A SPEECH ENHANCEMENT METHOD BASED ON KALMAN FILTERING

K.K. Paliwal and Anjan Basu

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

ABSTRACT

In this paper, the problem of speech enhancement when only corrupted speech signal is available for processing is considered. For this, the Kalman filtering method is studied and compared with the Wiener filtering method. Its performance is found to be significantly better than the Wiener filtering method. A delayed-Kalman filtering method is also proposed which improves the speech enhancement performance of Kalman filter further.

I. INTRODUCTION

In many situations of practical interest, the speech signal gets corrupted by the addition of white noise. Presence of noise affects the intelligibility of speech. An example is the communication between a pilot and an air traffic control tower, where speech is usually degraded by the addiion of engine noise. In such situations, it is desirable to enhance the quality and intelligibility of speech. In automatic speech and speaker recognition systems if a speech enhancement scheme is incorporated in a preprocessing stage, recognition becomes simpler and more reliable. Speech enhancement also plays an important role in speech coding applications.

The problem addressed in the present paper is to enhance speech when only the corrupted speech signal is available for processing. A large number of methods have been reported in the literature [1] for speech enhancement. The stationary Wiener filtering method is one of the important speech enhancement methods.

Since speech is nonstationary in nature, stationary Wiener filter does not perform very well. Therefore, methods based on short-time power-spectrum have been proposed. Recently, Paliwal [2] has proposed a nonstationary Wiener filtering method for speech enhancement, where the Wiener filter is designed for each short-time speech segment (duration= 20-30 msec) using a least-squares procedure.

Though the nonstationary Wiener filter is optimum for a given segment in a least-squares-error sense, it does not exploit the knowledge about speech production process. In the present paper, we propose Kalman filtering method which allows for the nonstationarity of speech and, at the same time, exploits speech production model. We also show that a delayed version of the same filter offers further improvement, though the computational complexity remains identical.

II. KALMAN FILTER FOR SPEECH ENHANCEMENT

A. Mathematical Formulation

Speech can be represented by an autoregressive (AR) process which is essentialy the output of an all-pole linear system driven by white noise sequence. Thus speech signal at k-th time instant, s(k), is given by:

$$s(k)=a_1 s(k-1)+\ldots.+a_p s(k-p)+u(k) \qquad \ldots.(1)$$

A little observation of the equation (1) reveals that it can be represented by the state-space model as shown below.

$$\begin{bmatrix} s(k-p+1) \\ s(k-p+2) \\ \vdots \\ s(k) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & & & & \vdots \\ -a_p & \ldots. & & & -a_1 \end{bmatrix} \begin{bmatrix} s(k-p) \\ s(k-p+1) \\ \vdots \\ s(k-1) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} u(k) \qquad \ldots.(2)$$

or $$X(k)= \Phi\, X(k-1)+ G\, u(k) \qquad (3)$$

where X(k), Φ and G are state vector, state transion matrix and input matrix, respectively. These are defined as follows:

CH2396-0/87/0000-0177 $1.00

$$X^T(k)=[s(k-p+1),\ldots,s(k-1),s(k)] \quad \ldots.(4)$$

$$\Phi = \begin{bmatrix} 0 & I \\ -a_p \ldots -a_1 \end{bmatrix} \quad (5)$$

$$G = [0 \ \ldots 0 \ 1] \quad (6)$$

When only the noise corrupted signal y(k) is available, the observation process can be written in the following form:

$$y(k)=s(k)+n(k) \quad (7)$$

This equation can be written in the matrix form as follows:

$$y(k)=H\,X(k)+n(k) \quad (8)$$

where X(k) is the state vector already defined by equation (4) and H is the observation matrix given by:

$$H = [0 \ 0 \ \ldots. \ 0 \ 1] \quad (9)$$

The noise sequences {u(k)} and {n(k)} are zero mean white noise processes and are uncorrelated. The observation noise n(k) is also uncorrelated to the state vector. For all k and l,we can write:

$$E\{u(k)\}=0 \qquad E\{u(k)u(l)\}=Q(k)\delta_k$$

$$E\{n(k)\}=0 \qquad E\{n(k)n(l)\}=R(k)\delta_k$$

$$E\{u(k)n(l)\}=0 \qquad E\{X(k)n(l)\}=0 \quad \ldots.(10)\text{-}(12)$$

It is also assumed that initial estimate of X is $\hat{X}(0)=X_0$ and is unbiased,i.e.

$$E\{X(0)-X_0\}=0 \quad (13)$$

The initial estimate of the error covariance matrix P_0, is known from the following relation

$$P_0=E\{[X(0)-X_0][X(0)-X_0]^T\} \quad (14)$$

The state and observation equations (3) and (8) clearly suggest that Kalman filter can readily be applied for an estimate of the state-vector X(k). It can be easily shown that the pairs $\{\Phi, G\}$ and $\{\Phi, H\}$ are controllable and observable, respectively. Hence, the Kalman filter based on this model will be 'stable' or 'robust' in the sense that the effects of initial errors and round-off and other computational errors will die out asymptotically.

The Kalman filter gives the minimum mean-square-error estimate of X(k) based on the observations {y(1),y(2),....,y(k)}, and this estimate is represented by $\hat{X}(k|k)$. The corresponding error covariance matrix is P(k|k). Similarly, the one step predicted estimate of X(k) is $\hat{X}(k|k-1)$ and associated error covariance matrix is P(k|k-1). Using these notations the Kalman filtering algorithm can be given by the following recursive relations:

$$\hat{X}(k|k)=\hat{X}(k|k-1)+K(k)[y(k)-H\,\hat{X}(k|k-1)]$$

$$\hat{X}(k|k-1)=\Phi\,\hat{X}(k-1|k-1), \text{ with } \hat{X}(0|0)=X_0$$

$$P(k|k)=[I-K(k)H]\,P(k|k-1) \quad \ldots.(15)\text{-}(17)$$

where

$$K(k)=P(k|k-1)H^T[H\,P(k|k-1)H^T+R(k)]^{-1}$$

$$P(k|k-1)=\Phi\,P(k-1|k-1)\Phi^T+G\,Q(k)G^T \quad \ldots.(18),(19)$$

Application of Kalman filtering for speech enhancement consists of two separate steps:

(1) Estimation of AR coefficients $\{a_1,a_2,\ldots,a_p\}$ and noise variances Q and R for each segment over which speech is assumed to be stationary. Different methods have been proposed in the literature for estimating these parameters [3-6].

(2) Apply the Kalman filtering algorithm using estimated parameter values. The last component of the state vector X (k)=[s(k-p+1) ....s(k)], i.e., $\hat{x}_p(k)=\hat{s}(k)$ gives the Kalman filtered estimate of speech signal s(k).

B. Delayed Kalman Filter

Further observation shows that the first component of the state vector, s(k-p+1) will give a better estimate of speech signal at (k-p+1)-th instant, since this estimate is using additional information in the form of (p-1) extra observation data {y(k-p+2),....,y(k)}. This phenomenon is reflected in the fact that the diagonal elements, $\{p_i(k|k), i=1,..,p\}$, of the error covariance matrix P(k|k) get arranged in their ascending order, as the filter reaches its steady state. Actually $\hat{x}_1(k)$ is the fixed-lag-smoothed estimate of s(k-p+1), where lag=p-1. This method delays the computation of $\hat{s}(k)$ untill (k+p-1). Hence, we have called this estimate the delayed Kalman filter estimate.

III. EXPERIMENTAL RESULTS

We have used 4 sec of continuous speech and evaluated the performance of the

Kalman filtering method at different signal-to-noise ratio (SNR) conditions. Performance is measured here in terms of output SNR and output segmental SNR (SEGSNR). In order to put the present method in proper perspective, we have compared its performance with that of the stationary and nonstationary Wiener filtering methods.

In the present study, we have used ideal values of parameters, $\{a_i, i=1,..,p, Q, R\}$. Their estimated values will be used in future and the sensitivity of the filter to different parameter estimation schemes will be reported later.

The experimental procedure is as follows. Kalman filter is initialized only for the first segment. In the subsequent segments, the state vector and error covariance matrix are initialized using the last values from the previous segment. For the first segment the filter state vector is initialised with the first p data points:

$$\hat{X}(0|0) = X_0 = [y(1), y(2), \ldots, y(p)]$$

and the error covariance matrix is accordingly set to

$$P(0|0) = P_0 = \mathrm{diag}[R, R, \ldots, R]$$

where R is the estimated observation noise variance for the first segment of speech. This filter starts from the (p+1)-th time instant and runs for the full data length. At the beginning of each segment Q, and R are replaced by their new estimated values.

The SNR and SEGSNR results are shown in Figs. 1 and 2, respectively. In terms of SEGSNR, Kalman filter offers an advantage of 4.5 dB over the nonstationary Wiener filtering method, and 7.4 dB over the stationary Wiener filtering method for input speech with SNR=0 dB. Results for delayed Kalman filter show an additional 1 dB improvement over the Kalman filtering method for 0 dB SNR case. For further illustration of our results, we show in Fig. 3 the 0 dB noisy speech processed by stationary Wiener filtering method and the delayed-Kalman filtering method. This figure clearly shows the superiority of the delayed-Kalman filtering method over the Wiener filtering method. Subjective listining tests have also confirmed these findings.

It might be noted here that the Wiener filtering methods provide improvements in terms of SNR only. These methods do not improve SEGSNR of the output speech. This is the reason that these methods have been found to show improvements in terms of speech quality but not in terms of speech intelligibility [8]. But Kalman filtering method offers improvement in terms of both SNR and SEGSNR. Therefore it is expected to improve speech quality as well as its intelligibility. However, we have not yet made formal intelligibility tests to confirm this conjecture.

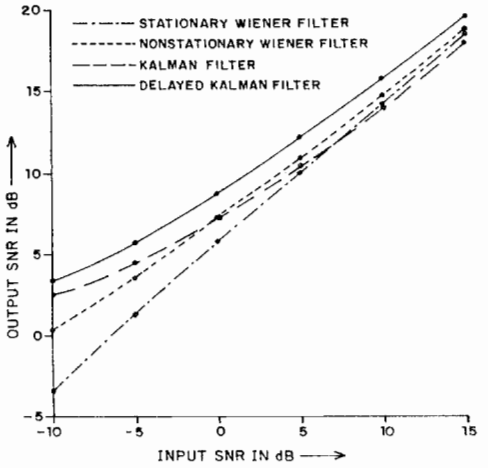

Fig. 1: Output SNR values (in dB) for different speech enhancement methods as a function of input SNR values (in dB).

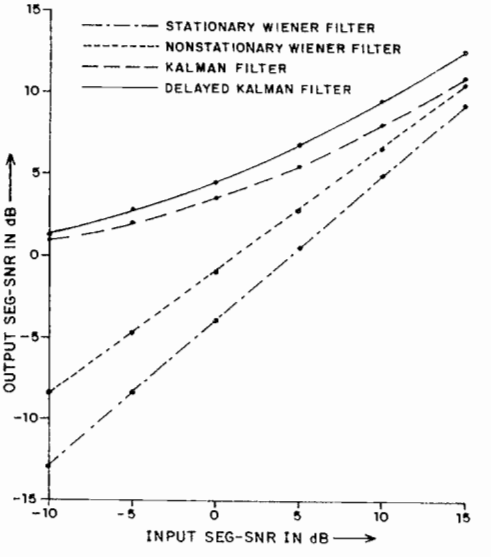

Fig. 2: Output SEGSNR values (in dB) for different speech enhancement methods as a function of input SNR values (in dB).

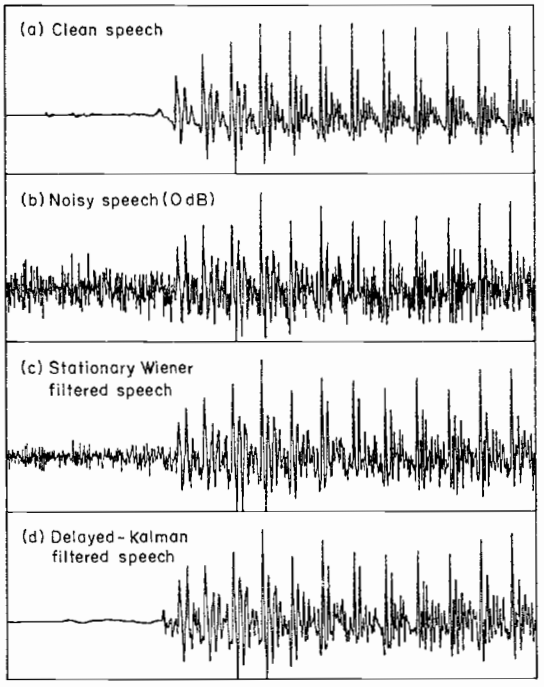

Fig. 3: Illustration of speech enhancement.

IV. COMPUTATIONAL COMPLEXITY

Kalman filtering method is undoubtedly more complicated computationally. Matrix-vector multiplications are needed at each iteration, resulting in an $O(p^2)$ number of operations. But, use of Fast Kalman algorithm [7], which relies upon some shift-invariant properties, reduces the computational complexity to $O(p)$ operations per iteration. Another interesting point is that for each segment, error covariance and Kalman gain matrices reach a steady state value after a few steps. After that point, steady state gain value can be used for the rest of the segment. Thus, a large saving in computation can be achieved.

V. CONCLUSION

In the present paper, a Kalman filtering method is proposed for speech enhancement and its performance is compared with that of the stationary and nonstationary Wiener filtering methods. Since Kalman filter exploits speech production model, it has been found to result in better performance (in terms of both SNR and SEGSNR) than the Wiener filtering method. A delayed Kalman filter has also been proposed which improves the speech enhancement performance of the Kalman filter further due to its inherent fixed-lag smoothing operation.

References

[1] J.S. Lim and A.V. Oppenheim, "Enhancement and bandwidth compression of Noisy Speech", Proc. IEEE, Vol. 67, No. 12, Dec. 1979, pp. 1587-1604.

[2] K.K. Paliwal, "A Linear-phase FIR filter design for speech enhancement", in: I.T. Young et al. (eds.), Signal Processing III: Theories and Applications, North-Holland, Sept. 1986.

[3] J.A. Cadzow, "Spectral estimation: An overdetermined rational model equation approach", Proc. IEEE, Vol. 70, No. 9, Sept. 1982, pp. 907-939.

[4] V.K. Jain and B.S. Atal, "Robust LPC analysis of speech by extended correlation matching", Proc. ICASSP, 1985, pp. 473-476.

[5] K.K. Paliwal, "A Constrained forward-backward correlation prediction method for AR spectral estimation of noisy signals.", in: I.T. Young et al. (eds.), Signal Processing III, EURASIP, 1986, North-Holland, pp. 295-298.

[6] K.K. Paliwal, "A noise-compensated long correlation matching method for AR spectral estimation of noisy signals", Proc. ICASSP, 1986, pp. 1369-1372.

[7] L. Ljung, M. Morf and D.D. Falconer, "Fast calculation of gain matrices for recursive estimation schemes." Int. J. Contr., Jan 1978, pp. 1-19.

[8] J.S. Lim, "Evaluation of a correlated subtraction method for enhancing speech degraded by additive white noise.", IEEE Trans. Acoustics, Speech and Signal Processing, vol ASSP-26, No.-5, 1978, pp. 471-472.