# Distinctive Methods for Speech Enhancement using Kalman Filtering

Chanchal Pandey
Dept. of Electronics and Communication Engg.
Technocrats Institute of Technology
Bhopal, India

Sandeep Nemad
Dept. of Electronics and Communication Engg.
Technocrats Institute of Technology
Bhopal, India

## ABSTRACT
In speech communication systems, it is mandatory to have noise free speech signal with high quality and clarity to obtain high performance. In real world it is very complicated to stockpile noise free speech signal all time for the speech communication system. It is found that speech signals get affected by background noise and tamper the system accuracy. It is very important to filter out the background noise form speech signal to enhance the performance of communication systems, it is also important to enhance the robustness of the speech code and also to enhance the listening ability. To filter out the background noise from the desired speech signal several speech filtering algorithms has been introduces in last few years. In this paper different speech enhancement systems have been examined and a Nobel method which is Second Ordered Fast Adaptive Extended Kalman Filter for speech enhancement has been proposed.

## General Terms
Speech Enhancement System, Background Noise, Speech Signal Denoising

## Keywords
Speech Enhancement, Speech Denoising, Speech Communication, Wiener Filtering, Kalman Filter, Fast Adaptive Kalman Filtering, Second Ordered Fast Adaptive Extended Kalman Filter.

## 1. INTRODUCTION
Speech is most regular manifestation of human correspondence. The observation of speech sign is typically measured as far as its quality and clarity. The quality is a subjective measure that shows the agreeableness or instinctive nature of the apparent speech. Clarity is a destination measure which predicts the rate of words that might be rightly recognized by audience members. Enhancement implies the change in the quality or nature of something. At the point when connected to speech, this basically implies the change in understandability and/or nature of a debased speech motion by utilizing indicator preparing devices. By speech enhancement, it alludes to noise lessening as well as to dereverberation and division of autonomous signs [1]. This is an extremely troublesome issue for two reasons. First and foremost, the nature and qualities of the noise indicators can change significantly in time and between provisions. It is additionally troublesome to discover algorithms that truly work in diverse viable situations. Second, the execution measure can likewise be characterized contrastingly for every requisition. Two criteria are regularly used to measure the execution: quality and coherence. It is tricky to fulfill both in the meantime.

A few procedures have been proposed for this reason .The essential method for speech enhancement is Spectral Subtraction Approach. It is extremely basic method and simple to execute. Different methods for Speech Enhancement are Iterative Wiener filtering analysis, Kalman filtering analysis, Linear Predictive Coding (LPC) analysis, Signal Subspace method [2]–[6]. The exhibitions of these methods rely on upon the quality and comprehensibility of the handled speech signal. The change in the speech signal-to-noise ratio (SNR) is the focus of most procedures [5]. Enhancement systems may be considered single channel and double channel or multi-channel enhancement methods. Single channel enhancement systems apply to circumstances in which one and only securing channel is accessible. The spectral subtraction method is a well-known single channel noise diminishment method [6]. The tried and true power spectral subtraction method considerably decreases the noise levels in the boisterous speech. Yet it presents an irritating distortion in the speech signal called background noise. So the distinctive approaches of spectral subtraction are spectral subtraction with over subtraction, non-straight otherworldly subtraction, Multiband unearthly subtraction and so forth. These methods are utilized to evacuate the background noise and enhance the quality and clarity of the signal. Assessment of otherworldly subtractive algorithms uncovered that these algorithms enhance speech quality and not influence substantially all the more on understandability of speech signal. So the other method called Signal Subspace method [7] is recommended that permits better and more concealment of the noise. The point of this method is to enhance the quality, while minimizing any misfortune in comprehensibility. This survey paper has goal to give a diagram of the verity of speech enhancement algorithms that has been propose to enhance the speech quality and comprehensibility.

In section II we have discuss deferent methods related to speech enhancement, in section III their experimental results is discussed, Section IV shows the concussion of the paper and propose a new method as future work to enhance the processing speed of the kalman filter related to speech enhancement.

## 2. SPEECH ENHANCEMENT METHODS
The fundamental destination of speech enhancement strategy is to enhance the quality and minimize the misfortune in clarity of the signal and audience weariness. The essential review is indicated in fig 1.

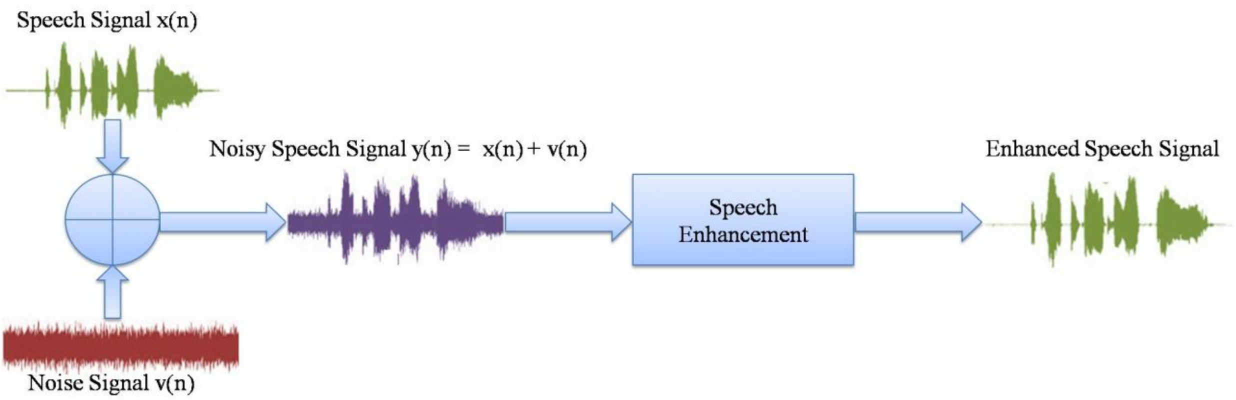


**Fig 1: speech enhancement system**

There are different speech enhancement methods proposed for noise lessening and to enhance the noise quality and clarity. Speech enhancement is possible in both time domain and transform domain [3]. Time domain methods utilize Finite Impulse Response (FIR) and Infinite Impulse Response (IIR) channels, straight prescient coefficients, Kalman Filtering, Hidden Markov Model and so on.

Transform domain strategies are procedures in which change is initially performed on loud speech before filtering took after by the comparing reverse conversion so as to restore the first signal [3]. The principle focal point of performing the noise filtering or lessening process in the transform domain lies in the relative simplicity of recognizing and evacuating noise from speech.

## 2.1 Spectral Subtraction Method

The exact essential method for speech enhancement is Spectral Subtraction method. In this segment, we give a concise outline on traditional unearthly subtraction [4]. Otherworldly subtraction accepts that a signal is made out of two added substance segments. The loud speech might be communicated as [15].

$$y(t) = s(t) + d(t) \tag{1}$$

Where is time, *s(t)* speaks to the uncorrupted speech signal, *d(t)* speaks to the added substance noise signal and is the defiled speech signal accessible for handling. The observed signal *y(t)* is isolated into covering casings utilizing the requisition of a window work and executed in the short-time Fourier transform (STFT) extent domain. In the recurrence domain this could be spoken to as in eq. (2):

$$Y(\omega) = S(\omega) + D(\omega) \tag{2}$$

Eq. (3) shows the estimation of power spectrum of noisy speech:

$$|Y(\omega)|^2 = |S(\omega)|^2 + \delta_n(\omega) \tag{3}$$

Where $\delta_n(\omega)$ are the statistical average values of $|D(\omega)|^2$ during non-speech period, So eq. (4) - (5) shows the enhanced speech signal amplitude.

$$\left|\hat{S}(\omega)\right| = [|Y(\omega)|^2 - E(|D(\omega)|^2)]^{\frac{1}{2}} \tag{4}$$

$$= [|Y(\omega)|^2 - \delta_n(\omega)]^{1/2} \tag{5}$$

Combined with the phase of the noisy signal to synthesize the signal again

$$S(\omega) = \left|\hat{S}(\omega)\right| e^{j\omega g[Y(\omega)]} \tag{6}$$

The reverse short-time Fourier transform is performed to transform the signals into time domain. Traditional spectral subtraction calculation assessing uproarious vitality throughout no speech stage, in any case, it can't upgrade noise throughout speech stage. Additionally the method obliges a VAD that may not work extremely well under low SNR.

## 2.2 Wiener Filtering

In signal preparing, the Wiener filter is used to process an appraisal of a imaginary or target arbitrary process by straight time-invariant filtering a watched uproarious procedure, accepting known stationary signal and noise spectra, and added substance noise [5]. This filter minimizes the mean square failure between the assessed arbitrary procedure and the sought methodology.

The goal of Wiener filter is to remove the noise from a corrupted signal. In general there are two processes which affect the signal that we want to measure:

First of all, it is a fact that every device introduces an error in the output when a signal is measured. If our original signal is $x_k$ and the response of the device is $h_k$ our signal in the output is:

$$y_k = x_k * h_k \leftrightarrow Y_j = X_j . H_J \tag{7}$$

Secondly, the signal outside has noise added due to the process.

$$\hat{y}_k = y_k + n_k \tag{8}$$

To solve this equation, if we don't have noise and we know the response, then the solution is easy to find:

$$X_j = \frac{Y_j}{H_j} \tag{9}$$

But if we have noise, we have to filter the output signal with a Wiener filter.

$$X_j = \frac{Y_j . W_j}{H_j} \tag{10}$$

For that, we should find the optimal Wiener filter. Norbert Wiener had proposed this filter during the 1940s. To reduce the amount of noise in the corrupted signal this filter is based on a statistical approach.

Normally, the filters are designed for a specific frequency, but in Wiener filters, first of all, we should have knowledge about the spectral properties of pure signal and noise, and after that, we have to find a LTI filter whose output would be as close as possible to the original signal. The Wiener filters are characterized by the assumption that, signal and additive noise are stationary linear stochastic processes with known spectral characteristics or known autocorrelation and cross-correlation. This filter must be realizable and causal in nature, the performance criteria for this filter is least mean-square error.

## 2.3 Linear Predictive Coding

Straight prescient coding (LPC) is a tool utilized, basically, in the sound signal and speech transforming to speak to the unearthly wrap of a speech computerized signal in a compacted manner [8]. This method is a standout amongst the most influential to break down the speech, and a standout amongst the most valuable methods for encoding with great quality at low rate.

LPC begins with the supposition that the speech signal is prepared by a buzz at the end of a tube, including, sometimes, murmuring and popping sounds. This model is a great close estimation to the actuality. The glottis prepares the buzz, which is portrayed by his force (clamor) and recurrence (pitch). The vocal tract creates a tube which is called formants. The murmurs and pops sounds are created by lips, tongue and throat. LPC dissects the speech signal utilizing the formants, evacuating their impact from the speech signal and evaluating the force and recurrence of the remaining speech signal buzz. The evacuating formants methodology is called converse filtering and the staying signal after the subtraction is called buildup. The numbers which depict the recurrence and power of the buzz, the formants and the deposit signal could be stored or transmitted.

The essential issue of the LPC system is to focus the formants from the first signal. The result is to express each one example as a direct blend of past specimens. This comparison is called straight predictor. The coefficients of the mathematical statement (the forecast coefficients) describe the formants, so we utilize the LPC system to gauge these coefficients [8].

## 2.4 DFT Based

These are most prevalent as they have less computational many-sided quality and simple execution. They utilize Short Time DFT (STDFT) and have been seriously examined; otherwise called Spectral Preparing Methods [7]. They are focused around the way that human speech recognition is not delicate to spectral stage however the clean spectral amplitude must be legitimately removed from the loud speech to have satisfactory quality speech at yield and subsequently they are called Short Time Spectral Amplitude (STSA) based methods.

## 2.5 Signal Subspace Method

The approach includes the utilization of a signal dependant transform to disintegrate an uproarious signal into two separate subspaces, the signal in addition to noise subspace, and the noise-just subspace. The transform utilized to perform this operation is the Karhuenen-Loeve Transform (KLT) [7].

This hypothesis expects that speech can just compass the signal in addition to noise subspace, for effortlessness called the Signal Subspace, while noise can compass the whole Euclidean space. Just the signal subspace is utilized when evaluating the clean signal.


**Fig 2: Block diagram of subspace speech enhancement system**

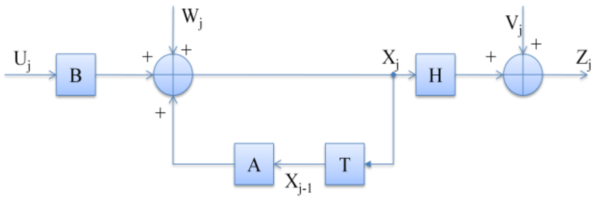


**Fig 3: State space model of extended first order kalman filter**

The KLT segments which speak to the noise just subspace are null, while the parts which speak to the loud signal are altered by an addition function. The upgraded signal is dead set from the converse KLT of the changed segments. The point here is to enhance the quality, while minimizing any misfortune in understandability.

The improved speech, transformed by the Signal Subspace with Adaptive Noise Estimation (SSANE) calculation, is of a great, regular-sounding quality and holds no discernable noise [7].

## 2.6 Kalman Filter

The Kalman filter, also called linear Quadratic Estimation (LQE), is a method that uses an arrangement of estimations saw about whether, holding noise (arbitrary varieties) and different mistakes, and produces appraisals of obscure variables that have a tendency to be more exact than those focused around a solitary estimation alone. All the more formally, the Kalman filter works recursively on streams of uproarious data information to generate a measurably ideal appraisal of the underlying system state. The filter is named for Rudolf (Rudy) E. Kálmán, one of the essential designers of its hypothesis [8].

The method works in a two-stage process. In the prediction step, the Kalman filter produces assessments of the current state variables, alongside their instabilities. Once the result of the following estimation (essentially defiled with some measure of slip, including irregular noise) is watched, these appraisals are overhauled utilizing a weighted normal, with more weight being given to gauges with higher conviction. On account of the calculation's recursive nature, it can run progressively utilizing just the present info estimations and the at one time ascertained state and its instability system; no extra past data is needed [9].

### *2.6.1 Extended First Order Kamlan Filter*

In the Extended Kalman Filter (EKF) as shown in fig. 3, the state move and discernment models oblige not be linear function of the state yet may rather be non-linear functions [10]. These functions are of differentiable sort as in comparison (11) and (12).

$$x_j = Ax_{j-1} + Bu_j + w_j \tag{11}$$

$$z_j = Hx_j + v_j \tag{12}$$

The function *f* could be used to process the anticipated state from the past evaluation and correspondingly the function *h* may be used to adjust the expected estimation from the anticipated state. Notwithstanding, *f* and *h* can't be associated with the covariance direct. Rather a network of incomplete subordinates (The Jacobian) is processed.

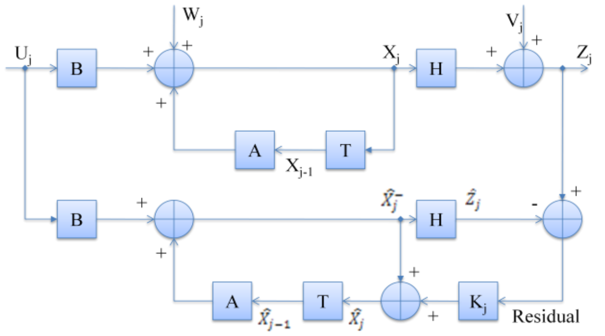


**Fig 4: State model of extended second order Kalman filter**

At each time step the Jacobian is surveyed with current expected states. These matrix may be used as a piece of the Kalman filter numerical articulations. This technique essentially laniaries the non-linear function around the current estimation.

### *2.6.2 Extended Second Order Kalman Filter*

The state space model of second order Kalman filter [11] is shown in Fig. 4. In this case the prediction equation can be expressed as equation (13)

$$\hat{X}_j^- = A\hat{X}_{j-1} + BU_j \tag{13}$$

and the correction equation can be expressed as equation (14)

$$\hat{X}_j = X_j^- + K_j\left(Z_j - H\hat{X}_j^-\right) \tag{14}$$

prior correction is expressed as in equation (15)

$$P_j^- = E\{e_j^- e_j^{-T}\} = E\left\{\left(X_J - \hat{X}_j^-\right)\left(X_J - \hat{X}_j^-\right)^T\right\} \tag{15}$$

and posteriori covariance can be expressed as in equation (16)

$$P_j = E\{e_j^- e_j^{-T}\} = E\left\{\left(X_J - \hat{X}_j\right)\left(X_J - \hat{X}_j\right)^T\right\} \tag{16}$$

Where the system has *p* inputs, *n* state variable, *m* yields, $X_j$ is *n* by 1 state network, $U_j$ is *p* by 1 information matrix, $Z_j$ is *m* by 1 yield vector, *A* is *n* by *n* satiate increase matrix, *B* is *n* by *p* data pick up matrix, $H_j$ is *m* by *n* yield pickup system, $K_j$ is *n* by *m* Kalman filter pick up matrix, $P_j$ is *n* by *n* back covariance grid, $P_j^-$ is *n* by *n* earlier covariance system and superscript *T* indicated the transpose of the matrix.

## 3. EXPERIMENTAL RESULTS AND PERFORMANCES

In [12] performed coherence and quality measure tests utilizing the Diagnostic Rhyme Test (DRT). Result demonstrated that spectral subtraction did not diminish speech clarity however enhance speech quality especially in the zone of enjoyableness and subtlety of the foundation noise.

In [13] examined different strategies to diminish the background noise from the uproarious speech signal. This background noise might be decreased to a certain cutoff by utilizing the spectral subtraction strategies utilizing regulation domain and geometric approach. At the point when both subjective and destination test were performed on the adjustment approach then we get the enhanced speech quality.

In [14] recommended that this paper shows and investigates another speech enhancement calculation focused around enhanced spectral subtraction. Enhanced spectral subtraction calculation exactly appraises the noise as indicated by that the amplitude spectral of narrowband white Gaussian noise obeys Rayleigh circulation, taking into account that all noise could be changed into AWGN. This calculation likewise embraces another speech action recognition innovation focused around recurrence band fluctuation to recognize signal movement. The emulation breaks down shows that the calculation in this paper is better suit for speech enhancement by evacuating the noise in correlation to standard spectral subtraction.

In [15] portrays diverse approaches of spectral subtraction method for improving the speech signal from the boisterous situations. The creators say that the clean signal's quality is debased by the added substance foundation noise. Around all the accessible methods the spectral subtraction calculation is historically one of the first calculations proposed for foundation noise diminishment. In this paper the author display the audit of essential spectral subtraction calculation, for example, spectral subtraction with over subtraction, non straight spectral subtraction and MMSE spectral subtraction focused around the perceptual properties that minimizes the limits of the fundamental methods.

In [16] depicts subspace filtering handles substantially less background noise than spectral subtraction does. Likewise, for enhanced speech distinguishment precision in boisterous situations, SVD-based speech enhancement ended up being profoundly aggressive with spectral subtraction. Generally, the MV estimator including its generalization to the TDC estimator and the SDC estimator demonstrated to give the best comes about.

In [17] found that KLT-based speech enhancement is to be favored over FFT-based (i.e., spectral subtraction) algorithms, despite the fact that the recent works at a lower computational burden. In [1] proposed calculation for signal subspace speech enhancement is executed and tried utilizing speech record tested at recurrence of 8 kH at 16 bits rate. The speech wave document is changed over into 16 bits ASCII values. The crude qualities are connected to Karhunen loeve transform to partition the speech and noise signal.

Table 1 shows the expected $SNR_{OUT}$ results of proposed Fast Adaptive Second Ordered Extended Kalman Filter (FASOEKF) for de-noising the noisy speech signal compared with the conventional second order extended Kalman filter, which shows that this proposed method de-noise the speech signal similarly as conventional SOEKF for deferent L (Length of transition matrix) and input SNRin. Table 2 show the expected processing time of proposed method FASOEKF compared with processing time of conventional SOEKF method for deferent L (Length of transition matrix) and input SNRin, which shows that proposed method de-noise the speech signal in less processing time.

**Table 1. Expected SNRout for noisy speech signal by proposed method**

| $SNR_{in}$ [dB] | L | $SNR_{out}$ | |
|---|---|---|---|
| | | **Conv. SOEKF** | **Proposed FASOEKF** |
| 0 | 20 | 2.38 | ≈ 3.66 |
| | 30 | 3.15 | |
| | 40 | 3.50 | |
| | 50 | 3.66 | |
| 10 | 20 | 6.67 | ≈ 8.88 |
| | 30 | 7.78 | |
| | 40 | 8.85 | |
| | 50 | 8.88 | |
| 20 | 20 | 10.11 | ≈ 12.38 |
| | 30 | 11.45 | |
| | 40 | 11.50 | |
| | 50 | 12.38 | |

**Table 2. Expected processing time for proposed method**

| $SNR_{in}$ [dB] | L | Processing Time [s] | |
|---|---|---|---|
| | | **Conv. SOEKF** | **Proposed FASOEKF** |
| 0 | 20 | 10.323 | ≈ 2.5 |
| | 30 | 32.356 | |
| | 40 | 35.012 | |
| | 50 | 56.456 | |
| 10 | 20 | 10.667 | ≈ 2.4 |
| | 30 | 17.784 | |
| | 40 | 38.805 | |
| | 50 | 58.784 | |
| 20 | 20 | 10.011 | ≈ 2.3 |
| | 30 | 34.256 | |
| | 40 | 38.457 | |
| | 50 | 55.841 | |

# 4. CONCLUSION

In this paper miscellaneous speech enhancement methods have examined. Spectral subtraction is the essential system for speech enhancement yet it has some serious disadvantages like inconceivability to use for non-stationary noise, Dependence on VAD (Voice Activity Detector) exactness, Background noise because of flawed noise estimation. The significant Drawback of Wiener filter is the altered recurrence reaction at all frequencies and the prerequisite to gauge the force spectral thickness of the clean signal and noise before filtering. The signal subspace approach is utilized for hearty and precise noise estimation in non-nature. Signal subspace speech enhancement has ended up being a compelling and extremely adaptable tool, both for expanding the speech understandability in speech correspondence requisitions and enhancing the exactness of automatic speech recognizers in added substance nature's domain. A standout amongst the most major contrasts between the Wiener filter and the Kalman filter is the function of the last to oblige non-stationary signals. By the recursive equations of the conventional second order extended Kalman filtering algorithm, we can find that equations contain a large number of matrix operations. Mainly the inverse matrix operations lead to an increase in the algorithm's complexity. If we can reduce the dimension of matrix or eliminate matrix operations, we can greatly reduce the complexity of the second order Kalman filter and enhance the processing speed of speech enhancement by implementing Fast Adaptive Extended Kalman Filter for speech enhancement is proposed.

# 5. ACKNOWLEDGMENTS

The author would like to express their sincere thanks to Dr. Shivendra Singh, Head of the Department of Electronics and Communication, all authors the reference for this research, All faculty members of the department and of all the persons who have supported in any way for this survey.

# 6. REFERENCES

[1] Prasad, G., and Surender. 2013 A Review of Different Approaches of Spectral Subtraction Algorithms for Speech Enhancement. Current Research in Engineering, Science and Technology (CREST) Journals. Vol. 01. Issue 02. 57-64.

[2] Ramalakshmi, K. 2013 speech enhancement with signal subspace filter based on perceptual post filtering. Dept of CSE Sri Ramakrishna Institute of Technology, Coimbatore. Vol. 2, Issue 1.

[3] Verteletskaya, E., and Simak, B. 2011 Noise Reduction Based on Modified Spectral Subtraction Method. IAENG International Journal of Computer Science. Vol. 38.

[4] SUI, L. Y., ZHANG, X. W., HUANG, J. J., and ZHOU, B. 2011 An Improved Spectral Subtraction Speech Enhancement Algorithm under Non-stationary Noise. Institute of Command Automation, PLAUST Nanjing, China.

[5] Abd El-Fattah, M. A., Dessouky, M. I., Diab, S. M., and Abd El-samie, F. E. 2008 Speech Enhancement using an Adaptive Wiener Filtering approach. Department of Electronics and Electrical communications, Menoufia University, Menouf, Egypt, Progress In Electromagnetics Research M. Vol. 4. 167–184.

[6] Udrea, R. M., Vizireanu, D. N. and Pirnog, I. 2007 A Perceptual Approach for Noise Reduction using Nonlinear Spectral Subtraction.

[7] Commins, B. 2005 Signal Subspace Speech Enhancement with Adaptive Noise Estimation. National University of Ireland, Galway.

[8] Kalman, R. E. 1900 A New Approach to Linear Filtering and Prediction Problems. ASME journal of basic engineering. Vol. 82. 35-45.

[9] Kalman, R. E., and Buch, R. S. 1961 New Results in Linear Filtering and Prediction Theory. ASME journal of basic engineering. 95-108.

[10] Lee, Y. K., Jung, G. W., and Kwon, O. W. 2013 Speech Enhancement by Kalman Filtering with a Particle Filter-Based Preprocessor. IEEE International Conference on Consumer Electronics (ICCE). 340-341.

[11] Julier, S. J., and Uhlmann, J. K. 2004 Unscented filtering and nonlinear estimation. Proceedings of the IEEE. 401–422.

[12] Hydari, M. 2009 Speech Signals Enhancement Using LPC Analysisbased on Inverse Fourier Methods. Department of Computer Engineering, Faculty of Engineering Noshirvani, Institute of Technology, Babol, Iran, Contemporary Engineering Sciences. Vol. 2.

[13] Goel, P., and Garg, A. 2011 Review of spectral subtraction Technique for speech enhancement. IJECT. Vol. 2.

[14] Gao, L., Guo, Y., Li, S., and Chen, F. 2009 Speech enhancement algorithm based on improved spectral subtraction. Intelligent Computing and Intelligent Systems. ICIS 2009. IEEE International Conference on, Shanghai. Vol. 3. 140-143.

[15] Fukane, A. R. 2011 Different Approaches of Spectral Subtraction method for Enhancing the Speech Signal in Noisy Environments. International Journal of Scientific & Engineering Research. Vol. 2. Issue 5.

[16] Ephraim, Y., and Van Trees, H. L. 1995 A signal subspace approach for speech enhancement. IEEE Transactions on Speech and Audio Processing. Vol. 3. No. 4. 251–266..

[17] Hermus, K., Wambacq, P., and Hamme, H. V. 2007 A Review of Signal Subspace Speech Enhancement and It's Application to Noise Robust Speech Recognition. Hindawi Publishing Corporation, EURASIP Journal on Advances in Signal Processing.