Comparisons of System Identification Methods in the Presence of High Noise Levels and Bandlimited Inputs

L. R. Rabiner,
R. E. Crochiere,
J. B. Allen
Bell Laboratories
Murray Hill, New Jersey 07974

ABSTRACT
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 is bandlimited. Both white and nonwhite random noise signals as well as speech signals are used as test signals to measure the performance of the system identification techniques. 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.

1. 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 studying the properties of a wide variety of algorithms for performing system identification. The purpose of this paper is to compare and contrast the performance of three well known system identification techniques for a class of signals which is characteristic of those obtained from speech waveform coders.

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

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

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

with

$$E[e(n)x(n)] = E[e(n)v(n)] = 0 \tag{3}$$

$$E[e(n)e(n-m)] = \sigma_e^2 \delta(m) \tag{4}$$

It is tacitly assumed in Eq. (2) that, the impulse response of the linear system, $h(n)$ is of finite duration (M samples) or can be effectively modelled with a finite 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 to 24 dB.

2. The input signal $x(n)$ is generally bandlimited and has a distinct spectral slope.

3. 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, in a known noise background. The purpose

of the simulation study was to measure the linear system estimation error as a function of N, the signal duration, $\hat{M}$, the impulse response duration, s/n the signal-to-noise ratio of the system, and the input signal characteristics.
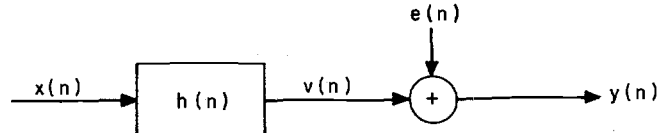

Fig. 1. Block diagram of the linear system model used for system identification.

II. System Identification Methods

The three system identification methods used in this study were the classical least squares analysis algorithm (LMS), the least mean 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 [3,6,7,8]) and the fact that each of these methods had distinct advantages in certain situations.

2.1 Least squares analysis (LSA)

Based on the model of Figure 1 we assume that the output $y(n)$ is related to the input $x(n)$ *exactly* by equations 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

$$y(n) = \sum_{m=0}^{\hat{M}-1} \hat{h}(m)x(n-m) + \hat{e}(n) . \tag{5}$$

The optimization criterion is to minimize the norm of $\hat{e}(n)$ over the set of coefficients of $\hat{h}(n)$. The solution to this minimization is compactly written as [4]

$$\sum_{m=0}^{\hat{M}-1} \hat{h}(m)\tilde{\phi}_{xx}(k,m) = \tilde{\phi}_{xy}(k) \qquad k = 0, 1, ..., \hat{M} - 1 \tag{6}$$

where

$$\tilde{\phi}_{xx}(k,m) = \sum_{n=0}^{N-1} x(n-k)x(n-m) \tag{7}$$

$$\tilde{\phi}_{xy}(k) = \sum_{n=0}^{N-1} y(n)x(n-k) \tag{8}$$

2.2 Least Mean Squares Adaptation Algorithm (LMS)

The least mean squares adaptation algorithm [4] is an iterative, minimum seeking method for determining the least squares solution (6). 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

$$\hat{h}_{i+1} = \hat{h}_i - u\nabla_i \tag{9}$$

where ∇ is the gradient of $\|\hat{e}(n)\|^2$ with respect to $\hat{h}$ and u is a constant. Basically $-\nabla$ 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 $\hat{h}$, 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 if

$$u < 1(\hat{M}\sigma_x^2) \qquad (10)$$

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

The LMS adaptation algorithm used the gradient of a single error

$$\nabla \approx -2x(n-m)\hat{e}(n) \qquad (11)$$

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

$$\hat{h}_{n+1}(m) = \hat{h}_n(m) + 2ux(n-m)\hat{e}(n) \qquad (12)$$

Since the choice of u depends on the variance of $x(n)$, a self-normalizing form of the LMS adaptation algorithm,

$$\hat{h}_{n+1}(m) = \hat{h}_n(m) + Kx(n-m)\hat{e}(n)/(\sigma_x^2\hat{M}) \qquad (13)$$

was used, where

$$\hat{M}\sigma_x^2 = \sum_{m=0}^{\hat{M}-1} x^2(n-m) \qquad (14)$$

and

$$u \cong K/(2M\sigma_x^2) \qquad (15)$$

and the algorithm converges if $K < 2$. 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.

2.3 Short-Time Spectral Analysis (SSA)

Based on the theory of short-time spectral analysis [8-11], the SSA procedure that was used for this study was to form the estimate

$$\hat{H}(z) = \frac{S_{xy}(z)}{S_{xx}(z)} = \frac{\sum_m X_m^*(z) Y_m(z)}{\sum_m X_m^*(z) X_m(z)} \qquad (16)$$

where $X_m(z)$ and $Y_m(z)$ are short-time spectra of $x(n)$ and $y(n)$ at time m, and the summation on m is for overlapping frames of the signal. It is readily shown that as m tends to infinity, the estimate converges to the true $H(z)$.

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 two performances measures for evaluating these methods. The Q measure is basically the ratio (expressed in dB) of the norm of the coefficient error vector or "misadjustment vector" [4,5] to the norm of the true coefficient vector. It is useful for characterizing how well the estimate $\hat{h}$ approximates the true h. The second measure, Q', is a frequency weighted measure which is useful for characterizing the performance of system identification methods for nonwhite inputs. It is also useful when estimates of $v(n)$ and $e(n)$ in the model of Fig. 1 are primarily desired and are obtained by first estimating h.

3.1 The Q and Q' Measure

The Q measure has the form

$$Q = 10 \log_{10} \left\{ \frac{\sum_{m=0}^{\hat{M}-1} [h(m)-\hat{h}(m)]^2}{\sum_{m=0}^{\hat{M}-1} h^2(m)} \right\} \qquad (17)$$

It can 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 the system parameters, namely N, $\hat{M}$, and the signal-to-noise ratio, $s/n = 10 \log(\sigma_v^2/\sigma_e^2)$, at the output of the system, and is of the form:

$$Q|_{\substack{white \\ input}} \cong 10 \log_{10} \left[\frac{\hat{M}}{N}\right] - s/n \quad (dB) \qquad (18)$$

Equation (18) predicts the performance of the least squares analysis system identification method for white uncorrelated inputs.

It is seen that Q 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, and it degrades (increases) with $\log \hat{M}$.

For the case of nonwhite inputs, it is not possible to express Q in a form as simple as (18). In general for nonwhite inputs the values of Q will be larger than (18) and in this sense (18) 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.

The modified Q measure, Q', that we propose applies frequency weighting which is equal to that of the frequency response of the filter $g(n)$ in Fig. 2 which is used to create the nonwhite sig-
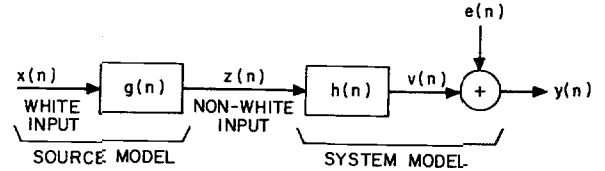

Fig. 2. Block diagram of the linear system model for non-white input signals.

nal $z(n)$ from the white signal $x(n)$. This weighting can conveniently be achieved by convolving $h(n)$ and $\hat{h}(n)$ by $g(n)$. This procedure serves to weight the performance measure by the frequency spectrum of the input signal. More formally, we define the measure Q', as

$$Q' = 10 \log_{10} \left[\frac{\sum_n [(h(n)-\hat{h}(n))*g(n)]^2}{\sum_n [h(n)*g(n))]^2} \right] \qquad (19)$$

Using Parsevals theorem Eq. (19) can be transformed to the frequency domain, giving

$$Q' = 10 \log_{10} \left[\frac{\int_{-\pi}^{\pi} |H(e^{j\omega}) - \hat{H}(e^{j\omega})|^2 |G(e^{j\omega})|^2 \, d\omega}{\int_{-\pi}^{\pi} |H(e^{j\omega})|^2 |G(e^{j\omega})|^2 \, d\omega} \right] \qquad (20)$$

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 unwhitening" filter, $g(n)$, affects the result. However it can be shown that, in most cases, the properties of Q' and Q are quite similar.

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 white Gaussian noise, bandlimited Gaussian noise, and speech. 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. 3. The other filter used in the simulations was a 25-point, linear phase, equiripple FIR lowpass filter. 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.
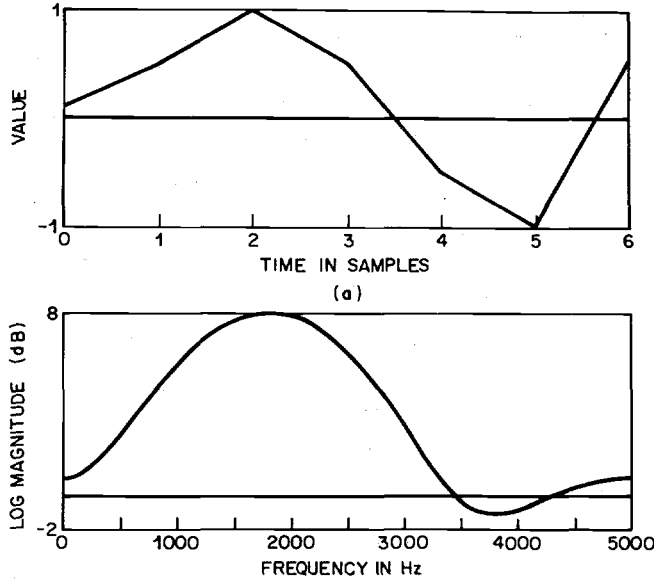

Fig. 3. Impulse response and log magnitude response of a simple filter used in the investigations.

4.1 White noise input

Figure 4 shows a series of curves of Q (on a log scale) versus N (on a log scale) for the filter of Fig. 3, 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. 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 lowpass filter (with $\hat{M}{=}25$ and 34) and the results were equivalent to those of Fig. 4 - i.e., close agreement between theory and measurement.

Figures 5 and 6 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 $L/4$ samples between adjacent sections [11], i.e., except for endpoint effects, each input sample was used in 4 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. Figure 5 shows a set of curves of Q versus N for the simple filter of Fig. 3, 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
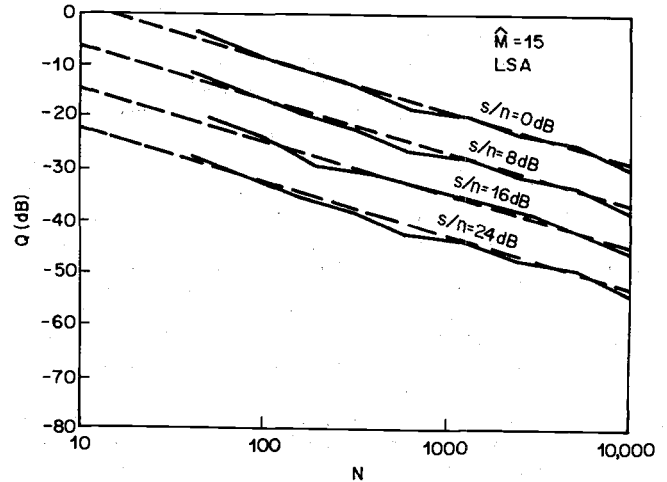

Fig. 4. Curves of Q versus N for $\hat{M} = 15$ and several values of s/n for the LSA method for a white input signal.
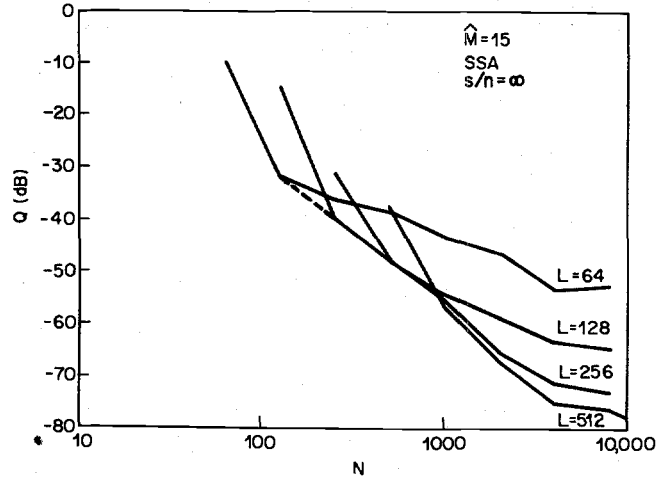

Fig. 5. Curves of Q versus N for $M = 15$, $s/n = \infty$, and several values of window size L for the SSA method for a white input signal.
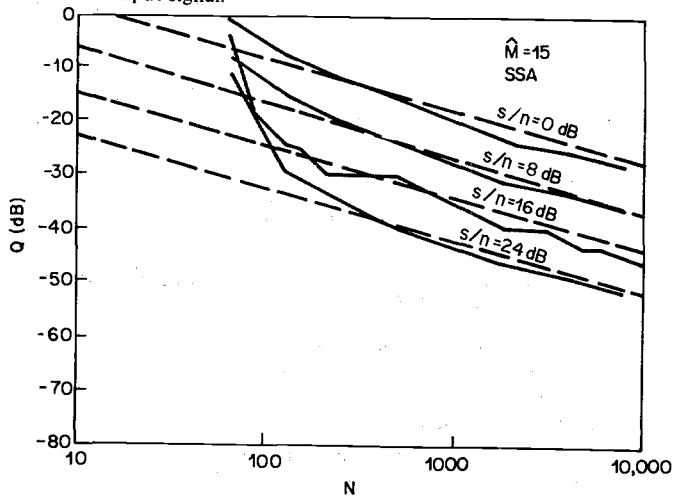

Fig. 6. Curves of the lower bound of Q versus N for $\hat{M} = 15$ and several values of s/n for the SSA method for a white input signal.

provide an increased number of short-time spectral estimates for averaging. This effect is related to the aliasing noise of the analysis 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.

Figure 6 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 don't 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. 7 and 8. Figure 7 shows a set of curves of Q versus N, for $\hat{M} = 15$, and $s/n = \infty$ for the filter of Fig. 3. 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 then we realize that if K is large, small errors in calculating the gradient of e can lead to large errors in estimating h as will be seen later in Fig. 8. Thus with noisy signals a tradeoff in choosing K is required. The curves in Fig. 7 give an indication of the value of N required to obtain a desired value of Q for the noise-free case.

Figure 8 shows curves of Q versus N for $s/n = 8$ dB, and the same parameters as Fig. 7. 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.
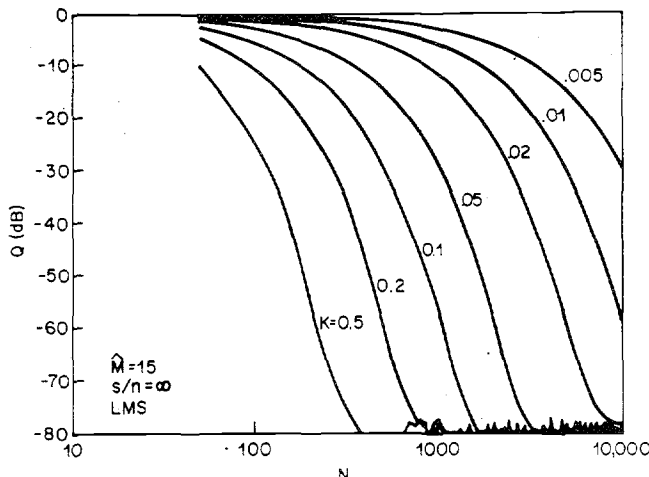

Fig. 7. Curves of Q versus N for $\hat{M} = 15$, $s/n = \infty$, and several values of K for the LMS adaptation algorithm for a white input signal.

4.2 Bandpass Input Signals

To evaluate the performance of each of the three system identification methods on bandlimited input signals, the system of Fig. 9 was simulated. The signal $z(n)$ was a lowpass signal whose frequency components were attenuated by at least 54 dB for frequencies above 0.2 F_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 $\hat{h}(n)$ relat-
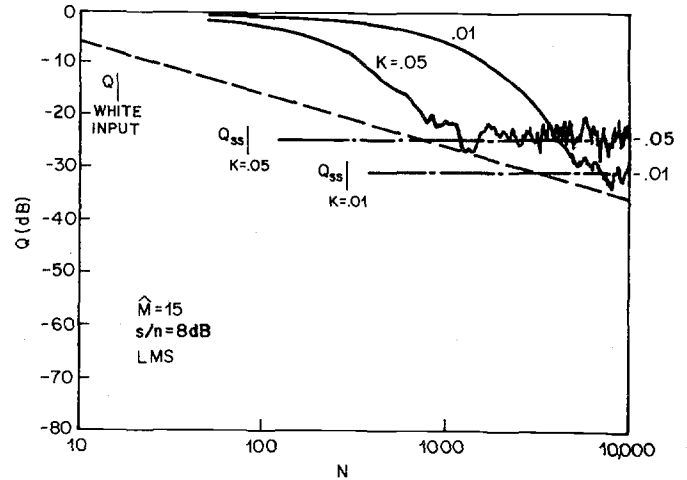

Fig. 8. Curves of Q versus N for $\hat{M} = 15$, $s/n = 8$ db for $K = 0.01$ and 0.05 for the LMS adaptation algorithm for a white input signal.
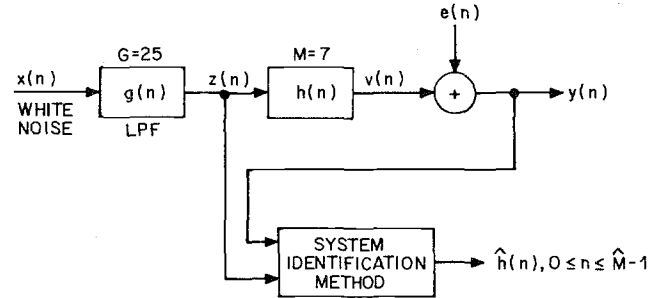

Fig. 9. Block diagram of actual system used to test the system identification algorithms for a lowpass input signal.

ing $y(n)$ to $z(n)$ was estimated.

Figure 10 shows a set of curves of Q' (the modified Q measure) versus N for $\hat{M} = 15$, and several values of s/n. 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. 10 due to the lack of high frequency information in the input signal. However when the Q' measure was used, the high frequency inaccuracies in $h(n)$ were given essentially zero weight by the "coloration" filter $g(n)$. Thus the curves of Q' versus N of Fig. 10 for the highly bandlimited input are essentially the same as the curves of Q versus N of Fig. 4 for the white input case. At the bottom of Fig. 10 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 $\hat{h}(n)$.

Results for both the SSA and LMS methods for the bandlimited input case using the Q' measure were essentially identical to those of the white input case with the Q measure, and are thus not shown here.

4.3 Speech Input Signal

The last test signal used to evaluate the three system identification techniques was an actual speech signal. The model for testing the systems using the speech signal was essentially that of Fig. 9 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 LSA analysis are given in Figure 11 for $\hat{M} = 15$. It is seen that the measured values of Q' are generally greater than the predictions 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. Results for the SSA and LMS methods were somewhat poorer for speech than those of the LSA method. However, they were still good enough to indicate that these methods could be used on such signals.
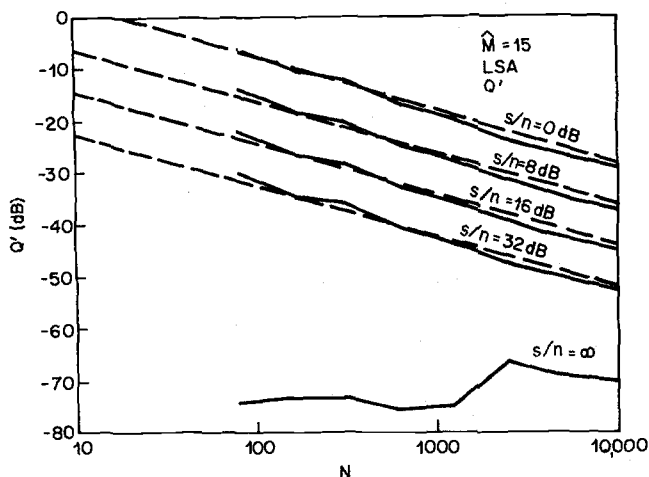

Fig. 10. Curves of Q' versus N for $\hat{M} = 15$ and several values of s/n for the LSA method with a bandlimited noise input signal.
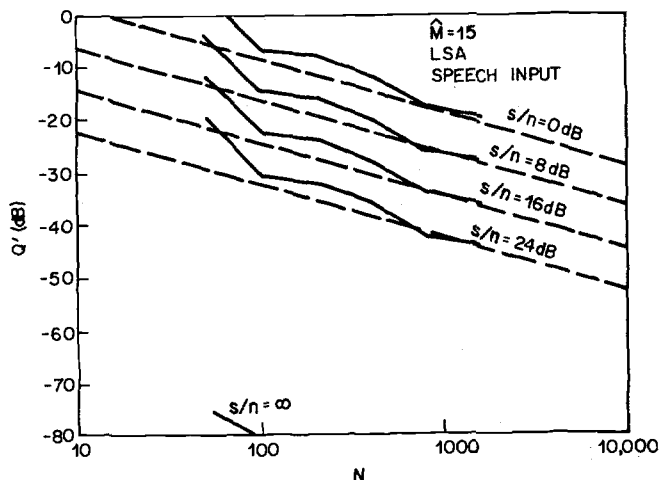

Fig. 11. Curves of Q' versus N for $\hat{M} = 15$, and several values of s/n for the LSA method for a speech input signal.

V. Discussion of the Results

Based on the theoretical discussions of Sections II and III, the results presented in Section IV, and information 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 bandlimited 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 equa-

tion (via an efficient recursion procedure). As such estimation of systems with large values of $\hat{M}$ (e.g. the speech echo canceller etc.) would generally not be practical using LSA. The recognition of this fact has led to the widespread use of the LMS adaptation method.

3. The SSA method was shown to perform almost as well as the LSA method for both white and non-white noise signals. The major disadvantage of the SSA method is the possibility of having severe aliasing distortion in the estimation of $\hat{h}(n)$ for nonwhite inputs due to dividing $S_{xy}(z)$ by $S_{xx}(z)$.

4. The big advantage of the SSA method is that the implementation is simple, can be used for large $\hat{M}$ values, and is readily amenable to either digital hardware or to array processors. 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 bandlimited inputs. Although the performance was not as good as the LSA method, the differences were not so large so as to make the method undesirable for virtually any application.

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

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 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.

References

[1] P. Eykhoff, *System Identification*, Wiley, N.Y., 1974.

[2] M. J. Levin, "Optimum Estimation of Impulse Response in the Presence of Noise," IRE Trans. on Circuit Theory, pp. 50-56, March 1960.

[3] B. Widrow, et al., "Adaptive Noise Cancelling: Principles and Applications,: Proc. IEEE, Vol. 63, No. 12, pp. 1692-1716, December 1975.

[4] B. Widrow, J. McCool, M. Larimore, and C. Johnson, "Stationary and Nonstationary Learning Characteristics of the LMS Adaptive Filter," Proc. IEEE, Vol. 64, No. 8, pp. 1151-1162, August 1976.

[5] M. M. Sondhi and D. Mitra, "New Results on the Performance of a Well Known Class of Adaptive Filters," Proc. IEEE, Vol. 64, No. 11, pp. 1583-1597, November 1976.

[6] M. M. Sondhi, "An Adaptive Echo Cancellor," Bell Syst. Tech. J., Vol. 46, No. 3, pp. 497-511, 1967.

[7] J. Makhoul, "Linear Prediction: A Tutorial Review," Proc. IEEE, Vol. 63, pp. 561-580, May 1974.

[8] J. B. Allen, "Short-Term Spectral Analysis and Synthesis and Modification by Discrete Fourier Transform," IEEE Trans. on Acoustics, Speech, and Signal Proc., Vol. ASSP-25, No. 3, pp. 235-238, June 1977.

[9] G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications*, Holden-Day Inc., San Francisco, 1968.

[10] R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra*, Dover Press, 1958.

[11] J. B. Allen and L. R. Rabiner, "A Unified Theory of Short-Time Spectrum Analysis and Synthesis," Proc. IEEE, November 1977.