FIR System Modeling and Identification in the Presence of Noise and with Band-Limited Inputs

LAWRENCE R. RABINER, FELLOW, IEEE, RONALD E. CROCHIERE, MEMBER, IEEE, AND JONT B. ALLEN, MEMBER, IEEE

Abstract—System identification, that is, the modeling and identification of a system from knowledge of its input and output signals, is a subject that is of considerable importance in many areas of signal and data processing. Because of the diversity of applications, a number of different methods for system identification with different advantages and disadvantages have been described and used in the literature. In this paper we investigate the performance of three well-known system identification methods based on an FIR (finite impulse response) model of the system. The methods will be referred to in this paper as the least squares analysis (LSA) method, the least mean squares adaptation algorithm (LMS), and the short-time spectral analysis (SSA) procedure.

Our particular interest in this paper concerns the performance of these algorithms in the presence of high noise levels and in situations where the input signal may be band-limited. Both white and nonwhite random noise signals as well as speech signals are used as test signals to measure the performance of each of the system identification techniques as a function of the signal-to-noise ratio of the systems output. Quantitative results in terms of an accuracy measure of system identification are presented and a simple analytical model is used to explain the measured results.

I. INTRODUCTION

THE AREA of system identification is one of the most important areas in engineering because of its applicability to a wide range of problems [1]-[6]. As such, a great deal of research has been carried out in studying the properties of a wide variety of algorithms for performing system identification [4]-[6]. As a result, system identification techniques are generally well known and understood for a wide variety of applications. The purpose of this paper is to compare and contrast the performance of three system identification techniques for a class of signals which is characteristic of those obtained from speech waveform coders [11]. Our intended application is to be able to characterize a digital waveform coder in terms of a time-varying, linear system (correlated signal component), and an additive uncorrelated noise component, in order to make meaningful objective evaluations of such coders. In this paper, however, we will restrict our attention to aspects of the system identification problem. Some results on speech waveform coder characterization are given in [20].

Fig. 1 shows the conventional system identification model. The input signal is $x(n)$ and the output signal is $y(n)$. The output is modeled as having been obtained via linear filtering of the input, followed by the addition of an uncorrelated white noise signal $e(n)$, i.e.,

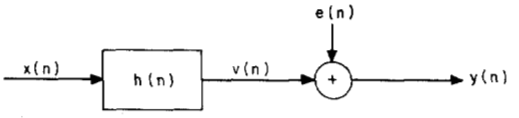

Fig. 1. Block diagram of the linear system model used for system identification.

$$y(n) = x(n) * h(n) + e(n) \tag{1}$$

$$= \sum_{m=0}^{M-1} h(m)\, x(n-m) + e(n) \tag{2}$$

with

$$\overline{e(n)\, x(n)} = \overline{e(n)\, v(n)} = 0 \tag{3}$$

$$\overline{e(n)\, e(n-m)} = \sigma_e^2 \delta(m). \tag{4}$$

It is tacitly assumed in (2) that the impulse response of the linear system $h(n)$, is of finite duration (M samples) or can be effectively modeled with a finite impulse response system.

The particular class of input signals in which we are interested has the following properties.

1) The noise level $e(n)$ at the output of the linear system is fairly large—i.e., we are interested in systems with signal-to-noise ratios in the range of 0-24 dB.

2) The input signal $x(n)$ is generally band-limited and has a distinct spectral slope. For speech coding $x(n)$ is usually band-limited to about 3 kHz, and sampled at rates from about 8 kHz to 16 kHz. Furthermore, the spectrum of speech is highly nonuniform and falls rapidly for frequencies above about 2 kHz. Also, for voiced speech the spectrum is a line spectrum containing significant energy only in a set of harmonics of the pitch period. Between harmonics the spectrum often falls as much as 20-40 dB.

3) The system $h(n)$ is generally time-varying and, in addition, it may be nonlinear. The time-varying nature of $h(n)$ is due to the nonstationarity of a speech signal (especially the signal level) which causes the operating region of speech coders to vary from slope overload to granularity with different talkers, transmission media, etc. As such the linear system characterization of the coder must adapt with time.

Manuscript received November 17, 1977; revised April 4, 1978.
The authors are with Bell Laboratories, Murray Hill, NJ 07974.

0096-3518/78/0800-0319$00.75

4) The duration of the impulse response, M, is generally unknown and could be relatively long in some cases.

Because of the importance of understanding both the limitations and advantages of the available system identification techniques, a study was performed on an artificially created signal with a known linear system and in a known noise background. The purpose of the simulation study was to measure the linear system estimation error as a function of the following:

1) Duration of the signal used to make the estimate, N;

2) Assumed duration of the linear system impulse response, $\hat{M}$;

3) Signal-to-noise ratio of the system, S/N, where

$$S/N = 10 \log_{10} \left[\frac{\sigma_v^2}{\sigma_e^2} \right] \tag{5}$$

where σ_v^2 is the variance of the signal $v(n)$ and σ_e^2 is the variance of the noise $e(n)$ (see Fig. 1);

4) Type of unknown system to be identified, $h(n)$;

5) Input signal characteristics—i.e., the bandwidth of the input signal, its spectral shape, etc.

In order to quantify the ideas to be presented here, an analytical measure of the accuracy of estimation is used which has the form[1]

$$Q = 10 \log_{10} \left[\frac{\sum_{n=0}^{\hat{M}-1} [h(n) - \hat{h}(n)]^2}{\sum_{n=0}^{\hat{M}-1} h^2(n)} \right] \tag{6}$$

where $\hat{h}(n)$ is the estimated linear system and $\hat{M}$ is its assumed length. This measure is the log of the normalized norm of the "misadjustment" or "misalignment" vectors, referred to, respectively, in [5] and [6], expressed in dB. It is shown that for white input signals, the measure of (6) provides a good description of the performance of a system identification method. For nonwhite input signals, a modified accuracy measure Q' is developed whose properties are analogous to those of the Q measure. Both analytical and measured curves of Q (and Q' when appropriate), as a function of the system parameters N, $\hat{M}$, S/N and $h(n)$, are presented.

The organization of this paper is as follows. In Section II we review the three system identification methods which were used in this study and explain why they were chosen for the intended application. In Section III we present both analytical and experimentally obtained results of the performance of the three system identification methods on the artificially created signals. In Section IV we expand the results to include actual speech inputs as well as band-limited noise. Finally, in Section V we compare and contrast the three methods and highlight the possible advantages and disadvantages of each algorithm.

[1]Throughout this paper we are assuming $\hat{M} \geq M$, i.e., a valid estimate of M is available and $\hat{M}$ is at least as large as this estimate.

II. System Identification Methods

The three system identification methods used in this study were the classical least squares analysis (LSA), the least squares adaptation algorithm (LMS), and a short-time spectral analysis (SSA) procedure. The reasons these three particular methods were chosen were because of their applicability to a wide range of problems (especially in the area of speech processing [4], [6], [8], [9], [12], [13]) and the fact that each of these methods had distinct advantages in certain situations. For example, the least squares analysis algorithm is a simple time-domain method for estimating a linear system from a block of data using efficient recursion methods to solve a matrix equation. Methods similar to this have found wide use in speech processing [12]. The least mean squares adaptation algorithm is a sample-by-sample adaptive method for recursively updating linear system estimates and is especially useful for efficiently estimating linear, slowly time-varying systems of large order [8]. It has been particularly useful for applications such as adaptive echo cancelers, adaptive line equalizers, etc. Spectral estimation methods attempt to identify the linear system from short-time spectral data, rather than from time-domain solutions. Potentially such methods have the capability of estimating high-order systems (i.e., systems with long impulse responses) without the need for recursive matrix inversions or successive update methods when the analysis is implemented using FFT techniques. Also, with the advent of high-speed inexpensive FFT chips (implemented in CCD technology), the potential low cost of such analysis methods makes them attractive.

In this section we review the specific algorithms implemented for this study. For completeness, we begin by describing the classical least squares analysis, followed by the least mean squares adaptation algorithm. We then describe the spectral estimation method.

A. Least Squares Analysis (LSA)

Based on the model of Fig. 1 we assume that the output $y(n)$ is related to the input $x(n)$ *exactly* by (1) and (2) where M is the true duration of the impulse response $h(n)$. For the least squares method we assume that $h(n)$, $e(n)$, and M are all unknown, and we wish to make an optimal estimate, $\hat{h}(n)$, of $h(n)$. The assumed duration of $\hat{h}(n)$ is $\hat{M}$ samples.

For LSA we form the estimate

$$y(n) = \sum_{m=0}^{M-1} \hat{h}(m)\, x(n-m) + \hat{e}(n) \tag{7}$$

or

$$\hat{e}(n) = y(n) - \sum_{m=0}^{M-1} \hat{h}(m)\, x(n-m). \tag{8}$$

The optimization criterion is to minimize the norm of $\hat{e}(n)$ over the set of coefficients of $\hat{h}(n)$, i.e.,

$$\min_{\hat{h}(n)} \|\hat{e}(n)\|^2 = \min_{\hat{h}(n)} \left[\sum_{n=0}^{N-1} [\hat{e}(n)]^2 \right] \tag{9}$$

where N is the frame of samples of $x(n)$ and $y(n)$ used in the system estimate. The solution to (9) is a classical one, obtained by differentiating (9) with respect to $\hat{h}(k)$, $k = 0, 1, 2, \cdots, \hat{M} - 1$, and setting the result to 0, leading to the set of equations

$$\sum_{m=0}^{\hat{M}-1} \hat{h}(m) \sum_{n=0}^{N-1} x(n-k)\,x(n-m) = \sum_{n=0}^{N-1} y(n)\,x(n-k) \qquad k = 0, 1, 2, \cdots, \hat{M} \quad (10)$$

or by defining

$$\tilde{\phi}_{xx}(k, m) = \sum_{n=0}^{N-1} x(n-k)\,x(n-m) \quad (11)$$

$$\tilde{\phi}_{xy}(k) = \sum_{n=0}^{N-1} y(n)\,x(n-k) \quad (12)$$

the solution is compactly written as

$$\sum_{m=0}^{\hat{M}-1} \hat{h}(m)\,\tilde{\phi}_{xx}(k, m) = \tilde{\phi}_{xy}(k) \qquad k = 0, 1, \cdots, \hat{M} - 1. \quad (13)$$

The set of equations given above is efficiently solved via a recursive procedure known as the Cholesky decomposition [12]. Equation (13) is sometimes referred to as the discrete Weiner-Hopf equation [14].

If we adopt a matrix notation to solve for the classical least squares solution to the system identification problem we gain some new insights into the solution. If we let

$$y^t = [y(0), y(1), \cdots, y(N-1)] \quad (14)$$

$$h^t = [\hat{h}(0), \hat{h}(1), \cdots, \hat{h}(\hat{M}-1)] \quad (15)$$

then $\|\hat{e}(n)\|^2$ can be expressed as

$$\|\hat{e}(n)\|^2 = (\hat{h} - \tilde{\phi}_{xx}^{-1}\tilde{\phi}_{xy})^t\,\tilde{\phi}_{xx}(\hat{h} - \tilde{\phi}_{xx}^{-1}\tilde{\phi}_{xy}) + y^t y - \tilde{\phi}_{xy}^t\tilde{\phi}_{xx}^{-1}\tilde{\phi}_{xy}. \quad (16)$$

This form clearly shows the quadratic nature of the problem and it is easily seen that to minimize $\|\hat{e}(n)\|^2$ with respect to $\hat{h}$, $\hat{h}$ must be

$$\hat{h} = \tilde{\phi}_{xx}^{-1}\tilde{\phi}_{xy} \quad (17)$$

giving for the minimum residual,

$$\|\hat{e}(n)\|_{\min}^2 = y^t y - \tilde{\phi}_{xy}^t\hat{h} \quad (18)$$

$$= \sum_{n=0}^{N-1} y^2(n) - \sum_{m=0}^{\hat{M}-1} \hat{h}(m) \sum_{n=0}^{N-1} y(n)\,x(n-m). \quad (19)$$

If $\tilde{\phi}_{xx}^{-1}$ does not exist, or is ill-conditioned, the pseudoinverse of $\tilde{\phi}_{xx}$ is used in the solution to (17).

B. Least Mean Square Adaptation Algorithm (LMS)

The least mean squares adaptation algorithm [2] is an iterative, minimum-seeking method for determining the least squares solution of (13). Assuming that $\hat{h}_i$ is an estimate of $\hat{h}$ at the ith iteration, the new estimate $\hat{h}_{i+1}$ is determined as

$$\hat{h}_{i+1} = \hat{h}_i - u\nabla_i \quad (20)$$

where ∇_i is the gradient of $\|\hat{e}(n)\|^2$ with respect to $\hat{h}$ and u is a constant. Basically $-\nabla_i$ determines the direction in which the correction is made for the $i+1$ iteration and u is a constant which controls the size of the step taken in that direction. Since $\|\hat{e}(n)\|^2$ is a quadratic function of $\hat{h}$, is a single minimum exists in the error surface and it can be shown (for the sampled data case) that the algorithm converges to this minimum (17) if

$$u < 1/(\hat{M}\sigma_x^2) \quad (21)$$

i.e., if the step size is not too large, where σ_x^2 is the variance of $x(n)$ (assuming zero mean) [4]. Generally, a much smaller value of u is chosen and frequently u is modified as the error decreases.

By differentiating (16) with respect to $\hat{h}$ and applying definitions (2), (3), (11), and (12), the gradient ∇ at the ith step can be shown to be [4], [5]

$$\nabla = \frac{\partial\|\hat{e}(n)\|^2}{\partial\hat{h}} = -2\tilde{\phi}_{xx}(h - \hat{h}). \quad (22)$$

With the aid of (2), (3), (7), and (11) the gradient of a single coefficient $\hat{h}(m)$ can be determined as [4], [5]

$$\frac{\partial\|\hat{e}(n)\|^2}{\partial\hat{h}(m)} = -2E[x(n-m)\,\hat{e}(n)] \quad (23)$$

where $E[\ ]$ denotes the expected value (ensemble average over random input).

The LMS adaptation algorithm frequently uses as an estimate of (23) the gradient of a single error

$$\frac{\partial\|\hat{e}(n)\|^2}{\partial\hat{h}(m)} \approx -2x(n-m)\,\hat{e}(n) \quad (24)$$

where $\hat{e}(n)$ is determined from (8). New estimates of $\hat{h}$ are then computed on a sample-by-sample basis as data samples $x(n)$ and $y(n)$ become available. The new estimate of the mth coefficient of $\hat{h}$ is then computed as

$$\hat{h}_{n+1}(m) = \hat{h}_n(m) + 2ux(n-m)\,\hat{e}(n). \quad (25)$$

Since the choice of u depends on the variance of $x(n)$ as shown by (21), we used in this work a self-normalizing form of the LMS adaptation algorithm,

$$\hat{h}_{n+1}(m) = \hat{h}_n(m) + Kx(n-m)\,\hat{e}(n)/(\sigma_x^2\hat{M}) \quad (26)$$

where

$$\hat{M}\sigma_x^2 = \sum_{m=0}^{\hat{M}-1} x^2(n-m). \quad (27)$$

Comparing (25) with (26) and (27) it is seen that

$$u \cong K/(2M\sigma_x^2) \quad (28)$$

and the algorithm converges if $K < 2$. It can be seen from (28)

that the effective value of u is normalized by σ_x^2 making the effective step size independent of the input signal level. In practice, this self-normalized form of the algorithm is not often used because of the added complexity of computing σ_x^2. This form is used here to make our results independent of signal level.

It should be noted that other variations of the LMS adaptation algorithm also exist which can achieve more accurate estimates of the gradient at a cost of more computation [5]. These methods have not been explored in this paper.

C. Short-Time Spectral Analysis (SSA)

Spectral procedures have been in use traditionally both for direct and indirect methods of estimating signal power spectra [15], [16]. Recent advances in the theory of short-time spectral analysis have provided a framework for implementing a system identification procedure entirely in the frequency domain [13], [17]. To illustrate this procedure, consider the linear system of Fig. 1. If we define the (infinite-time) z-transforms of $x(n)$, $y(n)$ and $e(n)$ as $X(z)$, $Y(z)$ and $E(z)$ (recalling that $E(z)$ is not well defined for $e(n)$, a noise sequence), then it is seen that

$$Y(z) = X(z)\,H(z) + E(z) \tag{29}$$

and thus an obvious procedure for estimating $H(z)$ would be to form the estimate

$$\hat{H}(z) = \frac{Y(z)}{X(z)} = H(z) + \frac{E(z)}{X(z)}. \tag{30}$$

The justification for the estimate of (30) is that for practical implementations using a finite section of data, the second term in (30) will average to 0, leaving the first term which is the true $H(z)$. Aside from the many practical issues concerning use of windows to implement the analysis, the estimate of (30) is notoriously slow in converging to the true $H(z)$, and is subject to extremely large errors (erroneous values of $H(z)$) when $X(z)$ is small for some value of z, e.g., when the input signal is band-limited. Recent unpublished work by M. M. Sondhi has also shown that (under some conditions) although the expected value of the estimate of (30) converges to the desired value (as viewed through the analysis window), the *variance of the estimate is infinite.*

A somewhat more robust and sophisticated analysis procedure is to use power spectrum estimation methods to estimate $\hat{H}(z)$. Consider the power spectrum of the input, $S_{xx}(z)$, defined as

$$S_{xx}(z) = X(z)\,X^*(z) \tag{31}$$

and the cross-power spectrum between x and y, $S_{xy}(z)$, defined as

$$S_{xy}(z) = Y(z)\,X^*(z). \tag{32}$$

It is readily shown that on an infinite-time basis, the cross-power spectrum satisfies the relationship

$$S_{xy}(z) = Y(z)\,X^*(z) = H(z)\,X(z)\,X^*(z) + E(z)\,X^*(z). \tag{33}$$

Short-time estimates of $\hat{S}_{xx}(z)$ and $\hat{S}_{xy}(z)$ can now be defined as

$$\hat{S}_{xx}(z) = X_r(z)\,X_r^*(z) \tag{34}$$

and

$$\hat{S}_{xy}(z) = Y_r(z)\,X_r^*(z) \tag{35}$$

where r corresponds to the rth block or windowed segment of samples of $x(n)$ and $y(n)$ from which estimates $X_r(z)$ and $Y_r(z)$ are obtained. A reasonable implementation of the spectral estimate, using short-time spectral estimation methods, is

$$\hat{H}(k) = \hat{H}(z)\Big|_{z=e^{j\omega_k}} = \frac{\sum_{r=1}^{L} Y_r(e^{j\omega_k})\,X_r^*(e^{j\omega_k})}{\sum_{r=1}^{L} X_r(e^{j\omega_k})\,X_r^*(e^{j\omega_k})} \tag{36}$$

where the summation on r includes L overlapping time segments of $x(n)$ and $y(n)$ [13], [16]. In the limit, as the number of time segments, L, becomes large, it can be seen, with the aid of (33) and (36), that

$$\lim_{L\to\infty} \hat{H}(z) = \lim_{L\to\infty} \frac{\sum_{r=1}^{L} H_r(z)\,X_r(z)\,X_r^*(z)}{\sum_{r=1}^{L} X_r(z)\,X_r^*(z)} + \lim_{L\to\infty} \frac{\sum_{r=1}^{L} E_r(z)\,X_r^*(z)}{\sum_{r=1}^{L} X_r(z)\,X_r^*(z)}. \tag{37}$$

It can be shown that the first term on the right of (37) converges to $H(z)$ (assuming $h(n)$ is not time varying) and the second term on the right of (37) converges to zero due to the fact that $e(n)$ and $x(n)$ are uncorrelated. Therefore, in the limit, as L becomes large the estimate $\hat{H}(z)$ in (36) converges to $H(z)$.

The formal definition of the short-time spectrum of a signal $x(n)$ at the time sample n is

$$X_n(e^{j\omega_k}) = \sum_m x(m)\,w(n-m)\,e^{-j\omega_k m} \tag{38}$$

where the finite duration low-pass window $w(n)$ determines both the temporal and spectral resolution of the estimates. Unfortunately for the procedure used here, the window also manifests itself directly in the estimation of $\hat{H}(k)$ in a complex manner. By way of example, if $y(n)$ is a pure delay of $x(n)$, i.e.,

$$y(n) = x(n - k_0) \tag{39}$$

for a fixed value of k_0, i.e., $h(n) = 1$ for $n = k_0$, and 0 otherwise, then it can easily be shown that using (36) leads to the estimate (see Appendix I)

$$\hat{H}(\omega) = \frac{\mathbf{F}[R_x(k_0 - n)\,R_w(n)]}{\mathbf{F}[R_x(-n)\,R_w(n)]} \tag{40}$$

where $R_x(n)$ and $R_w(n)$ represent autocorrelation functions of the input signal and the window, respectively, and F represents the Fourier transform. Thus (as in the case of white noise), if we assume that

$$R_x(m) = \sigma_e^2 \delta(m) \tag{41}$$

then (40) says that

$$\hat{H}(\omega) = e^{-j\omega k_0} \frac{R_w(k_0)}{R_w(0)} \tag{42}$$

whereas for the assumed system

$$H(\omega) = e^{-j\omega k_0}. \tag{43}$$

Thus, in the time domain, estimates of $\hat{h}(n)$ are weighted by the normalized autocorrelation function of the window.

Based on the above discussion, the algorithm chosen to estimate $\hat{h}(n)$ was to estimate $\hat{H}(k)$ using (36), inverse Fourier transform the estimate to give $\tilde{h}(n)$, and then normalize the result by using (42).[2] It should be noted that (42) provides only a first-order correction to the effects of the window.

III. Performance Measures and Error Models for System Identification Methods

In the preceding section we outlined three distinct methods which can be used to estimate a linear system whose output is corrupted by noise, given the input and output of the system. In this section we define and discuss two performance measures for evaluating these methods. They will be denoted as the Q measure and the modified Q or Q' measure. The Q measure is basically the ratio (expressed in dB) of the norm of the coefficient error vector or "misadjustment vector" [5], [6] and the norm of the true coefficient vector. It is useful for characterizing how well the estimate $\hat{\boldsymbol{h}}$ approximates the true $\boldsymbol{h}$. The second measure, Q', is a frequency weighted measure which is useful for characterizing the performance of system identification methods for nonwhite inputs and it is also useful when estimates of $v(n)$ and $e(n)$ in the model of Fig. 1 are desired.

A. The Q Measure

The Q measure has the form

$$Q = 10 \log_{10} \left\{ \frac{\sum_{m=0}^{\hat{M}-1} \Delta h^2(m)}{\sum_{m=0}^{\hat{M}-1} h^2(m)} \right\} \tag{44}$$

$$= 10 \log \left[\frac{\Delta \boldsymbol{h}^t \Delta \boldsymbol{h}}{\boldsymbol{h}^t \boldsymbol{h}} \right] \tag{45}$$

where

$$\Delta h(m) = h(m) - \hat{h}(m) \qquad m = 0, 1, 2, \cdots, \hat{M} - 1 \tag{46}$$

and $\Delta \boldsymbol{h}$ and $\boldsymbol{h}$ are vectors of length $\hat{M}$ and are defined in the same manner as $\boldsymbol{h}$ in (15). It will be shown that for a white input signal $x(n)$ and with uncorrelated white noise $e(n)$, the quantity Q is a simple function of three system parameters, namely, the measurement interval N, over which the linear system estimate is made, the estimated number of impulse response terms $\hat{M}$, and the signal-to-noise ratio S/N, at the output of the system. Q is seen to be a weak function of the system $h(n)$ in that the signal-to-noise ratio is dependent on $h(n)$.

For the least squares analysis method, it is relatively simple to derive an analytical expression for Q. From (12), (13), and (2) we derive the result

$$\sum_{m=0}^{\hat{M}-1} \hat{h}(m)\, \tilde{\phi}_{xx}(k, m) = \sum_{n=0}^{N-1} \left[\sum_{m=0}^{\hat{M}-1} h(m)\, x(n-m) + e(n) \right] \cdot x(n-k) \tag{47}$$

$$= \sum_{m=0}^{\hat{M}-1} h(m) \sum_{n=0}^{N-1} x(n-m)\, x(n-k) + \sum_{n=0}^{N-1} e(n)\, x(n-k) \tag{48}$$

$$= \sum_{m=0}^{\hat{M}-1} h(m)\, \tilde{\phi}_{xx}(k, m) + \tilde{\phi}_{xe}(k) \tag{49}$$

or

$$\sum_{m=0}^{\hat{M}-1} [\hat{h}(m) - h(m)]\, \tilde{\phi}_{xx}(k, m) = \tilde{\phi}_{xe}(k) \tag{50}$$

where

$$\tilde{\phi}_{xe}(k) = \sum_{n=0}^{N-1} e(n)\, x(n-k). \tag{51}$$

Equation (50) says that the finite time correlation between $x(n)$ and $e(n)$ (due to the finite measurement interval N) leads to an error $\Delta h(n)$ in estimating the true $h(n)$ in such a way that the estimated component $\hat{h}(n)$ produces a noise signal $\hat{e}(n)$ which is finite time uncorrelated with $x(n)$. This fundamentally important result is illustrated by the error model in Fig. 2 which shows a parallel path in which $x(n)$ is convolved with $\Delta h(n)$ (which is a hypothetical filter that represents the finite time correlation between $e(n)$ and $x(n)$), and the noise sequence $\hat{e}(n)$ is added to the result. The true noise sequence $e(n)$ satisfies the relation

$$e(n) = \hat{e}(n) + \Delta h(n) * x(n). \tag{52}$$

Equations (46) and (50) can now be used to show that

$$\Delta \boldsymbol{h} = \boldsymbol{h} - \hat{\boldsymbol{h}} = \tilde{\boldsymbol{\phi}}_{xx}^{-1} \tilde{\boldsymbol{\phi}}_{xe} \tag{53}$$

and that

$$\Delta \boldsymbol{h}^t \Delta \boldsymbol{h} = \tilde{\boldsymbol{\phi}}_{xe}^t (\tilde{\boldsymbol{\phi}}_{xx}^{-1})^t \tilde{\boldsymbol{\phi}}_{xx}^{-1} \tilde{\boldsymbol{\phi}}_{xe}. \tag{54}$$

For a stationary, white input signal, the matrix $\tilde{\boldsymbol{\phi}}_{xx}$ assumes the form

$$\tilde{\phi}_{xx}(k, l) = \sum_{n=0}^{N-1} x(n-k)\, x(n-l) \tag{55}$$

$$\approx \sigma_x^2 \cdot N \cdot \delta(k-l) \tag{56}$$

i.e., $\tilde{\boldsymbol{\phi}}_{xx}$ is a constant times an identity matrix. Thus, (54) becomes

[2] The correction used is to scale $h(n)$ by $Rw(0)/Rw(n)$. Clearly this correction is valid only for small values of $\hat{M}$.

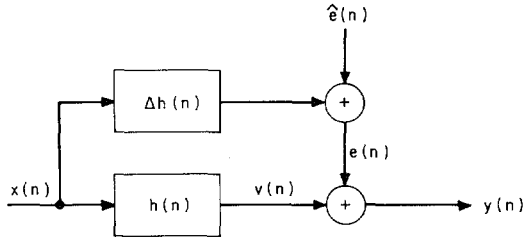

Fig. 2. A block diagram interpretation of the system identification model in terms of a misalignment filter and a modified error signal.

$$\Delta \boldsymbol{h}^t \Delta \boldsymbol{h} \cong \frac{1}{N^2 \sigma_x^4} \tilde{\phi}_{xe}^t \tilde{\phi}_{xe} \tag{57}$$

$$= \frac{1}{N^2 \sigma_x^4} \sum_{n=0}^{N-1} \sum_{k=0}^{N-1} e(n)\, e(k) \cdot \sum_{m=0}^{\hat{M}-1} x(n-m)\, x(k-m). \tag{58}$$

Assuming that $x(n)$ is uncorrelated from sample to sample, the last term in (58) is approximately

$$\sum_{m=0}^{\hat{M}-1} x(n-m)\, x(k-m) \cong \hat{M} \sigma_x^2 \delta(n-k). \tag{59}$$

Thus, (58) reduces to

$$\Delta \boldsymbol{h}^t \Delta \boldsymbol{h} \approx \frac{\hat{M}}{N^2 \sigma_x^2} \sum_{n=0}^{N-1} e^2(n) \tag{60}$$

$$\approx \frac{\hat{M} \sigma_e^2}{N \sigma_x^2} \tag{61}$$

We define the term σ_v^2 (where $v(n)$ is defined in Fig. 1) as

$$\sigma_v^2 = \frac{1}{N} \sum_{n=0}^{N-1} v^2(n) \tag{62}$$

$$= \frac{1}{N} \sum_{n=0}^{N-1} \sum_{k=0}^{\hat{M}-1} h(k)\, x(n-k) \sum_{m=0}^{\hat{M}-1} h(m)\, x(n-m). \tag{63}$$

By the same argument used to derive (61), (63) can be written as

$$\sigma_v^2 \cong \sigma_x^2 \cdot \boldsymbol{h}^t \boldsymbol{h}. \tag{64}$$

Combining (45), (61), and (64) gives for Q

$$Q\Big|_{\substack{\text{white} \\ \text{input}}} \cong 10 \log_{10} \left[\frac{\hat{M}}{N} \frac{\sigma_e^2}{\sigma_v^2} \right]$$

and by using the definition of (5) we get

$$Q\Big|_{\substack{\text{white} \\ \text{input}}} \cong 10 \log_{10} \left[\frac{\hat{M}}{N} \right] - S/N(\text{dB}). \tag{65}$$

Equation (65) predicts the performance of the least squares analysis system identification method for white uncorrelated inputs.

It is seen that Q in (65) is directly dependent on the signal-to-noise ratio at the output of the system. It improves (decreases) by 3 dB per doubling of the block size N of data used in the estimation of $\boldsymbol{h}$ and it degrades (increases) with $\log \hat{M}$ where $\hat{M}$ is the assumed size of the system (and where it is assumed that $\hat{M} > M$).[3] Furthermore, it can be seen that Q in (65) is independent of the actual filter $\boldsymbol{h}$ that is being estimated.

For the case of nonwhite inputs the matrix $\tilde{\phi}_{xx}$ is no longer diagonal and it is not possible to express Q in a form as simple as (65). In general, for nonwhite inputs the values of Q will be larger than (65) and in this sense (65) represents a lower bound on the expected value of Q. That is, a white uncorrelated input signal is the best form of input signal to use in the system identification problem.

B. Modified Q Measure for Nonwhite Inputs

For the case of nonwhite inputs, the overall model can be expressed in the form in Fig. 3(a). We assume that the input to the system $z(n)$, can be modeled as the output of a linear system $g(n)$, with spectrally flat input signal $x(n)$. For simplicity, we assume $g(n)$ is an FIR filter with impulse response duration G samples, i.e., $g(n)$ is nonzero only for $0 \leqslant n \leqslant G-1$. Fig. 3(b) shows the system identification model for this problem.

Whenever the input signal does not have a flat spectrum, the problems in system identification can become greatly magnified. This can be seen by considering $z(n)$ to be a bandpass signal. For this case there is almost no information in the output signal $y(n)$, about the behavior of the system $h(n)$, in frequency ranges where the input signal is greatly attenuated. As such, reliable identification, using any system identification procedure, is quite difficult. Formally, one could express this problem in terms of ill-conditioned matrices that need to be inverted, etc.

In some applications, however, it is sufficient to have a good estimate of the system $\boldsymbol{h}$ at frequencies *only* where the input signal energy $Z(e^{j\omega})$ is large. This applies particularly in cases where estimates of $v(n)$ and $e(n)$ are primarily desired. Such estimates can be obtained by first solving the system identification problem to obtain $\hat{\boldsymbol{h}}$. The estimate $\hat{v}(n)$ is then obtained by convolving the known input $z(n)$ with $\hat{\boldsymbol{h}}$, and $\hat{e}(n)$ is obtained by subtracting $\hat{v}(n)$ from the known output $y(n)$. At frequencies where $Z(e^{j\omega})$ is small, it is not as important in this case to have very good estimates of $H(e^{j\omega})$. In applications such as this a frequency weighted measure of performance is desired which emphasizes the importance of those frequencies where $X(e^{j\omega})$ (or $G(e^{j\omega})$) is large and deemphasizes the importance of those frequencies where $X(e^{j\omega})$ (or $G(e^{j\omega})$) is small. In this section we propose such a measure, Q'.

The modified Q measure Q', that we propose, applies a frequency weighting which is equal to that of the frequency response of the filter $g(n)$ in Fig. 3(a) which is used to create the nonwhite signal $z(n)$ from the white signal $x(n)$. This weighting can conveniently be achieved by convolving $h(n)$ and $\hat{h}(n)$ by the $g(n)$ as shown in Fig. 4. This procedure, illus-

[3]For $\hat{M} < M$, the error in estimating $\hat{h}(n)$ for $n > \hat{M}$ clearly can get very large since no value of $\hat{h}(n)$ is obtained.

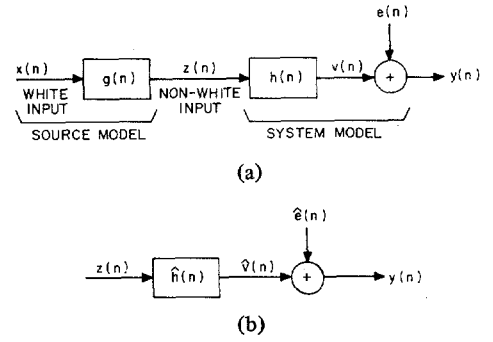

Fig. 3. Block diagram of the linear system model for nonwhite input signals.

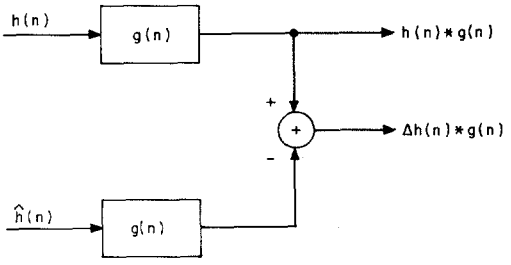

Fig. 4. Block diagram of the processing to give the signals used in the Q' measure.

trated in Fig. 4, serves to weight the performance measure by the frequency spectrum of the input signal. More formally, we define the measure Q', as

$$Q' = 10 \log_{10} \left[\frac{\sum_n [(h(n) - \hat{h}(n)) * g(n)]^2}{\sum_n [h(n) * g(n)]^2} \right]. \quad (66)$$

Using Parsevals theorem (66) can be transformed to the frequency domain, giving

$$Q' = 10 \log_{10} \left[\frac{\int_{-\pi}^{\pi} |H(e^{j\omega}) - \hat{H}(e^{j\omega})|^2 |G(e^{j\omega})|^2 \, d\omega}{\int_{-\pi}^{\pi} |H(e^{j\omega})|^2 |G(e^{j\omega})|^2 \, d\omega} \right] \quad (67)$$

which explicitly shows the frequency weighting of the Q' measure.

The properties of Q' as a function of N, $\hat{M}$, and S/N are somewhat more complicated than those of the Q measure, since the "coloring or unwhitening" filter, $g(n)$, affects the result. If we denote Q' as $Q'(N, \hat{M}, G, S/N)$, and similarly denote Q of (65) as $Q(N, \hat{M}, S/N)$ then it can be shown [see Appendix II] that

$$Q'(N, \hat{M}, G, S/N) < Q(N, \hat{M} + G - 1, S/N) \quad (68a)$$

i.e., the estimate of $h(n)$ is no worse than an equivalent Q estimate with an impulse response which is $\hat{M} + G - 1$ points long (i.e., duration of the convolution of $g(n)$ and $h(n)$). Furthermore, it is anticipated (but not rigorously proved) [see Appendix II] that

$$Q'(N, \hat{M}, G, S/N) \approx Q(N, \hat{M}, S/N) \quad (68b)$$

i.e., the variation of Q' with S/N and N is essentially identical to that of Q for white inputs as discussed earlier.

An important consideration in implementing the Q' measurement of (66) is how one obtains the $g(n)$ for a signal whose frequency response is not flat, such as a speech waveform. For such signals the techniques of linear prediction have been successfully used to give a good approximation to a linear system which can be excited by a flat spectrum input (either pulses or noise), and whose spectrum is a least squares estimate of the signal spectrum. As such, the linear system $g(n)$ is obtained directly from linear prediction techniques as the impulse response of the linear prediction filter.

We now present examples which demonstrate how the above measures worked on both synthetic and actual signal for various types of linear systems and various levels of noise.

IV. Experimental Results

To validate the models of the previous sections, a digital simulation of the model of Fig. 1 was made. Three types of input signal, $x(n)$, were used. These included: 1) white Gaussian noise; 2) band-limited Gaussian noise; and 3) speech signals. For the linear system $h(n)$, two examples were used. One was a simple 7-point FIR filter ($M = 7$) whose impulse and log magnitude responses are given in Fig. 5 and in Table I. As seen in this figure, the log magnitude response was smooth, and varied about 10 dB across the entire frequency band. The other filter used in the simulations was a 25-point, linear phase, equiripple FIR low-pass filter. The impulse and log magnitude responses of this filter are given in Fig. 6 and Table II. Independent Gaussian noise ($e(n)$) was added to the filtered input to give signal-to-noise ratios (S/N) of 0, 8, 16, 24, and infinite dB.

The three system identification methods of section II were used to estimate the known $h(n)$ for several combinations of the above system parameters. For each example the quantity Q (or Q') of section III was measured. In this section we present typical curves of Q as a function of N (the analysis frame duration), $\hat{M}$, $h(n)$, and S/N.

A. White Noise Input

Fig. 7 shows a series of curves of Q (on a log scale) versus N (on a log scale) for the filter of Fig. 5, with $\hat{M} = 15$, and various values of S/N, for the LSA method. The solid curves show the measured values of Q and the dashed curves show the predicted values of Q as given by (65). The agreement between the computations and the predicted values is well within the expected statistical variations for these cases. For the infinite S/N case, the measured curve of Q versus N is below the -80 dB cutoff level of the plots, and is thus not included here. Equivalent computations were made for other values of $\hat{M}$ (notably 7 and 25), and for the low-pass filter (with $\hat{M} = 25$ and 34) and the results were equivalent to those of Fig. 7, i.e., close agreement between theory and measurement.

Figs. 8 and 9 show a set of comparable curves for the SSA method. For this method a Hamming window of size L samples was used in the analysis, and the window was moved by

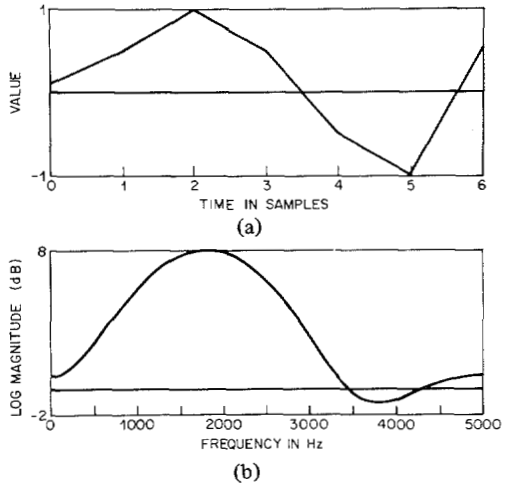

Fig. 5. Impulse response and log magnitude response of a simple filter used in the investigations.

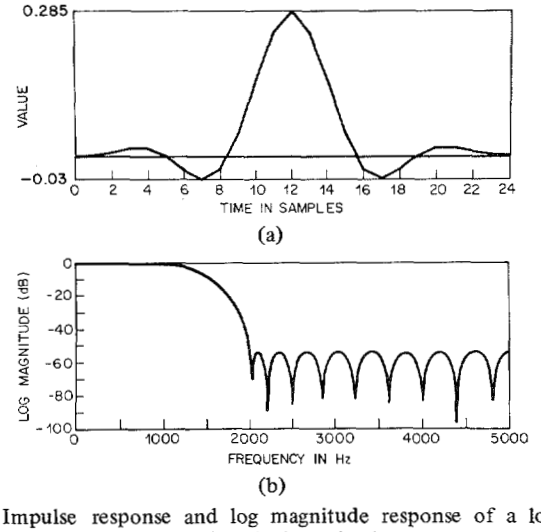

Fig. 6. Impulse response and log magnitude response of a low-pass filter used in the investigations.

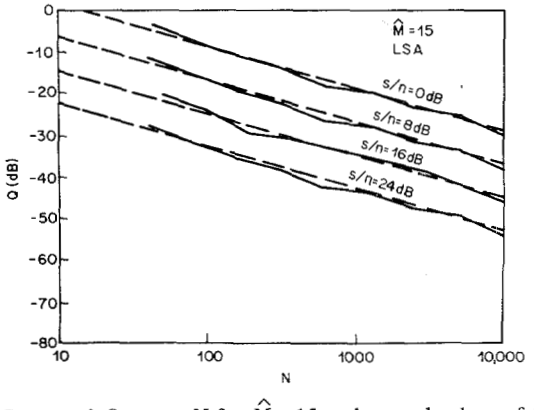

Fig. 7. Curves of Q versus N for $\hat{M} = 15$ and several values of S/N for the LSA method for a white input signal.

TABLE I
VALUES OF THE IMPULSE RESPONSE FOR THE FILTER OF FIG. 5

n	$h(n)$
0	0.1
1	0.5
2	1.0
3	0.5
4	-0.5
5	-1.0
6	0.5

TABLE II
VALUES OF THE IMPULSE RESPONSE FOR THE FILTER OF FIG. 6

n	$h(n)$	n	$h(n)$
0	-0.0016	13	0.2456
1	0.0007	14	0.1490
2	0.0070	15	0.0428
3	0.0145	16	-0.0280
4	0.0143	17	-0.0470
5	-0.0013	18	-0.0287
6	-0.0287	19	-0.0013
7	-0.0280	20	0.0143
8	-0.0280	21	0.0145
9	0.0428	22	0.0070
10	0.0149	23	0.0007
11	0.2456	24	-0.0016
12	0.2847		

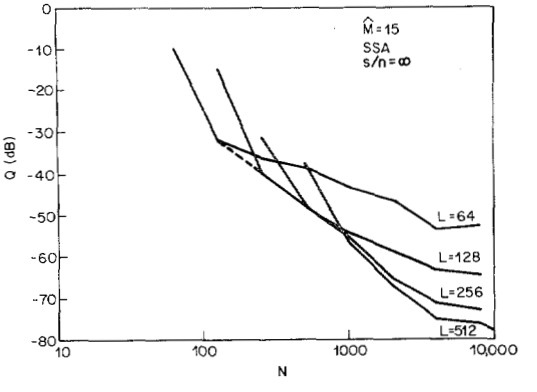

Fig. 8. Curves of Q versus N for $\hat{M} = 15$, $S/N = \infty$, and several values of window size L for the SSA method for a white input signal.

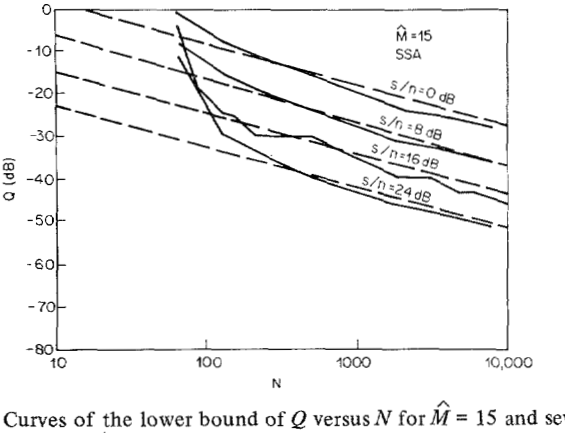

Fig. 9. Curves of the lower bound of Q versus N for $\hat{M} = 15$ and several values of S/N for the SSA method for a white input signal.

$L/4$ samples between adjacent sections [13], i.e., except for endpoint effects, each input sample was used in four distinct short-time spectral estimates. Thus, in presenting results for the SSA method, the window length L is an additional analysis parameter whose effects must be considered. Thus, Fig. 8 shows a set of curves of Q versus N for the simple filter of Fig. 5, with $\hat{M} = 15$, with $L = 64, 128, 256$, and 512, and with $S/N = \infty$ (i.e., no additive noise). A complete analysis of these curves is beyond the scope of this paper. However, several key points about this method of analysis can be seen from this figure. First, it is seen that for smaller values of N it is preferable to use shorter windows to reduce the end effects and to provide an increased number of short-time spectral estimates for averaging. However, for longer values of N, the longer the window duration the lower the value of Q which is achieved. This effect is related to the aliasing noise of the analysis (obtained as a result of performing the division in (36)), which is reduced with increased window length. Finally, it is seen that for large values of N, the value of Q approaches the machine accuracy of about -80 dB, thus showing that the method will eventually converge to the least squares estimate.

Fig. 9 shows a set of curves of Q versus N for several values of S/N for the SSA method. Based on the discussion above, the "lower bound" of the curves is drawn as the solid curves shown in the figure. The dashed curves again show the theoretical predictions for these cases. It is seen in this figure that for S/N in the range 0 to 24 dB, the SSA method can provide filter estimates that are fairly close to the optimum, except for small values of N where the end effects still dominate. The aliasing effects for large N do not occur here because the additive noise for these cases is significantly greater than the aliasing noise.

Curves of the results obtained for the LMS adaptation method are given in Figs. 10 and 11. Fig. 10 shows a set of curves of Q versus N, for $\hat{M} = 15$, and $S/N = \infty$ for the filter of Fig. 5. The parameter for the individual curves is K, the step-size multiplier of the adaptation algorithm. As seen in these curves, the values of Q decrease monotonically to the computation noise floor. The rate at which these curves decrease is determined by the value of K. Thus, for small values of K the convergence is slow; for large values of K it is much faster. At first thought such curves would seem to imply that one should use large values of K. However, if we recall that K is the correction term multiplier (26), then we realize that if K is large, small errors in calculating the gradient of e can lead to larger errors in estimating h as will be seen later in Fig. 11. Thus, with noisy signals a tradeoff in choosing K is required. The curves in Fig. 10 give an indication of the value of N required to obtain a desired value of Q for the noise-free case.

Fig. 11 shows curves of Q versus N for $S/N = 8$ dB, and the same parameters as Fig. 10. Values of K of 0.01 and 0.05 are used to show the convergence properties of the algorithm. For these cases a steady-state noise floor (due to the gradient calculation of the noisy signal) limits the value of Q which can be obtained. An expression for this steady-state noise floor can be obtained based on the work of Widrow (see [4, equation D.20]). Based on results by Widrow it can be shown that

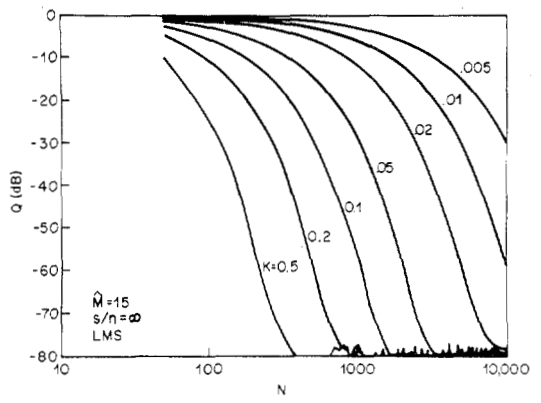

Fig. 10. Curves of Q versus N for $\hat{M} = 15$, $S/N = \infty$, and several values of K for the LMS adaptation algorithm for a white input signal.

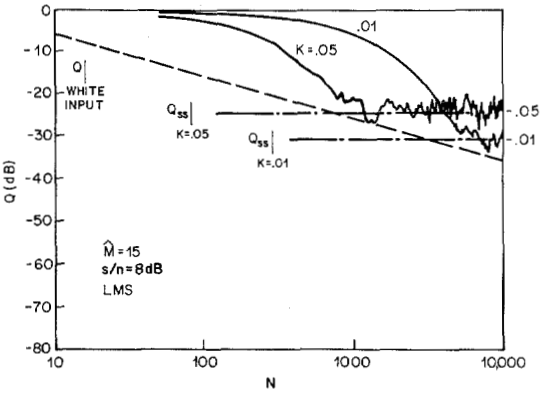

Fig. 11. Curves of Q versus N for $\hat{M} = 15$, $S/N = 8$ dB for $K = 0.01$ and 0.05 for the LMS adaptation algorithm for a white input signal.

$$\sum_{m=0}^{\hat{M}-1} [\hat{h}(m) - h(m)]^2 = u\hat{M}\sigma_e^2 \tag{69}$$

$$= \frac{K}{2}\frac{\sigma_e^2}{\sigma_x^2} \tag{70}$$

and since

$$\sigma_x^2 = \frac{\sigma_v^2}{\sum_{m=0}^{\hat{M}-1} h^2(m)} \tag{71}$$

then

$$Q\Big|_{\substack{LMS \\ SS}} = 10 \log\left(\frac{K}{2}\right) - S/N(\text{dB}). \tag{72}$$

Plots of $Q\big|_{SS}$ for $K = 0.01$ and 0.05 are shown in Figure 11, as well as the curve of $Q\Big|_{\text{white input}}$ versus N for optimal estimation. The point at which these two curves intersect provides a *lower limit* on the value of N required for convergence of the LMS adaptation algorithm. Combining (72) and (65) provides an expression for this lower limit of the form

$$N\big|_{\min} \geq 2\hat{M}/K. \tag{73}$$

It should be emphasized that one would expect a value of N

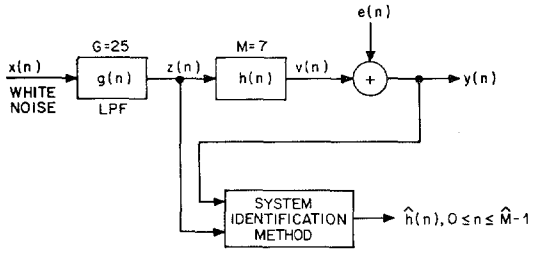

Fig. 12. Block diagram of actual system used to test the system identification algorithms for a low-pass input signal.

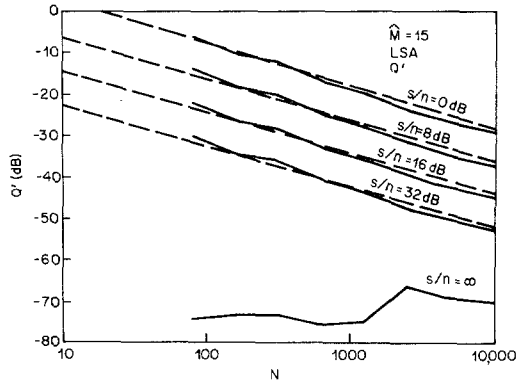

Fig. 13. Curves of Q' versus N for $\hat{M} = 15$ and several values of S/N for the LSA method with a band-limited noise input signal.

on the order of two or more times greater than the value of (73) in most practical cases. As seen in Fig. 11, once the convergence to the noise floor of (72) is obtained, larger values of N provide no improvement in the estimation.

B. Bandpass Input Signals

To evaluate the performance of each of the three system identification methods on band-limited input signals, the system of Fig. 12 was simulated. The filters $g(n)$ and $h(n)$ corresponded to the filters of Figs. 6 and 5, respectively. Thus, the signal $z(n)$ was a low-pass signal whose frequency components were attenuated by at least 54 dB for frequencies above $0.2F_s$, where F_s was the sampling rate of the system. Independent additive Gaussian noise $e(n)$ was again used to provide the signal $y(n)$ from which the system function $\hat{h}(n)$ relating $y(n)$ to $z(n)$ was estimated.

Fig. 13 shows a set of curves of Q' (the modified Q measure) versus N for $\hat{M} = 15$, and several values of S/N for the LSA method. The solid lines are the measured values of Q', whereas the dotted lines show the theoretical curves of Q versus N for the same set of conditions. As discussed earlier, the measured curves of Q versus N were vastly different from those shown in Fig. 13 due to the lack of high-frequency information in the input signal. However, when the Q' measure was used, the high-frequency inaccuracies in $\hat{h}(n)$ were given essentially zero weight by the "coloration" filter $g(n)$. Thus, the curves of Q' versus N of Fig. 13 for the highly band-limited input are essentially the same as the curves of Q versus N of Fig. 7 for the white-input case. At the bottom of Fig. 13 is shown the curve of Q' versus N for infinite S/N. In this case the value of Q' is about -70 dB, reflecting the residual error in estimating the high frequency behavior of $\hat{h}(n)$.

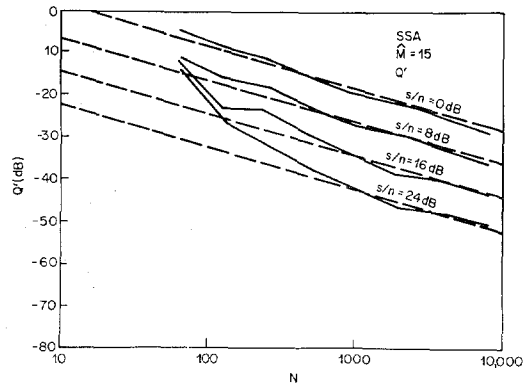

Fig. 14. Curves of Q' versus N for $\hat{M} = 15$ and several values of S/N for the SSA method with a band-limited noise input signal.

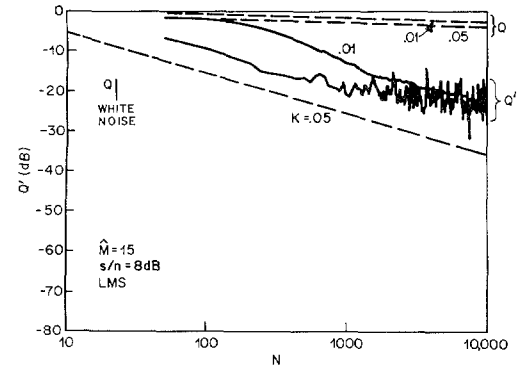

Fig. 15. Curves of Q and Q' versus N for $\hat{M} = 15$, $S/N = 8$ dB for $K = 0.01$ and 0.05 for the LMS adaptation algorithm with a band-limited noise input signal.

Results on the band-limited input signal for the SSA method are given in Fig. 14. The curves plotted in this figure again represent the lower bound of the individual curves for different window lengths, L. The shapes of the curves are essentially identical to those of Fig. 9 in that for small values of N, the curves are significantly above the theoretical estimates of the Q measure (shown as the dotted curves) due to the end effects in the SSA method. For larger values of N the values of Q' are essentially equal to the theoretical values. One very important point must be made concerning the way in which Q' was computed for these curves. For the SSA method, when the input signal has essentially no energy in a band, the estimate in frequency is essentially unconstrained at these frequencies. As such, taking the inverse DFT of the spectral estimates leads to an impulse response with a large amount of time aliasing—i.e., an unconstrained set of frequency samples cannot possibly guarantee a finite duration time response. Therefore, straightforward linear filtering of $\hat{h}(n)$, the time response from the SSA estimate by $g(n)$, the unwhitening filter, is not adequate to eliminate the aliasing noise. However, it can readily be shown that circular convolution of $\hat{h}(n)$ with $g(n)$ will indeed eliminate the aliasing noise. The proof of this statement is straightforward and is based on the fact that circular convolution commutes with time-aliasing. Since the convolved signal is time-limited, no aliasing is present when this procedure is used.

Fig. 15 shows a set of results obtained on the band-limited input using the LMS adaptation algorithm for $\hat{M} = 15$, $S/N = 8$ dB, and for $K = 0.01$ and 0.05. Included in this figure as curves of Q versus N, Q' versus N, and the theoretical white noise curve of Q versus N. As discussed above, the measurements of Q versus N are highly in error for these cases, whereas the measurements of Q' versus N approach the theoretical least square curves. As seen in this figure, the shape of the Q' versus N curve differs somewhat from that of Fig. 11; however, the differences are of little consequence. The general behavior of the curves is essentially the same with a region of convergence, followed by a statistical fluctuation about the steady-state noise floor. Fig. 16 shows a comparison between the estimated $\hat{H}(e^{j\omega})$ and the actual $H(e^{j\omega})$ for one case using the LMS adaptation algorithm. The simulation was done as in Fig. 12 with $\hat{M} = 15$, $K = 0.02$, $g(n)$ in Fig. 6 and $h(n)$ in Fig. 5. Fig. 16(a) shows a comparison of the log magnitude responses; Fig. 16(b) shows the group delay responses; and Fig. 16(c) shows the impulse responses (for $N = 2000$). It is clearly seen from this figure that most of the error in estimation occurs for frequencies above the passband cutoff of the $g(n)$ filter, as anticipated ($Q = -2.9$ dB and $Q' = -37$ dB).

C. Speech Input Signals

The last test signal used to evaluate the three system identification techniques was an actual speech signal. Fig. 17(a) shows a 400 sample section of voiced speech (quasi-periodic waveform) weighted by a Hamming window, with a log magnitude spectrum as shown in Fig. 17(b), and with LPC fit to the spectrum shown in Fig. 17(c). As seen in this figure the speech signal spectrum is essentially a harmonic spectrum showing valleys which are 20–40 dB lower in magnitude than adjacent harmonic peaks. The overall shape of the spectral envelope (as seen from the LPC fit) is that of a cascade of resonators. As such a 60 dB variation in the spectral envelope occurs between the first prominent peak, and the valley at 5 kHz. Thus this input signal would be expected to provide a good test of the overall capabilities of the different system identification methods.

The model for testing the systems using the speech signal was essentially that of Fig. 12 with one major exception. If we denote the speech signal as $z(n)$, then the "coloration" linear system $g(n)$ is not known exactly. Thus to provide analytical estimates of Q' for the speech input, the system $g(n)$ also had to be estimated from $z(n)$. For this problem standard LPC techniques were used. As such the range of values of N which was considered was from 50 to 1000. Outside this range the LPC estimates were sufficiently inaccurate to greatly degrade the Q' measure. It should be noted that even within this range the LPC estimates of $g(n)$ are not exact; as such, the Q' computations were somewhat affected.

The curves of Q' versus N for various S/N values for the three systems are given in Figs. 18–20. Fig. 18 shows the results for the LSA method for $\hat{M} = 15$. It is seen that the measured values of Q' are generally greater than the predictions of (65) for the white input; however the differences, except for small N, are about 5 dB or less. Thus the LSA method is seen to work quite well on this section of speech.

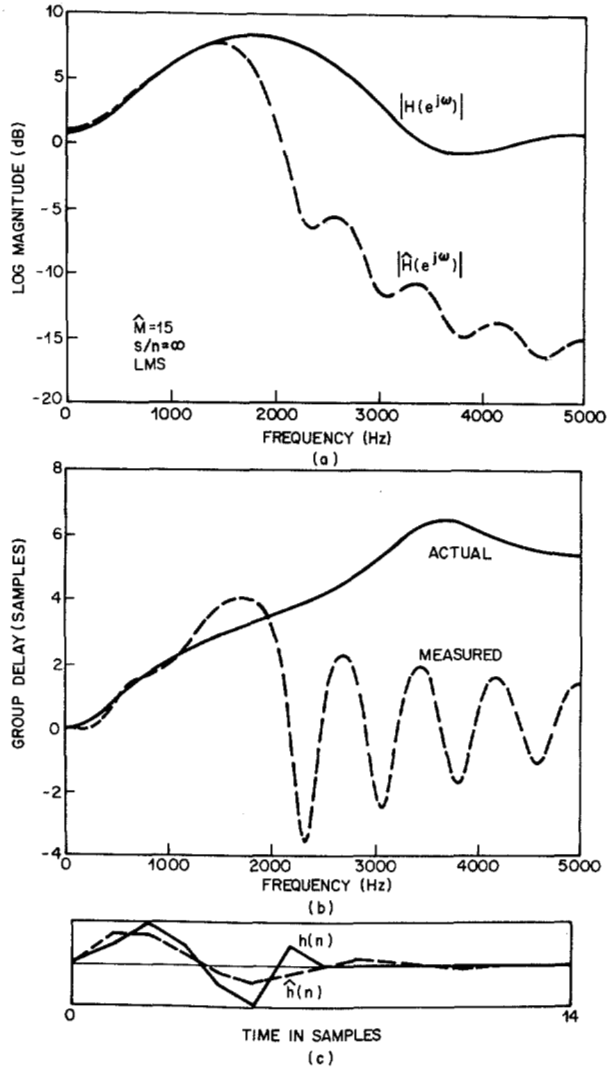

Fig. 16. Curves of log magnitude, group delay, and impulse responses for both the estimated and actual systems for the LMS adaptation algorithm for $\hat{M} = 15$, $S/N = \infty$ and $N = 2000$.

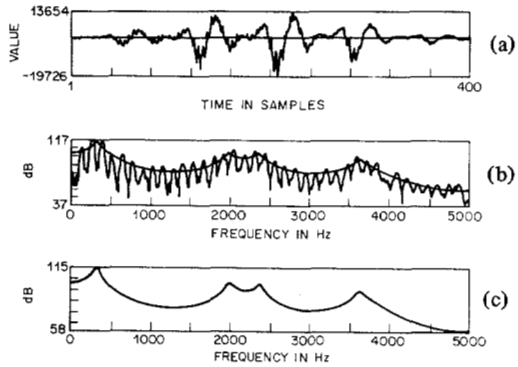

Fig. 17. Curves of a speech section of 400 samples (windowed by a Hamming window), its log magnitude spectral, and the LPC log magnitude fit.

Fig. 19 shows the results for the SSA method for $\hat{M} = 15$. The effects of having extremely low signal level in parts of the frequency band are seen in that the measured curves are about 4–12 dB above the theoretical curves for the white noise case. As discussed in the previous section, Q' estimates for the SSA

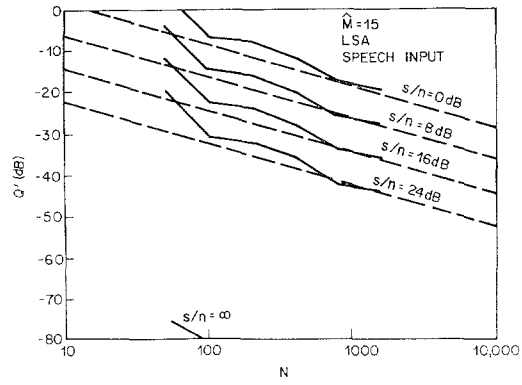

Fig. 18. Curves of Q' versus N for $\hat{M} = 15$, and several values of S/N for the LSA method for a speech input signal.

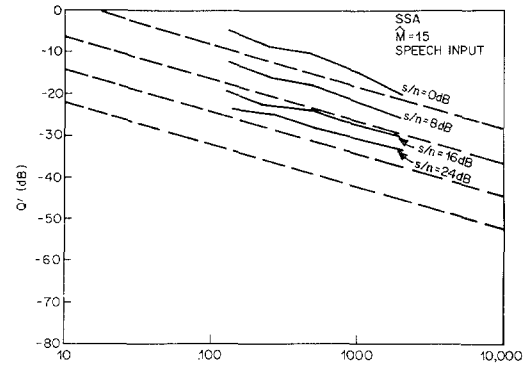

Fig. 19. Curves of Q' versus N for $\hat{M} = 15$, and several values of S/N for the SSA method for a speech input signal.

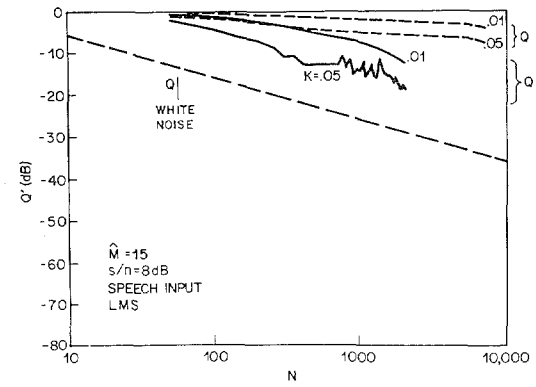

Fig. 20. Curves of Q and Q' versus N for $\hat{M} = 15$, $S/N = 8$ dB, for $K = 0.01$, and 0.05, for the LMS adaptation system for a speech input signal.

method are more sensitive to exact values of $g(n)$ than the LSA method since $g(n)$ must also remove the aliasing noise from the estimate. Since $g(n)$ was estimated from LPC methods, the Q' estimates for the SSA method were somewhat poorer than the LSA method.

Finally, Fig. 20 shows the results for the LMS adaptation method for $\hat{M} = 15$, and for $S/N = 8$ dB. The measured values of Q' converge to the steady-state noise floor as in the previous examples. For this case the performance of the LMS adaptation algorithm is comparable to the SSA method but not as good as the LSA, as seen by the difference between the theoretical curve for the white input (the dotted line in Fig. 20) and the actual Q' measurements.

V. Discussion of the Results

The purpose of this work was to investigate the advantages and disadvantages of three system identification methods using a given class of input signals. Based on the theoretical discussions of Sections II and III, the results presented in Section IV, and informal observations from the simulations, the following conclusions are drawn.

1) As expected, the LSA method was the most robust method of the three, providing excellent estimates of $h(n)$ for both white and band-limited signals, and across a wide range of signal-to-noise ratios.

2) The only possible disadvantage of the LSA method is that the implementation effectively solves an $\hat{M}$th order matrix equation (via an efficient recursion procedure). Using double precision arithmetic, inaccuracies in both the computation of $\tilde{\phi}_{xx}(i,j)$ and $\hat{h}(n)$ started to become measureable for $\hat{M}$ on the order of 50.[4] As such, estimation of systems with large values of $\hat{M}$ (e.g. the speech echo canceler, etc.) would generally not be practical using LSA. The recognition of this fact has led to the widespread use of the LMS adaptation method. In addition, the storage for $\tilde{\phi}_{xx}(i,j)$ grows as $\hat{M}^2$, again making the method impractical for large values of $\hat{M}$.

3) The SSA method was shown to perform almost as well as the LSA method for both white and nonwhite noise signals in the presence of noise. The major disadvantage of the SSA method is the possibility of having severe aliasing distortion in the estimation of $\hat{h}(n)$ which would have to be removed in any practical system. Thus, for speech signals, removal of the aliasing by filtering $\hat{h}(n)$ with $\hat{g}(n)$, the LPC estimate of the "speech filter" is not completely adequate because of errors in obtaining $\hat{g}(n)$. As such the SSA estimates were somewhat poorer than the LSA estimates for speech signals.

4) The big advantage of the SSA method is that the implementation is simple, can be used for large $\hat{M}$ values, readily amenable to either digital hardward (e.g., CCD implementations of the DFT's) or to array processors which are currently becoming more widespread in use. The storage for this method grows linearly with $\hat{M}$ and thus estimation for values of $\hat{h}(n)$ for $\hat{M}$ on the order of 512 or more is entirely practical.

5) The LMS adaptation algorithm provides a robust alternative to the LSA method, and is useful for both white inputs, as well as band-limited inputs. Although the performance was not as good as the LSA method, the differences were not so as to make the method undesirable for virtually any application.

6) Generally the convergence rate of the LMS adaptation algorithm is affected by band-limited inputs and by high signal-to-noise ratios. It requires a much larger number of samples, N, compared to the LSA method and, therefore, is limited to applications where $h(n)$ varies very slowly.[5]

7) An important advantage of the LMS adaptation algorithm is that the implementation is simple and can be used for large values of $\hat{M}$. In addition, the storage for the LMS adaptation algorithm grows linearly with $\hat{M}$, as with the SSA method.

[4] Alternative slower matrix solution methods can make the LSA method more accurate for larger values of $\hat{M}$. However, the computation time can become a problem.

[5] Multiple passes of the LMS adaptation algorithm on the same data with adaptive step sizes can increase the speed of convergence of this method.

VI. Summary

In this paper we have studied three distinct methods for identifying a linear system in the presence of noise. We have shown the advantages and disadvantages of each of the methods for the class of signals which we studied. The results of this investigation will hopefully help a user of such methods to make efficient use of each of these three techniques as warranted by the individual problems.

Appendix I

SSA Analysis of a Pure Delay

Consider an arbitrary signal $x(n)$ as input to a linear system with output $y(n) = x(n - k_0)$, i.e., a pure delay. Applying the definition of the short-time spectrum (38) to $y(n)$ we see that

$$Y_n(e^{j\omega}) = \sum_m y(m)\, w(n-m)\, e^{-j\omega m} \tag{A.1}$$

$$= \sum_m x(m-k_0)\, w(n-m)\, e^{-j\omega m} \tag{A.2}$$

$$= e^{-j\omega k_0} \sum_l x(l)\, w(n-k_0-l)\, e^{-j\omega l}. \tag{A.3}$$

Now we consider the definition of the SSA estimate (36). In particular, if we examine the numerator of (36) [call this $\hat{N}(\omega_q)$] we get

$$\hat{N}(\omega_q) = e^{-j\omega_q k_0} \sum_r \sum_l x(l)\, w(r-k_0-l)\, e^{-j\omega_q l} \cdot \sum_m x(m)\, w(r-m)\, e^{+j\omega_q m} \tag{A.4}$$

where we have assumed both x and w are real signals. Equation (A.4) can be put in the form

$$\hat{N}(\omega_q) = e^{-j\omega_q k_0} \sum_l \sum_m x(l)\, x(m)\, e^{-j\omega_q (l-m)} \cdot R_w(k_0 + l - m) \tag{A.5}$$

where

$$R_w(k_0 + l - m) = \sum_r w(r-k_0-l)\, w(r-m) \tag{A.6}$$

is the autocorrelation function of the window. If we taken the DFT of (A.5) [calling the result $\hat{N}(n)$] we get

$$\hat{N}(n) = \frac{1}{P} \sum_q \hat{N}(\omega_q)\, e^{j\omega_q n}$$

$$= \frac{1}{P} \sum_l \sum_m x(l)\, x(m)\, R_w(k_0 + l - m) \cdot \sum_q e^{-j\omega_q (k_0 + l - m - n)} \tag{A.7}$$

Recognizing that the last summation gives the result

$$\frac{1}{P} \sum_q e^{-j\omega_q (k_0 + l - m - n)} = \delta(k_0 + l - m - n) \tag{A.8}$$

(where P is the size DFT used throughout the analysis), (A.7) becomes

$$\hat{N}(n) = \sum_l x(l)\, x(k_0 + l - n)\, R_w(n) \tag{A.9}$$

$$= R_x(k_0 - n)\, R_w(n). \tag{A.10}$$

Thus the original numerator $\hat{N}(\omega_q)$ can be written as

$$\hat{N}(\omega_q) = \mathbf{F}[R_x(k_0 - n)\, R_w(n)] \tag{A.11}$$

where $\mathbf{F}$ represents a Fourier transform of the sequence. Similarly, the denominator of (36) can trivially be shown to be of the form

$$\hat{D}(\omega_q) = \mathbf{F}[R_x(-n)\, R_w(n)]. \tag{A.12}$$

Thus, the short-time spectral estimate of a pure delayed signal with delay k_0 gives

$$\hat{H}(\omega_q) = \frac{\hat{N}(\omega_q)}{\hat{D}(\omega_q)} = \frac{\mathbf{F}[R_x(k_0 - n)\, R_w(n)]}{\mathbf{F}[R_x(-n)\, R_w(n)]}. \tag{A.13}$$

Appendix II

Properties of the Q' Measure

As shown in Section III, the Q' measure is defined as

$$Q' = 10 \log \left[\frac{\Delta r^t \Delta r}{r^t r} \right] \tag{B.1}$$

where

$$\Delta r(n) = \Delta h(n) * g(n) \tag{B.2}$$

and

$$r(n) = h(n) * g(n) \tag{B.3}$$

and where it is assumed that the length of $g(n)$ is G, the length of $h(n)$ is M (and it can be extended to length $\hat{M}$ by appending zeros) and the length of $\hat{h}(n)$ is $\hat{M}$. The lengths of both $r(n)$ and $\Delta r(n)$ are then $R = G + \hat{M} - 1$. In vector form (B.2) and (B.3) can be expressed in the form

$$\Delta r = F \Delta h \tag{B.4}$$

and

$$r = Fh \tag{B.5}$$

where F is the R by $\hat{M}$ convolution matrix

$$F = \begin{bmatrix} g(0) & 0 & 0 & \cdots & 0 \\ g(1) & g(0) & 0 & \cdots & 0 \\ g(2) & g(1) & g(0) & & \\ \vdots & & & & g(0) \\ g(G-1) & & & & \\ 0 & & & & \vdots \\ \vdots & & & & \\ 0 & 0 & \cdots & & g(G-1) \end{bmatrix} \tag{B.6}$$

where

$$R = G + \hat{M} - 1 \tag{B.7}$$

and

$$\Delta r = [\Delta r(0), \Delta r(1), \cdots, \Delta r(R-1)] \tag{B.8}$$

$$r = [r(0), r(1), \cdots, r(R-1)]. \tag{B.9}$$

Furthermore, in a manner similar to the derivation of (17) and (53) and with aid of (B.4) and (B.5) it can be shown that

$$\Delta r = F\tilde{\phi}_{zz}^{-1}\tilde{\phi}_{ze} \tag{B.10}$$

$$r = F\tilde{\phi}_{zz}^{-1}\tilde{\phi}_{zy} \tag{B.11}$$

where the model of Fig. 3 and Fig. 4 is assumed. Also, by recognizing that

$$\tilde{\phi}_{ze}(k) = \sum_{n=0}^{N-1} e(n)\, z(n-k) \tag{B.12}$$

$$\tilde{\phi}_{zz}(k,l) = \sum_{n=0}^{N-1} z(n-k)\, z(n-l) \tag{B.13}$$

and

$$z(n) = \sum_{m=0}^{G-1} g(m)\, x(n-m) \tag{B.14}$$

and with definition (B.6) it can be shown that

$$\tilde{\phi}_{ze} = F^t\tilde{\phi}_{xe}. \tag{B.15}$$

It should be noted that $\tilde{\phi}_{ze}$ is $\hat{M} \times 1$ in size and $\tilde{\phi}_{xe}$ is defined to be $R \times 1$ in size.

From the definition of $\tilde{\phi}_{zz}$ in (B.13), and with the aid of (B.14) and the diagonal approximation as in (56), $\tilde{\phi}_{zz}$ can be shown to have the form

$$\tilde{\phi}_{zz} \approx N\sigma_x^2 F^t F \tag{B.16}$$

where $\tilde{\phi}_{zz}$ is an $\hat{M} \times \hat{M}$ matrix. From (B.10) we now see that $\Delta r^t \Delta r$ has the form

$$\Delta r^t \Delta r = [F\tilde{\phi}_{zz}^{-1}\tilde{\phi}_{ze}]^t\,[F\tilde{\phi}_{zz}^{-1}\tilde{\phi}_{ze}]. \tag{B.17}$$

Noting that $\tilde{\phi}_{zz} = \tilde{\phi}_{zz}^t$ we then get

$$\Delta r^t \Delta r = \tilde{\phi}_{ze}^t\tilde{\phi}_{zz}^{-1}F^tF\tilde{\phi}_{zz}^{-1}\tilde{\phi}_{ze}. \tag{B.18}$$

Applying (B.15) and (B.16) then gives

$$\begin{aligned}\Delta r^t \Delta r &\approx \frac{1}{N^2\sigma_x^4}\tilde{\phi}_{xe}^t F(F^tF)^{-1}F^tF(F^tF)^{-1}F^t\tilde{\phi}_{xe} \\ &= \frac{1}{N^2\sigma_x^4}\tilde{\phi}_{xe}^t F(F^tF)^{-1}F^t\tilde{\phi}_{xe} \\ &= \frac{1}{N^2\sigma_x^4}\tilde{\phi}_{xe}^t A\tilde{\phi}_{xe}\end{aligned} \tag{B.19}$$

where

$$A = F(F^tF)^{-1}F^t. \tag{B.20}$$

In a manner similar to the derivation of (64) we get

$$\sigma_v^2 = \sigma_x^2 \cdot r^t r \tag{B.21}$$

and therefore the Q' measure can be written in the form

$$Q' = 10\log\left[\frac{1}{N^2\sigma_v^2\sigma_x^2}\tilde{\phi}_{xe}^t A\tilde{\phi}_{xe}\right] \tag{B.22}$$

where $\tilde{\phi}_{xe}^t$ is an $R \times 1$ column vector and A is an $R \times R$ matrix. The matrix A has several interesting properties. First, it can be shown that

$$A = A^tA. \tag{B.23}$$

From this property if follows that if A is of rank $\hat{M}$, then it has $\hat{M}$ eigenvalues that are equal to 1, and $G - 1$ eigenvalues that are equal to 0. The eigenvalues of A satisfy the relation

$$A\phi = \lambda\phi \tag{B.24}$$

where ϕ is an eigenvector of A and λ is an eigenvalue. Premultiplying by A^t and noting that $A = A^t$ gives

$$A^tA\phi = \lambda A^t\phi = \lambda A\phi.$$

Applying (B.23) and (B.24) leads to

$$A\phi = \lambda A\phi$$

$$\lambda\phi = \lambda^2\phi$$

or

$$\lambda = \lambda^2 = 0, 1 \tag{B.25}$$

i.e., λ must be either 0 or 1. Since F is an $R \times \hat{M}$ matrix where $R \geqslant \hat{M}$ it is assumed that the rank of F, and therefore A, is $\hat{M}$. Therefore $\hat{M}$ eigenvalues of A are equal to one and the remaining $R - \hat{M} = G - 1$ eigenvalues are equal to zero. A can now be written, by means of a singular value decomposition [19], in the form

$$A = V\Lambda U^t \tag{B.26}$$

where Λ is a diagonal matrix of eigenvalues of the form

$$\Lambda = \begin{bmatrix} 1 & & & & 0 \\ & 1 & & & \\ & & 1 & & \\ & & & 0 & \\ 0 & & & & 0 \end{bmatrix} \tag{B.27}$$

and V and U are $R \times R$ unitary matrices. Applying this property to (B.22) gives

$$\tilde{\phi}_{xe}^t A\tilde{\phi}_{xe} = (\tilde{\phi}_{xe}^t V)\,\Lambda\,(\tilde{\phi}_{xe}^t U)^t. \tag{B.28}$$

In this form it can be seen that the presence of the zero eigenvalues in Λ tend to "dilute" the value $\tilde{\phi}_{xe}^t A\tilde{\phi}_{xe}$ by a factor of approximately $(G - 1)/R$.

Another way of looking at this property is to recognize that $(F^tF)^{-1}F$ is the generalized inverse of the matrix F [19]. That is if we define

$$F^* = (F^tF)^{-1}F^t \tag{B.29}$$

and

$$s = F^*\tilde{\phi}_{xe} \tag{B.30}$$

then s is the solution to the equation

$$u = \tilde{\phi}_{xe} - Fs \tag{B.31}$$

such that u^tu is a minimum. Using (B.31), u^tu can be expressed as

$$u^tu = (\tilde{\phi}_{xe} - Fs)^t(\tilde{\phi}_{xe} - Fs). \tag{B.32}$$

Applying (B.30) and expanding terms gives

$$\boldsymbol{u}^t\boldsymbol{u} = \tilde{\phi}_{xe}^t\tilde{\phi}_{xe} - 2\tilde{\phi}_{xe}^t\boldsymbol{F}\boldsymbol{F}^*\tilde{\phi}_{xe} + (\boldsymbol{F}\boldsymbol{F}^*\tilde{\phi}_{xe})^t(\boldsymbol{F}\boldsymbol{F}^*\tilde{\phi}_{xe}). \quad \text{(B.33)}$$

Using the definition of $\boldsymbol{F}^*$ in (B.29), the definition of A in (B.20) and canceling terms gives

$$\boldsymbol{u}^t\boldsymbol{u} = \tilde{\phi}_{xe}^t\phi_{xe} - \tilde{\phi}_{xe}^t A\tilde{\phi}_{xe} \quad \text{(B.34)}$$

or

$$\tilde{\phi}_{xe}^t A\tilde{\phi}_{xe} = \tilde{\phi}_{xe}^t\tilde{\phi}_{xe} - \boldsymbol{u}^t\boldsymbol{u}. \quad \text{(B.35)}$$

Thus, it is again seen that the matrix A effectively "dilutes" the inner product $\tilde{\phi}_{xe}^t\tilde{\phi}_{xe}$ by some positive value $\boldsymbol{u}^t\boldsymbol{u}$. Based on the derivation of the Q measure in (57) through (61) we see that

$$\tilde{\phi}_{xe}^t\tilde{\phi}_{xe} \cong RN\sigma_x^2\sigma_e^2$$
$$\approx \hat{M}N\sigma_x^2\sigma_e^2 + (G-1)N\sigma_x^2\sigma_e^2. \quad \text{(B.36)}$$

Thus the Q' measure can be written in the form

$$Q' \cong 10\log\left[\frac{\hat{M}\sigma_e^2}{N\sigma_v^2} + \left(\frac{(G-1)}{N}\frac{\sigma_e^2}{\sigma_v^2} - \frac{\boldsymbol{u}^t\boldsymbol{u}}{N^2\sigma_v^2\sigma_x^2}\right)\right]. \quad \text{(B.37)}$$

As G goes to 1 the matrix Λ becomes the identity matrix and $\boldsymbol{u}^t\boldsymbol{u}$ goes to zero because $\boldsymbol{F}$ has been assumed to have rank $\hat{M}$. In this case the Q' measure degenerates to the Q measure and the second two terms in (B.37) go to zero. Based on the experimental evidence for the LSA method in Section IV and from (B.28), it is anticipated that the second two terms in (B.37) always cancel, approximately giving

$$Q' \approx 10\log\left[\frac{\hat{M}}{N}\frac{\sigma_e^2}{\sigma_v^2}\right].$$

Thus, we expect that the Q' measure has the same theoretical expected value as the Q measure for white noise in (65).

ACKNOWLEDGMENT

The authors gratefully acknowledge the guidance and help provided by Dr. M. Sondhi in investigating several of the analytical problems discussed in the paper. In addition, Dr. D. Mitra and Dr. S. Levinson provided consulation on some of the work in Appendix II of this paper. The authors also wish to thank Dr. J. Tribolet for helpful comments on the manuscript.

REFERENCES

[1] P. Eykhoff, *System Identification.* New York: Wiley, 1974.
[2] B. Widrow, "Adaptive filters," in *Aspects of Network and System Theory.* R. Kalman and N. DeClaris, Eds., Holt, Rinehart and Winston, 1971, pp. 563–587.
[3] M. J. Levin, "Optimum estimation of impulse response in the presence of noise," *IRE Trans. Circuit Theory*, vol. CT-7, pp. 50–56, Mar. 1960.
[4] B. Widrow *et al.*, "Adaptive noise cancelling: Principles and applications," *Proc. IEEE*, vol. 63, pp. 1692–1716, Dec. 1975.
[5] B. Widrow, J. McCool, M. Larimore, and C. Johnson, "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proc. IEEE*, vol. 64, pp. 1151–1162, Aug. 1976.
[6] M. M. Sondhi and D. Mitra, "New results on the performance of a well known class of adaptive filters," *Proc. IEEE*, vol. 64, pp. 1583–1597, Nov. 1976.
[7] R. W. Lucky and W. R. Rudin, "An automatic equalizer for general-purpose communication channels," *Bell Syst. Tech. J.*, vol. 46, pp. 2179–2208, Nov. 1967.
[8] M. M. Sondhi, "An adaptive echo cancellor," *Bell Syst. Tech. J.*, vol. 46, no. 3, pp. 497–511, 1967.
[9] J. D. Gibson, S. V. Jones, and J. L. Melsa, "Sequentially adaptive prediction and coding of speech signals," *IEEE Trans. Comm.*, vol. COM-22, pp. 1789–1796, Nov. 1974.
[10] L. J. Griffiths, "A simple adaptive algorithm for real-time processing in antenna arrays," *Proc. IEEE*, vol. 57, pp. 1696–1704, Oct. 1969.
[11] N. S. Jayant, "Digital coding of speech waveforms: PCM, DPCM, and DM quantizers," *Proc. IEEE*, vol. 62, pp. 611–632, May 1974.
[12] J. Makhoul, "Linear prediction: a tutorial review," *Proc. IEEE*, vol. 63, pp. 561–580, May 1974.
[13] J. B. Allen, "Short-term spectral analysis and synthesis and modification by discrete Fourier transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-25, pp. 235–238, June 1977.
[14] M. Schwartz and L. Shaw, *Signal Processing: Discrete Spectral Analysis, Detection, and Estimation.* New York: McGraw-Hill, 1975, p. 278.
[15] G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications.* San Francisco: Holden-Day, 1968.
[16] R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra.* New York: Dover, 1958.
[17] J. B. Allen and L. R. Rabiner, "A unified approach to short-time Fourier analysis and synthesis," *Proc. IEEE*, vol. 65, pp. 1558–1564, Nov. 1977.
[18] M. M. Sondhi, private communication.
[19] G. W. Stewart, *Introduction to Matrix Computations.* New York: Academic, 1973.
[20] R. E. Crochiere, L. R. Rabiner, N. S. Jayant, and J. M. Tribolet, "A study of objective measures for speech waveform coders," in *Proc. Zurich Seminar on Comm.*, Mar. 1978, pp. H1.1–H1.7.