World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:2, No:5, 2008

Optimum Cascaded Design for Speech Enhancement Using Kalman Filter

T. Kishore Kumar

***Abstract*—**Speech enhancement is the process of eliminating noise and increasing the quality of a speech signal, which is contaminated with other kinds of distortions. This paper is on developing an optimum cascaded system for speech enhancement. This aim is attained without diminishing any relevant speech information and without much computational and time complexity. LMS algorithm, Spectral Subtraction and Kalman filter have been deployed as the main de-noising algorithms in this work. Since these algorithms suffer from respective shortcomings, this work has been undertaken to design cascaded systems in different combinations and the evaluation of such cascades by qualitative (listening) and quantitative (SNR) tests.

***Keywords*—**LMS, Kalman filter, Speech Enhancement and Spectral Subtraction.

I. Introduction

VARIOUS speech processing systems have found their way in our everyday life through their vivid use in voice communication, speech and speaker recognition, aid for the hearing impaired, and numerous other applications. However, in most of the cases the ambient environment is noisy which degrades the performance of the speech processing systems drastically. This is particularly true in hands-free speech communication systems, where the background noise is often annoying for persons in conversation. Therefore, speech enhancement has been a challenging topic of research for many years. Speech enhancement is the term used to describe algorithms or devices whose purpose is to improve some perceptual aspects of speech for the human listener or to improve the speech signal so that it may be better exploited by other speech processing algorithms. Approaches to retrieve enhanced speeches are plentiful.

Recently, development of a speech enhancement algorithm which works well in low signal-to-noise ratio (SNR) environments has been desired. Although many speech enhancement methods have been proposed, we may have to distinguish one achieving high quality of the enhanced speech with low computational complexity. Among them the spectral subtraction methods are the most widely used due to the simplicity of implementation and also due to low computational load, making them the primary choice for real time applications.

In general, using the family of subtraction-type algorithms, the enhanced speech spectrum is obtained by subtracting an average noise spectrum from the noisy speech spectrum. The phase of the noisy speech is kept unchanged, since it is assumed that the phase distortion is not perceived by human ear. However, the subtraction-type algorithms have a serious draw back in that the enhanced speech is accompanied by unpleasant musical noise artifact which is characterized by tones with random frequencies. Apart from being extremely annoying to the listeners, the musical noise also hampers the performance of the speech-coding algorithms to a great extent.

In this paper different Speech Enhancement algorithms have been presented that can be pipelined with Spectral Subtraction algorithm to counter the musical noise effect and to increase the overall effectiveness of the Speech enhancement system. However we have to bear this constraint in mind that the enhancement must be prominent in both quantitative as well as qualitative manner and at the same time, we should not be overlooking the complexity and the ease with which these algorithms can be implemented on hardware platforms.

II. Least Mean Square Algorithm

An adaptive filter is a filter that self-adjusts its transfer function according to an optimizing algorithm. Because of the complexity of the optimizing algorithms, most adaptive filters are digital filters that perform digital signal processing and adapt their performance based on the input signal.

Least mean squares (LMS) algorithms is such an adaptive filter used to find the filter coefficients that relate to producing the least mean squares of the error signal (difference between the desired and the actual signal). It is a stochastic gradient descent method in that the filter is only adapted based on the error at the current time.

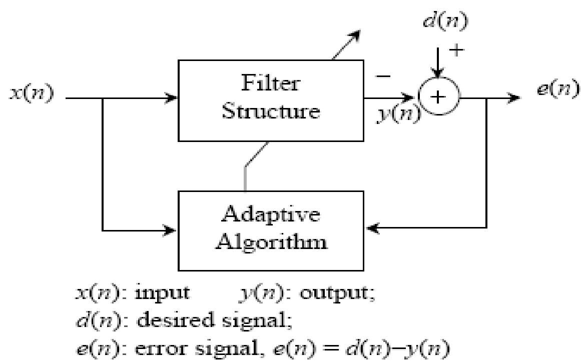

Fig. 1 LMS Algorithm block diagram

The least mean square (LMS) algorithm uses the statistical properties of the data signals. The objective of this method is to minimize the mean square error. The LMS algorithm is widely used in the adaptive algorithm because of its simplicity in structure and its robustness for numerical analysis. A general formulation of the LMS algorithm for an N-tap adaptive filter (Fig. 1) can be described by the following equations:

$$y(n)=\sum w_k(n)x(n-k)=W^T(n)X(n) \text{ ,summation from } k=0 \text{ to } k=N-1 \quad (1)$$

$$w_k(n+1) = w_k(n) + \mu e(n)x(n-k), 0 \le k \le N-1 \quad (2)$$

III. Spectral Subtraction

Spectral subtraction is a method to enhance the perceived quality of single channel speech signals in the presence of additive noise. It is assumed that the noise component is relatively stationary. Specifically, the spectrum of the noise component is estimated from the pauses that occur in normal human speech. Fig. 2 shows the simplified structure of basic spectral subtraction systems.

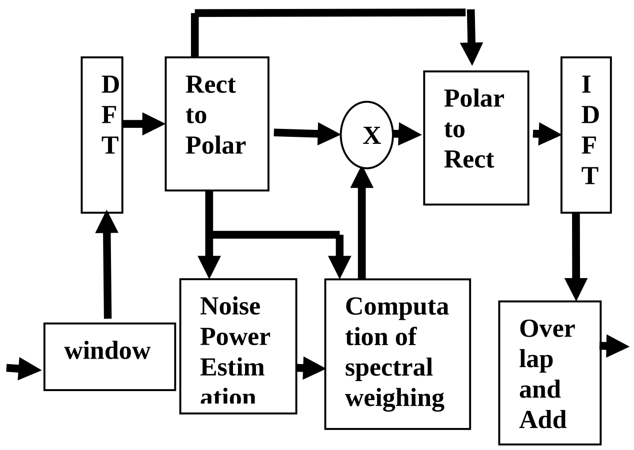

Fig. 2 General Spectral Subtraction Flow chart

In spectral subtraction, input signal is segmented into frames and multiplied with hamming window. We obtain the DFT of these frames and separate magnitude and phase from speech. The noise power estimation and computation of spectral weighting takes place on magnitude. Once the noise estimate is subtracted from speech spectrum magnitude, it is recombined with the original phase of the noisy signal. IDFT is taken and outputs is obtained by overlap and add method The first detailed treatment of spectral subtraction was performed by Boll.[8] Later papers expanded and generalized Boll's method [7] to power subtraction, Wiener Filtering and maximum likelihood envelope estimation.

Speech which is “contaminated" by noise can be expressed as

$$x(n) = s(n) + v(n); \quad (3)$$

where x(n) is the speech with noise, s(n) is the “clean" speech signal and v(n) is the noise process, all in the discrete time domain. What spectral subtraction attempts to do is to estimate s(n) from x(n). If the noise process is represented by its power spectrum estimate $|W\hat{}(f)|^2$, the power spectrum of the speech estimate $|S\hat{}(f)|^2$ can be written as

$$S\hat{}(f)|^2 = |X\hat{}(f)|^2 - |W\hat{}(f)|^2 \quad (4)$$

since the power spectrum of two uncorrelated signals is additive. By generalizing the exponent from 2 to a, Eq. (4) becomes

$$S\hat{}(f)|^a = |X\hat{}(f)|^a - |W\hat{}(f)|^a \quad (5)$$

The speech phase θ S(f) is estimated directly from the noisy signal phase θ X(f).

$$\theta S(f) = \theta X(f) \quad (6)$$

Thus a general form of the estimated speech in frequency domain can be written as

$$S\hat{}(f) = (\max(|X(f)|^a - k|W\hat{}(f)|^a, 0))^{1/a} . e^{j\theta x(f)} \quad (7)$$

Where k > 1 is used to overestimate the noise to account for the variance in the noise estimate, as explained below. The inner term $|X(f)|^a - k|W\hat{}(f)|^a$ is limited to positive values, since it is possible for the overestimated noise to be greater than the current signal.

IV. Kalman Filter

Kalman Filter is an adaptive least square error filter that provides an efficient computational recursive solution for estimating a signal in presence of Gaussian noises. It is an algorithm which makes optimal use of imprecise data on a linear (or nearly linear) system with Gaussian errors to continuously update the best estimate of the system's current state [10].

Kalman filter theory is based on a state-space approach in which a state equation models the dynamics of the signal generation process and an observation equation models the noisy and distorted observation signal. For a signal $x(k)$ and noisy observation $y(k)$, equations describing the state process model and the observation model are defined as

$$x(k) = Ax(k-1)+w(k-1) \quad (8)$$

$$y(k) = Hx(k)+n(k) \quad (9)$$

where, $x(k)$ is the P-dimensional signal vector, or the state parameter, at time k, A is a $P \times P$ dimensional state transition matrix that relates the states of the process at times $k-1$ and k, $w(k)$ (*process noise*) is the P-dimensional uncorrelated input excitation vector of the state equation. $w(k)$ is assumed to be a normal (Gaussian) process $p(w(k))\sim N(0, Q)$, Q being the $P \times P$ covariance matrix of $w(k)$ or *process noise covariance*. $y(k)$ is the M dimensional noisy observation vector, H is a $M \times P$ dimensional matrix which relates the observation to the state vector. $n(k)$ is the M-dimensional noise vector, also known as measurement noise, $n(k)$ is assumed to have a normal distribution $p(n(k))\sim N(0, R)$) and R is the $M \times M$ covariance matrix of $n(k)$ (*measurement noise covariance*).

The most important part of the algorithm is the computation of AR parameters by solving the Yule-Walker equations directly on the noisy signal in the function lpc.m. lpc determines the coefficients of a forward linear predictor by minimizing the prediction error in the least-squares sense.

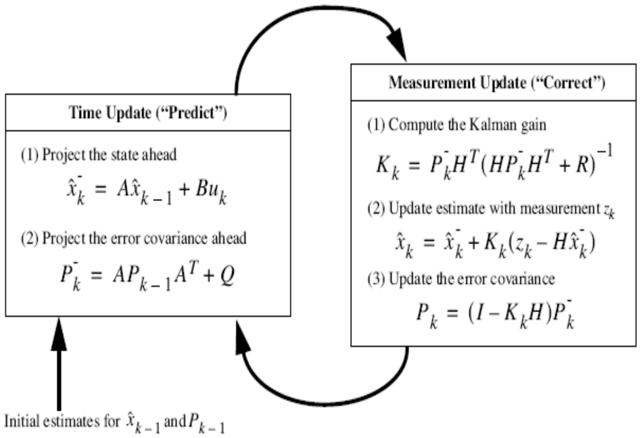

Fig. 3 Main equations of Kalman Filter

V. Cascade

The LMS algorithm provides only a trivial enhancement, Spectral Subtraction has got the problem of musical tones, whereas, Kalman Filters' performance declines rapidly with increasing noise. Hence we connected the aforementioned algorithms in different combinations. After evaluating them, we come to a conclusion that LMS-KF-SS is the optimum cascade for speech enhancement. The LMS would provide some processing in the form of echo and noise cancellation, and Kalman filter would estimate the clean speech signal, the introduction of spectral subtraction at this stage will remove the residual amount of external noise.

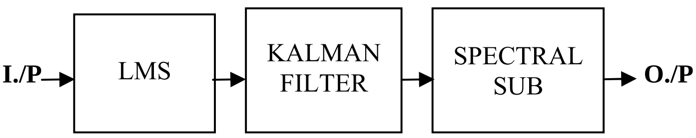

Fig. 4 Block diagram representing Cascading of LMS with Kalman filter and spectral subtraction

VI. Simulation Results

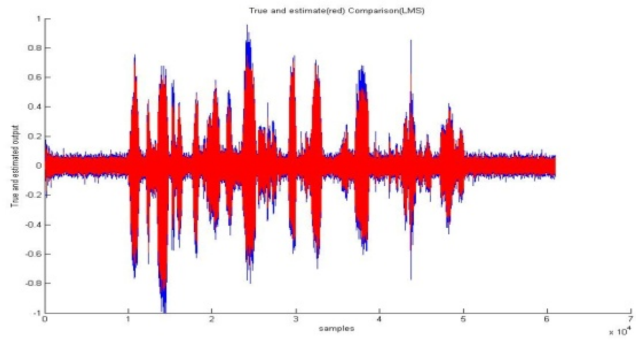

Fig. 5 Time-domain comparison of input signal with the output of LMS Block. (Red is corrected waveform)

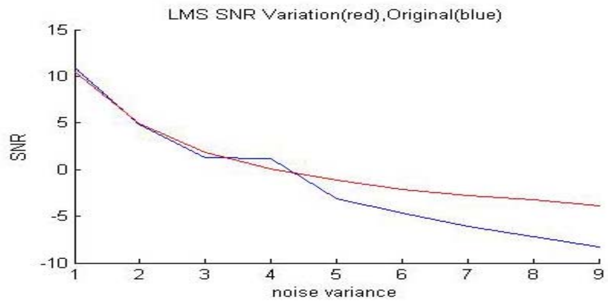

Fig. 6 Variation of SNR with increasing noise in the input (LMS) (Red is corrected waveform)

From Fig. 5 and Fig. 6 we observe that LMS only provides trivial enhancement which is not pretty high in terms of SNR improvement.

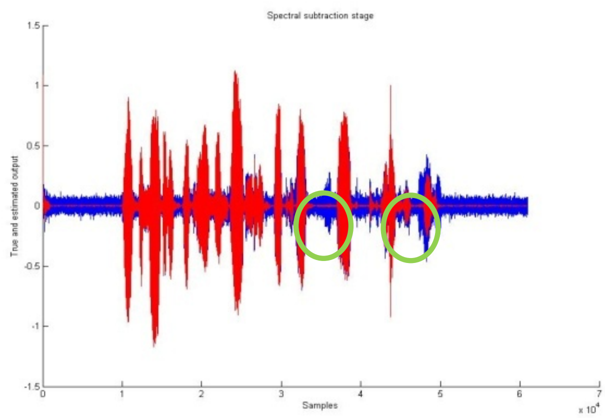

Fig 7 Time domain comparison of input signal and the output of Spectral Subtraction Block. (Red is corrected waveform)

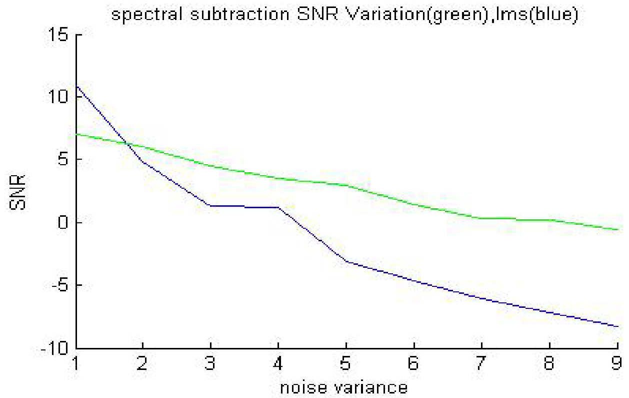

Fig. 8 Variation of SNR with increasing noise in the input (Spectral Subtraction)

Though Spectral Subtraction a large part of the external noise, it suffers from the undesirable musical noise effect which can be seen in Fig. 7 as well as in the listening tests.

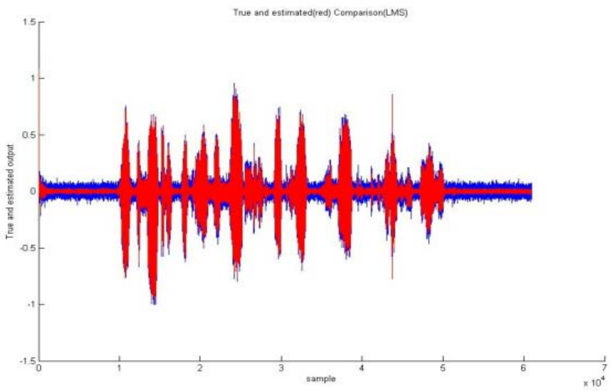

Fig. 9 Time domain comparison of input signal and the output of Kalman Filter Block. (Red is corrected waveform)

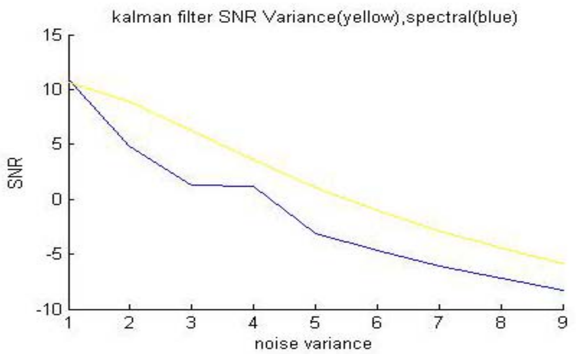

Fig. 10 Variation of SNR with increasing noise in the input (Kalman Filter)

Kalman filter has proven to be the most efficient algorithm by now. From the time domain comparison of input and output for the Kalman filter stage, we observe that Kalman filter effectively removes noise without removing any part of the relevant speech signal.

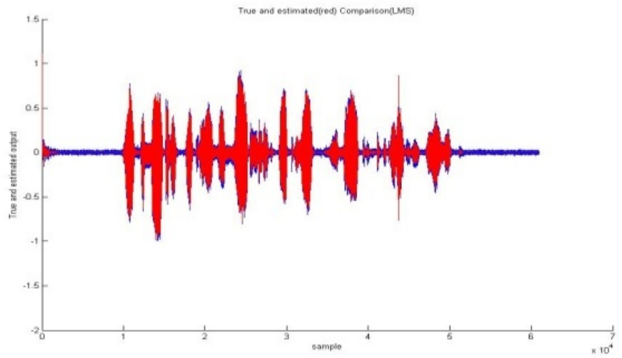

Fig. 11 Time domain comparison of input signal and the output of LMS-Kalman Filter-Spectral Subtraction Cascade

TABLE I
SNR VALUES (DB) OF DIFFERENT SPEECH ENHANCEMENT TECHNIQUES DISCUSSED

(noise) Sigmav	Original	Individual performance			Cascaded performance		
		LMS	Spectral subtraction	Kalman Filter	LK	LSK	LKS
0.01	10.8861	10.4552	7.0367	10.6739	10.1004	6.5261	7.0909
0.02	4.7987	4.8891	6.0432	8.8905	7.7131	5.9092	6.8397
0.03	1.254	1.857	4.4878	6.2384	5.4107	4.4396	5.9774
0.04	1.1618	0.10323	3.4752	3.6305	3.5337	3.2207	5.3655
0.05	-3.1495	-1.1123	2.9387	1.0953	2.1467	2.7125	4.5270
0.06	-4.7101	-2.1322	1.3801	-0.98115	0.95753	1.2653	3.5172

From Table I it was observed that the LKS stage provides considerable enhancement to the input signal. For lower levels of noise, cascading is of not much of use; however as the noise ratio in the input increases, the LMS-Kalman Filter-Spectral Subtraction yields better results as shown in figure.

VII. CONCLUSION

- During the course of experiments we have found that SNR tests alone can't reflect the effectiveness of a de-noising system, results are to be confirmed with Listening tests.

- LMS algorithm can't be used alone as it provides trivial improvement. However it can be used as a preprocessing algorithm owing to its echo cancellation and channel equalization features.

- Spectral subtraction eliminates a good deal of noise however the Residual noise is heavy and undesirable as a good amount of vital speech information sometimes get subtracted.

- Kalman Filter is very effective for speech Enhancement and can be used even alone
- From the SNR curves of individual algorithms, we learn that SNR decreases drastically with increasing noise even while using Kalman Filter.
- By cascaded systems we can eliminate large amounts of noise whereas for lower levels of noise, any single algorithm (preferably Kalman Filter) will do.
- By pipelining LMS with Kalman Filter, we slightly improve the efficiency of the system. Providing stability with increasing noise proportions.
- When LMS-SS-KF sequence was used, Spectral Subtraction rendered a lot of residual noise to the signal, and since Kalman filter was unable to correct or retrieve the lost information, this cascading isn't quite useful. Though it may show appreciable SNR improvements.
- LMS-KF-SS Configuration gives quite good results up to Sigmav=0.03, up to this stage even LMS-KF and KF alone can be used. However for Sigmav more than this, LMS-KF-SS is far better than any other algorithm. Moreover, the efficiency of this system varies only a little with increasing noise. (Red line).

REFERENCES

[1] K. Paliwal and A. Basu, A Speech Enhancement Method Based on Kalman Filtering, proceedings of IEEE Int.Conf. Acoust. Speech, 1987.
[2] P. Pollak and P. Sovka, Spectral Subtraction and Speech/Pause Detection, internal report R95-1, ˇCVUT, Praha, 1995.
[3] V. Sharma et al., Adaptive Speech Enhancement Using a Sub-band Based Kalman Filter, research paper.
[4] Udar Mittal, Nam Phamdo, "Signal/ Noise KLT Based Approach for Enhancing Speech Degraded by Colored Noise", IEEE Trans. on Speech and Audio Processing, Vol.8, No.2, March 2000, pp.159- 167.
[5] H. G. Hirsch, C. Ehrlicher, "Noise Estimation Techniques for Robust Speech Recognition", Proceedings of ICASSP-95, pp.153-156, 1995.
[6] P.Satyanarayana, "Short Segment Analysis of Speech for Enhancement", Ph.D. Thesis - February 1999, Department of Electrical Engineering, IIT Madras.
[7] Steven F. Boll, A Spectral Subtraction Algorithm for Suppression of Acoustic Noise in Speech, Computer Science Department University of Utah Salt Lake City, Utah 84112.
[8] Chi-Chou Kao, Design of Echo Cancellation and Noise Elimination for Speech Enhancement, IEEE transactions, Vol. 49, No. 4, November 2003, Information Technology Department, National Pingtung Institute of Commerce, 900, Taiwan (e-mail: cckao@npic.edu.tw).
[9] Bc. Jan Kybic, Kalman Filtering and Speech Enhancement, http://cmp.felk.cvut.cz/˜kybic/dipl
[10] Wan and A. Nelson, Neural Dual Extended Kalman Filtering: Applications in Speech Enhancement and Monaural Blind Signal Separation, IEEE Workshop and Neural Networks and Signal Processing, 1997.
[11] Acoustic Noise Suppression for Speech Signals using Auditory Masking Effects, Joachim Thiemann, Department of Electrical & Computer Engineering McGill University Montreal, Canada..

Kishore Kumar. T (Member IETE,INDIA) received the B.Tech Degree in Electronics & Communication Engineering amd Master of Techonology in Digital Systems and Computer Electronics from Sri Venkateswara University and Jawaharlal Nehru Technological University India in the year 1992 and 1996 respectively and obtained Doctrate degree in Digital Signal Processing from Jawaharlal Nehru Technological University,Hyd India in 2004.

From 1999 he was associated with NIT Warangal in the position of Assitant Professor, prior to this he was selected for Cabinet Secretariat Prime Minister office Gov INDIA as Technical officer(Tele). His current areas of interest are signal processing and speech processing. At present he is guiding two Phd Scholars and two M.Tech Students. He has published several papers in the areas of Signal & Speech Processing. He has participated in various courses such as Signal processing, VLSI and Curriculum development which are organized by reputed Institutions in India such as IIT Khargpur, Intel Bangalore, IUCEE at Infosys, Mysore. He has undertaken two UG projects in collaboration with Analog Devices, Bangalore in connection with Analog Devices USA sponsored BF 533 DSP Kits to the department of ECE,NIT Warangal- India.