ESTIMATION OF NOISE VARIANCE FROM THE NOISY AR SIGNAL AND ITS APPLICATION IN SPEECH ENHANCEMENT

K. K. PALIWAL

Computer Systems and Communications Group
Tata Institute of Fundamental Research
Homi Bhabha Road, Bombay 400005

Abstract. In a number of applications involving the processing of noisy signals, it is desirable to know a priori the noise variance. We propose here a method of estimating the noise variance from the autoregressive (AR) signal corrupted by the additive white noise. This method first estimates the AR parameters from the high-order Yule-Walker equations and then uses these AR parameters to estimate the noise variance from the low-order Yule-Walker equations. The method is studied for a number of examples of noisy AR signals and its performance is found to be close to the Cramer-Rao lower bound for high signal-to-noise ratios. It is also used in a speech enhancement application where its performance is studied for stationary as well as nonstationary noise conditions. The results are found to be encouraging.

1. Introduction

In a number of signal processing applications, the signal available for analysis is noisy; i.e., it is corrupted by the addition of background noise which can be assumed to be white in nature for many practical problems of interest. In these applications, it is desirable to know the variance of the additive white noise process. For example, for speech enhancement there is a class of methods (based on short-time Fourier transform magnitude) for which it is necessary to know a priori the white noise variance [1]. Similarly, some of the spectral estimation methods for noisy signals [2] require a priori knowledge of the noise variance.

In the speech enhancement literature [1,3], the white noise variance is estimated from the preceding silent portions of speech. However, this method which will be referred hereafter as the conventional method has two major drawbacks. Firstly, detection of silent segments of speech is a difficult problem [4]. When speech is noisy with signal-to-noise ratio (SNR) as low as 0 dB or lower, it becomes almost impossible to detect these silent segments. Secondly, the noise variance estimated from the preceding silent segments can be used in the processing of the present speech segment only when the noise is stationary in nature. However, in practice, it might be varying with time and the noise-variance estimate from the preceding silent segments may not give good results for the present segment. In the present paper, we propose a method for estimating the noise variance. This method estimates the noise variance from the same segment of the noisy signal which is available for processing.

2. Noise-variance estimation method

In this section, we describe the method of estimating noise variance from the given segment of the noisy signal. Let the noisy signal to be analysed be

$$y(n) = x(n) + v(n), \qquad n=1,2,----,N,$$

where $\{x(n)\}$ is the uncontaminated signal and the $\{v(n)\}$ the zero-mean white noise of variance σ_v^2. The aim here is to estimate the noise variance σ_v^2 from the observed noisy signal $\{y(n)\}$.

In order to solve this problem, we assume that the uncontaminated signal $\{x(n)\}$ follows the pth order AR model

$$H(z) = 1/[1+\sum_{i=1}^{p} a_i z^{-i}],$$

whose parameters $\{a_i\}$ satisfy the following set of Yule-Walker equations:

$$\sum_{k=1}^{p} a_k R_x(|i-k|)=-R_x(i), \qquad i>0,$$

where $\{R_x(i)\}$ are the autocorrelation coefficients of the uncontaminated signal $\{x(n)\}$.

Since the additive noise $\{v(n)\}$ is white, the autocorrelation coefficients $\{R_x(i)\}$ of the uncontaminated signal $\{x(n)\}$ are related to the autocorrelation coefficients $\{R_y(i)\}$ of the noisy signal $\{y(n)\}$ as follows:

$$R_x(0)=R_y(0)-\sigma_v^2,$$

and

$$R_x(i)=R_y(i), \qquad i>0.$$

With these preliminaries, we now

CH2396-0/87/0000-0297 $1.00

present a three-step procedure for estimating the noise variance σ_v^2. These steps are outlined below.

Step 1: From the observed (noisy) signal $\{y(n)\}$, unbiased estimates of the autocorrelation coefficients $\{\hat{R}_y(i)\}$ are obtained.

Step 2: In this step, the AR parameters are computed from an overdetermined set of q (>p) high-order Yule-Walker equations using a least-squares procedure [6].

Step 3: In this step, we propose to estimate the noise variance by solving an overdetermined set of p low-order Yule-Walker equations using a least-squares procedure. This leads to the following estimate of the noise variance:

$$\hat{\sigma}_v^2 = [\sum_{i=1}^{p} a_i \{\hat{R}_y(i) + \sum_{k=1}^{p} a_k \hat{R}_y(|i-k|)\}] / \sum_{i=1}^{p} a_i^2.$$

3. Simulation results for the AR signals

In this section, we study the present noise-variance estimation method for the noisy AR signals and compare its performance with the Cramer-Rao (CR) lower bound. The CR bound is computed for the root-mean-square (RMS) error in estimating the noise variance, using the formula given by Pagano [7]. Here, we illustrate the results of this method in terms of two examples, the first example corresponds to the narrowband AR process and the other to the broadband AR process.

Example 1: Narrowband AR process

In this example, we consider a narrowband AR process. Here, the signal is generated by exciting the second order AR model (with parameters a_1=-1.4 and a_2=0.95) by a zero-mean white Gaussian noise signal of variance 0.0472. The noise variance estimation method is tested at different SNR conditions with p=2 and q=158. In each case, 100 independent signal and noise realizations of length N=1024 points are generated. The RMS error in estimating the noise variance is computed for each SNR condition. Results are shown in Fig. 1 alongwith the CR lower bounds. We can see from this figure that for SNR greater than 0 dB the RMS error in estimating the noise variance is close to the CR lower bound. However, for SNR less than 0 dB the RMS error departs considerably from the CR lower bound. In the literature [8], this is a common phenomenon occuring in estimation problems. This phenomenon has been called as the 'threshold effect' and the value of SNR below which the estimation error increases dramatically has been termed as the 'threshold SNR'. Thus, the value of threshold SNR for the present noise variance estimation method is 0 dB for this example.

Example 2: Broadband AR process

In this example, we consider a broadband AR process. Here, the signal is generated by exciting the second order AR system (with parameters a_1=-0.45 and a_2=0.55) by a zero- mean white Gaussian noise sequence of variance 0.6387. The test conditions used in this example are same as those in Example 1, except that q = 78 in this example. The RMS error in estimating the noise variance is shown in Fig. 2 along with the CR lower bound as a function of the SNR of the noisy AR signal. Here also we can observe the threshold effect and the value of threshold SNR is 0 dB for this example. For SNR values greater than 0 dB, the performance of the present noise-variance estimation method is close to the CR lower bound.

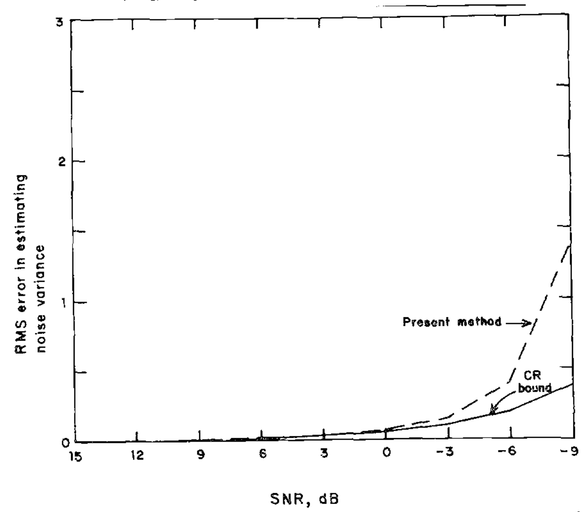

Fig. 1. RMS error in estimating the noise variance as a function of input SNR. Narrowband AR process.

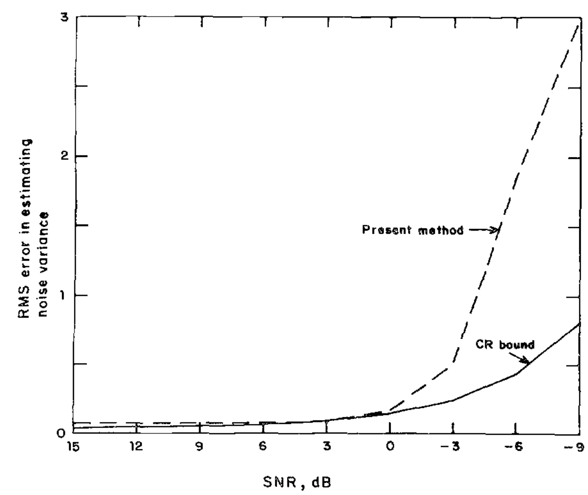

Fig. 2. RMS error in estimating the noise variance as a function of input SNR. Broadband AR process.

4. Speech enhancement application

In this section, we apply the noise variance estimation method proposed in the present paper for speech enhancement and study its performance for stationary as well as nonstationary noise conditions. The spectrum subtraction method [5] is used in the present study for speech enhancement. Implementation of this method as done in the present paper is briefly described in sub-section A. Speech enhancement results are described in sub-section B for stationary noise conditions and in sub-section C for nonstationary noise conditions.

A. Spectrum subtraction method

Here, we present a brief description of the spectrum subtraction method of speech enhancement. More details can be found in [5].

The noisy speech signal sampled at 8 kHz is analysed here segment by segment. Duration of each segment is taken here to be 256 samples (i.e., 32 ms). Each segment overlaps with the preceding segment by 128 samples.

The noise variance σ_v^2 is estimated from the observed (noisy) speech signal $\{y(n), n=1,2,---,256\}$ by using the noise-variance estimation method described in the preceding section. The order p of the AR model is taken here to be 10 and q=15 high-order Yule-Walker equations are employed in the present implementation.

For speech enhancement, the noisy speech signal $\{y(n)\}$ is weighted by a triangular window and an N=256 point discrete Fourier transform $Y_w(f)$ of the windowed signal $\{y_w(n)\}$ is computed. Spectral subtraction is performed on the Fourier transform magnitude as follows:

$$|\hat{X}_w(f)|^a = |Y_w(f)|^a - |V(f)|^a,$$

where a is the appropriately chosen positive constant. $|V(f)|^2$ is the power spectrum of the additive noise signal $\{v(n)\}$. Since the noise is white, its power specturm is given by

$$|V(f)|^2 = CN\sigma_v^2,$$

where C is the constant whose value depends on the type of window function used and on the value of N. For a triangular window with N=256, the value of constant C is 0.335. If the Fourier transform magnitude $|\hat{X}_w(f)|$ of the enhanced signal is negative, it is set to zero. The Fourier transform magnitude $|\hat{X}_w(f)|$ is combined with the phase of the noisy signal and inverse-Fourier transformed to get the enhanced-windowed signal $\{\hat{x}_w(n)\}$. The output (enhanced) signal can be obtained from the windowed signal by adding it to the windowed signal of the preceding segment with an overlap of 128 samples. In the present paper, we use a=1 since this value has shown best performance in earlier studies reported by Lim [5].

B. Results for stationary white noise

Here we study the speech enhancement performance of the present noise-variance estimation method for the speech signal corrupted by the addition of stationary white Gaussian noise. We compare the speech enhancement results of this method with those obtained using the ideal noise variance estimates. The ideal noise variance estimate is taken to be the actual variance value of the white noise used for the generation of noisy speech.

Noisy speech $\{y(n)\}$ is generated here by adding to the uncontaminated speech signal $\{x(n)\}$ the white Gaussian noise of variance σ_v^2. The noise variance σ_v^2 is adjusted to get the desired input SNR. In the present study, continuous speech data of 4 sec duration is used. Speech enhancement performance of the noise-variance estimation method is studied for input SNR values ranging from -10 dB to 15 dB. Results in terms of output SNR of the enhanced speech signal are shown in Fig. 3. Results obtained with ideal noise variance estimates are also shown in this figure. It can be seen from Fig. 3 that the present method of noise variance estimation results in an improvement in SNR for all the input SNR values. For example, for noisy speech

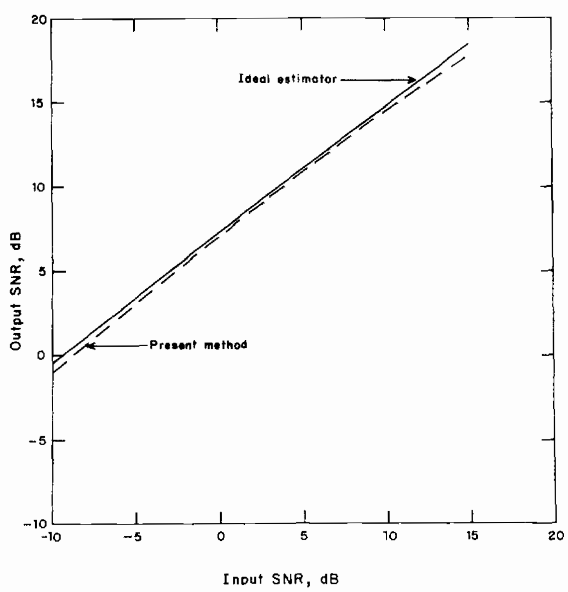

Fig. 3. SNR of the output (enhanced) speech as a function of input SNR.

with input SNR=0 dB, this method results in an SNR improvement of 7.4 dB. It can also be seen from Fig. 3 that the results from the present method are inferior to those obtained with the ideal noise variance estimates by only 0.6 dB. Subjective listening of the output speech has shown that the present method improves the quality of speech for all input SNR values and there is no audible difference between the output speech obtained through the present noise-variance estimation method and that obtained with the ideal noise variance estimates.

C. Results for nonstationary white noise

In this sub-section, we study the speech enhancement performance of the present noise variance estimation method for the speech signal corrupted by the addition of nonstationary white Gaussian noise. We compare its performance with that of the conventional method [3] which computes the noise variance from the silent intervals preceding the speech segments. The speech data used in the present study consists of a few silent segments in the beginning. The conventional method uses here these silent segments to compute the noise variance.

In order to generate noisy speech with nonstationary white Gaussian noise, the noise variance σ_v^2 is first calculated to get the desired input SNR. The speech data is then corrupted by the addition of zero-mean white Gaussian noise whose variance varies linearly with time with the variance being $(1-g)\sigma_v^2$ at the begining and $(1+g)\sigma_v^2$ at the end of the speech data. The parameter g defines here the measure of nonstationarity of the white noise process. For g=0, this process becomes stationary. By increasing g, the nonstationarity of the white noise process can be increased. In this paper, we will be referring this parameter as the 'nonstationarity index'.

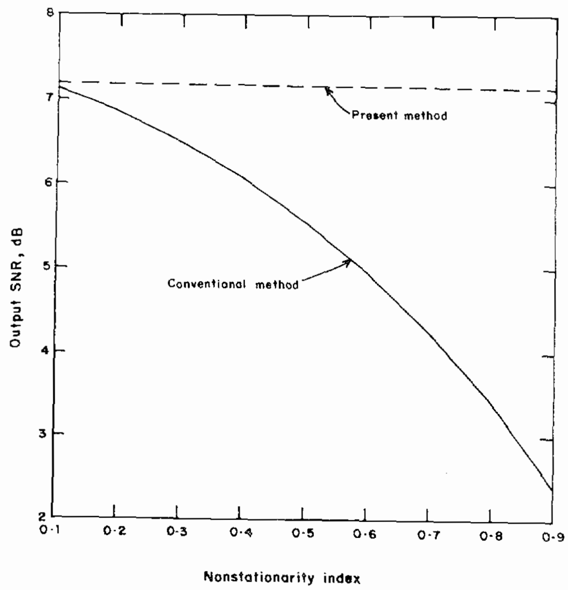

Fig. 4. SNR of the output (enhanced) speech as a function of nonstationarity index.

Here, we study the noise-variance estimation method for input SNR=0 dB and evaluate its speech enhancement performance as a function of nonstationarity index. The SNR results for this method are shown in Fig. 4. The results for the conventional method are also shown in this figure. It can be seen from this figure that as the nonstationarity of the additive white noise process increases, performance of the conventional method deteriorates; while that of the present method remains steady throughout. For the nonstationarity index g equal to 0.9, the present method of noise-variance estimation gives an advantage of about 5 dB in output SNR. Subjective listening of the output speech also confirms these observations.

5. Conclusions

In this paper, a noise-variance estimation method is proposed. This method is studied for the noisy AR signals and its performance is found to be close to the CR lower bound for the SNR values greater than 0 dB. This method is applied to the problem of speech enhancement where the spectrum subtraction scheme [5] is used. Speech enhancement performance of this method is studied for speech corrupted by stationary as well as nonstationary additive white Gaussian noise processes and encouraging results are obtained. This method has also been applied to the problem of spectral estimation for noisy AR signals and the results are reported elsewhere [9].

References

[1] J.S. Lim and A.V. Oppenheim, Proc. IEEE, Vol. 67, No. 12, pp. 1586-1604, Dec. 1979.

[2] S.M. Kay, IEEE Trans. ASSP-28, No. 3, pp. 292-303, June 1980.

[3] R.J. McAulay and M.L. Malpass, IEEE Trans. ASSP-28, No. 2, pp. 137-145, Apr. 1980.

[4] L.R. Rabiner and M.R. Sambur, Bell Syst. Tech. J., Vol. 54, pp. 297-315, Feb. 1975.

[5] J.S. Lim, IEEE ASSP-26, No. 5, pp. 471-472, Oct. 1978.

[6] J.A. Cadzow, Proc. IEEE, Vol. 70, No. 9, pp. 907-939, Sept. 1982.

[7] M. Pagano, Annals of Statistics, Vol. 2, No. 1, pp. 99-108, 1974.

[8] S.W. Lang and J.H. McClellan, IEEE Trans. ASSP-28, No. 6, pp. 716-724, Dec. 1980.

[9] K.K. Paliwal, Proc. ICASSP, pp. 1369-1372, Apr. 1986.