

# Short-Time Fourier Analysis Techniques for FIR System Identification and Power Spectrum Estimation

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

**Abstract**—A wide variety of methods have been proposed for system modeling and identification. To date, the most successful of these methods have been time domain procedures such as least squares analysis, or linear prediction (ARMA models). Although spectral techniques have been proposed for spectral estimation and system identification, the resulting spectral and system estimates have always been strongly affected by the analysis window (biased estimates), thereby reducing the potential applications of this class of techniques. In this paper we propose a novel short-time Fourier transform analysis technique in which the influences of the window on a spectral estimate can essentially be removed entirely (an unbiased estimator) by linearly combining biased estimates. As a result, section (FFT) lengths for analysis can be made as small as possible, thereby increasing the speed of the algorithm without sacrificing accuracy. The proposed algorithm has the important property that as the number of samples used in the estimate increases, the solution quickly approaches the least squares (theoretically optimum) solution. The method also uses a fixed Fourier transform length independent of the amount of data being analyzed, allowing the estimate to be recursively updated as more data is made available. The method assumes that the system is a finite impulse response (FIR) system.

## LIST OF SYMBOLS

$x(n)$  = input to system  
 $y(n)$  = output from system  
 $X(e^{j\omega}) = F[x]$   
 $Y(e^{j\omega}) = F[y]$   
 $\omega = 2\pi f$   
 $f$  = frequency  
 $w(n)$  = window function  
 $W(e^{j\omega}) = F[w]$   
 $x_{mR}(n) = w(mR - n)x(n)$   
 $y_{mR}(n) = w(mR - n)y(n)$   
 $R$  = window decimation period  
 $R_w(n)$  = autocorrelation of  $w(n)$   
 $\phi_{nm} = \sum x_{nR}x_{mR}$   
 $\phi = [\phi_{nm}]$  matrix  
 $n$  = time index  
 $\hat{S}_{xy}(e^{j\omega})$  = spectral estimate of  $YX^*$   
 $\hat{S}_{xx}(e^{j\omega})$  = spectral estimate of  $XX^*$   
 $X_m(e^{j\omega}) = F[x_{mR}]$   
 $Y_m(e^{j\omega}) = F[y_{mR}]$   
 $X^*$  = complex conjugate of  $X$

$\Phi = F[\phi]$   
 $L$  = length of  $w(n)$   
 $N$  = length of DFT  $F[\ ]$   
 $N'$  = number of data points  
 $N_1$  = lower limit on  $\phi$  sum  
 $N_2$  = upper limit on  $\phi$  sum  
 $M$  = length of  $h$   
 $\hat{M}$  = estimate of  $M$   
 $h(n)$  = systems impulse response  
 $\hat{h}(n)$  = estimate of  $h(n)$   
 $F[\ ]$  = Fourier transform operator  
 $F^{-1}[\ ]$  = inverse Fourier transform  
 $q_1$  = number of  $\phi$  diagonals below main  
 $q_2$  = number of  $\phi$  diagonals above main  
 $\hat{H}(e^{j\omega}) = F[\hat{h}]$   
 $H(e^{j\omega}) = F[h]$   
 $Q$  = misalignment of  $\hat{h}$  (dB scale)  
 $\epsilon(n)$  = noise at system output  
 $\epsilon(e^{j\omega}) = F[\epsilon(n)]$   
 $\delta(l) = \begin{cases} 1 & \text{for } l = 0 \\ 0 & \text{for } l \neq 0 \end{cases}$   
 $E(x)$  = statistical expectation

## I. INTRODUCTION

ALTHOUGH time-domain methods for system identification and modeling have been in use since at least the time of Wiener [1]-[10], no simple, robust procedure has been proposed which operates entirely in the frequency domain. Classical spectral analysis methods are generally inadequate for all but the simplest cases because of their unsatisfactory properties [8]. Recent results on the theory of short-time spectral analysis have suggested a framework for an "efficient" and accurate frequency-domain procedure for spectral estimation and system identification [11], [12]. It is the purpose of this paper to describe this new algorithm and present several examples demonstrating its improved effectiveness.

Fig. 1 shows a block diagram of the general system identification model. The input signal  $x(n)$  is assumed to be a white Gaussian noise with variance  $\sigma_x^2$  and zero mean. The linear system  $h(n)$  is assumed to be a finite impulse response (FIR) system of duration  $M$  samples,

$$h(n) = 0 \quad n < 0, \quad n > M - 1. \quad (1)$$

Manuscript received June 8, 1978; revised September 25, 1978.

The authors are with the Acoustics Research Department, Bell Laboratories, Murray Hill, NJ 07974.

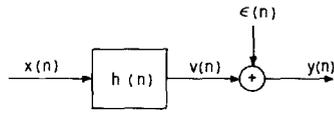


Fig. 1. Block diagram of the system identification model.

At the output of the linear system, an independent white Gaussian noise  $\epsilon(n)$  is added to  $v(n)$  to give the output signal  $y(n)$ . Thus,

$$y(n) = x(n) * h(n) + \epsilon(n) \quad (2a)$$

$$= \sum_{m=0}^{M-1} h(m) x(n-m) + \epsilon(n). \quad (2b)$$

The signal-to-noise ratio (SNR) at the output of the system is defined as

$$\text{SNR} = 10 \log_{10} \left( \frac{E(v^2(n))}{E(\epsilon^2(n))} \right) = 10 \log_{10} \left( \frac{\sigma_v^2}{\sigma_\epsilon^2} \right). \quad (3)$$

The system identification problem is one of finding an estimate  $\hat{h}(n)$  of  $h(n)$ , given only  $N'$  samples of the input  $x(n)$  and the output  $y(n)$ . In general,  $N' \gg \hat{M}$ . It is assumed that the duration of  $\hat{h}(n)$  (call this  $\hat{M}$ ) satisfies the relation

$$\hat{M} \geq M, \quad (4)$$

i.e., that we have knowledge of, or can accurately bound the duration of the system impulse response. If (4) is not satisfied, the estimation errors are strongly dependent on  $h(n)$  and it becomes very difficult to assess system performance. We will not be concerned with such cases. Therefore, assuming that the constraint of (4) is obeyed, one reasonable measure of system performance is the quantity [6]

$$Q = 10 \log_{10} \left[ \frac{\sum_{m=0}^{\hat{M}-1} [h(m) - \hat{h}(m)]^2}{\sum_{m=0}^{\hat{M}-1} h^2(m)} \right] \text{ (dB)}. \quad (5)$$

The quantity  $Q$  is called the "misadjustment" or "misalignment" between  $h(n)$  and  $\hat{h}(n)$ , expressed in decibels. It has been shown that for *white* input signals the measure of (5) provides a good description of the performance of a system identification method [7].

Several classes of "optimal" techniques in the time domain have been developed for solving the system identification problem. Included among them are the classical least squares analysis (LSA), and the least mean squares (LMS) adaptation algorithm. In the frequency domain, however, only very simple, suboptimal techniques have been proposed for solving the system identification problem, and even these techniques have proven not to be entirely adequate for any reasonable class of problems [7]. In this paper we drive a new frequency-domain approach to the system identification problem which alleviates many of the problems encountered using previously proposed frequency-domain methods. High quality frequency-domain techniques are interesting for many reasons, but two notable ones are that: 1) very large systems may be dealt with

accurately ( $\hat{M} \approx 10^3$  points), and 2) FFT methods are numerically very efficient. Furthermore, with the advent of high-speed array processors, algorithmic procedures which use FFT's are frequently easily programmed.

In Section II we review the ideas of short-time spectral analysis, and show how they can be applied to the problem at hand to give a frequency-domain solution to the problem. In Section III we show results of simulations using the algorithm obtained in Section II. In Section IV we discuss the properties of the algorithm and point to areas in which subsequent investigations would appear to be fruitful.

## II. LEAST SQUARES SOLUTION IN THE FREQUENCY DOMAIN

### A. Short-Time Fourier Transform Relations

When the signals have finite energy we may define the Fourier transform (on an infinite time interval) of (2a) to give the relation

$$Y(e^{j\omega}) = X(e^{j\omega}) \cdot H(e^{j\omega}) + \epsilon(e^{j\omega}) \quad (6)$$

where  $\epsilon(e^{j\omega}) = F[\epsilon(n)]$ . For noise signals, however, the Fourier transform does not exist. The classical method for obtaining a realizable Fourier transform from a noise signal is to Fourier transform a rectangular window of length  $L$ , divide by the window length  $L$ , and then take the limit as  $L$  goes to  $\infty$ . The alternative that we propose here is to utilize the theory of short-time Fourier analysis [11], [12] to provide insights on how to deal with noise-like (constant power) signals in the frequency domain.

We recall that  $X(e^{j\omega})$ , the short-time Fourier transform of the signal  $x(n)$ , at time  $mR$ , is defined as

$$X_m(e^{j\omega}) = \sum_{l=mR-L+1}^{mR} x(l) w(mR-l) e^{-j\omega l} \quad (7a)$$

$$= F[x(l) w(mR-l)] \quad (7b)$$

$$= F[x_{mR}(l)] \quad (7c)$$

where  $R$  is the period (in samples) between adjacent estimates of the short-time transform of the signal,  $w(n)$  is a causal (i.e.,  $w(n) = 0$  for  $n < 0$ ) low-pass FIR filter of duration  $L$  samples which determines both the temporal and spectral resolution of the transform and  $F[\ ]$  denotes the operation of taking a Fourier transform of the sequence within the brackets. If we express the windowed signal at time  $mR$  as

$$x_{mR}(n) = x(n) w(mR-n) \quad (8a)$$

$$= F^{-1}[X_m(e^{j\omega})] \quad (8b)$$

where  $F^{-1}[\ ]$  denotes the inverse Fourier transform of the function within the brackets, then it has been shown [11] that for any "bandlimited" window  $w(n)$ , if the short-time transform is obtained at a sufficiently high rate (i.e., at or above the "Nyquist" decimation rate of the window), then the original signal  $x(n)$  can be recovered within a negligible aliasing error by "overlap adding" the windowed signal  $x_{mR}(n)$ , i.e.,

$$x(n) = \frac{1}{D} \sum_{m=-\infty}^{\infty} x_{mR}(n) \quad (9)$$

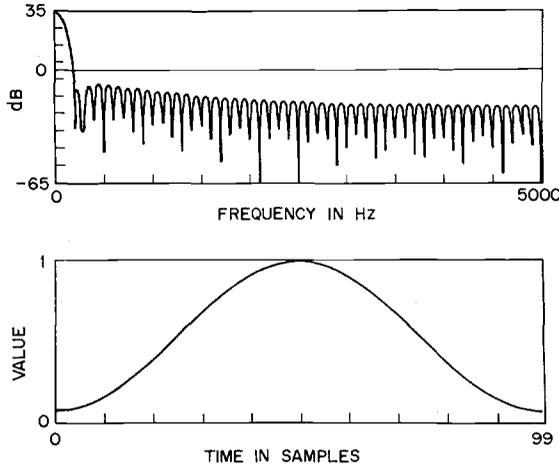


Fig. 2. A Hamming window of length 100 samples and its Fourier transform. The sampling frequency is 10 kHz.

$$= \frac{1}{D} \sum_{m=-\infty}^{\infty} x(n) w(mR - n)$$

where

$$D = \frac{W(e^{j0})}{R} \quad (10)$$

and  $W(e^{j0})$  is the zero frequency value of the Fourier transform of the window.

By way of example consider an  $L$ -point Hamming window, as shown in Fig. 2, with  $L = 100$ . The bandwidth of this window (for a 10 kHz sampling rate) is approximately 200 Hz, i.e., for frequencies above 200 Hz the log magnitude response of the window falls below -42 dB and never rises above this level. The Nyquist rate of the window is 400 Hz, i.e., samples of the short-time transform must be taken at least 400 times per second, or at least once every 25 samples. Thus  $R$ , for the Hamming window, is an integer in the range  $1 \leq R \leq 25$ . Typically, a value of  $R = 25$ , i.e., one quarter of the window duration, is used for short-time spectral analysis with a Hamming window.

In a similar manner, the short-time transform of the output  $Y_m(e^{j\omega})$  is defined as

$$Y_m(e^{j\omega}) = \sum_{l=mR-L+1}^{mR} y(l) w(mR-l) e^{-j\omega l} \quad (11a)$$

$$= F[y(l) w(mR-l)] \quad (11b)$$

$$= F[y_{mR}(l)]. \quad (11c)$$

Again, defining a windowed output signal  $y_{mR}(n)$  as

$$y_{mR}(n) = y(n) w(mR-n) \quad (12a)$$

$$= F^{-1}[Y_m(e^{j\omega})] \quad (12b)$$

then, as in (9),  $y(n)$  can be expressed as

$$\begin{aligned} y(n) &= \frac{1}{D} \sum_{m=-\infty}^{\infty} y_{mR}(n) \\ &= \frac{1}{D} \sum_{m=-\infty}^{\infty} F^{-1}[Y_m(e^{j\omega})] \end{aligned}$$

$$= \frac{1}{D} F^{-1} \left[ \sum_{m=-\infty}^{\infty} Y_m(e^{j\omega}) \right]. \quad (13)$$

Since linear filtering can be implemented exactly with short-time Fourier transforms [11], it has been shown that  $v(n)$  of Fig. 1 can be obtained as

$$\begin{aligned} v(n) &= \frac{1}{D} \sum_{m=-\infty}^{\infty} F^{-1}[X_m(e^{j\omega}) H(e^{j\omega})] \\ &= \frac{1}{D} F^{-1} \left[ \sum_{m=-\infty}^{\infty} X_m(e^{j\omega}) H(e^{j\omega}) \right] \end{aligned} \quad (14)$$

where  $H(e^{j\omega})$  is the Fourier transform of  $h(n)$  of Fig. 1.

If we temporarily consider the case where  $\epsilon(n)$  is 0 (i.e., no noise), then from (13) and (14) we get

$$\sum_{m=-\infty}^{\infty} X_m(e^{j\omega}) H(e^{j\omega}) = \sum_{m=-\infty}^{\infty} Y_m(e^{j\omega}). \quad (15)$$

It should be emphasized that it is *not* true that for any  $m$

$$X_m(e^{j\omega}) H(e^{j\omega}) = Y_m(e^{j\omega}). \quad (16)$$

Equation (16) has been the basis of some spectral estimation procedures where it is assumed (without formal justification [8]) that the results on an infinite time basis (6) can be extended to the short-time (windowed) case.

Equation (15), although theoretically exact, is only valid when the summation is for all  $m$ . When a finite range on  $m$  is used, however, the result is still approximately true. (In the time domain it can be shown that  $\hat{y}(n) = \hat{x}(n) * \hat{h}(n)$ , where  $\hat{x}$  and  $\hat{y}$  are obtained by truncating the sum on  $m$ , except for a small number of points at the beginning and the end of the analysis interval. Thus, it is anticipated that as the number of analysis intervals (range of  $m$ ) increases, the quality of the approximation improves.) Thus, if we define the error signal  $e(n)$  as

$$e(n) = y(n) - x(n) * \hat{h}(n) \quad (17)$$

or in frequency as

$$E(e^{j\omega}) = \sum_{m=-\infty}^{\infty} [Y_m(e^{j\omega}) - X_m(e^{j\omega}) \hat{H}(e^{j\omega})] \quad (18)$$

then we have a quantity  $(E(e^{j\omega}))$  which has the following components.

1) A component due to the difference between  $h(n)$  and  $\hat{h}(n)$  (or  $H(e^{j\omega})$  and  $\hat{H}(e^{j\omega})$ ) (misalignment error).

2) An aliasing component due to the imperfect bandlimiting ability of the window (aliasing error).

In any practical implementation of estimates for  $\hat{H}$ , component 1) will consist of two misalignment components: a) an error due to the finite range on  $m$  (end misalignment), and b) an error due to the effects of the noise  $\epsilon$  on our estimate of  $\hat{h}$  (noise misalignment). Note that these components are essentially independent.<sup>1</sup>

<sup>1</sup>The noise error depends on the signal to noise ratio SNR, the end effect on the ratio of the number of points in the estimate  $N'$  to the length of the window  $L$ , and the aliasing error on the window quality and sampling rate  $R$ .

We begin by minimizing a functional  $I$  with respect to  $\hat{h}(n)$ ,  $0 \leq n \leq \hat{M} - 1$ , defined as

$$I = \sum_{n=-\infty}^{\infty} e^2(n) \tag{19}$$

giving a smallest value of  $I$  (call this  $I_{\min}$ ), defined as

$$I_{\min} = \min_{\hat{h}} \sum_{n=-\infty}^{\infty} e^2(n). \tag{20}$$

A solution to (20) may be obtained using standard least squares techniques as reviewed in Appendix A.

**B. Solution for  $\hat{H}(e^{j\omega})$**

To implement the procedure given in the preceding section, we first find the solution to (20) in the time domain and then transform the result to the short-time Fourier transform domain by the use of relations (8), (9), (12), and (13). We then study some important properties of the resulting equation.

The solution to (20) is given in the Appendix. For clarity the final result is repeated here as

$$\sum_{n'=0}^{\hat{M}-1} \hat{h}(n') \phi(n' - l) = r(l) \quad 0 \leq l \leq \hat{M} - 1 \tag{21}$$

or in short-hand notation

$$\phi * \hat{h} = r. \tag{22}$$

Note that (21) is an  $\hat{M} \times \hat{M}$  matrix equation. In (21),  $\phi$  and  $r$  are defined as

$$\phi(l) = \sum_{n'=-\infty}^{\infty} x(n') x(n' - l) \tag{23}$$

$$r(l) = \sum_{n'=-\infty}^{\infty} y(n') x(n' - l). \tag{24}$$

By use of our expansions in terms of windowed signals, (9), (13), we may express  $\phi$  and  $r$  in terms of windowed signals.

$$\phi(l) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \phi_{nm}(l) \tag{25}$$

$$r(l) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} r_{nm}(l) \tag{26}$$

where we have defined

$$\phi_{nm}(l) = \frac{1}{D^2} \sum_{n'=-\infty}^{\infty} x_{nR}(n') x_{mR}(n' - l) \tag{27}$$

$$r_{nm}(l) = \frac{1}{D^2} \sum_{n'=-\infty}^{\infty} y_{nR}(n') x_{mR}(n' - l). \tag{28}$$

Thus (21), the solution to (20), becomes

$$\sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \phi_{nm}(l) * \hat{h}(l) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} r_{nm}(l) \tag{29}$$

$$0 \leq l \leq \hat{M} - 1.$$

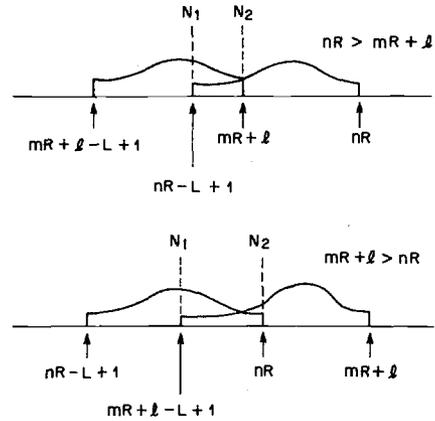


Fig. 3. Relative position of the windows for the matrix element  $\phi_{nm}$ .

To understand the properties of (29), we first examine the matrix of terms  $\phi_{nm}(l)$  in more detail. Define matrix  $\phi(l)$  as

$$\phi(l) = \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} & \cdots \\ \phi_{21} & \phi_{22} & \cdots & \\ \vdots & \vdots & \cdots & \\ \vdots & \vdots & \cdots & \end{bmatrix}. \tag{30}$$

In actual fact, this matrix is an infinite matrix and we are only considering a piece of the full  $\phi_{nm}$  here.

Since a typical element of  $\phi$  is

$$\phi_{nm}(l) = \frac{1}{D^2} \sum_{n'=-\infty}^{\infty} x_{nR}(n') x_{mR}(n' - l) \tag{31a}$$

$$= \frac{1}{D^2} \sum_{n'=N_1}^{N_2} w(nR - n') x(n') w(mR + l - n') x(n' - l) \tag{31b}$$

the sum on  $n'$  is finite and depends on  $nR$  and  $l + mR$  due to the time truncation of the window  $w(n)$ . For a causal window  $L$  points long (Fig. 2), the lower limit  $N_1$  and the upper limit  $N_2$  of the  $n'$  sum are given by

$$N_1 = \max(l + mR, nR) + 1 - L \tag{32a}$$

$$N_2 = \min(l + mR, nR). \tag{32b}$$

This may be seen by studying Fig. 3. If  $N_2$  is less than  $N_1$ , the windows do not overlap and  $\phi_{nm}$  is identically zero. We may determine the matrix elements  $\phi_{nm}$ , which are identically equal to zero (because of the no overlap condition), if we examine the limits on the range of  $l$  which is determined by the length of  $\hat{h}$ , i.e.,

$$0 \leq l \leq \hat{M} - 1.$$

For the lower diagonals of  $\phi_{nm}$  where  $n - l/R > m$  the limits are

$$N_1 = nR + 1 - L$$

$$N_2 = mR + l,$$

while for the upper diagonals where  $m > n - l/R$  they are

$$N_1 = mR + l + 1 - L$$

$$N_2 = nR.$$

For the lower diagonals  $\phi_{nm}$  will be identically zero when  $N_1 > N_2$ , or

$$-l + nR > mR + L - 1$$

which for  $l = (\hat{M} - 1)$  becomes (the worst case)

$$(n - m) > (L + \hat{M} - 2)/R. \quad (33)$$

For the upper diagonals,  $\phi_{nm}$  will be identically zero for  $N_1 > N_2$  or

$$l + 1 - L + mR > nR$$

which, for  $l = 0$ , gives

$$(m - n) > (L - 1)/R. \quad (34)$$

By way of example, suppose that  $\hat{M}$  is chosen equal to the window length  $L$  and that the window is a Hamming window so that  $L = 4R$ . Then the matrix  $\phi$  is an eleven diagonal band matrix

$$\phi = \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} & \phi_{14} & 0 & 0 \\ \phi_{21} & \phi_{22} & \phi_{23} & \phi_{24} & \phi_{25} & \\ \phi_{31} & \phi_{32} & \phi_{33} & & & \\ \mathbf{0} & & & & & \end{bmatrix}. \quad (35)$$

The sum on  $r_{nm}$  in (29) may be treated in an identical manner.

Because of the nature of the zero diagonals of  $\phi$ , it is useful to rewrite the sums over  $\phi_{nm}$  and  $r_{mn}$  by a transformation of variables to diagonal coordinates. In this form our least squares equation (29) becomes

$$\sum_{m=-\infty}^{\infty} \sum_{q=q_1}^{q_2} \hat{h}(l) * \phi_{m,m+q}(l) = \sum_{m=-\infty}^{\infty} \sum_{q=q_1}^{q_2} r_{m,m+q}(l) \quad 0 \leq l \leq \hat{M} - 1 \quad (36)$$

where

$$q_1 = -[(L + \hat{M} - 2)/R] \quad (37a)$$

$$q_2 = [(L - 1)/R] \quad (37b)$$

where  $[x]$  denotes the integer part of  $x$ . For our example above,  $q_1 = -7$  and  $q_2 = 3$ .

We would like to point out that the matrix elements at the beginning and end of each row are very small relative to those in the center because the window is small at its ends and because  $N_2 - N_1$  becomes small. The ratio of a typical center term  $\phi_{mm}$  to an end term  $\phi_{m,m+q_2}$  is on the order of

$$\frac{\phi_{m,m+q_2}(0)}{\phi_{m,m}(0)} = \frac{w(0)w(L)}{\sum_{n=0}^{L-1} w^2(n)} \quad (38)$$

where we have assumed  $x(n) = 1$ , all  $n$ , and  $q_2$  is such that the  $m$ th and the  $(m + q_2)$ th windows overlap at a single point.

When  $q_1 = q_2 = 0$ , (29) reduces to the classical method well known in the literature [8]. The sums along off diagonal rows of the matrix correspond to estimates of  $\hat{h}$  with somewhat different window properties. That this is true may be demonstrated by isolating the  $q$ th diagonal sum of  $\hat{h}$  for the case of

white noise input ( $\phi(l) = \sigma_x^2 \delta(l)$ ), namely we look at a modified version of (36) which is defined by ignoring the summation on  $q$  and considering a single value of  $q$ , i.e.,

$$\sum_{m=-\infty}^{\infty} \hat{h}(l) * \phi_{m,m+q}(l) = \sum_{m=-\infty}^{\infty} r_{m,m+q}(l). \quad (39)$$

If we define the autocorrelation of the window  $R_w(n)$  as

$$R_w(n) = \sum_{m=-\infty}^{\infty} w(mR)w(mR+n), \quad (40)$$

then (39) may be expressed in terms of  $R_w$  as

$$\sum_p \hat{h}(l-p) R_w(p+qR) \phi(p) = R_w(l+qR) r(l) \quad (41)$$

which, when we use the white noise assumption, reduces to

$$\hat{h}(l) R_w(qR) \sigma_x^2 = R_w(l+qR) r(l). \quad (42)$$

The equation for  $r(l)$  may be written in terms of  $h$  since

$$r(l) = h(l) \sigma_x^2. \quad (43)$$

Combining (43) with (42) we find

$$\hat{h}(l) = \frac{R_w(l+qR)}{R_w(qR)} h(l). \quad (44)$$

Equation (44) tells us that a sum along the  $q$ th diagonal would give us an estimate  $\hat{h}$  of  $h$  which is modified by the autocorrelation of the window  $R_w$  shifted by  $qR$ .

By summing over  $q$ , the bias effect of the window as seen in (44) is removed because, for any low-pass function,  $R_w$  in this case, and for a properly chosen decimation period  $R$ ,

$$\sum_{q=-\infty}^{\infty} R_w(l+qR) = DW(e^{j\omega}) \quad (45)$$

which is independent of  $l$ , as shown in [11].

To illustrate the implications of the above discussion, Fig. 4 caricaturizes an impulse response sequence  $h(n)$  which is non-zero for  $M = 10$  terms, and a window autocorrelation for a 16-point window. For values of  $q$  near 0 (i.e.,  $q = 0, \pm 1$ ) the overlap between  $h(n)$  and  $R_w(n+qR)$  is substantial. However, for  $q < -4$  and for  $q > 3$ , there is no overlap between  $h(n)$  and  $R_w(n+qR)$ . Thus, using such terms in the summations of (36) leads to adding no information to the spectral estimate. In fact, even the  $q = \pm 3, -4$  terms tend to have more noise than signal since the impulse response is multiplied by the tails of the autocorrelation of the window, as seen from (44).

Equation (36), or its Fourier transform, tells us that the value of the spectral estimate, in the case of white noise, is exactly the true impulse response when a sufficient number of sums of off diagonal matrix elements (values of  $q$ ) are included in the summation.

Based on the above discussion, our proposed spectral estimate of  $H(e^{j\omega})$  is found by Fourier transforming (36) and is of the form

$$\hat{H}(e^{j\omega}) = \hat{S}_{xy}(e^{j\omega}) / \hat{S}_{xx}(e^{j\omega}) \quad (46)$$

where, after truncating the sum on  $m$  to  $P$  terms, gives

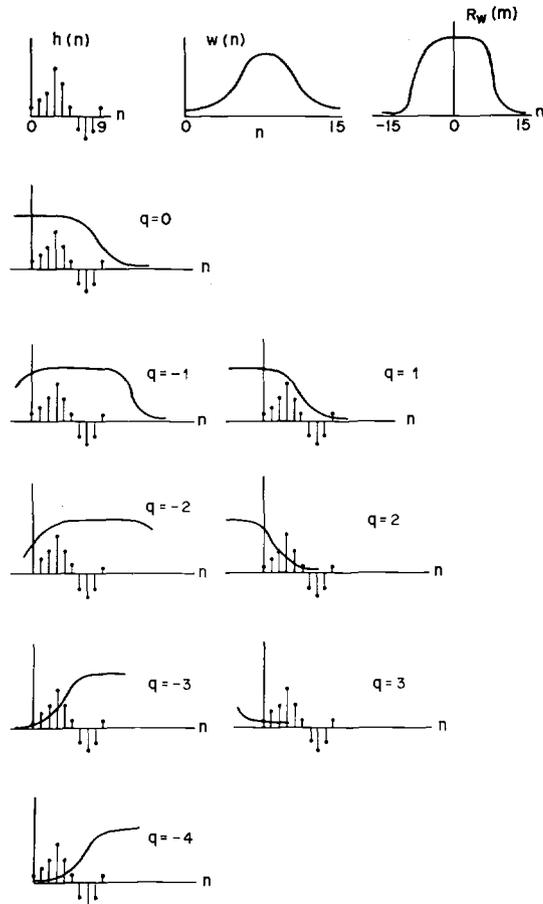


Fig. 4. Sequence of caricatures showing the overlap between the 10-point impulse response sequence  $h(n)$  and the windowed autocorrelation  $R_w(n)$  for several values of relative shift  $q$ .

$$\hat{S}_{xx}(e^{j\omega}) = \sum_{m=0}^{P-1} \sum_{q=q_1}^{q_2} X_m(e^{j\omega}) X_{m+q}^*(e^{j\omega}) \quad (47)$$

$$\hat{S}_{xy}(e^{j\omega}) = \sum_{m=0}^{P-1} \sum_{q=q_1}^{q_2} Y_m(e^{j\omega}) X_{m+q}^*(e^{j\omega}). \quad (48)$$

Algorithmically, the procedure for implementing the spectral estimate of (46) using discrete Fourier transforms to give the spectral estimates of (47) and (48) is as follows.

1) Choose a section (window) length  $L$ . The section length should be at least as large as the estimated impulse response duration  $\hat{M}$  of the system being estimated.

2) Choose the sampling period  $R$  for computation of the short-time spectra. As discussed previously, the value of  $R$  is chosen based on the window characteristics. For a Hamming window,  $R \leq L/4$ . We assume  $R = L/4$  for illustrative purposes.

3) Choose the total number of data samples  $N'$  used in the analysis. Previously published data provide a reasonable estimate of the required length  $N'$  for a given performance criterion  $Q$  at a given signal-to-noise ratio (3) [7]. The number of analysis sections  $P$  is then given as

$$P = \left\lfloor \frac{(N' - L + R)}{R} \right\rfloor \quad (49)$$

where  $\lfloor x \rfloor$  is the greatest integer less than or equal to  $x$ .

4) Choose the FFT size  $N$  such that

$$N \geq \hat{M} + L - 1 \quad (50)$$

and such that  $N$  is compatible with the FFT program used to compute the short-time spectra (e.g.,  $N$  a power of 2).

5) Compute short-time spectra of input and output signals for times  $n = mR + L$ ,  $m = 0, 1, \dots, P - 1$  by windowing the  $m$ th section of signal by  $w(n)$ , padding the  $L$ -point sequence with  $(N - L)$  zero valued samples, and taking an  $N$ -point FFT. If we number the samples as  $n = 1, \dots, N'$ , then the  $m$ th frame is defined from

$$\begin{aligned} x_{mR+L}(mR - n) &= x(mR + L - n) w(n) & 0 \leq n < L - 1 \\ & & 0 \leq m \leq P - 1 \end{aligned} \quad (51a)$$

$$\begin{aligned} y_{mR+L}(mR - n) &= y(mR + L - n) w(n), & 0 \leq n \leq L - 1 \\ & & 0 \leq m \leq P - 1. \end{aligned} \quad (51b)$$

6) Compute the numerator terms of the spectral estimate [from (48)] for each value of  $q$  used in the estimate and sum the results and divide by the sum of the denominator terms (47). Both the numerator and denominator terms are ob-

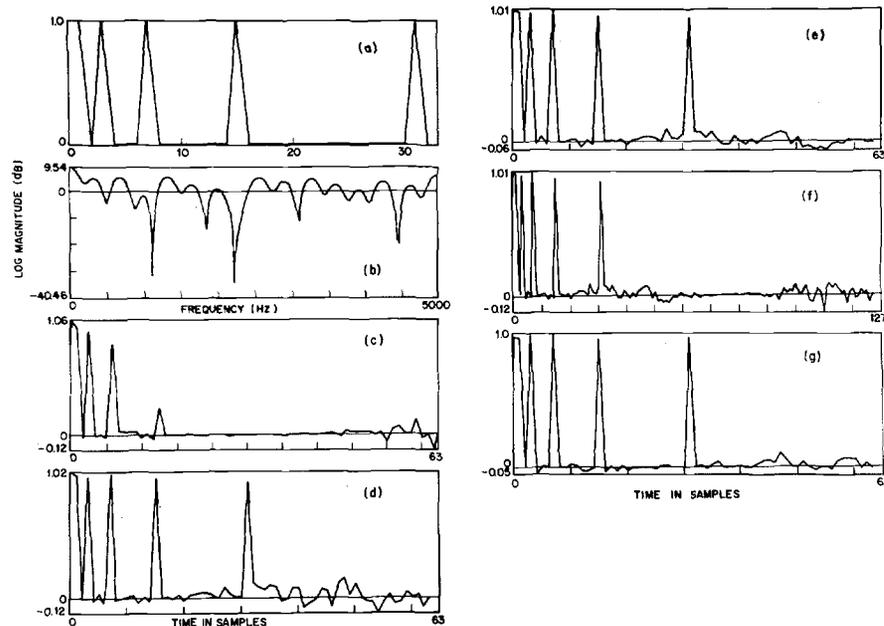


Fig. 5. An example illustrating the spectral estimates for several values of the analysis parameters.

tained (via an FFT computation) at the discrete set of frequencies  $\omega_k = (2\pi/N)k$ , giving the estimate for  $\hat{H}(k)$  rather than  $\hat{H}(e^{j\omega})$ . (Compensation for the shift of  $qR$  samples between windows must be made in the DFT implementation.)

7) Inverse transform  $\hat{H}(k)$  to give  $\hat{h}(n)$  on the interval  $0 \leq n \leq \hat{M} - 1$ .

In the next section we present results which illustrate the operation of the above algorithm for a range of values of  $L$  and  $N'$ .

### III. EXPERIMENTAL RESULTS

To investigate the performance of this new spectral technique for system identification by specific examples, we implemented the system of Fig. 1. The input ( $x(n)$ ) was a Gaussian white noise signal, and the impulse response ( $h(n)$ ) was one of two systems. The first system is shown in Fig. 5. Fig. 5(a) shows the impulse response of a filter which was of the form

$$\begin{aligned} h(n) &= 1 & n = 0, 1, 3, 7, 15, 31 \\ &= 0 & \text{otherwise} \end{aligned} \quad (52)$$

and Fig. 5(b) shows the frequency response of the filter. For this example  $M = 32$ .

Fig. 5(c)-(g) shows the results of applying (46) to this system (we assume here that  $\epsilon(n) = 0$ , i.e., infinite signal-to-noise ratio) to estimate  $\hat{h}(n)$  as the inverse transform of  $\hat{H}(k)$ . It should be noted that different time scales are used in parts (a) and (f) than in parts (c), (d), and (e) of this figure. For all sections of Fig. 5 the short-time spectra  $Y_l$  and  $X_l$  were computed using a Hamming window with a shift of  $R = L/4$  samples between adjacent short-time spectral estimates. The examples of Fig. 5 differ in the number of matrix diagonals (i.e., values of  $q$ ) used in the system estimate and in the size FFT

( $N$ ) used in the spectral analysis. In all the experimental studies we used  $q_2 = q_0$  and  $q_1 = -q_0$ . Fig. 5(c) shows the results obtained with a section size of  $L = 32$  samples, with  $N' = 512$  data points,  $q_0 = 0$  (i.e., the "classical" way of doing spectral estimation), and an FFT size of 512.<sup>2</sup> The resulting estimate of the impulse response  $\hat{h}(n)$  shown in Fig. 5(c) displays the problem associated with the classical analysis procedure, i.e., the true impulse response is weighted by the autocorrelation of the analysis window, leading to a very poor estimate of  $h(n)$  (bias in the frequency estimate). The impulse at  $n = 15$  is strongly attenuated, and the impulse at  $n = 31$  is hidden in the noise. We can alleviate these problems with a larger value of  $q_0$ . Fig. 5(d) shows the results for  $L = 32$ ,  $N' = 512$ ,  $q_0 = 3$  (e.g.,  $q_1 = -3$ ,  $q_2 = 3$ ), and an FFT size of 64 points. It can be seen that a section size  $L$ , comparable to the impulse response duration  $M$ , is entirely adequate as shown in this figure. The impulses in this figure are all essentially of equal amplitude (the bias has been eliminated). However, it can be seen that there is a considerable amount of noise throughout  $\hat{h}(n)$  due to the spectral division in (46). This noise can be reduced by using a larger FFT size, as shown in Fig. 5(e) where all parameters of the example of Fig. 5(d) were held constant except the FFT size which was doubled to 128 points. The noise of this example is approximately half that of the preceding case. Fig. 5(f) shows the estimate  $\hat{h}(n)$  plotted for all 128 points, showing that the error is most severe near  $n = 128$  (i.e., negative time samples of the DFT) and least severe near  $n = 64$ , i.e., at half the FFT size. Finally, Fig. 5(g) shows the results of doubling the section size  $L$ , but keeping the total number of samples  $N'$  the same value. Again, the noise is reduced be-

<sup>2</sup>More recent, unpublished work has shown how to use smaller FFT sizes with no less accuracy in the result.

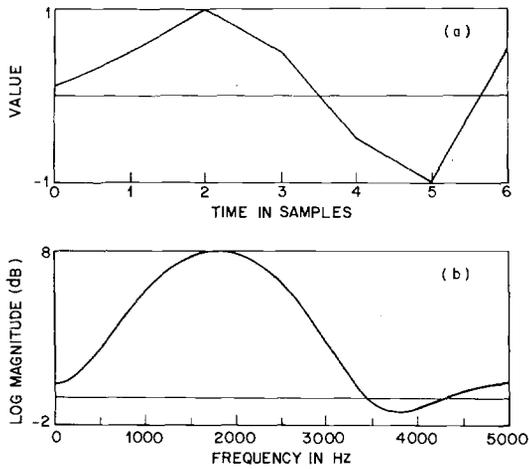


Fig. 6. Impulse response and frequency response of a simple 7-point FIR system used to evaluate the spectral method. The sampling frequency is 10 kHz.

cause the amount of coherent spectral information is increased, relative to the fixed noise component.

The preceding example showed that when a sufficient number of matrix diagonals were used in the computation, it was possible to obtain a good estimate of the system impulse response that was independent of the analysis window. We now examine the individual terms that go into the summations of (47), (48). To do this we use the system with impulse response shown in Fig. 6(a) and frequency response shown in Fig. 6(b). The impulse response duration in this case is  $M = 7$  samples [7].

Figs. 7 and 8 show sequences of plots of the time functions  $g_1(n)$  and  $g_2(n)$  defined as

$$g_1(n) = F^{-1}[\hat{S}_{xx}(e^{j\omega})] \tag{53}$$

and

$$g_2(n) = F^{-1}[\hat{S}_{xy}(e^{j\omega})] \tag{54}$$

where we have set  $-q_1 = q_2 = q$  in (47), (48).  $g_1$  and  $g_2$  are then the inverse transforms of sums along the  $q$ th diagonal alone, as discussed in (39)-(44). Fig. 7(a)-(d) shows plots of  $g_1(n)$  for  $q = 0, 1, 2,$  and  $3,$  respectively. As shown earlier [left-hand side of (42)] the expected value of  $g_1(n)$  is an impulse (for the white noise case) weighted by the autocorrelation function of the window shifted by  $qR$  samples. The examples of Fig. 7 were computed using a section length of  $L = 64$  samples, and a total number of samples  $N' = 128$  using a 128-point FFT. A Hamming window was used throughout with a shift of  $R = L/4 = 16$  samples. Fig. 7(a)-(d) shows that as  $q$  increases, the effective noise level of the computation increases due to the decrease in amplitude of the impulse at  $n = 0$ . For  $q = 3,$  the peak noise level is almost three times the level of the impulse at  $n = 0$ .

Fig. 8(a)-(h) shows plots of  $g_2(n)$  for  $q = 0, +1, -1, +2, -2, +3, -3,$  and  $-4,$  respectively. It is clear that the weighting effect of the window has its greatest effect for  $q = +2, \pm 3,$  and  $-4$  as the window autocorrelation tails align with  $h(n)$ . It can also be seen that the worst noise in the estimate occurs for

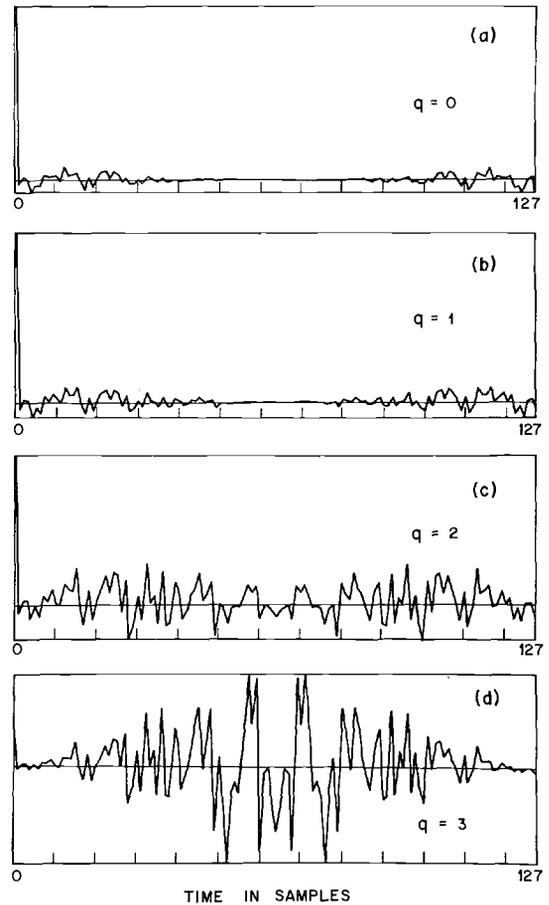


Fig. 7. Plots of the function  $g_1(n)$  for several values of  $q$ . In this example  $q$  indicates the diagonal of the matrix  $\phi_{nm}$ . All plots are normalized to full scale.

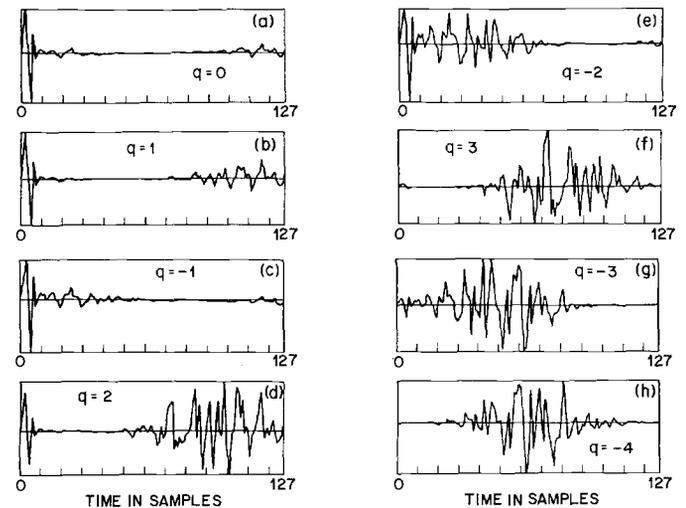


Fig. 8. Plots of the function  $g_2(n)$  for several values of  $q$ , where  $q$  indicates the diagonal used in the sum. For this example  $M = 7, L = 64, N = 128$ . All plots are normalized to full scale.

large  $n$  for  $q$  negative, and for small  $n$  for  $q$  positive. This is due to the window autocorrelation function which is centered on large  $n$  for  $q$  negative, and small  $n$  for  $q$  positive, and thus increases the noise level at these points. For  $q = 0,$  the auto-

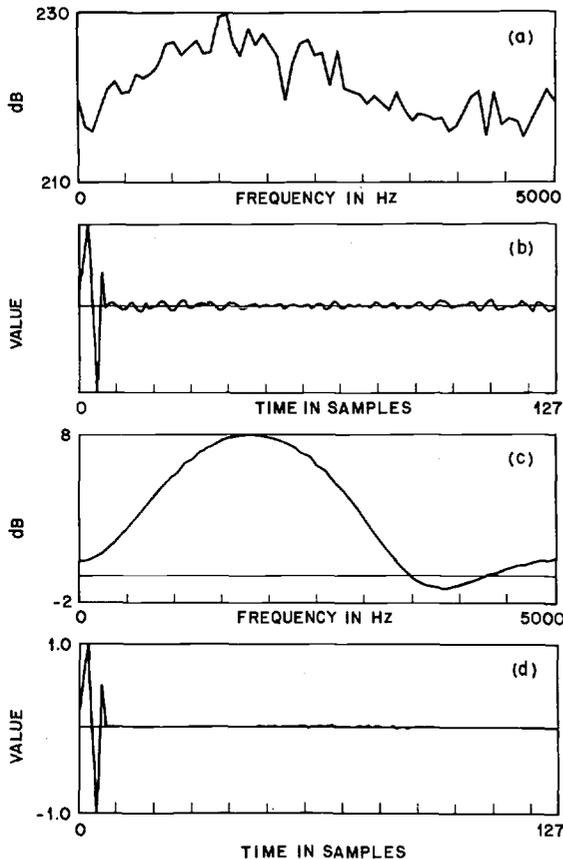


Fig. 9. Example illustrating the importance of the denominator term of the spectral estimate for  $\hat{H}$ , (46). Panels (a), (b) are  $\hat{H}$  and  $\hat{h}$  as estimated from  $\hat{S}_{xy}$  while (c), (d) are  $\hat{H}$  and  $\hat{h}$  from (46). The sampling frequency is 10 kHz.

correlation is symmetric around  $n = 0$ , as seen by the noise terms in Fig. 8(a).

#### A. Effect of Denominator Term $\hat{S}_{xx}$

Before presenting quantitative data on the performance of our method (46), we show one final interesting example of how the procedure worked without and with the denominator term  $\hat{S}_{xx}(e^{j\omega})$  of (46). The issue here is whether or not the denominator terms are required, especially for a white input signal, since in that case the expectation of the denominator is known exactly and is just  $\sigma_x^2$  for all frequencies  $k$ . Without the denominator term, (46) is essentially the Fourier transform of a cross correlation. Fig. 9(a)-(d) illustrates this point. Fig. 9(a) is a plot of the log magnitude response of the system of Fig. 6, estimated using only the numerator term  $\hat{S}_{xy}(e^{j\omega})$  of (46). (The value of  $q_0$  was 3, namely,  $q_2 = 3, q_1 = -3$ .) Fig. 9(b) shows the resulting impulse response estimate. A large amount of noise is seen in both the log spectrum and in the impulse response estimate. Fig. 9(c) and (d) shows similar plots when the denominator of (46) was included in the estimate. The dramatic improvements in the responses for this example are striking. (The scale difference is, of course, due to the necessity of dividing the estimate of part  $a$  by  $\sigma_x^2$ .) As in other system identification procedures (e.g., least squares), the input correlation  $\phi$  (or input spectral estimate) provides significant information on the short-time correlation (or spec-

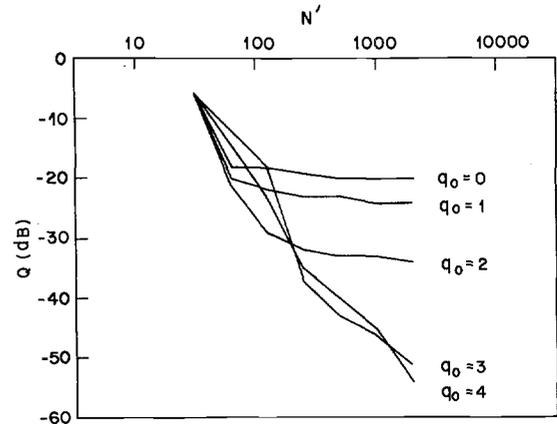


Fig. 10. Curves of  $Q$  versus  $N'$  for several values of  $q_0$  for a white noise input exciting the linear system of Fig. 6 with no output additive noise. Here  $q_1 = -q_0$  and  $q_2 = q_0$ .

trum) of the input signal. Clearly, the denominator term is an important correction to  $\hat{H}(e^{j\omega})$  even though it has been derived from a windowed time signal.

#### B. Performance with White Noise Input

To quantitatively investigate the performance of the spectral estimation algorithm of (46) as a function of  $q_0$  (the number of off-diagonal spectral terms included in the summation), and  $P$  (the number of short-time spectra which are used in the computation), we used the system of Fig. 6 as the linear system, and excited it with zero-mean, white Gaussian noise. For the case where  $\epsilon(n) = 0$  the system impulse response was estimated as the inverse Fourier transform of the frequency response estimate of (46). As a performance measure, we used the  $Q$  measure of (5) which is a good performance measure for system identification methods when the input is white noise [6], [7].

Fig. 10 shows a plot of  $Q$  versus  $\log N'$  (the total number of samples of  $x(n)$  and  $y(n)$  used in the analysis) for a section length of  $L = 32$  samples<sup>3</sup> and for several values of  $q_0$  ( $q_2 = -q_1 = q_0$ ). The value  $q_0 = 0$  corresponds to the "standard" spectral estimation technique, whereas the values  $q_0 = 3$  and 4 are in the range of the generalized spectral estimates according to the analysis of the preceding section. It is readily seen that for  $q_0 = 0, 1, \text{ and } 2$ , the values of  $Q$  flatten off as  $N'$  increases, since in those cases the theoretical value of  $Q$  is limited by the window bias, as shown previously. However, for  $q_0 = 3$  and 4, the value of  $Q$  monotonically decreases to about  $-52$  dB for  $N' = 2048$ . For these cases the limiting effect is the contribution of the first and last analysis sections (end effect error) to the overall spectral estimate, since for these sections of data the analysis leading to (46) is not entirely correct, namely, since the sum on  $m$  has been truncated to the interval  $[0, P - 1]$ . The samples for the first and last  $(L/R - 1)$  sections (3 for a Hamming window) of the truncated interval rely on samples outside the analysis interval in order for the relation

<sup>3</sup>The section length was chosen larger than  $M$  (7) since the estimated value of  $M$  ( $\hat{M}$ ) was chosen as 25 for this example.

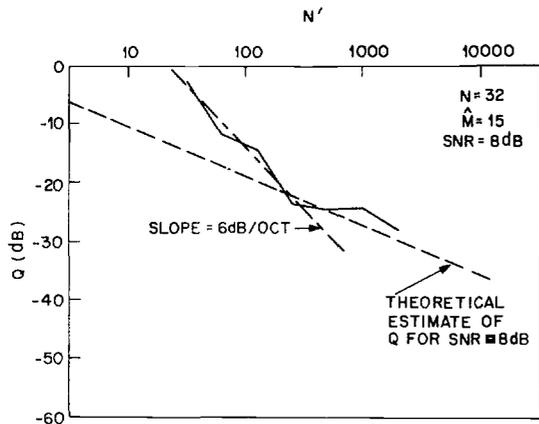


Fig. 11. Curve  $Q$  versus  $N'$  for a signal-to-noise ratio of 8 dB. The dashed line is the theoretical expected curve based on the time-domain least squares estimation procedure. The dash-dot line is a line having 6 dB/octave  $N'$  slope showing the characteristic slope of the end effect error.

$$\sum_m w(mR - n) = D \tag{55}$$

to be valid independent of  $n$ . Clearly, as  $N'$  gets larger, the end effect error becomes relatively smaller at a rate of 6 dB per doubling of  $N'$ , as seen in Fig. 10. This is a result of the decreasing fractional number of points in error, namely,  $2(L - R)/N'$  as  $N'$  increases.

C. Performance in the Presence of Noise

As a final example, Fig. 11 shows a plot of  $Q$  versus  $\log N'$  for the case  $L = 32$ ,  $q_0 = 3$ , with  $\epsilon(n)$  included, such that the SNR ratio was 8 dB. For this case the curve of  $Q$  versus  $\log N'$  is limited by the additive noise (as discussed in [7]) so that as  $N'$  gets large, the values of  $Q$  only decrease by 3 dB as  $N'$  doubles. The dotted line shows the curve of the expected value of  $Q$  for a white noise signal input using the least squares method of system identification [7]. Values of  $Q$  from our spectral estimation method are clearly comparable to the expected values from the least squares method when  $N'$  is sufficiently large (i.e.,  $N' \geq 300$ ).

IV. DISCUSSION

The major significance of the results presented in the preceding section is that they showed that a spectral method may be implemented for system identification with the following properties.

- 1) The section (window) length  $L$  can be comparable to the duration of the unknown impulse response being estimated.
- 2) The effects of the analysis window used in making the short-time spectral estimate may be entirely compensated in the spectral estimate for a white noise input signal.
- 3) The spectral estimate is related (and is superior) to previous methods of spectral estimation.
- 4) The size of the FFT used in the analysis is required to be on the order of the impulse response duration plus the window length. However, as the FFT size increased further, the error monotonically decreases.
- 5) For the case of a white input signal with no additive noise, the estimate of  $h(n)$  improved monotonically as the

number of samples of  $x(n)$  and  $y(n)$  used in the estimate increased. This effect was shown, experimentally, to be due to the "endpoint" problems created by performing a finite size analysis (finite  $N'$ ). It was shown, experimentally, that when the analysis size became sufficiently large, the magnitude of the error in the estimate of  $h(n)$  due to this effect could be made arbitrarily small (at a rate of about 6 dB per octave  $N'$ ). Thus, the estimate is ultimately limited by the poorly understood aliasing error of the window.

6) For the case of a white input signal with additive, uncorrelated white noise, the estimate of  $h(n)$  using the spectral method was comparable to the estimate of  $h(n)$  obtained from the classical least squares analysis method (i.e., the end effect is negligible for large  $N'$  and the crossover point is SNR dependent).

7) The denominator term  $\hat{S}_{xx}(e^{j\omega})$  of our estimate greatly improved the quality of the estimate of  $\hat{H}(e^{j\omega})$ . Therefore, (46) for  $\hat{H}(e^{j\omega})$  is far superior to a simple cross-correlation method (i.e.,  $\hat{H}(e^{j\omega}) = \hat{S}_{xy}(e^{j\omega})$ ) with white noise input (where  $S_{xx}(e^{j\omega}) = 1$  is a priori knowledge).

Thus, based on the above results, it is seen that the proposed spectral estimation procedure for system identification does indeed have some useful capabilities under some conditions. However, our analysis is not yet complete in that we have not yet developed a theory (nor performed experimentation) for the important cases when the following is true.

- 1) The input is a noise signal which is not white, i.e., it has some spectral shaping. Of greatest interest are bandpass signals where both a low frequency and a high frequency band are greatly attenuated. For such cases special attention must be paid to those frequencies where the denominator of (46) gets small, as discussed for example in [7].
- 2) The input is a nonnoise signal (e.g., speech) with a given spectral shape or a nonstationary noise signal. For these cases it is generally quite difficult to analytically describe the spectral shape of the input signal, except in some parametric form (e.g., as a linear prediction spectrum). Thus, problems in handling such cases are somewhat more difficult than those discussed above.

Other issues related to the current algorithm which warrant further consideration include the following.

- 1) An estimate of the expected variation of the spectral estimate  $\hat{H}(k)$  around the mean value, i.e., some insight into the statistical variation of  $\hat{H}(k)$  given  $N'$  points of a random input signal.
- 2) A theoretical model to analytically account for the "end-point" noise due to finite analysis intervals.
- 3) A theoretical model to account for the effects of aliasing due to "imperfect windows" on the resulting spectral estimate.

Thus, although the algorithm presented here is promising, a great deal of work must be performed before it can be considered a candidate for system identification spanning a broad range of applications.

V. SUMMARY

We have presented in this paper a new algorithm for system identification based on short-time unbiased spectral estimation and Fourier transform methodology. We have shown how this

method, while creating many questions, corrects many of the shortcomings of previous spectral estimation procedures. The new algorithm has the potential for application to a wide variety of problems.

#### APPENDIX

We wish to minimize the quantity

$$\begin{aligned} I &= \|e(n)\|^2 \\ &= \sum_{n=-\infty}^{\infty} e^2(n) \\ &= \sum_{n=-\infty}^{\infty} (y(n) - \hat{y}(n))^2 \end{aligned}$$

where

$$y(n) = \sum_{n'=0}^{M-1} h(n') x(n-n') = h * x$$

and

$$\hat{y}(n) = \sum_{n'=0}^{\hat{M}-1} \hat{h}(n') x(n-n') = \hat{h} * x.$$

If we take partials of  $I$  with respect to  $\hat{h}(l)$  for  $0 \leq l \leq \hat{M} - 1$  we find that

$$\sum_{n=-\infty}^{\infty} (y(n) - \hat{y}(n)) x(n-l) = 0 \quad 0 \leq l \leq \hat{M} - 1$$

or in terms of  $x$  and  $y$

$$\begin{aligned} \sum_{n'=0}^{\hat{M}-1} h(n') \sum_{n=-\infty}^{\infty} x(n-n') x(n-l) &= \sum_{n=-\infty}^{\infty} y(n) x(n-l) \\ 0 \leq l \leq \hat{M} - 1. \end{aligned}$$

We now define

$$\begin{aligned} \phi(n'-l) &= \sum_{n=-\infty}^{\infty} x(n-n') x(n-l) \\ &= \sum_{m=-\infty}^{\infty} x(m+l-n') x(m) \end{aligned}$$

and

$$r(l) = \sum_{n=-\infty}^{\infty} y(n) x(n-l)$$

giving

$$\sum_{n'=0}^{\hat{M}-1} \hat{h}(n') \phi(n'-l) = r(l).$$

This may be written as

$$\hat{h} * \phi = r$$

which after Fourier transforming becomes

$$H(e^{j\omega}) \Phi(e^{j\omega}) = R(e^{j\omega}).$$

#### 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, Ed. New York: 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] L. R. Rabiner, R. E. Crochiere, and J. B. Allen, "FIR system modeling and identification in the presence of noise and band-limited inputs," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, Aug. 1978.
- [8] G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications*. San Francisco, CA: Holden-Day, 1968.
- [9] R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra*. New York: Dover, 1958.
- [10] P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, vol. AU-15, pp. 70-73, June 1967.
- [11] J. B. Allen, "Short-term spectral analysis, synthesis, and modification by discrete Fourier transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-25, pp. 235-238, June 1977.
- [12] 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.