Experimental Analysis of Kalman Filter in Speech Enhancement

[1]Pravitha Prakash, [2]Deepa Elizabeth George
[1,2]Dept. of ECE, Toc H Institute of science & Technology, Kerala, India

Abstract
Speech enhancement is concerned with the processing of corrupted or noisy speech signal in order to improve the quality or intelligibility of the signal. There are so many applications of speech still to be far from reality just because of lack of efficient and reliable noise removal mechanism and preserving or improving the intelligibility for the speech signals. The aim of the speech enhancement techniques is to provide noiseless communication. To filter out the background noise from the desired speech signal several speech filtering algorithms has been introduced in last few years. In this paper Kalman Filter for speech enhancement has been proposed. The Kalman filter has been tuned to get a suitable value of Q by defining the robustness and sensitivity metrics and then applied on noisy speech signals.

Keywords
Kalman Filter, Robustness Metrics, Sensitivity Metric

I. Introduction

Communication via speech is one of the essential functions of human beings. Humans possess varied ways to retrieve information from the outside world or to communicate with each other and the three most important sources of information are speech, images and written text. The most common problem in speech processing is the effect of interfering of noise in the speech signals. The noise masks the speech signal reduces the quality and the speech is greatly affected by presence of surroundings noise. This makes the listening task hard for both circuitous and straight spectators and gives poor performance in some of speech processing. Restraining or reducing such backdrop noise and improving the perceptual quality and intelligibility of a speech without disturbing the speech signal quality is very tough to hold. The algorithms of speech enhancement for noise reduction are Kalman filter, wiener filter and spectral subtraction.

II. Kalman Filter

The Kalman filter is an optimal recursive data processing algorithm or an optimized quantitative expression. The Kalman filter is a set of mathematical equations that provides an efficient computational (recursive) means to estimate the state of a process, in a way that minimizes the mean of the squared error. It is recursive in the sense that it doesn't need to store all previous measurements and reprocess all data each time step [1]. The filter is very powerful in several aspects: it supports estimations of past, present and even future states. Kalman filter which operates through a prediction and correction mechanism.

The first group of equations are used to initialize the state by taking into reference of the previous state and the intermediate state update of the covariance matrix of that state. The second group of equations have to take care of the feedback which adds new information to the previous estimation; so that the proposed estimated state is achieved. The time equations which are updated from time to time are treated as prediction equations, and these equations will generate and add new information to the correction equations. This type of estimation algorithm is called prediction-correction algorithm and is used to solve many problems [2]. Hence, Kalman filter works with projection and correction mechanism and to predict the new state and its uncertainty and correct the projection with the new measure. The two most important parameters in the Kalman filter are the process noise covariance matrix, Q and the measurement noise covariance matrix, R.

III. Determination of Q and R

Speech can be modelled as the output of a linear time-varying filter, excited by either quasi periodic pulses or noise. A closer inspection of this system shows that speech can be modelled as a pth order autoregressive process, where the present sample, x(k) depends on the linear combination of past p samples added with a stochastic or random component that represents noise. In other words, it is an all-pole FIR filter with Gaussian noise as input.

$$X(k) = -\sum_{i=1}^{p} a_i\, x(k-i) + u(k) \tag{1}$$

a_i - linear prediction coefficients (LPCs) u(k) - the process noise
σ_u^2 - Variance of zero-mean Gaussian noise
R, the measurement noise covariance is simply the variance of sample values in the frame. To determine the value for Q, let two terms A_k and B be defined for a particular frame as

$$A_K = H(\emptyset P(K-1|K-1)\emptyset^T)H^T \tag{2}$$

$$B = H(GQG^T)H^T = \sigma_u^2 = Q^f \tag{3}$$

The term A_K denotes the k^{th} instant of the a priori state estimation error covariance [3]. B represents the k^{th} instant estimate of the process noise covariance in the measured output. In our case A_K, B and R are all scalars. The two performance metrics J_1, J_2 and a controlling parameter, n_q. J_1 and J_2 are defined as,

$$J_1 = [(A_K + B + R)^{-1}R] = \frac{\sigma_\omega^2}{A_K + \sigma_U^2 + \sigma_\omega^2} \tag{4}$$

$$J_2 = [(A_K + B)^{-1}B] = \frac{\sigma_\omega^2}{A_K + \sigma_U^2 + \sigma_\omega^2} \tag{5}$$

Any mismatch between the assumed process noise covariance σ_u^2 and the actual process noise covariance is due to error in modelling, hence J_2, which is dependent on σ_u^2 is termed as the robustness metric [4]. Similarly, any mismatch between actual R of the measurement and assumed R adversely affects the a posteriori estimate[5]. Since it is reflected in J_1, it is termed as the sensitivity metric

$$n_q = log_{10}(B) = log_{10}(\sigma_u^2) \tag{6}$$

The method of determining Q is summarised as below. We have denoted the process noise variance, for a frame as Q_f. For each frame of speech, a nominal value of $Q_f = Q_{f\text{-norm}}$ is taken for initial calculation. This Q_f is then varied as $Q_{f\text{-norm}}$x 10n where nϵZ. Hence, n_q = n x log10 Q_f and so, in this case, the metrics are obtained in terms of changing n instead of n_q. For each value of n, corresponding Q_f, J_1 and J_2 values are determined. If the value of Q_f is increased such that it exceeds R substantially, we can say that J_1 reduces to zero while J_2 is high. On the other hand if Q_f is decreased to a small value, then J_2 reduces to zero and J_1 is high, as evident in the graph [6]. Thus, robust filter performance may be expected for large values of Q_f, where as small values of Q_f give sensitive filter performance [7]. A trade-off between the two can be achieved by taking the working value of Q_f as the intersection point of J_1 and J_2 [8-9]. Q_c is the value of Q_f at intersection of J_1 and J_2.

Noise removal from corrupted speech is performed by Kalman filter using these estimated values of R and Q.

Time Update Equations:

$$X(k) = \emptyset(k-1) + Gu(k) \tag{7}$$

where
$X(k)$ = (p x 1) state vector matrix

$\emptyset$ = (p x p) state transition matrix

$$X(K|K-1) = \emptyset X(K-1|K-1) \tag{8}$$

$$P(k|k-1) = \emptyset P(k-1|k-1)\emptyset^T + GQG^T \tag{9}$$

Where
$X(K|K-1)$ = a priori estimate of the current state vector $X(k)$

$P(k|k-1)$ = error covariance matrix of the a priori estimate
Q = process noise covariance matrix
R = measurement noise covariance matrix

$$K(k) = P(k|k-1)H^T(HP(K|K-1) + R)^{-1} \tag{10}$$

$$X(k|k) = X(k|k-1) + K(k)(y(k) - HX(k|k-1)) \tag{11}$$

$$P(k|k) = (I - K(k)H)P(k|k-1) \tag{12}$$

Where
$K(k)$ = Kalman gain
P(k|k) = error covariance matrix of the a posterior estimate
$X(k|k)$ = a posteriori estimate of the state vector.

IV. Simulation Results

Model order p is fixed at 20, For each frame, the pth order LPC coefficients are calculated from noisy speech. The state transition matrix $\emptyset$ is determined from these coefficients. The prediction error covariance from LPC estimation is taken to be the nominal process noise covariance $Q_{f\text{-norm}}$. Process noise variance Q_f is varied as 10n $Q_{f\text{-norm}}$ as mentioned.

The last a posterior error covariance matrix of the previous frame is taken as P(k-1|k-1) for the calculation of A_k. J_1 and J_2 are calculated. Ideally, for most balanced performance, $Q_f = Q_c$ should be selected at the point of intersection of J_1 and J_2 curves.

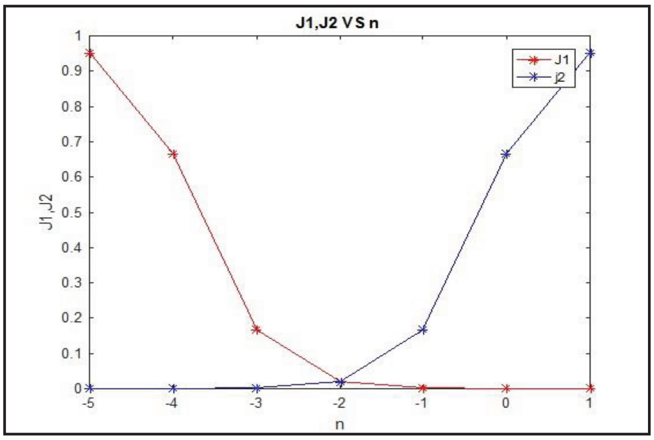

Fig. 1: J1, J2 plot - Estimated Q

4.1 Case 1: Experiment has been done by assuming process noise covariance,Q and measurement noise covariance,R. In this experiment the noisy speech signal has been simulated using a measurement noise variance of R=2.
The Kalman filter output for 100 samples of the noisy input speech signal is as shown in fig. 2.

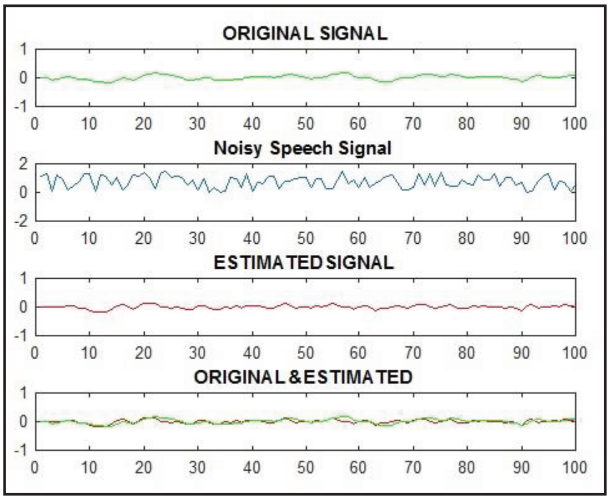

Fig. 2: Filtering Result – Fixed

Case 2: This case is similar to case 1, the only difference is that filtering operation is performed on the input speech signal having 8 x10^4 samples.The original, noisy and estimated speech signal are as shown in fig. 3.

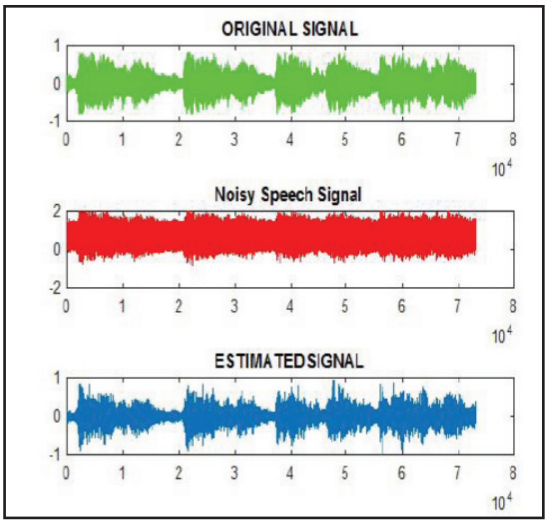

Fig. 3: Filtering Result for More Number of sample- Fixed Q

ISSN : 2230-7109 (Online) | ISSN : 2230-9543 (Print)

Case 3: The Experiment was performed using the estimated value of process noise covariance, Q and measurement noise covariance, R. The original and estimated signal are shown in fig. 4.

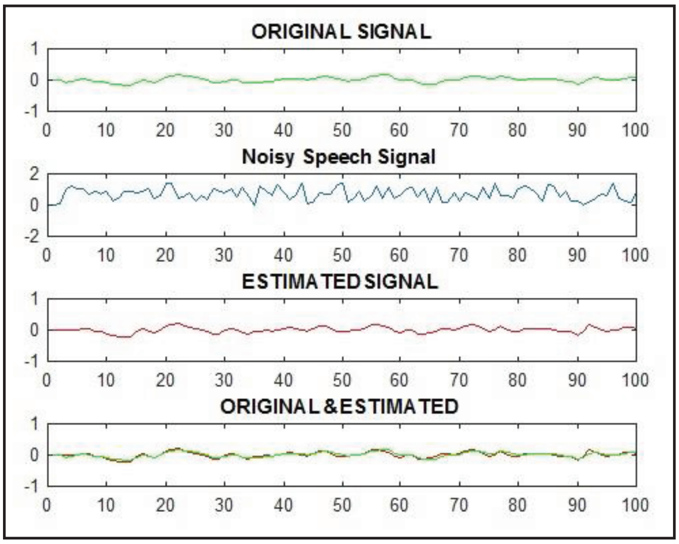

Fig. 4: Filtering result - Estimated Q

Case 4: This is similar to case 3, the estimated value of process noise covariance and measurement noise covariance, is used by the Kalman filter for filtering the input signal of 8 x 10^4 samples. The simulation results are as show in fig. 5.

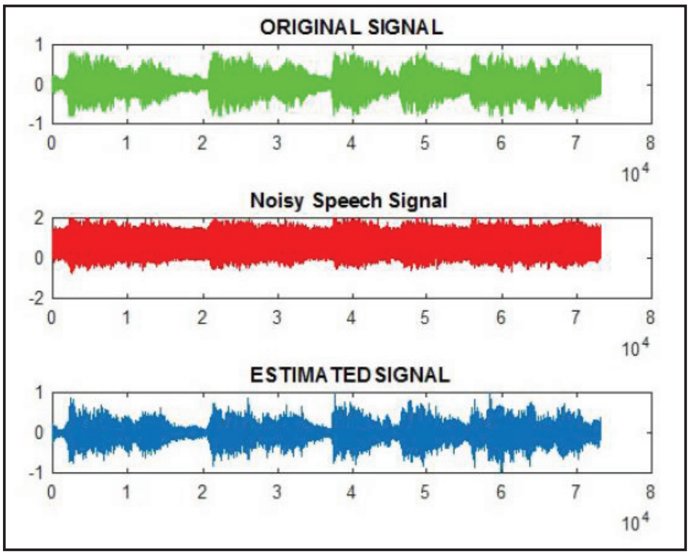

Fig. 5: Filtering Result for More Number of Samples - Estimated Q

V. Conclusion

This paper experiments Kalman filter for speech enhancement. A new method has been verified, where the sensitivity and robustness metrics and their role in choosing a suitable process noise covariance, Q_f for each frame has been explained. It has been shown that toggling between two values of Q_f for the silent and the voiced frames gives us the best results. Kalman filter improve the quality of speech signal. The accurate estimation by Kalman filter on noisy speech would enhance and reduce the noise.

References

[1] K.K. Paliwal, Anjan Basu,"A Speech Enhancement Method Based On Kalman Filtering", IEEE Conference of ICASSP 87 Vol. 12, pp. 12-177-180 May 1987.

[2] Jerry D. Gibson, Boneung Koo, Steven D.,"Gray Filtering of Colored Noise for Speech Enhancement and Coding", IEEE Transactions on Signal Processing. Vol. 39, No. 8. August 1991.

[3] Ching-Ta Lu,"Enhancement of Single Channel Speech Using Perceptual-Decision-Directed Approach", At Sciencedirect, Speech Communication 53 (2011) pp. 495-507.

[4] Vyass Ramakrishnan, Karthik Shetty, P Awan Kumar G, Chandra Sekhar Seelamantula,"Efficient Post-Processing Techniques For Speech Enhancement", 2011 IEEE.

[5] Arun Narayanan, Deliang Wang,"A Casa-Based System For Long-Term Snr Estimation", IEEE Transactions on Audio, Speech and Language Processing, Vol. 20, No. 9, November 2012.

[6] Nima Youse_An, Philipos C. Loizou,"A Dual-Microphone Algorithm That Can Cope With Competing-Talker Scenarios", IEEE Transactions on Audio, Speech and Language Processing, Vol. 21, No. 1, January 2013.

[7] Daniel P. K. Lun, Tak-Wai Shen, K. C. Ho,"A Novel Expectation- Maximization Framework For Speech Enhancement In Non - Stationary Noise Environments", IEEE/ACM Transactions On Audio, Speech and Language Processing, Vol. 22, No. 2, February 2014.

[8] Jesper Jensen, Cees H. Taal,"Speech Intelligibility Prediction Based On Mutual Information", IEEE/ACM Transactions On Audio, Speech and Language Processing, Vol. 22, No. 2, February 2014.

[9] Redha Bendoumia, Mohamed Djendi,"Two-Channel Variable-Step-Size Forward and Backward Adaptive Algorithms For Acoustic Noisereduction and Speech Enhancement", University of Blida1, Signal Processing Andimage Laboratory (LATSI), Route De Soumaa, B.P.270, Blida 09000, Algeria 27 September 2014.