper [22], the authors point out that a major problem in the recording of electrocardiograms (ECG's) is "the appearance of unwanted 60-Hz interference in the output." They analyze the various causes of such power-line interference, including magnetic induction, displacement currents in leads or in the body of the patient, and equipment interconnections and imperfections. They also describe a number of techniques that are useful for minimizing it and that can be effected in the recording process itself, such as proper grounding and the use of twisted pairs. Another method capable of reducing 60-Hz ECG interference is adaptive noise cancelling, which can be used separately or in conjunction with more conventional approaches.

Fig. 10 shows the application of adaptive noise cancelling in electrocardiography. The primary input is taken from the ECG preamplifier; the 60-Hz reference input is taken from a wall outlet. The adaptive filter contains two variable weights, one applied to the reference input directly and the other to a version of it shifted in phase by 90°. The two weighted versions of the reference are summed to form the filter's output, which is subtracted from the primary input. Selected combinations of the values of the weights allow the reference waveform to be changed in magnitude and phase in any way required for cancellation. The two variable weights, or two

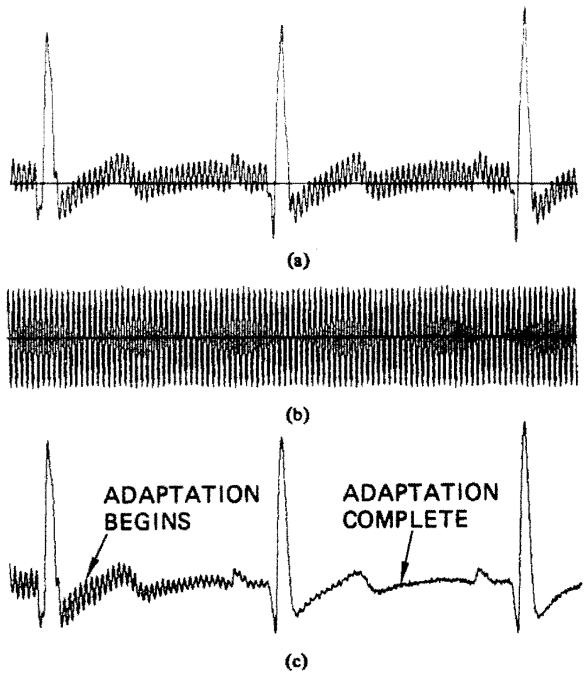

Fig. 11. Result of electrocardiographic noise cancelling experiment. (a) Primary input. (b) Reference input. (c) Noise canceller output.

"degrees of freedom," are required to cancel the single pure sinusoid.

A typical result of a group of experiments performed with a real-time computer system is shown in Fig. 11. Sample size was 10 bits and sampling rate 1000 Hz. Fig. 11(a) shows the primary input, an electrocardiographic waveform with an excessive amount of 60-Hz interference, and Fig. 11(b) shows the reference input from the wall outlet. Fig. 11(c) is the noise canceller output. Note the absence of interference and the clarity of detail once the adaptive process has converged.

B. Cancelling the Donor ECG in Heart-Transplant Electrocardiography

The electrical depolarization of the ventricles of the human heart is triggered by a group of specialized muscle cells known as the atrioventricular (AV) node. Though capable of independent, asynchronous operation, this node is normally controlled by a similar complex, the sinoatrial (SA) node, whose depolarization initiates an electrical impulse transmitted by conduction through the atrial heart muscle to the AV node. The SA node is connected through the vagus and sympathetic nerves to the central nervous system, which by controlling the rate of depolarization controls the frequency of the heartbeat [23], [24].

The cardiac transplantation technique developed by Shumway of the Stanford University Medical Center involves the suturing of the "new" or donor heart to a portion of the atrium of the patient's "old" heart [25]. Scar tissue forms at the suture line and electrically isolates the small remnant of the old heart, containing only the SA node, from the new heart, containing both SA and AV nodes. The SA node of the old heart remains connected to the vagus and sympathetic nerves, and the old heart continues to beat at a rate controlled by the central nervous system. The SA node of the new heart, which is not connected to the central nervous system because the severed vagus nerve cannot be surgically reattached, generates a spontaneous pulse that causes the new heart to beat at a separate self-pacing rate.

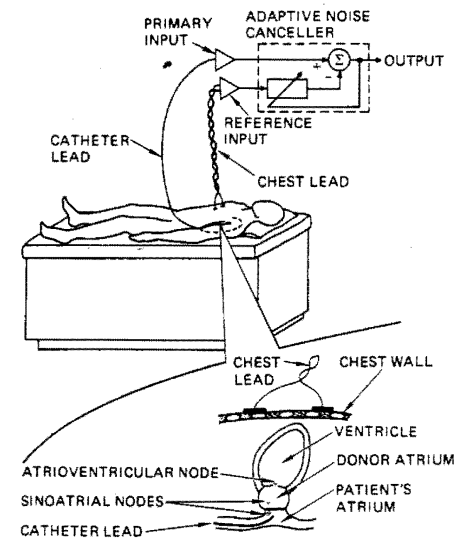

Fig. 12. Deriving and processing ECG signals of a heart-transplant patient.

It is of interest to cardiac transplant research, and to cardiac research in general, to be able to determine the firing rate of the old heart and, indeed, to be able to see the waveforms of its electrical output. These waveforms, which cannot be obtained by ordinary electrocardiographic means because of interference from the beating of the new heart, are readily obtained with adaptive noise cancelling.

Fig. 12 shows the method of applying adaptive noise cancelling in heart-transplant electrocardiography. The reference input is provided by a pair of ordinary chest leads. These leads receive a signal that comes essentially from the new heart, the source of interference. The primary input is provided by a catheter consisting of a small coaxial cable threaded through the left brachial vein and the vena cava to a position in the atrium of the old heart. The tip of the catheter, a few millimeters long, is an exposed portion of the center conductor that acts as an antenna and is capable of receiving cardiac electrical signals. When it is in the most favorable position, the desired signal from the old heart and the interference from the new heart are received in about equal proportion.

Fig. 13 shows typical reference and primary inputs and the corresponding noise canceller output. The reference input contains the strong QRS waves that, in a normal electrocardiogram, indicate the firing of the ventricles. The primary input contains pulses that are synchronous with the QRS waves of the reference input and indicate the beating of the new heart. The other waves seen in this input are due to the old heart, which is beating at a separate rate. When the reference input is adaptively filtered and subtracted from the primary input, one obtains the waveform shown in Fig. 13(c), which is that of the old heart together with very weak residual pulses originating in the new heart. Note that the pulses of the two hearts are easily separated, even when they occur at the same instant. Note also that the electrical waveform of the new heart is steady and precise, while that of the old heart varies significantly from beat to beat.

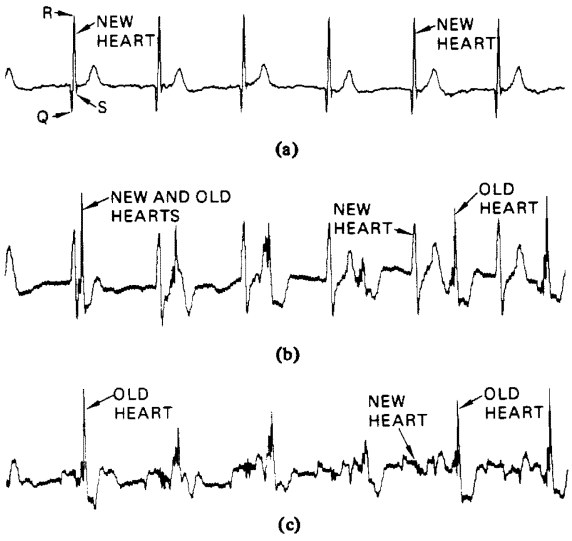

Fig. 13. ECG waveforms of heart-transplant patient. (a) Reference input (new heart). (b) Primary input (new and old heart). (c) Noise canceller output (old heart).

For this experiment the noise canceller was implemented in software with an adaptive transversal filter containing 48 weights. Sampling rate was 500 Hz.

C. *Cancelling the Maternal ECG in Fetal Electrocardiography*

Abdominal electrocardiograms make it possible to determine fetal heart rate and to detect multiple fetuses and are often used during labor and delivery [26]-[28]. Background noise due to muscle activity and fetal motion, however, often has an amplitude equal to or greater than that of the fetal heartbeat [29]-[31]. A still more serious problem is the mother's heartbeat, which has an amplitude two to ten times greater than that of the fetal heartbeat and often interferes with its recording [32].

In the spring of 1972, a group of experiments was performed to demonstrate the usefulness of adaptive noise cancelling in fetal electrocardiography. The objective was to derive as clear a fetal ECG as possible, so that not only could the heart rate be observed but also the actual waveform of the electrical output. The work was performed by Marie-France Ravat, Dominique Biard, Denys Caraux, and Michel Cotton, at the time students at Stanford University.[13]

Four ordinary chest leads were used to record the mother's heartbeat and provide multiple reference inputs to the canceller.[14] A single abdominal lead was used to record the combined maternal and fetal heartbeats that served as the primary input. Fig. 14 shows the cardiac electric field vectors of mother and fetus and the positions in which the leads were placed. Each lead terminated in a pair of electrodes. The chest and abdominal inputs were prefiltered, digitized, and recorded on tape. A multichannel adaptive noise canceller,

[13] A similar attempt to cancel the maternal heartbeat had previously been made by Walden and Bimbaum [33] without the use of an adaptive processor. Some reduction of the maternal interference was achieved by the careful placement of leads and adjustment of amplifier gain. It appears that substantially better results can be obtained with adaptive processing.

[14] More than one reference input was used to make the interference filtering task easier. The number of reference inputs required essentially to eliminate the maternal ECG is still under investigation.

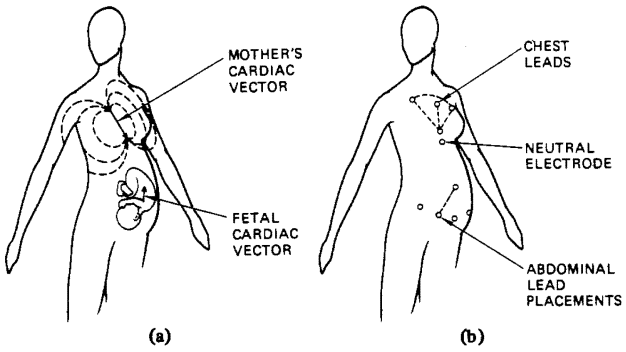

Fig. 14. Cancelling maternal heartbeat in fetal electrocardiography. (a) Cardiac electric field vectors of mother and fetus. (b) Placement of leads.

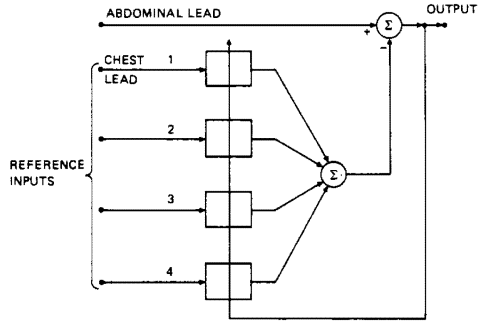

Fig. 15. Multiple-reference noise canceller used in fetal ECG experiment.

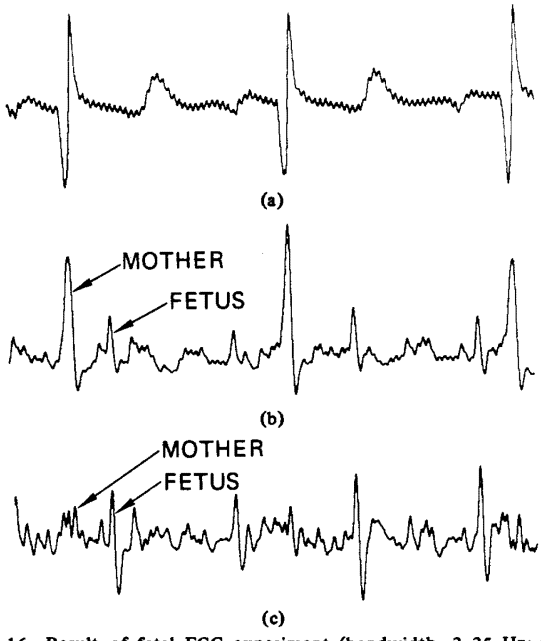

Fig. 16. Result of fetal ECG experiment (bandwidth, 3-35 Hz; sampling rate, 256 Hz). (a) Reference input (chest lead). (b) Primary input (abdominal lead). (c) Noise canceller output.

shown in Fig. 15 and described theoretically in Appendix C, was used. Each channel had 32 taps with nonuniform (log periodic) spacing and a total delay of 129 ms.

Fig. 16 shows typical reference and primary inputs together with the corresponding noise canceller output. The prefilter-

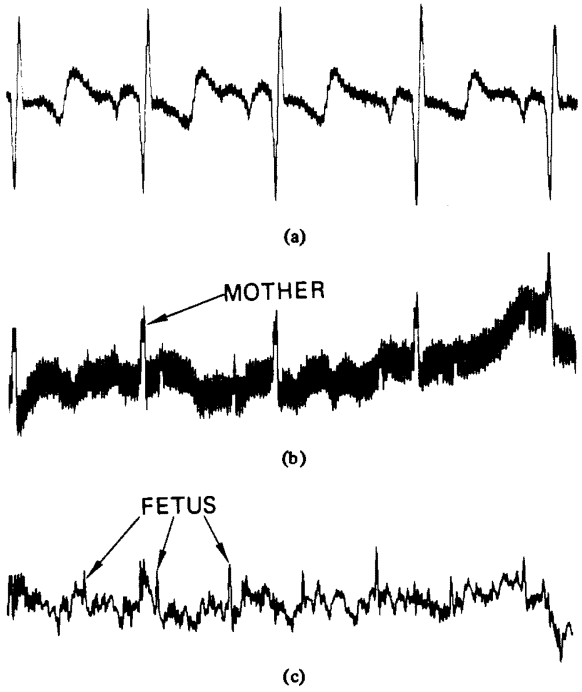

Fig. 17. Result of wide-band fetal ECG experiment (bandwidth, 0.3-75 Hz; sampling rate, 512 Hz). (a) Reference input (chest lead). (b) Primary input (abdominal lead). (c) Noise canceller output.

ing bandwidth was 3 to 35 Hz and the sampling rate 256 Hz. The maternal heartbeat, which dominates the primary input, is almost completely absent in the noise canceller output. Note that the voltage scale of the noise canceller output, Fig. 16(c), is approximately two times greater than that of the primary input, Fig. 16(b).

Fig. 17 shows corresponding results for a prefiltering bandwidth of 0.3 to 75 Hz and a sampling rate of 512 Hz. Baseline drift and 60-Hz interference are clearly present in the primary input, obtained from the abdominal lead. The interference is so strong that it is almost impossible to detect the fetal heartbeat. The inputs obtained from the chest leads contained the maternal heartbeat and a sufficient 60-Hz component to serve as a reference for both of these interferences. In the noise canceller output both interferences have been significantly reduced, and the fetal heartbeat is clearly discernible.

Additional experiments are currently being conducted with the aim of further improving the fetal ECG by reducing the background noise caused by muscle activity. In these experiments various averaging techniques are being investigated together with new adaptive processing methods for signals derived from an array of abdominal leads.

D. Cancelling Noise in Speech Signals

Consider the situation of a pilot communicating by radio from the cockpit of an aircraft where a high level of engine noise is present. The noise contains, among other things, strong periodic components, rich in harmonics, that occupy the same frequency band as speech. These components are picked up by the microphone into which the pilot speaks and severely interfere with the intelligibility of the radio transmission. It would be impractical to process the transmission with

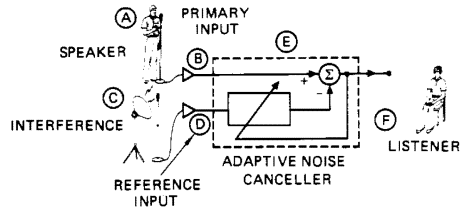

Fig. 18. Cancelling noise in speech signals.

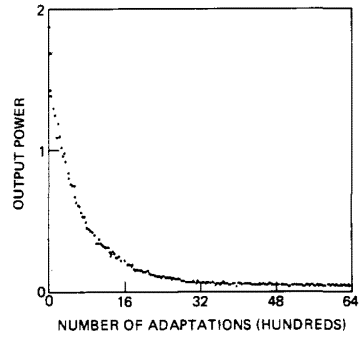

Fig. 19. Typical learning curve for speech noise cancelling experiment.

a conventional filter because the frequency and intensity of the noise components vary with engine speed and load and position of the pilot's head. By placing a second microphone at a suitable location in the cockpit, however, a sample of the ambient noise field free of the pilot's speech could be obtained. This sample could be filtered and subtracted from the transmission, significantly reducing the interference.

To demonstrate the feasibility of cancelling noise in speech signals a group of experiments simulating the cockpit noise problem in simplified form was conducted. In these experiments, as shown in Fig. 18, a person (*A*) spoke into a microphone (*B*) in a room where strong acoustic interference (*C*) was present. A second microphone (*D*) was placed in the room away from the speaker. The output of microphones (*B*) and (*D*) formed the primary and reference inputs, respectively, of a noise canceller (*E*), whose output was monitored by a remote listener (*F*). The canceller included an adaptive filter with 16 hybrid analog weights whose values were digitally controlled by a computer. The rate of adaptation was approximately 5 kHz. A typical learning curve, showing output power as a function of number of adaptation cycles, is shown in Fig. 19. Convergence was complete after about 5000 adaptations or one second of real time.

In a typical experiment the interference was an audiofrequency triangular wave containing many harmonics that, because of multipath effects, varied in amplitude, phase, and waveform from point to point in the room. The periodic nature of the wave made it possible to ignore the difference in time delay caused by the different transmission paths to the two sensors. The noise canceller was able to reduce the output power of this interference, which otherwise made the speech unintelligible, by 20 to 25 dB, rendering the interference barely perceptible to the remote listener. No noticeable distortion was introduced into the speech signal. Convergence times were on the order of seconds, and the processor was readily able to readapt when the position of the microphones was changed or when the frequency of the interference was varied over the range 100 to 2000 Hz.

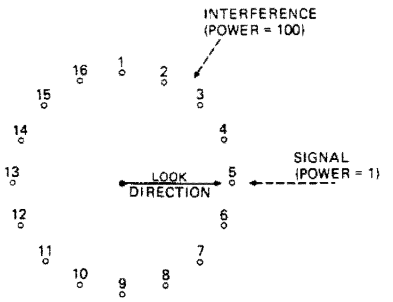

Fig. 20. Array configuration for adaptive sidelobe cancelling experiment.

E. Cancelling Antenna Sidelobe Interference

Strong unwanted signals incident on the sidelobes of an antenna array can severely interfere with the reception of weaker signals in the main beam. The conventional method of reducing such interference, adaptive beamforming [6], [18], [19], [34]-[37], is often complicated and expensive to implement. When the number of spatially discrete interference sources is small, adaptive noise cancelling can provide a simpler and less expensive method of dealing with this problem.

To demonstrate the level of sidelobe reduction achievable with adaptive noise cancelling, a typical interference cancelling problem was simulated on the computer. As shown in Fig. 20, an array consisting of a circular pattern of 16 equally spaced omnidirectional elements was chosen. The outputs of the elements were delayed and summed to form a main beam steered at a relative angle of 0°. A simulated "white" signal consisting of uncorrelated samples of unit power was assumed to be incident on this beam. Simulated interference with the same bandwidth and with a power of 100 was incident on the main beam at a relative angle of 58°. The array was connected to an adaptive noise canceller in the manner shown in Fig. 5. The output of the beamformer served as the canceller's primary input, and the output of element 4 was arbitrarily chosen as the reference input. The canceller included an adaptive filter with 14 weights; the adaptation constant in the LMS algorithm was set at $\mu = 7 \times 10^{-6}$.

Fig. 21 shows two series of computed directivity patterns, one representing a single frequency of $\frac{1}{4}$ the sampling frequency and the other an average of eight frequencies of from $\frac{1}{8}$ to $\frac{3}{8}$ the sampling frequency. These patterns indicate the evolution of the main beam and sidelobes as observed by stopping the adaptive process after the specified number of iterations. Note the deep nulls that develop in the direction of the interference. At the start of adaptation all weights were set at zero, providing a conventional 16-element beam pattern.

The signal-to-noise ratio at the system output, averaged over the eight frequencies, was found after convergence to be +20 dB. The signal-to-noise ratio at the single array element was -20 dB. This result bears out the expectation arising from (37) that the signal-to-noise ratio at the system output would be the reciprocal of the ratio at the reference input, which is derived from a single element.

A small amount of signal cancellation occurred, as evidenced by the changes in sensitivity of the main beam in the steering direction. These changes were not unexpected, since the main-lobe pattern was not constrained by the adaptive process. A method of LMS adaptation with constraints that could have been used to prevent this loss of sensitivity has been developed by Frost [37].

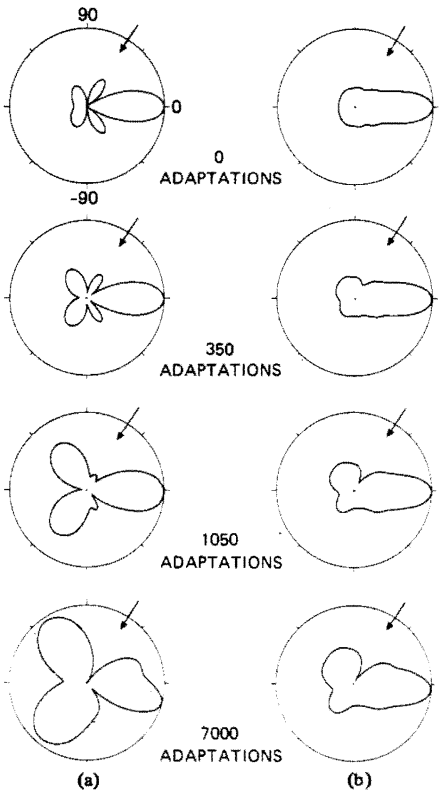

Fig. 21. Results of adaptive sidelobe cancelling experiment. (a) Single frequency (0.5 relative to folding frequency). (b) Average of eight frequencies (0.25 to 0.75 relative to folding frequency).

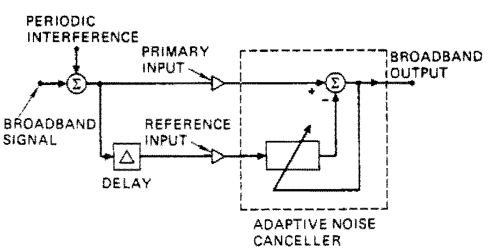

Fig. 22. Cancelling periodic interference without an external reference source.

F. Cancelling Periodic Interference without an External Reference Source

There are a number of circumstances where a broad-band signal is corrupted by periodic interference and no external reference input free of the signal is available. Examples include the playback of speech or music in the presence of tape hum or turntable rumble. It might seem that adaptive noise cancelling could not be applied to reduce or eliminate this kind of interference. If, however, a fixed delay Δ is inserted in a reference input drawn directly from the primary input, as shown in Fig. 22, the periodic interference can in many cases be readily cancelled.[15] The delay chosen must be of sufficient length to cause the broad-band signal components in the reference input to become decorrelated from those in

[15] The delay Δ may be inserted in the primary instead of the reference input if its total length is greater than the total delay of the adaptive filter. Otherwise, the filter will converge to match it and cancel both signal and interference.

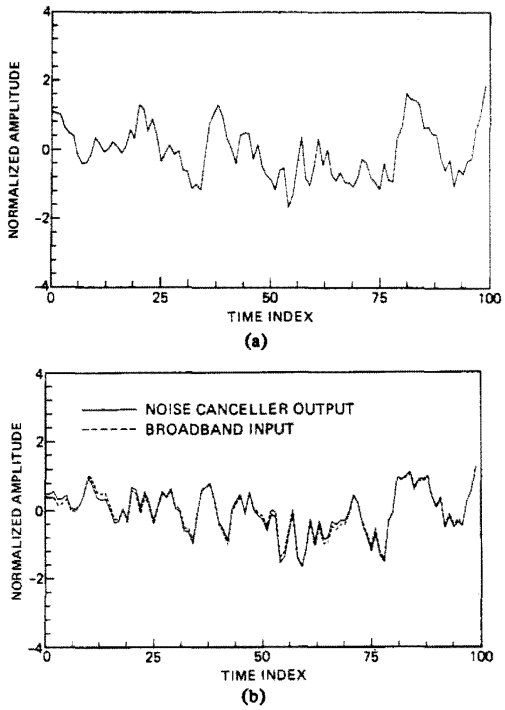

Fig. 23. Result of periodic interference cancelling experiment. (a) Input signal (correlated Gaussian noise and sine wave). (b) Noise canceller output (correlated Gaussian noise).

the primary input. The interference components, because of their periodic nature, will remain correlated with each other.

Fig. 23 presents the results of a computer simulation performed to demonstrate the cancelling of periodic interference without an external reference. Fig. 23(a) shows the primary input to the canceller. This input is composed of colored Gaussian noise representing the signal and a sine wave representing the interference. Fig. 23(b) shows the noise canceller's output. Since the problem was simulated, the exact nature of the broad-band input was known and is plotted together with the output. Note the close correspondence in form and registration. The correspondence is not perfect only because the filter was of finite length and had a finite rate of adaptation.

G. Adaptive Self-Tuning Filter

The previous experiment can also be used to demonstrate another important application of the adaptive noise canceller. In many instances where an input signal consisting of mixed periodic and broad-band components is available, the periodic rather than the broad-band components are of interest. If the system output of the noise canceller of Fig. 22 is taken from the adaptive filter, the result is an adaptive self-tuning filter capable of extracting a periodic signal from broad-band noise.

Fig. 24 shows the adaptive noise canceller as a self-tuning filter. The output of this system was simulated on the computer with the input of sine wave and correlated Gaussian noise used in the previous experiment and shown in Fig. 23(a). The resulting approximation of the input sine wave is shown in Fig. 25 together with the actual input sine wave. Note once again the close agreement in form and registration. The error is a small-amplitude stochastic process.

Fig. 26 shows the impulse response and transfer function of the adaptive filter after convergence. The impulse response,

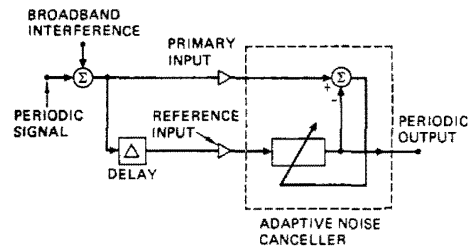

Fig. 24. The adaptive noise canceller as a self-tuning filter.

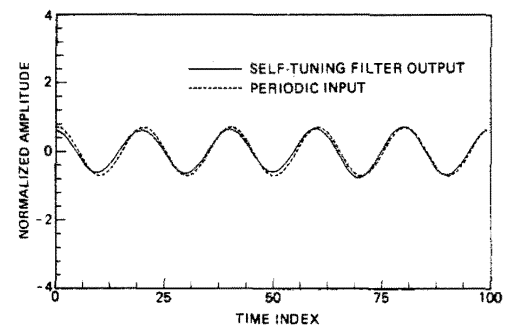

Fig. 25. Result of self-tuning filter experiment.

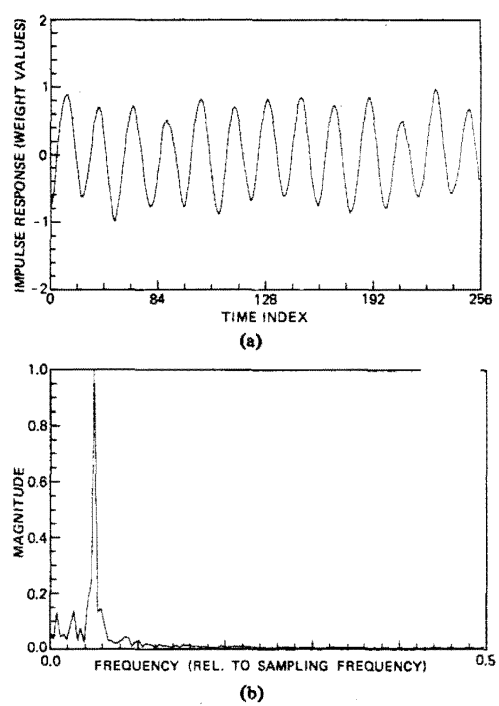

Fig. 26. Adaptive filter characteristics in self-tuning filter experiment. (a) Impulse response of adaptive filter after convergence. (b) Magnitude of transfer function of adaptive filter after convergence.

shown in Fig. 26(a), is somewhat different from but bears a close resemblance to a sine wave. If the broad-band input component had been white noise, the optimal estimator would have been a matched filter, and the impulse response would have been sinusoidal.

The transfer function, shown in Fig. 26(b), is the digital Fourier transform of the impulse response. Its magnitude at

the frequency of the interference is nearly one, the value required for perfect cancellation. The phase shift at this frequency is not zero but when added to the phase shift caused by the delay Δ forms an integral multiple of 360°.

Similar experiments have been conducted with sums of sinusoidal signals in broad-band stochastic interference. In these experiments the adaptive filter developed sharp resonance peaks at the frequencies of all the spectral line components of the periodic portion of the primary input. The system thus shows considerable promise as an automatic signal seeker.

Further experiments have shown the ability of the adaptive self-tuning filter to be employed as a line enhancer for the detection of extremely low-level sine waves in noise. An introductory treatment of this application, which promises to be of great importance, is provided in Appendix D.

IX. Conclusion

Adaptive noise cancelling is a method of optimal filtering that can be applied whenever a suitable reference input is available. The principal advantages of the method are its adaptive capability, its low output noise, and its low signal distortion. The adaptive capability allows the processing of inputs whose properties are unknown and in some cases nonstationary. It leads to a stable system that automatically turns itself off when no improvement in signal-to-noise ratio can be achieved. Output noise and signal distortion are generally lower than can be achieved with conventional optimal filter configurations.

The experimental data presented in this paper demonstrate the ability of adaptive noise cancelling greatly to reduce additive periodic or stationary random interference in both periodic and random signals. In each instance cancelling was accomplished with little signal distortion even though the frequencies of the signal and the interference overlapped. The experiments described indicate the wide range of applications in which adaptive noise cancelling has potential usefulness.

Appendix A
The LMS Adaptive Filter

This Appendix provides a brief description of the LMS adaptive filter, the basic element of the adaptive noise cancelling systems described in this paper. For a full description the reader should consult the extensive literature on the subject, including the references cited below.

A. Adaptive Linear Combiner

The principal component of most adaptive systems is the adaptive linear combiner, shown in Fig. 27.[16] The combiner weights and sums a set of input signals to form an output signal. The input signal vector X_j is defined as

$$X_j \triangleq \begin{Bmatrix} x_{0j} \\ x_{1j} \\ \cdot \\ \cdot \\ \cdot \\ x_{nj} \end{Bmatrix}. \tag{A.1}$$

The input signal components are assumed to appear simultaneously on all input lines at discrete times indexed by the subscript j. The component x_{0j} is a constant, normally set to the value +1, used only in cases where biases exist among the inputs (A.1) or in the desired response (defined below). The weighting coefficients or multiplying factors $w_0, w_1, \cdots, w_n$ are adjustable, as symbolized in Fig. 27 by circles with arrows through them. The weight vector is

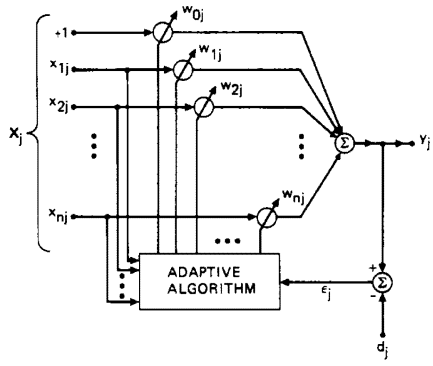

Fig. 27. The adaptive linear combiner.

$$W = \begin{Bmatrix} w_0 \\ w_1 \\ \cdot \\ \cdot \\ \cdot \\ w_n \end{Bmatrix} \tag{A.2}$$

where w_0 is the bias weight.

The output y_j is equal to the inner product of X_j and W:

$$y_j = X_j^T W = W^T X_j. \tag{A.3}$$

The error ϵ_j is defined as the difference between the desired response d_j (an externally supplied input sometimes called the "training signal") and the actual response y_j:

$$\epsilon_j = d_j - X_j^T W = d_j - W^T X_j. \tag{A.4}$$

In most applications some ingenuity is required to obtain a suitable input for d_j. After all, if the actual desired response were known, why would one need an adaptive processor? In noise cancelling systems, however, d_j is simply the primary input.[17]

B. The LMS Adaptive Algorithm

It is the purpose of the adaptive algorithm designated in Fig. 27 to adjust the weights of the adaptive linear combiner to minimize mean-square error. A general expression for mean-square error as a function of the weight values, assuming that the input signals and the desired response are statistically stationary and that the weights are fixed, can be derived in the following manner. Expanding (A.4) one obtains

$$\epsilon_j^2 = d_j^2 - 2d_j X_j^T W + W^T X_j X_j^T W. \tag{A.5}$$

Taking the expected value of both sides yields

$$E[\epsilon_j^2] = E[d_j^2] - 2E[d_j X_j^T]\, W + W^T E[X_j X_j^T]\, W. \tag{A.6}$$

Defining the vector P as the cross correlation between the

[16] This component is linear only when the weighting coefficients are fixed. Adaptive systems, like all systems whose characteristics change with the characteristics of their inputs, are by their very nature nonlinear.

[17] The actual desired response is the primary noise n_0, which is not available apart from the primary input $s + n_0$. The converged weight vector solution is easily shown to be the same when either n_0 or $s + n_0$ serves as the desired response.

desired response (a scalar) and the X vector then yields

$$P \triangleq E[d_j X_j] = E \begin{Bmatrix} d_j x_{0j} \\ d_j x_{1j} \\ \cdot \\ \cdot \\ \cdot \\ d_j x_{nj} \end{Bmatrix}. \tag{A.7}$$

The input correlation matrix R is defined as

$$R \triangleq E[X_j X_j^T] = E \begin{bmatrix} x_{0j}x_{0j} & x_{0j}x_{1j} & x_{0j}x_{2j} & \cdots \\ x_{1j}x_{0j} & x_{1j}x_{1j} & x_{1j}x_{2j} & \cdots \\ x_{2j}x_{0j} & x_{2j}x_{1j} & x_{2j}x_{2j} & \cdots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdots & \cdots & \cdots & x_{nj}x_{nj} \end{bmatrix}. \tag{A.8}$$

This matrix is symmetric, positive definite, or in rare cases positive semidefinite. The mean-square error can thus be expressed as

$$E[\epsilon_j^2] = E[d_j^2] - 2P^T W + W^T R W. \tag{A.9}$$

Note that the error is a quadratic function of the weights that can be pictured as a concave hyperparaboloidal surface, a function that never goes negative. Adjusting the weights to minimize the error involves descending along this surface with the objective of getting to the "bottom of the bowl." Gradient methods are commonly used for this purpose.

The gradient ∇ of the error function is obtained by differentiating (A.9):

$$\nabla \triangleq \begin{Bmatrix} \dfrac{\partial E[\epsilon_j^2]}{\partial w_0} \\ \cdot \\ \cdot \\ \cdot \\ \dfrac{\partial E[\epsilon_j^2]}{\partial w_n} \end{Bmatrix} = -2P + 2RW. \tag{A.10}$$

The optimal weight vector W^*, generally called the Wiener weight vector, is obtained by setting the gradient of the mean-square error function to zero:

$$W^* = R^{-1}P. \tag{A.11}$$

This equation is a matrix form of the Wiener–Hopf equation [1], [2].

The LMS adaptive algorithm [7], [8], [19], [20] is a practical method for finding close approximate solutions to (A.11) in real time. The algorithm does not require explicit measurements of correlation functions, nor does it involve matrix inversion. Accuracy is limited by statistical sample size, since the weight values found are based on real-time measurements of input signals.

The LMS algorithm is an implementation of the method of steepest descent. According to this method, the "next" weight vector W_{j+1} is equal to the "present" weight vector W_j plus a change proportional to the negative gradient:

$$W_{j+1} = W_j - \mu \nabla_j. \tag{A.12}$$

The parameter μ is the factor that controls stability and rate of convergence. Each iteration occupies a unit time period. The true gradient at the jth iteration is represented by ∇_j.

The LMS algorithm estimates an instantaneous gradient in a crude but efficient manner by assuming that ϵ_j^2, the square of a single error sample, is an estimate of the mean-square error and by differentiating ϵ_j^2 with respect to W. The relationships between true and estimated gradients are given by the following expressions:

$$\nabla_j \triangleq \begin{Bmatrix} \dfrac{\partial E[\epsilon_j^2]}{\partial w_0} \\ \cdot \\ \cdot \\ \cdot \\ \dfrac{\partial E[\epsilon_j^2]}{\partial w_n} \end{Bmatrix}_{W=W_j} \qquad \hat{\nabla}_j = \begin{Bmatrix} \dfrac{[\partial \epsilon_j^2]}{\partial w_0} \\ \cdot \\ \cdot \\ \cdot \\ \dfrac{\partial \epsilon_j^2}{\partial w_n} \end{Bmatrix}_{W=W_j} = 2\epsilon_j \begin{Bmatrix} \dfrac{\partial \epsilon_j}{\partial w_0} \\ \cdot \\ \cdot \\ \cdot \\ \dfrac{\partial \epsilon_j}{\partial w_n} \end{Bmatrix}_{W=W_j} \tag{A.13}$$

The estimated gradient components are related to the partial derivatives of the instantaneous error with respect to the weight components, which can be obtained by differentiating (A.5). Thus the expression for the gradient estimate can be simplified to

$$\hat{\nabla}_j = -2\epsilon_j X_j. \tag{A.14}$$

Using this estimate in place of the true gradient in (A.12) yields the Widrow–Hoff LMS algorithm:

$$W_{j+1} = W_j + 2\mu\epsilon_j X_j. \tag{A.15}$$

This algorithm is simple and generally easy to implement. Although it makes use of gradients of mean-square error functions, it does not require squaring, averaging, or differentiation.

It has been shown [18], [19] that the gradient estimate used in the LMS algorithm is unbiased and that the expected value of the weight vector converges to the Wiener weight vector (A.11) when the input vectors are uncorrelated over time (although they could, of course, be correlated from input component to component).[18] Starting with an arbitrary initial weight vector, the algorithm will converge in the mean and will remain stable as long as the parameter μ is greater than 0 but less than the reciprocal of the largest eigenvalue λ_{max} of the matrix R:

$$1/\lambda_{max} > \mu > 0. \tag{A.16}$$

Fig. 28 shows a typical individual learning curve resulting from the use of the algorithm. Also shown is an ensemble

[18] Adaptation with correlated input vectors has been analyzed by Senne [38] and Daniell [39]. Extremely high correlation and fast adaptation can cause the weight vector to converge in the mean to something different than the Wiener solution. Practical experience has shown, however, that this effect is generally insignificant. See also Kim and Davisson [40].

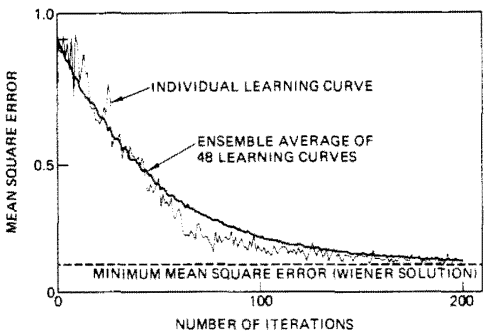

Fig. 28. Typical learning curves for the LMS algorithm.

average of 48 learning curves. The ensemble average reveals the underlying exponential nature of the individual learning curve. The number of natural modes is equal to the number of degrees of freedom (number of weights). The time constant of the pth mode is related to the pth eigenvalue λ_p of the input correlation matrix P and to the parameter μ by

$$\tau_{p\,\mathrm{mse}} = \frac{1}{4\mu\lambda_p}. \tag{A.17}$$

Although the learning curve consists of a sum of exponentials, it can in many cases be approximated by a single exponential whose time constant is given by (A.17) using the average of the eigenvalues of R:

$$\lambda_{\mathrm{av}} = \frac{\lambda_0 + \lambda_1 + \cdots + \lambda_p + \cdots + \lambda_n}{(n+1)} = \frac{\mathrm{tr}\,R}{(n+1)}. \tag{A.18}$$

Accordingly, the time constant of an exponential roughly approximating the mean-square error learning curve is

$$\tau_{\mathrm{mse}} = \frac{(n+1)}{4\mu\,\mathrm{tr}\,R} = \frac{(\text{number of weights})}{(4\mu)(\text{total input power})}. \tag{A.19}$$

The total input power is the sum of the powers incident to all of the weights.

Proof of these assertions and further discussion of the characteristics and properties of the LMS algorithm are presented in [19], [20], and [41].

C. The LMS Adaptive Filter

The adaptive linear combiner may be implemented in conjunction with a tapped delay line to form the LMS adaptive filter shown in Fig. 29, where the bias weight has been omitted for simplicity. Fig. 29(a) shows the details of the filter, including the adaptive process incorporating the LMS algorithm. Because of the structure of the delay line, the input signal vector is

$$X_j = \begin{Bmatrix} x_j \\ x_{j-1} \\ \cdot \\ \cdot \\ \cdot \\ x_{j-n+1} \end{Bmatrix}. \tag{A.20}$$

The components of this vector are delayed versions of the input signal x_j. Fig. 29(b) is the representation adopted to symbolize the adaptive tapped-delay-line filter.

This kind of filter permits the adjustment of gain and phase at many frequencies simultaneously and is useful in adaptive broad-band signal processing. Simplified design rules, giving

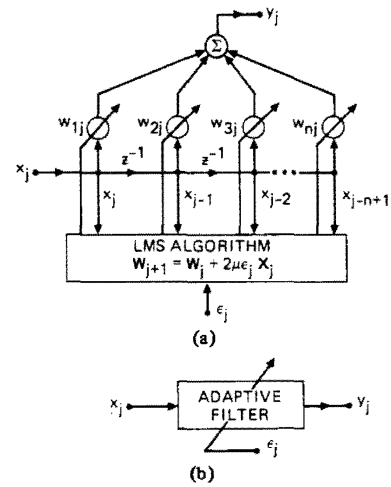

Fig. 29. The LMS adaptive filter. (a) Block diagram. (b) Symbolic representation.

the tap spacings and number of taps (weights), are the following: The tap spacing time must be at least as short as the reciprocal of twice the signal bandwidth (in accord with the sampling theorem). The total real-time length of the delay line is determined by the reciprocal of the desired filter frequency resolution. Thus, the number of weights required is generally equal to twice the ratio of the total signal bandwidth to the frequency resolution of the filter. It may be possible to reduce the number required in some cases by using nonuniform tap spacing, such as log periodic. Whether this is done or not, the means of adaptation remain the same.

Appendix B
Finite-Length, Causal Approximation of the Unconstrained Wiener Noise Canceller

In the analyses of Sections IV and V questions of the physical realizability of Wiener filters were not considered. The expressions derived were ideal, based on the assumption of an infinitely long, two-sided (noncausal) tapped delay line. Though such a delay line cannot in reality be implemented, fortunately its performance, as shown in the following paragraphs, can be closely approximated.

Typical impulse responses of ideal Wiener filters approach amplitudes of zero exponentially over time. Approximate realizations are thus possible with finite-length transversal filters. The more weights used in the transversal filter, the closer its impulse response will be to that of the ideal Wiener filter. Increasing the number of weights, however, also slows the adaptive process and increases the cost of implementation. Performance requirements should thus be carefully considered before a filter is designed for a particular application.

Noncausal filters, of course, are not physically realizable in real-time systems. In many cases, however, they can be realized approximately in delayed form, providing an acceptable delayed real-time response. In practical circumstances excellent performance can be obtained with two-sided filter impulse responses even when they are truncated in time to the left and right. By delaying the truncated response it can be made causal and physically realizable.

Fig. 30 shows an adaptive noise cancelling system with a delay Δ inserted in the primary input. This delay causes an equal delay to develop in the unconstrained optimal filter

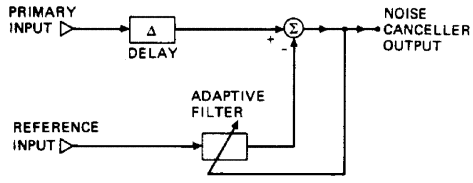

Fig. 30. Adaptive noise canceller with delay in primary input path.

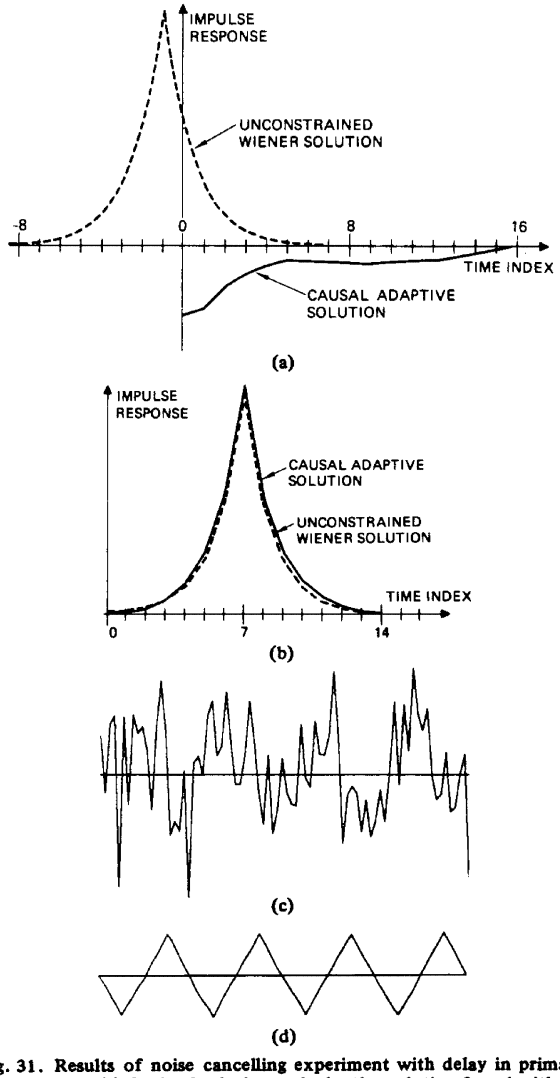

Fig. 31. Results of noise cancelling experiment with delay in primary input path. (a) Optimal solution and adaptive solution found without time delay. (b) Optimal solution and adaptive solution found with delay of eight time units. (c) Noise canceller output without delay. (d) Noise canceller output with delay.

impulse response, which remains otherwise unchanged. In practical, finite-length adaptive transversal filters, on the other hand, the optimal impulse response generally changes shape with changes in the value of Δ, which is chosen to cause the peak of the impulse response to center along the delay line.

Experience has shown that the value of Δ is not critical within a certain optimal range; that is, the curve showing minimum mean-square error as a function of Δ generally has a very broad minimum. A value typically equal to about half the time delay of the adaptive filter produces the least minimum output noise power.

Fig. 31 shows the results of a computer-simulated noise cancelling experiment with an unconstrained optimal filter response that was noncausal. The primary input consisted of a triangular wave and additive colored noise. The reference input consisted of colored noise correlated with the primary noise.[19] The unconstrained Wiener impulse response and the causal, finite time adaptive impulse response obtained without a delay in the primary input are plotted in Fig. 31(a). The large difference in these impulse responses indicates that the noise canceller output will be a poor approximation of the signal. The corresponding Wiener and adaptive impulse responses obtained with a delay of eight time units (half the length of the adaptive filter) are shown in Fig. 31(b). These solutions are similar, indicating that performance of the adaptive filter will be close to optimal. Typical noise canceller outputs with and without delay are shown in Fig. 31(c) and Fig. 31(d). The waveform obtained with the delay is very close to that of the original triangular-wave signal, whereas that obtained with no delay still contains a great amount of noise.

Appendix C
Multiple-Reference Noise Cancelling

When there is more than one noise or interference to be cancelled and a number of linearly independent reference inputs containing mixtures of each can be obtained, it is usually advantageous to use a multiple-reference noise cancelling system. Such a system may be considered a generalization of the single-reference noise cancellers analyzed in this paper. In the model shown in Fig. 32 the ψ_i represent mutually uncorrelated sources of either input signal or noise. The transfer functions $\mathcal{G}_i(z)$ represent the propagation paths from these sources to the primary input. The $\mathcal{F}_i(z)$ similarly represent the propagation paths to the reference inputs and allow for cross-coupling. This model permits treatment not only of multiple noise sources but also of signal components in the reference inputs and uncorrelated noises in the reference and primary inputs. In other words, it is a general representation of an adaptive noise canceller.

The unconstrained Wiener transfer function of the multiple-reference canceller is the matrix equivalent of (13) and is derived in the following manner. The source spectral matrix of ψ_i is defined as

$$[\delta_{\psi\psi}(z)] = \begin{bmatrix} \delta_{\psi_1\psi_1}(z) & & & \bigcirc \\ & \delta_{\psi_2\psi_2} & & \\ & & \ddots & \\ \bigcirc & & & \delta_{\psi_m\psi_m}(z) \end{bmatrix}. \tag{C.1}$$

The spectral matrix of the k reference inputs to the adaptive

[19] Except for the delay in the primary input, the simulated noise cancelling system was identical with the system shown above in Fig. 3. The transfer function $\mathcal{H}(z)$ was a nonminimum phase, low-pass transversal filter with two zeros and no poles $[\mathcal{H}(z) = 2z^{-1}(1 - \frac{1}{2}z^{-1}) \cdot (1 - \frac{1}{2}z)]$. The optimal unconstrained adaptive filter solution, in this case given by (18), is the reciprocal of $\mathcal{H}(z)$. It has one pole inside and one pole outside the unit circle in the Z plane. A stable realization of $\mathcal{D}^{\bullet}(z)$ must, therefore, be two sided.

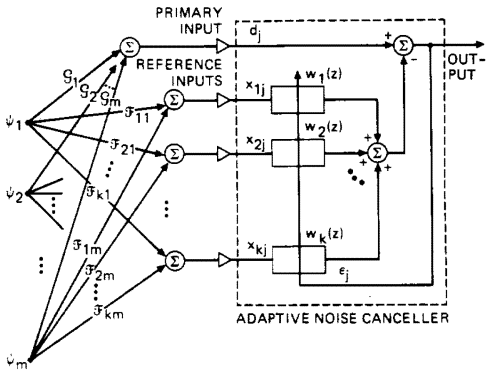

Fig. 32. Generalized multiple-reference adaptive noise canceller.

filters is then

$$[\mathcal{S}_{xx}(z)] = [\mathcal{F}(z^{-1})]\ [\mathcal{S}_{\psi\psi}(z)]\ [\mathcal{F}(z)]^T \tag{C.2}$$

where

$$[\mathcal{F}(z)] = \begin{bmatrix} \mathcal{F}_{11}(z) & \mathcal{F}_{12}(z) & \cdots & \mathcal{F}_{1m}(z) \\ \mathcal{F}_{21}(z) & & & \\ \vdots & & \ddots & \\ \mathcal{F}_{k1}(z) & & \cdots & \mathcal{F}_{km}(z) \end{bmatrix} \tag{C.3}$$

and $\mathcal{F}_{il}(z)$ is the transfer function from input source l to reference input i.

The cross-spectral vector between the reference inputs and the primary input is given by

$$\{\mathcal{S}_{xd}(z)\} = [\mathcal{F}(z^{-1})]\ [\mathcal{S}_{\psi\psi}(z)]\ \{\mathcal{G}(z)\} \tag{C.4}$$

where

$$\mathcal{G}(z) = \begin{bmatrix} \mathcal{G}_1(z) \\ \mathcal{G}_2(z) \\ \vdots \\ \mathcal{G}_m(z) \end{bmatrix} \tag{C.5}$$

and $\mathcal{G}_i(z)$ is the transfer function from input source i to the primary input.

The Wiener optimal weight vector is then

$$\begin{aligned} \mathcal{W}^*(z) &= [\mathcal{S}_{xx}(z)]^{-1}\ \mathcal{S}_{xd}(z) \\ &= [[\mathcal{F}(z^{-1})]\ [\mathcal{S}_{\psi\psi}(z)]\ [\mathcal{F}(z)]^T]^{-1} \\ &\quad \cdot [\mathcal{F}(z^{-1})]\ [\mathcal{S}_{\psi\psi}(z)]\ \{\mathcal{G}(z)\}. \end{aligned} \tag{C.6}$$

If $[\mathcal{F}(z)]$ is square, at those frequencies for which $[\mathcal{F}(z)]$ is invertible (55) simplifies to

$$\{\mathcal{W}^*(z)\} = [\mathcal{F}(z)^T]^{-1}\ \{\mathcal{G}(z)\} \tag{C.7}$$

which is the matrix equivalent of (17).

These expressions can be used to derive steady-state Wiener solutions to multiple-source, multiple-reference noise cancelling problems more general than those of Sections IV and V. An example of a multiple-reference problem is given in Section VIII.

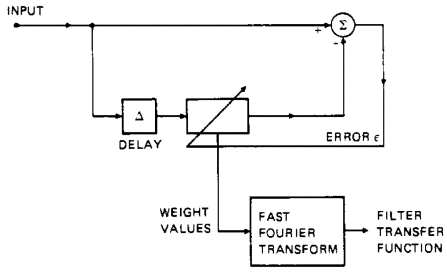

Fig. 33. The adaptive line enhancer.

Appendix D
Adaptive Line Enhancer

A classical detection problem is that of finding a low-level sine wave in noise. The adaptive self-tuning filter, whose capability of separating the periodic and stochastic components of a signal was illustrated above (where these components were of comparable level), is able to serve as an "adaptive line enhancer" for the detection of extremely low-level sine waves in noise. The adaptive line enhancer becomes a competitor of the fast Fourier transform algorithm as a sensitive detector and has capabilities that may exceed those of conventional spectral analyzers when the unknown sine wave has finite bandwidth or is frequency modulated.

The method is illustrated in Fig. 33. The input consists of signal plus noise. The output is the digital Fourier transform of the filter's impulse response. Detection is accomplished when a spectral peak is evident above the background noise. The same method, with minor differences, has been proposed by Griffiths for "maximum entropy spectral estimation" [42], [43].

It should be noted that the filter output signal is also available. This signal could be used directly or as an input to a spectral analyzer or phase-lock loop. The method of Fig. 33 could further be used for the simultaneous detection of multiple sine waves. None of these possibilities is considered here. Only the detection of single low-level sine waves in noise is treated.

A. Optimal Transfer Function

Fig. 34 shows the ideal impulse response and transfer function of the adaptive line enhancer for a given input spectrum. It is assumed that the input noise is white, with a total power of ν^2, and that the input signal has a power of $C^2/2$ at frequency ω_0. The ideal impulse response, equivalent to the matched filter response, is a sampled sinusoid whose frequency is ω_0.[20] The phase shift of this response at frequency ω_0 when added to that of the delay is an integral multiple of 360°. If the peak value of the transfer function is a, the peak value of the weights is to a close approximation $2a/n$, where n is the number of weights.

The adaptive process minimizes the mean square of the error. The error power is the sum of three components, the primary input noise power, the noise power at the output of the adaptive filter, and the sinusoidal signal power. Accordingly, the error power may be expressed as

$$\text{error power} = \nu^2 + (\nu^2/2)\,(2a/n)^2\, n + (C^2/2)\,(1-a)^2. \tag{D.1}$$

[20] This assertion is proved analytically for arbitrary input signal-to-noise ratio in J. R. Zeidler and D. M. Chabries, "An analysis of the LMS adaptive filter used as a spectral line enhancer," Naval Undersea Center, Tech. Note 1476, Feb. 1975.

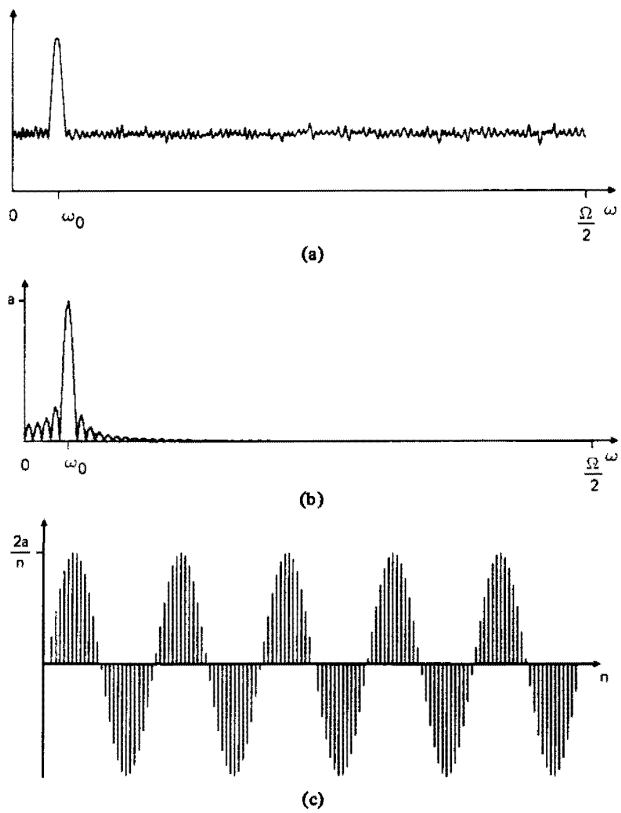

Fig. 34. Ideal adaptive filter impulse response and transfer function of adaptive line enhancer for a given input spectrum. (a) Input spectrum. (b) Transfer function magnitude. (c) Impulse response.

We have used the facts that 1) the output noise power of a digital filter with a white input equals the input power multiplied by the sum of the squares of the impulse values of the impulse response and 2) the primary and filter output sinusoidal components combine coherently at the summing junction. The signal gain from input to error is $(1 - a)$.

The optimal value of a that minimizes error power a^* is obtained by setting the derivative of (D.1) to zero:

$$a^* = \frac{\left(\frac{C^2/2}{\nu^2}\right)\left(\frac{n}{2}\right)}{1 + \left(\frac{C^2/2}{\nu^2}\right)\left(\frac{n}{2}\right)} = \frac{(\mathrm{SNR})\,(n/2)}{1 + (\mathrm{SNR})\,(n/2)}. \tag{D.2}$$

At high signal-to-noise ratios, $a^* \cong 1$. At low signal-to-noise ratios, $a^* < 1$. Low signal-to-noise conditions can be dealt with by using a large number of adaptive weights, although other problems could result because of weight-vector noise.

B. Noise in the Weight Vector

The ability to detect peaks in the transfer function due to the presence of sinusoidal signals is limited by the presence of spurious peaks caused by noise in the weight vector. One thus needs to know the nature of weight-vector noise and its effects on the transfer function.

The gradient estimate $\hat{\nabla}_j$ used by the LMS algorithm, given by (A.14), may be expressed as

$$\hat{\nabla}_j = -2\epsilon_j X_j = \nabla_j + \mathfrak{N}_j \tag{D.3}$$

where ∇_j is the true gradient and $\mathfrak{N}_j$ is the zero-mean gradient estimation noise. At the minimum point of the quadratic mean-square-error surface the true gradient is zero. The gradient estimate at this point is thus equal to the gradient estimation noise:

$$\hat{\nabla}_j = \mathfrak{N}_j = -2\epsilon_j X_j. \tag{D.4}$$

If one assumes that the input signal vector X_j is uncorrelated over time,[21] then $\mathfrak{N}_j$ is also uncorrelated over time. In addition, when the weight vector W_j is equal to the optimal weight vector W^*, Wiener filter theory shows that the error ϵ_j and the input vector X_j are uncorrelated. If one now assumes that ϵ_j and X_j are Gaussian, then these terms are statistically independent and the covariance of $\mathfrak{N}_j$ is

$$\begin{aligned}\mathrm{cov}\,[\mathfrak{N}_j] &= E[\mathfrak{N}_j \mathfrak{N}_j^T] = 4E[\epsilon_j^2 X_j X_j^T] = 4E[\epsilon_j^2] \\ &\quad \cdot E[X_j X_j^T] = 4E[\epsilon_j^2]\, R\end{aligned} \tag{D.5}$$

where R is the input correlation matrix. Since at the minimum point of the mean-square-error surface $E[\epsilon_j^2] = \xi_{\min}$, (D.5) can be expressed as

$$\mathrm{cov}\,[\mathfrak{N}_j] = 4\xi_{\min} R. \tag{D.6}$$

In the vicinity of the minimum point the covariance of the gradient noise is closely approximated by (D.6), and the gradient noise is statistically stationary and uncorrelated over time.

For the purpose of the following analysis it is more convenient to work in "primed coordinates." The correlation matrix R may be expressed in normal form as

$$R = Q\Lambda Q^{-1} \tag{D.7}$$

where Q is an orthonormal modal matrix, Λ is a diagonal matrix of eigenvalues, and Q^{-1} is equal to Q^T. The gradient noise in the primed coordinates is then

$$\mathfrak{N}_j' = Q^{-1}\mathfrak{N}_j \tag{D.8}$$

and the covariance of the gradient noise is

$$\begin{aligned}\mathrm{cov}\,[\mathfrak{N}_j'] &= E[\mathfrak{N}_j' \mathfrak{N}_j'^T] = E[Q^{-1}\mathfrak{N}_j \mathfrak{N}_j^T Q] = Q^{-1} E[\mathfrak{N}_j \mathfrak{N}_j^T]\, Q \\ &= Q^{-1}\,\mathrm{cov}\,[\mathfrak{N}_j]\, Q = 4\xi_{\min} Q^{-1} R Q = 4\xi_{\min}\Lambda.\end{aligned} \tag{D.9}$$

It should be noted that the components of $\mathfrak{N}_j'$ are mutually uncorrelated and proportional to the respective eigenvalues.

The effect of gradient noise on the weight vector can now be determined as follows. The LMS algorithm with a noisy gradient estimate can be expressed in accordance with (A.12) as

$$W_{j+1} = W_j + \mu(-\hat{\nabla}_j) = W_j + \mu(-\nabla_j + \mathfrak{N}_j). \tag{D.10}$$

Reexpressing (D.10) in terms of V_j, where V_j is defined as $W_j - W^*$, yields

$$V_{j+1} = V_j + \mu(-2RV_j + \mathfrak{N}_j). \tag{D.11}$$

Projecting into the primed coordinates by premultiplying both sides by Q^{-1} yields

$$V_{j+1}' = V_j' - 2\mu\Lambda V_j' + \mu\mathfrak{N}_j' = (I - 2\mu\Lambda)\, V_j' + \mu\mathfrak{N}_j'. \tag{D.12}$$

Note once again that, since the components of $\mathfrak{N}_j'$ are mutually uncorrelated and since (D.12) is diagonalized, the components of noise in V_j' are mutually uncorrelated.

[21] This common assumption is not strictly correct in this case but greatly simplifies the analysis and yields results that work well in practice.

Near the minimum point of the error surface, in steady state after adaptive transients have died out, the mean of V_j' is zero, and the covariance of the weight-vector noise may be obtained as follows. Postmultiplying both sides of (D.12) by their transposes and taking expected values yields

$$E[V_{j+1}' V_{j+1}'^T] = E[(I - 2\mu\Lambda) V_j' V_j'^T (I - 2\mu\Lambda)] + \mu^2 E[\mathfrak{N}_j' \mathfrak{N}_j'^T] + \mu E[\mathfrak{N}_j' V_j'^T (I - 2\mu\Lambda)] + \mu E[(I - 2\mu\Lambda) V_j' \mathfrak{N}_j'^T]. \quad (D.13)$$

It has been assumed that the input vector X_j is uncorrelated over time; the gradient noise $\mathfrak{N}_j$ is accordingly uncorrelated with the weight vector W_j, and therefore $\mathfrak{N}_j'$ and V_j' are uncorrelated. Equation (D.13) can thus be expressed as

$$E[V_{j+1}' V_{j+1}'^T] = (I - 2\mu\Lambda) E[V_j' V_j'^T] (I - 2\mu\Lambda) + \mu^2 E[\mathfrak{N}_j' \mathfrak{N}_j'^T]. \quad (D.14)$$

Furthermore, if V_j' is stationary, the covariance of V_{j+1}' is equal to the covariance of V_j', which may be expressed as

$$\text{cov}\,[V_j'] = (I - 2\mu\Lambda)\,\text{cov}\,[V_j']\,(I - 2\mu\Lambda) + \mu^2\,\text{cov}\,[\mathfrak{N}_j']. \quad (D.15)$$

Since the noise components of V_j' are mutually uncorrelated, (D.15) is diagonal. It can thus be rewritten as

$$\text{cov}\,[V_j'] = (I - 2\mu\Lambda)^2\,\text{cov}\,[V_j'] + \mu^2 (4\xi_{\min}\Lambda) \quad (D.16)$$

or

$$(I - \mu\Lambda)\,\text{cov}\,[V_j'] = \mu\xi_{\min}. \quad (D.17)$$

When the value of the adaptive constant μ is small (as is consistent with a converged solution near the minimum point of the error surface), it is implied that

$$\mu\Lambda \ll I. \quad (D.18)$$

Equation (D.17) thus becomes

$$\text{cov}\,[V_j'] = \mu\xi_{\min} I. \quad (D.19)$$

The covariance of V_j can now be expressed as follows:

$$\text{cov}\,[V_j] = E[V_j V_j^T] = E[Q V_j' V_j'^T Q^{-1}] = Q\,\text{cov}\,[V_j']\,Q^{-1} = \mu\xi_{\min} I \quad (D.20)$$

where the components of the weight-vector noise are all of the same variance and are mutually uncorrelated. This derivation of the covariance depends on the assumptions made above. It has been found by experience, however, that (D.20) closely approximates the exact covariance of the weight-vector noise under a considerably wider range of conditions than these assumptions imply. A derivation of bounds on the covariance based on fewer assumptions has been made by Kim and Davisson [40].

C. Noise in the Transfer Function

The filter weights, comprising the impulse response, undergo digital Fourier transformation to yield the transfer function. The noise in each of the weights is uncorrelated over time, uncorrelated from weight to weight, and of variance $\mu\xi_{\min}$. At the jth instant the impulse response has n samples, $w_{0j}, w_{1j}, \cdots, w_{kj}, \cdots, w_{n-1j}$, and their transform is

$$H_j(l) = \sum_{k=0}^{n-1} w_{kj} \exp(-i2\pi kl/n) \quad (D.21)$$

where l is the frequency index. For a single value of l, $H_j(l)$ is a linear combination of all the weights, each weighted by a phasor of unit magnitude. Since $H_j(l)$ is complex, the power of this noise is the sum of its "real" and "imaginary" power and equals the sum of the noise power in the weights themselves. Thus at each frequency l, the noise power in $H_j(l)$ is

$$n\mu\xi_{\min}. \quad (D.22)$$

In spectral analysis, "ensemble averaging" techniques are commonly used. The same approach could be used here, averaging the weights over time before transforming. Although the gradient noise is essentially uncorrelated over j, the weight-vector noise is generally highly correlated over time. Averaging with each adaptive iteration could be done but is not necessary; averaging the weight vector at intervals corresponding to about four adaptive time constants ($4\tau_{\text{mse}}$) would assure noise independence and would be appropriate in gathering the information contained in the time history of the weights. On this basis, averaging N weight vectors would produce, at the lth frequency, a noise power in $\overline{H(l)}$, the averaged transfer function, with the following value:

$$(n/N)\,\mu\xi_{\min}. \quad (D.23)$$

This expression for weight-vector noise can be put in more usable form by relating $\xi_{\min}$ to the physical line enhancing process shown in Fig. 33. The noise power at the filter output will always be negligible compared to the input noise power, since the optimized filter transfer function will be small in magnitude except at the peaks whose value is a^*. When signal power is low compared to noise power, which is the case of interest in the present context, the error power is essentially equal to the input noise power. Thus

$$\xi_{\min} = \nu^2. \quad (D.24)$$

The noise power in $\overline{H(l)}$ at the lth frequency is accordingly

$$(n/N)\,\mu\nu^2. \quad (D.25)$$

D. Detectability of Sine Waves by Adaptive Line Enhancing

Detection of a signal is dependent on identification of its adaptive filter transfer function peak (of value a^*) as distinct from other peaks due to weight-vector noise. For this purpose one could compare the value of a^* with the standard deviation of the noise in $\overline{H(l)}$. A still better procedure is to work with signal and noise power by comparing the squares of these quantities. "Detectability" for the adaptive line enhancer (ALE) is accordingly defined as follows:

$$D_{\text{ALE}} \triangleq \frac{(a^*)^2}{\text{noise power in } \overline{H(l)}}. \quad (D.26)$$

This measure must typically be one or greater to achieve signal detection. Using (D.2) and (D.22), equation (D.26) can be re-expressed as

$$D_{\text{ALE}} = \frac{(\text{SNR})^2 (n^2/4)}{(n/N)\,\mu\nu^2 [1 + (\text{SNR})(n/2)]^2}. \quad (D.27)$$

The power of the adaptive filter input is essentially that of the noise, equal to ν^2. Since the filter input is essentially white, the input correlation matrix can be well represented by

$$R = \nu^2 I. \quad (D.28)$$

All eigenvalues are equal to ν^2. The trace of R is equal to $n\nu^2$.

Using (A.19) of Appendix A, one may thus write

$$\tau_{\text{mse}} = \frac{n}{4\mu \operatorname{tr} R} = \frac{n}{4\mu n \nu^2} = \frac{1}{4\mu\nu^2}. \tag{D.29}$$

This is the time constant of the mean-square error learning curve. Note that the line enhancer does not have a bias weight and that the number of weights is thus n rather than $n + 1$. Equation (D.27) may now be expressed in more useful form as follows:

$$D_{\text{ALE}} = \left(\frac{4N\tau_{\text{mse}}}{n}\right)\left[\frac{(\text{SNR})\,(n/2)}{1 + (\text{SNR})\,(n/2)}\right]^2. \tag{D.30}$$

For high signal-to-noise ratios—that is, for $(\text{SNR})\,(n/2) \gg 1$—equation (D.30) becomes

$$D_{\text{ALE}} \cong \frac{4N\tau_{\text{mse}}}{n}. \tag{D.31}$$

For low signal-to-noise ratios—that is, for $(\text{SNR})\,(n/2) \ll 1$—equation (D.30) shows that

$$D_{\text{ALE}} \cong (N\tau_{\text{mse}})\,(\text{SNR})^2\, n. \tag{D.32}$$

Intermediate values must be independently calculated.

Choice of the number of weights has an influence on the value of D_{ALE} for a given input signal-to-noise ratio. Differentiating (D.30) with respect to n and setting the derivative to zero yields the following expression for the optimal value of n:

$$n^* = 2/\text{SNR}. \tag{D.33}$$

Substituting (D.33) into (D.30) then yields the optimal value[22] of D_{ALE}:

$$D^*_{\text{ALE}} = (N\tau_{\text{mse}})\,(\text{SNR}/2). \tag{D.34}$$

It is interesting to note that, when n is so optimized,

$$a^* = 1/2. \tag{D.35}$$

E. Detectability of Sine Waves by Spectral Analysis

Let the power spectrum of a signal in white Gaussian noise be derived from an L-point digital Fourier transform. The frequency of the signal is assumed to be at the center of a spectral bin. Input signal power is assumed to be $C^2/2$ and noise power to be ν^2. At the signal frequency, the component of the power spectrum due to the signal will have the value $C^2L^2/4$. Each spectral bin will have an identical average noise power of $\nu^2 L$.

For the signal to be detected its spectral peak must be distinguishable from noise peaks that are deviations about the mean noise power. The variance of the noise power about the mean can be reduced by ensemble averaging; that is, by averaging N power spectra, each derived from L data points. With Gaussian noise the variance of the noise power about its mean in any spectral bin can be shown [23] to be $(2/N)$ (average noise power)^2, which is equivalent to

$$(2/N)\,(\nu^2 L)^2. \tag{D.36}$$

[22] The exact value of n is not critical; it may be as much as 8 times larger or smaller than n^* and D_{ALE} will remain within approximately 50 percent of D^*_{ALE}.

[23] The variance in the estimate of variance from N samples of a zero-mean process equals (mean fourth − [mean square]^2)/N.

It is reasonable to compare the average signal power in the selected spectral bin with the standard deviation of the noise power fluctuations that occur in each spectral bin; that is, with the square root of (D.36). We thus define "detectability" for spectral analysis as

$$D_{\text{DFT}} \triangleq \frac{C^2L^2/4}{(2/N)^{1/2}\nu^2 L} = \frac{1}{2}\,(\text{SNR})\,L\,(N/2)^{1/2}$$
$$= (\text{SNR})\,L\,(N/8)^{1/2}. \tag{D.37}$$

The motive for this definition is derived from the early work of Woodward [44], Skolnick [45], Swerling [46], Marcum [47], and others.

F. Comparison of Adaptive Line Enhancing and Spectral Analysis

Fig. 35 illustrates the definitions of the detectability of a sine wave by adaptive line enhancing and spectral analysis given in (D.30) and (D.37). It is useful to compare Fig. 35(a) with Fig. 35(b). Note that in the former case the measure of detectability is based on the magnitude of the adaptive filter transfer function, whereas in the latter it is based on the digital power spectrum. Since the measure of detectability is different for the two techniques, in a sense one is comparing "apples and oranges." Yet both D_{ALE} and D_{DFT} are ratios of signal power to noise power.

Fig. 36 presents experimental results, obtained by computer simulation, showing the performance of the adaptive line enhancing and spectral analysis techniques for three values of D_{ALE} and D_{DFT}. Visual examination indicates that D_{ALE} and D_{DFT} do provide a reasonable basis for comparing the performance of the two techniques.

Equation (D.34) describes the detectability of a sine wave by the adaptive technique when n is optimized. This equation can be rewritten as

$$D^*_{\text{ALE}} = (4N\tau_{\text{mse}})\,(\text{SNR}/8). \tag{D.38}$$

Since weight vectors are taken for ensemble averaging at $4\tau_{\text{mse}}$ intervals and N vectors are averaged, $4N\tau_{\text{mse}}$ represents the total number of input data samples. Note that the time constant τ_{mse} is not expressed in seconds but in number of adaptive iterations, which is equivalent to number of input data samples. Thus (D.38) can be rewritten as

$$D^*_{\text{ALE}} = (\text{number of data samples})\,(\text{SNR}/8). \tag{D.39}$$

The detectability of a sine wave by spectral analysis is given by (D.37). Since N sample spectra are ensemble-averaged, and since each requires L data points, the number of data samples required is the product of N and L. Equation (D.37) can thus be rewritten as

$$D_{\text{DFT}} = (\text{number of data samples})\,(\text{SNR}/[8N]^{1/2}). \tag{D.40}$$

The ratio of detectabilities is, therefore,

$$\frac{D^*_{\text{ALE}}}{D_{\text{DFT}}} = \left(\frac{N}{8}\right)^{1/2}. \tag{D.41}$$

Accordingly, spectral analysis is advantageous as long as the number of ensemble members is less than eight. Adaptive line enhancing would be advantageous when the number of ensemble members required for spectral analysis is greater than eight.

For the comparative experiment represented by Fig. 36, input signal-to-noise ratio in each case was 0.01562. The number of data samples used with spectral analysis was the same as

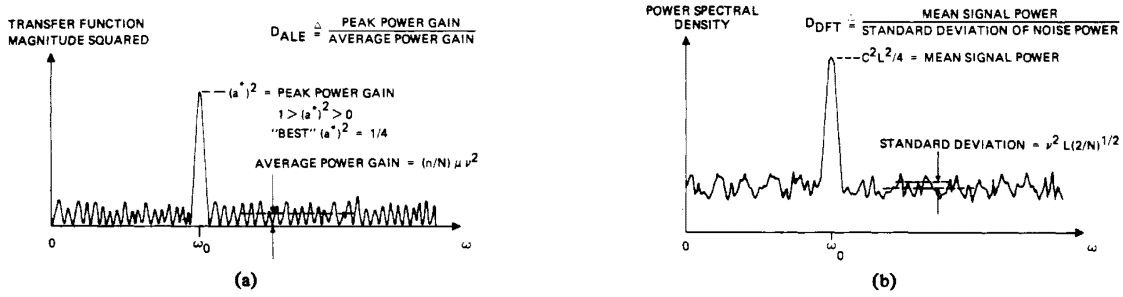

Fig. 35. Definition of detectability D of a sine wave in noise. (a) With adaptive line enhancing. (b) With spectral analysis.

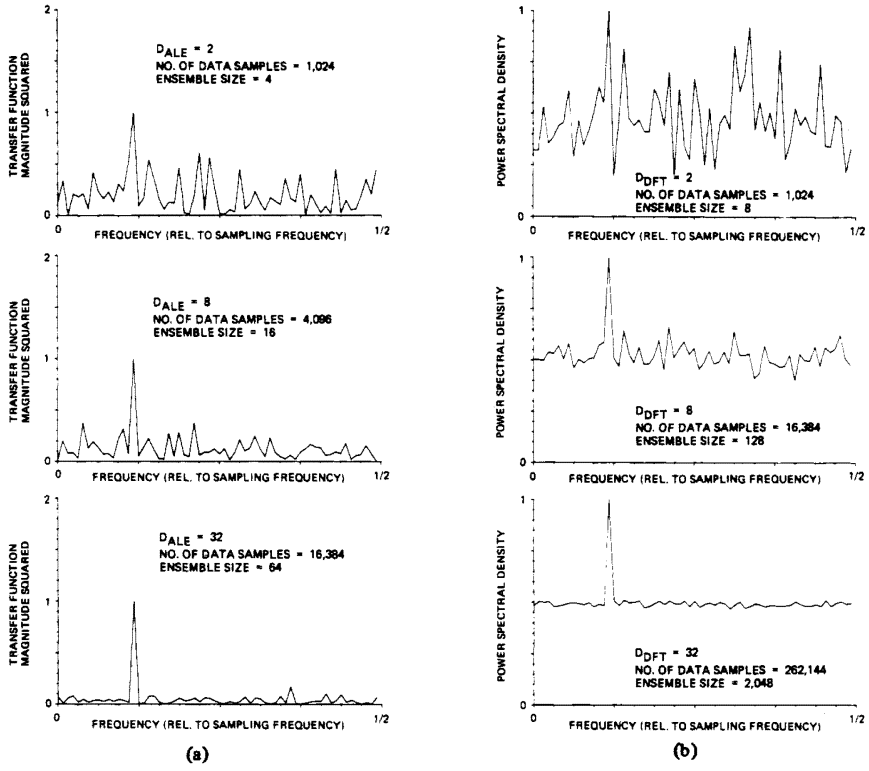

Fig. 36. Experimental comparison of adaptive line enhancing and spectral analysis for three values of detectability D; input signal-to-noise ratio, 0.01562; number of weights and transform points, 128. (a) Adaptive line enhancing. (b) Spectral analysis.

the number used with adaptive line enhancing when the value of D_{ALE} and D_{DFT} was 2 but became 16 times greater when the value of D_{ALE} and D_{DFT} was 32.

With adaptive line enhancing one could freely trade N for τ_{mse}. Their product is all that is important. Ensemble averaging may not even be required, since τ_{mse} can be made large by making μ small (although this may cause one to go to "double precision" arithmetic). With spectral analysis, on the other hand, ensemble averaging cannot be avoided in most cases. The size of L may be limited by cost considerations, computer speed, or in the case where the signal is an imperfect or modulated sine wave by signal bandwidth. Large values of N are required when input signal-to-noise ratio is low, and values in the thousands are not uncommon.

The reason that adaptive methods may be superior to spectral methods in certain cases, especially those of low signal-to-noise ratio, can be stated as follows. Averaging within the digital Fourier transform itself provides coherent signal enhancement. Thus the detectability D_{DFT} of the signal is proportional to L. Since ensemble averaging is incoherent ("postdetection averaging"), however, the detectability D_{DFT} is proportional only to the square root of N. The adaptive process, on the other hand, provides coherent signal averaging, making the detectability D_{ALE} proportional to τ_{mse}. It is equally coherent in averaging the weight vector ensemble, making D_{ALE} proportional also to N.

An analytical comparison of the computational requirements of the two techniques has not yet been made, but it appears that the adaptive process will provide a simpler implementation when spectral analysis involves large values of L. The adaptive process has the advantage of being a smooth, steadily flowing process, whereas spectral analysis is performed with consecutive time segments of data.

The subject of signal detection by adaptive filtering is relatively new, and the analysis presented here should be regarded as preliminary. The formulas derived have been verified by

simulation and experiment, but the concepts they describe have not been in existence long enough to provide an adequate perspective. It is hoped that this work can be extended in the future.

Acknowledgment

Many people have contributed support, assistance, and ideas to the work described in this paper. The authors especially wish to acknowledge the contributions of Prof. T. Kailath, Prof. M. Hellman, Dr. H. Garland, J. Treichler, and M. Larimore of Stanford University; Prof. L. Griffiths of the University of Colorado; Dr. D. Chabries and M. Ball of the Naval Undersea Center; Dr. O. Frost of Argo Systems, Inc.; Dr. M. Hoff of the Intel Corp.; and the students in two classes in the Department of Electrical Engineering at Stanford University: EE 280, Computer Applications Laboratory, and EE 373, Adaptive Systems. Special thanks are also due to R. Fraser of the Naval Undersea Center, who assisted in editin the paper; his efforts led to significant improvements in i ; organization and clarity.

References

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

[2] H. Bode and C. Shannon, "A simplified derivation of linear least squares smoothing and prediction theory," *Proc. IRE*, vol. 38, pp. 417-425, Apr. 1950.

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

[4] 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.

[5] T. Kailath, "A view of three decades of linear filtering theory," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 145-181, Mar. 1974.

[6] P. Howells, "Intermediate frequency side-lobe canceller," U.S. Patent 3 202 990, Aug. 24, 1965.

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

[8] J. Koford and G. 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.

[9] F. Rosenblatt, "The Perceptron: A perceiving and recognizing automaton, Project PARA," Cornell Aeronaut. Lab., Rep. 85-460-1, Jan. 1957.

[10] —, *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms.* Washington, D.C.: Spartan Books, 1961.

[11] N. Nilsson, *Learning Machines.* New York: McGraw-Hill, 1965.

[12] D. Gabor, W. P. L. Wilby, and R. Woodcock, "A universal nonlinear filter predictor and simulator which optimizes itself by a learning process," *Proc. Inst. Elec. Eng.*, vol. 108B, July 1960.

[13] R. Lucky, "Automatic equalization for digital communication," *Bell Syst. Tech. J.*, vol. 44, pp. 547-588, Apr. 1965.

[14] R. Lucky *et al.*, *Principles of Data Communication.* New York: McGraw-Hill, 1968.

[15] J. Kaunitz, "Adaptive filtering of broadband signals as applied to noise cancelling," Stanford Electronics Lab., Stanford Univ., Stanford, Calif., Rep. SU-SEL-72-038, Aug. 1972 (Ph.D. dissertation).

[16] M. Sondhi, "An adaptive echo canceller," *Bell Syst. Tech. J.*, vol. 46, pp. 497-511, Mar. 1967.

[17] J. Rosenberger and E. Thomas, "Performance of an adaptive echo canceller operating in a noisy, linear, time-invariant environment," *Bell Syst. Tech. J.*, vol. 50, pp. 785-813, Mar. 1971.

[18] R. Riegler and R. Compton, Jr., "An adaptive array for interference rejection," *Proc. IEEE*, vol. 61, pp. 748-758, June 1973.

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

[20] —, "Adaptive filters," in *Aspects of Network and System Theory*, R. Kalman and N. DeClaris, Eds. New York: Holt, Rinehart, and Winston, 1971, pp. 563-587.

[21] J. Glover, "Adaptive noise cancelling of sinusoidal interferences," Ph.D. dissertation, Stanford Univ., Stanford, Calif., May 1975.

[22] J. C. Huhta and J. G. Webster, "60-Hz interference in electrocardiography," *IEEE Trans. Biomed. Eng.*, vol. BME-20, pp. 91-101, Mar. 1973.

[23] W. Adams and P. Moulder, "Anatomy of heart," in *Encycl. Britannica*, vol. 11, pp. 219-229, 1971.

[24] G. von Anrep and L. Arey, "Circulation of blood," in *Encycl. Britannica*, vol. 5, pp. 783-797, 1971.

[25] R. R. Lower, R. C. Stofer, and N. E. Shumway, "Homovital transplantation of the heart," *J. Thoracic and Cardiovascular Surgery*, vol. 41, p. 196, 1961.

[26] T. Buxton, I. Hsu, and R. Barter, "Fetal electrocardiography," *J.A.M.A.*, vol. 185, pp. 441-444, Aug. 10, 1963.

[27] J. Roche and E. Hon, "The fetal electrocardiogram," *Amer. J. Obst. and Gynecol.*, vol. 92, pp. 1149-1159, Aug. 15, 1965.

[28] S. Yeh, L. Betyar, and E. Hon, "Computer diagnosis of fetal heart rate patterns," *Amer. J. Obst. and Gynecol.*, vol. 114, pp. 890-897, Dec. 1, 1972.

[29] E. Hon and S. Lee, "Noise reduction in fetal electrocardiography," *Amer. J. Obst. and Gynecol.*, vol. 87, pp. 1087-1096, Dec. 15, 1963.

[30] J. Van Bemmel, "Detection of weak foetal electrocardiograms by autocorrelation and crosscorrelation of envelopes," *IEEE Trans. Biomed. Eng.*, vol. BME-15, pp. 17-23, Jan. 1968.

[31] J. R. Cox, Jr., and L. N. Medgyesi-Mitschang, "An algorithmic approach to signal estimation useful in fetal electrocardiography," *IEEE Trans. Biomed. Eng.*, vol. BME-16, pp. 215-219, July 1969.

[32] J. Van Bemmel, L. Peeters, and S. Hengeveld, "Influence of the maternal ECG on the abdominal fetal ECG complex," *Amer. J. Obst. and Gynecol.*, vol. 102, pp. 556-562, Oct. 15, 1968.

[33] W. Walden and S. Birnbaum, "Fetal electrocardiography with cancellation of maternal complexes," *Amer. J. Obst. and Gynecol.*, vol. 94, pp. 596-598, Feb. 15, 1966.

[34] J. Capon, R. J. Greenfield, and R. J. Kolker, "Multidimensional maximum likelihood processing of a large aperture seismic array," *Proc. IEEE*, vol. 55, pp. 192-211, Feb. 1967.

[35] S. P. Applebaum, "Adaptive arrays," Special Projects Lab., Syracuse Univ. Res. Corp., Rep. SPL 769.

[36] L. J. Griffiths, "A simple adaptive algorithm for real-time processing in antenna arrays," *Proc. IEEE*, vol. 57, pp. 1696-1704, Oct. 1969.

[37] O. L. Frost, III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, pp. 926-935, Aug. 1972.

[38] K. Senne, "Adaptive linear discrete-time estimation," Stanford Electronics Lab., Stanford Univ., Rep. SEL-68-090, June 1968 (Ph.D. dissertation).

[39] T. Daniell, "Adaptive estimation with mutually correlated training samples," Stanford Electronics Lab., Stanford Univ., Rep. SEL-68-083, Aug. 1968 (Ph.D. dissertation).

[40] J. K. Kim and L. D. Davisson, "Adaptive linear estimation for stationary M-dependent processes," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 23-31, Jan. 1975.

[41] B. Widrow, "Adaptive filters 1: Fundamentals," Stanford Electronics Lab., Stanford Univ., Rep. SU-SEL-66-126, Dec. 1966.

[42] L. J. Griffiths, "Rapid measurement of instantaneous frequency," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-23, pp. 209-222, Apr. 1975.

[43] J. P. Burg, "Maximum entropy spectral analysis," presented at the 37th Annual Meeting, Soc. Exploration Geophysicists, Oklahoma City, Okla., 1967.

[44] P. M. Woodward, *Probability and Information Theory with Applications to Radar*, 2nd ed. London: Pergamon Press, 1964.

[45] M. I. Skolnik, *Introduction to Radar Systems.* New York: McGraw-Hill, 1962.

[46] P. Swerling, "Probability of detection for fluctuating targets," *IRE Trans. Inform. Theory*, vol. IT-6, pp. 269-308, Apr. 1960.

[47] J. I. Marcum, "A statistical theory of target detection by pulsed radar: Mathematical appendix," *IRE Trans. Inform. Theory*, vol. IT-6, pp. 145-267, Apr. 1960.