International Journal of Innovative Research in Electronics and Communications (IJIREC)
Volume 3, Issue 2, 2016, PP 15-23
ISSN 2349-4042 (Print) & ISSN 2349-4050 (Online)
www.arcjournals.org

Comparative Performance Analysis of Speech Enhancement Methods

Deepa Srinivasa[1], Dr. P A Vijaya[2]

[1]M.Tech Student, [2] Professor and Head,
Department of Electronics and Communication,
BNM Institute of Technology, Bangalore, India
[1]deepaanvekar15@gmail.com, [2]pavmkv@gmail.com

Abstract: *The implementation of Speech Enhancement using Spectral Subtraction method and Kalman Filtering method are presented here. Spectral subtraction mainly removes the background noise from noisy speech spectrum. Original speech is obtained by subtracting noise spectrum from noisy speech spectrum. The main disadvantage of spectral subtraction is the presence of musical noise in enhanced speech signal. Secondly, 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. It is observed that spectral subtraction method is very efficient for noise elimination in high SNR condition but performance degrades with reduction in SNR. It is observed that Kalman Filter is efficient in reduction of noise particularly at low SNR conditions. The two designs are modelled using MATLAB for seven different noise types and the SNR are compared.*

Keywords: *Speech Enhancement, Spectral Subtraction, Kalman filter, Musical noise*

1. Introduction

Speech enhancement is used to improve intelligibility and overall perceptual quality of degraded speech using various algorithms and audio signal processing techniques. The aim of speech enhancement is to improve quality of speech.Quality here refers to clarity, intelligibility, pleasantness or compatibility. It is very difficult for any mathematical algorithms to measure intelligibility and pleasantness. The methods for enhancing speech are removal of background noise, noise suppression, adaptive filtering of noise etc.

The speech band in telephone networks range between 300-3400Hz. The markets are now dominated by fourth generation phones with higher frequency ranges. It is very crucial not to harm or garble the speech signal whenever the background noise is suppressed. The background noise sounds more comfortable than the quiet unnatural noise. Comfortness is not an issue if the speech signal is given to speech recognizer than to human ear.

Background noise suppression has number of applications like using telephone in a noisy environment, air-ground communication etc. The demand is more for good speech enhancement algorithms for hearing aid. Operation of Speech Recognition systems rely on features extracted from speech which are in turn disturbed by extra noise sounds. In active noise suppression method an anti-noise is produced into listener’s ear to cancel the noise. The delay must be small so as to avoid producing more noise instead of cancellingthe existing noise. Therefore most of methods for active noise suppression are analog.

Noise Reduction Principles:

The main requirements of any noise reduction system for speech enhancement are:-

1. Intelligibility and naturalness of enhanced signal
2. Improvement of Signal-to-noise ratio (SNR)
3. Short Signal delay
4. Computational simplicity

Naturalness refers to how natural is the speech after enhancement is applied and how well it is reflecting original speech signal.The Signal-to-noise ratio must be improved so as to prove the use of speech enhancement algorithms suppressed noise to a considerable level.The delay of the enhanced signal is another important aspect. The delay of enhanced signal must be as short as possible.The complexity of applying these algorithms, speech enhancement techniques must be relatively less as possible. The entire process and techniques must be as simple as possible.

2. Literature Survey

Berouti in his implementation proposed a method that even reduces the effect of musical noise in the enhanced signal. Here as in case of conventional subtraction not only noise spectrum is subtracted from noisy speech spectrum but also this method subtracts an over-estimate of noise spectrum so that output will not go beyond noise floor [2].

Yi Zhang and Yunxin Zhao proposed a new algorithm that subtracts real and imaginary parts separately. Thus both real and imaginary parts are processed and enhanced separately. The restored signal is more effective[3].

Y.Ephraim and D.Malah [4] proposed work on short time spectral amplitude (STSA) of speech. In this work, spectral components of noise and speech are modelled as statistically independent Gaussian random variables. Both the complex phase of noise signal and MMSE STSA estimator are combined to get enhanced signal.

At times when noise itself is contaminated, its peak will shift; thus it would be difficult to differenciate between clean speech and noise which in turn may cause overlap of speech and noise spectrums; with respect to above problem, Peng Dai and Ing Yann proposed an algorithm that moves the noise peak and decreases the overlap of spectrums [5].

Research in field of unvoiced speech separation has also been done. Ke Hu and Dehiang Wang in their algorithm segregated unvoiced speech from the speech signal. This algorithm removes periodic interference [6].

The use of averaging factor to estimate the a-priori SNR [7] was proposed by Md.Kamral Hasan, SayeefSalahuddin and M.Rezwan Khan. The aforesaid method reduced musical noise to larger extent.

Another technique called time-frequency block thresholding was implemented for audio denoising proposed by Gvoshen Yu, Stephane Mallat and Emmanuel Bacry [8].

Kalman, R. E. proposed a new approach to Linear Filtering and Prediction Problems by introducing prediction and correction algorithm [10].Kalman, R. E., and Buch, R. S. worked jointly to find new and efficient results in linear filtering and prediction theory[11].Mrs. Ganga K Moorthyand Mrs. MeenaPriyaDharshini proposed a method of implementing spectral subtraction on FPGA thus making use of VLSI architecture for speech enhancement [1].

3. Spectral Subtraction

At times noise is present on separate channel apart from noisy signal. Therefore the noise spectrum estimate can be subtracted from noisy speech spectrum. It is very difficult to remove noise completely but we can reduce its effects. The additive noise increases mean and variance of spectrum. The variance which is due to random nature of noise is very difficult to cancel out. The increase in mean can be nullified by subtracting mean of noise spectrum noisy speech. In this domain, the noisy signal is represented as in below equation

$$\mathbf{y(m) = x(m) + n(m)} \tag{1}$$

where y(m) is noisy signal, x(m) is clean speech signal and n(m) is noise signal. The frequency domain representation of the above equation is given by

$$\mathbf{Y(f) = X(f) + N(f)} \tag{2}$$

where Y(f), X(f) and N(f) are Fourier transform of noisy signal, speech signal and noise signal respectively. The input noisy speech signal is first buffered and divided into segments of sample length N. Each segment undergoes windowing by Hanning Window. Then each windowed segment is transformed into frequency domain by using Fast Fourier Transform (FFT).

Windowed Signal is represented by

$$\mathbf{y(m) = w(m)\, y(m)} \tag{3}$$

$$= w(m)\,[\, x(m) + n(m)] \tag{4}$$

$$= x_w(m) + y_w(m) \tag{5}$$

In frequency domain, the above equation is expressed as

$$\mathbf{Y_w(f) = W(f) * Y(f)} \tag{6}$$

$$= W(f)\,[\, X(f) + N(f)] \tag{7}$$

$$= X_w(f) + N_w(f) \tag{8}$$

Thus spectral subtraction is defined asin equation

$$|\, \hat{X}(f)\, |^{b} = |Y(f)|^{b} - \alpha\overline{|N(f)|^{b}} \tag{9}$$

where$|\, \hat{X}(f)\, |^{b}$is estimate of original speech signal $\overline{|N(f)|^{b}}$ is time averaged noise estimate.

The magnitude of clean speech is combined with noisy phase to obtain the restored signal. Inverse Fourier transform is performed to convert signal back to time domain.

4. Kalman Filtering

Kalman filter is named after Rudolf .E. Kalman[10]. The basic operation of Kalman filter is based on recursive process and provided solution for problem based on linear filtering for discrete data. Many research work is going on in the area of State-Space models.The aim here is to estimate the signal through least square process recursively.

Kalman filter provided very good results and is applied to many applications like Missiles Search, Navigation and Economy. Kalman filter is based on concepts of Wiener filter. Kalman filter allows both stationary and nonstationary parts of speech and is also capable to estimate errors more accurately than any other filters.It is also called Linear Quadratic Estimation(LQE) because it uses estimations of holding noise of arbitrary varieties and various different mistakes and produces observation of unknown variables that are more exact than those focused around confined solitary estimation.

Kalman filter recursively works on streams of data to generate ideal appraisal of the system state. The use of this filter for Speech enhancement was first presented by Paliwal. It is best suitable for white noise reduction from a signal and also fulfills the Kalman Filter assumptions. In order to derive Kalman filter equations we assumed additive noise is uncorrelated and has normal distribution.

Speech signal is considered stationary during each frame in order to prove that AR model of speech remains same along the entire segment.

State vector of Kalman filter to fit one dimensional Speech Signal into State Space model is given by:-

$$\mathbf{x(k)=([x(k-p+1)x(k-p+2)x(k-p+3)\ldots\; x(k)])^{T}} \tag{10}$$

Where, x(k) is the input speech signal at time k and consider that speech signal is corrupted by additive white noise n(k) then the summation of these two form y(k).

$$\mathbf{y(k)=x(k)+n(k)} \tag{11}$$

5. Design Flow and Methodology

5.1. Design I using Spectral Subtraction

The time domain input signal is taken first. As processing is simple and easier in frequency domain time domain signal is converted to frequency domain. This is done by Fast Fourier Transform (FFT) block. We go for FFT as it is computationally efficient than Discrete Fourier Transform (DFT).The three main stages in the design are preprocessing stage, main processing stage and the final stage.

The Preprocessing stage mainly does the framing of signal that is the time domain signal is divided into frames using Hamming window and then converted into frequency domain using Fast Fourier Transform (FFT).Next in main processing stage the decomposition of signal, estimating the magnitude and then finally subtracting the magnitude of the signal with noise magnitude is done. In final stage of the design the deframing is done in which signal divided into frames is combined into one and then the inverse FFT is performed.

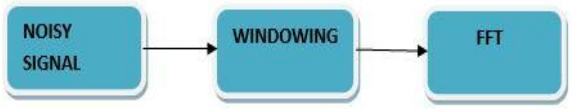


Figure1. *Preprocessing stage*

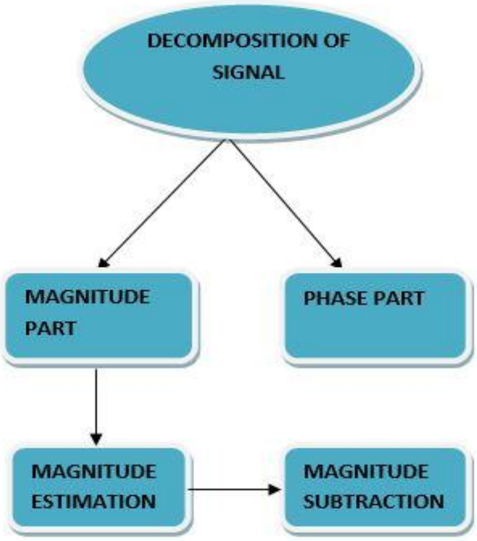


Figure2. *Main Processing block of design*

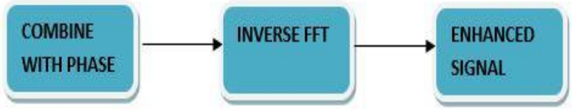


Figure3. *Final stage of the design*

The frequency domain signal is decomposed into magnitude and phase blocks.As the phase block doesn't undergo processing and is retained as such and then finally added to processed signal. The noise estimation - subtraction block initially finds out estimate of noise spectrum and then noise spectrum is subtracted from noisy spectrum.

The phase block divides phase into sine and cosine parts. These sine and cosine parts are inturn given to multiplier along with magnitude of clean signal. The output of these multipliers is given to Inverse FFT block which converts frequency domain signal back to time domain. The output of Inverse FFT block is enhanced signal.

5.2. Design II using Kalman Filter

Kalman method is a two stage process. One is Prediction step and second is Correction step. In prediction step, filter produces considerations of current state variables with their instabilities. Because of recursive nature of calculation the system can run progressively utilizing the present data estimations and with no extra past data requirement[9].

The Kalman filter is designed to calculate the previous process using a feedback control. The process is estimated over the time and then it gets the feedback through observed data.This filter derives the possibilities into two groups. In step one derive the equation in order to update the time or prediction. In step two, update the observed data.

Firstly initialize the state by taking reference of previous state and intermediate state update of covariance matrix of that particular state. Secondly, take care of feedback which adds new information to previous estimation to achieve proposed estimation.

The time based equations are updated from time to time and are called prediction equations. These equations will generate and add new information to correction equations. This method of estimation algorithm is known as Prediction Correction algorithm. This cycle of mechanism is shown below

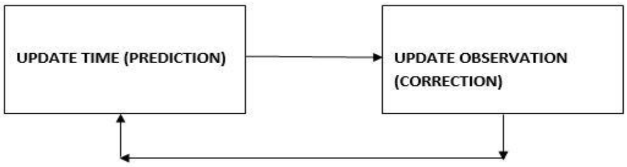


Figure4. *Block diagram of prediction and correction algorithm*

As theory of Kalman filter is based on State-Space approach, the state equation models the dynamics of generating signal and an observation equation models the distorted and noisy observation signal.

The state equation is given by $x_{k+1}=F_k x_k + G_k w_k$ (12)

The observation equation is given by $y_k = H_k x_k + v_k$ (13)

Where w_k and $v_k$are independent zero mean Gaussian white noises. The $\sum w_k$ and $\sum v_k$ are covariance matrices.$F_k$is transition matrix$G_k$is input matrix and $H_k$is output matrix.

Kalman filter interms of equation is given below

$\hat{x}_{k|k} =\hat{x}_{k|k-1} + K_k(y_k -H_k\hat{x}_{k|k-1})$ (14)

$K_k$is Kalman gain matrix,$\hat{x}_{k|k}$ is called estimated value of x_k at time k and $\hat{x}_{k|k-1}$ is called predicted value of x_k at time k-1

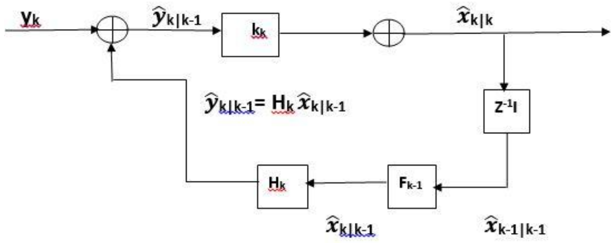
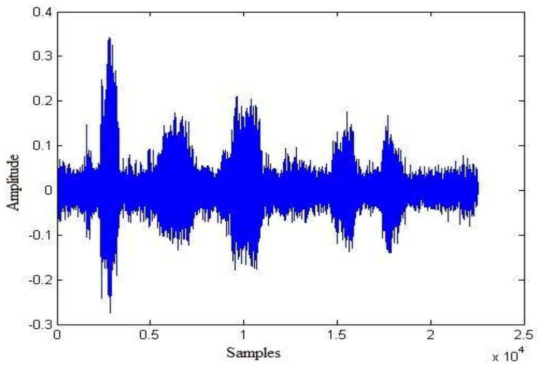


Figure5. *Block diagram of Kalman filter*

6. Results and Performance Analysis

The simulation results of both the Speech enhancement methods are plotted below.The SNR obtained for both the methods with seven different noise types are compared. Below are the simulation results for signal corrupted with carnoise in Spectral Subtraction and Kalman filtering methods.

The input and output waveforms of the Spectral Subtraction method is shown in the following simulation results.The spectrogram representation of noisy signal and enhanced signal is plotted below. All the waveforms and spectrograms are related to speech signal corrupted by Car noise at 5db SNR.



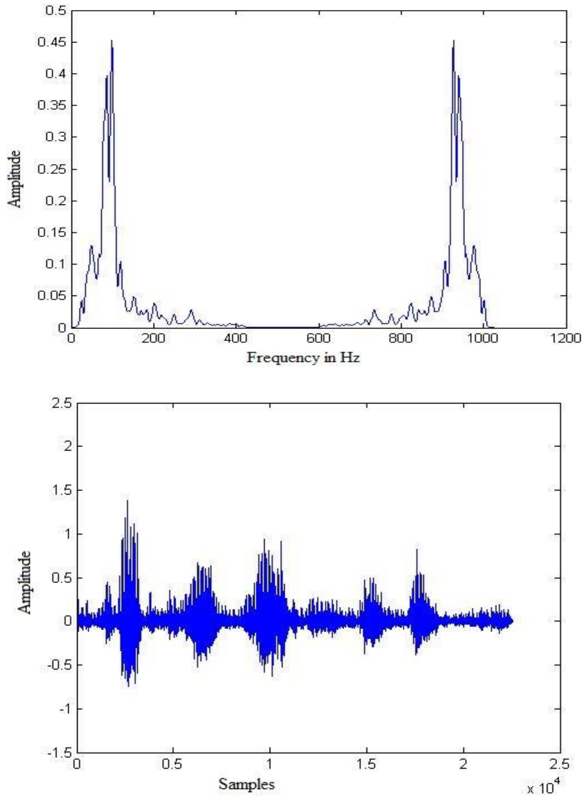


Figure6. *Simulation Results using Spectral Subtraction method -Noisy Speech, Noise Estimates, Restored Speech*

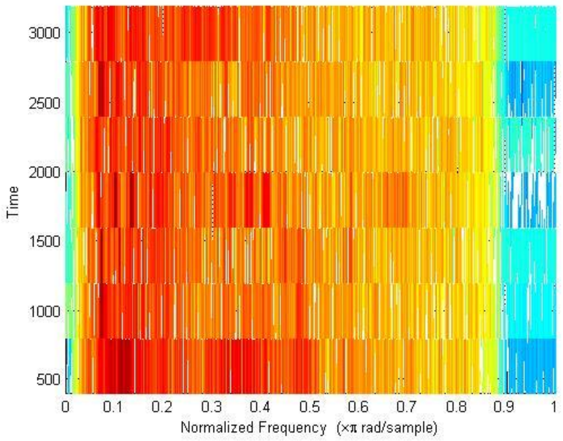


Figure7. *Spectrogram of Noisy Speech signal using Spectral Subtraction method*

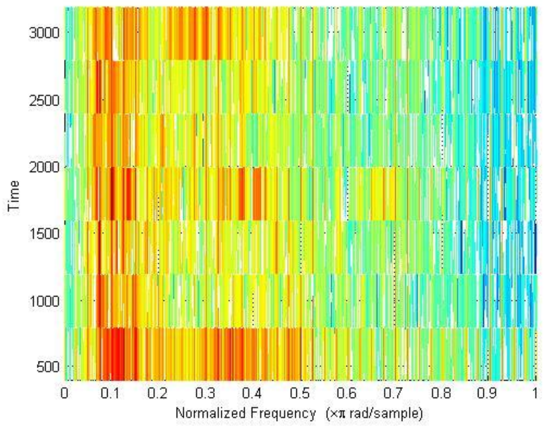


Figure8. *Spectrogram of Enhanced Speech signal using Spectral Subtraction method*

The input and output waveforms of the Kalman Filtering method is shown in the following simulation results.The spectrogram representation of noisy signal and enhanced signal is plotted below. All the waveforms and spectrograms are related to speech signal corrupted by Car noise at 5db SNR.

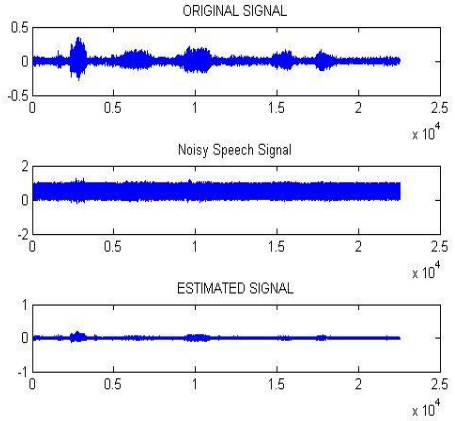


Figure9. *Simulation Results using Kalman filter method - Original signal, Noisy Speech, Restored Speech*

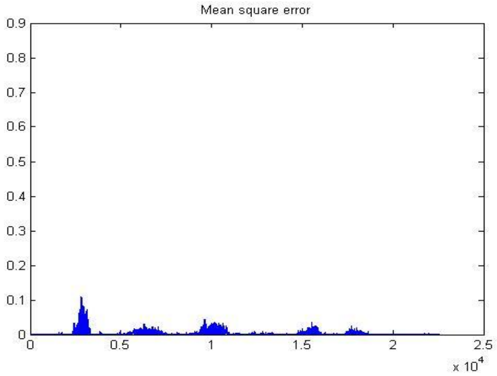


Figure10. *Estimated mean square error*

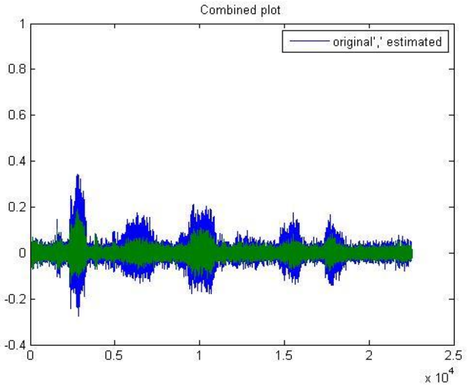


Figure11. *Combined plot of original and estimated signals*

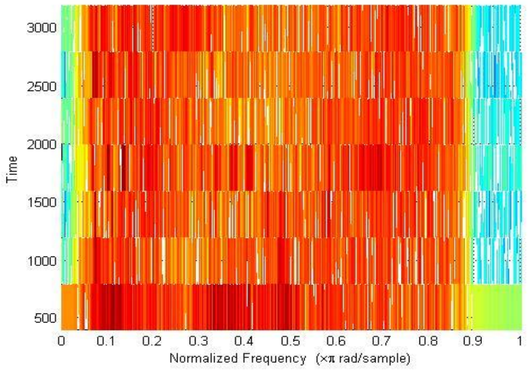


Figure12. *Spectrogram of input original signal using Kalman filter method*

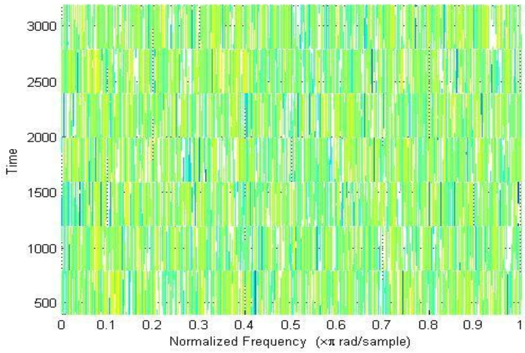


Figure13. *Spectrogram of Noisy signal using Kalman filter method*

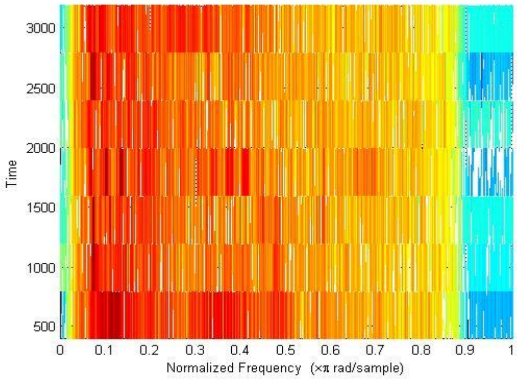


Figure14. *Spectrogram of output enhanced signal using Kalman filter method*

Table1. *Comparing SNR of Spectral Subtraction and Kalman Filter methods of Speech Enhancement*

Types of noise	Signal-to-noise ratio(SNR in dB) using Spectral Subtraction	Signal-to-noise ratio(SNR in dB) using Kalman filter
Airport noise	8.6025	3.1482
Babble noise	8.3255	4.4355
Car noise	8.1904	4.7489
Exhibition noise	9.1980	3.5338
Restaurant noise	8.3999	4.5431
Street noise	8.8574	3.0056
Train noise	8.6876	3.3507

7. Conclusion

Thus comparing the SNR values obtained in both the above methods, Spectral subtraction method provides high SNR than Kalman Filter method. But the naturalness and intelligibility is high in Kalman filter based Speech Enhancement method. The performance spectral subtraction method at high SNR is very efficient and degrades as SNR reduces. Thus we can use the Kalman filter based Speech Enhancement at low SNR as it provides best performance in low SNR conditions

Acknowledgement

I wish to express my thanks to Dr. P A Vijaya, staff members and my friends for their suggestion and encouragement.

References

[1] Mrs. Ganga K Moorthy and Mrs. MeenaPriyaDharshini,” High Throughput and Efficient VLSI Architecture for Speech Enhancement”, IRG-IJEEE, vol .1,issue.1,2015.

[2] Berouti, M. and Schwartz, R. and Makhoul, J.,” Enhancement of speech corrupted by acoustic noise” Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP ’79., vol. 4., pp. 208– 211, doi= 10.1109/ICASSP.1979.1170788,1979

[3] Yi Z. and YunxinZ.,” Real and imaginary modulation spectral subtraction for speech enhancement,” in Speech Communication, vol.55, no.4, pp.509–522. doi=http://dx.doi.org /10.1016/j.specom.2012.09.005,url=”http://www.sciencedirect.com/science/article/pii/S0167639 312001276”,ISSN=0167-6393,2013

[4] Ephraim, Y. and Malah, D., "Speech enhancement using a minimum mean square error short-time spectral amplitude estimator," in Acoustics, Speech and Signal Processing, IEEE Transactions on, pp. 11091121, doi=10.1109/TASSP.1984.1164453, ISSN=0096-3518,1984

[5] Peng Dai and Ing Yann, ," Robust speech recognition by using spectral subtraction with noise peak shifting" in Signal Processing, IET, Volume: 7,pp 684-692,2013

[6] Ke Hu and DeLiang Wang," Unvoiced Speech Segregation From Non-speech Interference via CASA and Spectral Subtraction", Audio, Speech and Language Processing, IEEE Transactions on Volume:19, pp 1600-1609,2011

[7] Md. Kamrul Hasan, SayeefSalahuddin and M. Rezwan Khan," A Modified A Priori SNR for Speech Enhancement Using Spectral Subtraction", Signal Processing Letters, IEEE, Volume:11, pp 450-453,2004

[8] Guoshen Y. and Mallat, S. and Bacry, E.,"AudioDenoising by Time-Frequency Block Thresholding," in Signal Processing, IEEE Transactions on Volume: 56, pp 1830-1839, 2008

[9] M.S. Grewal and A.P. Andrews, Kalman Filtering Theory and Practice Using MATLAB 2nd edition, John Wiley & Sons, Canada, 2001

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

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

AUTHORS' BIOGRAPHY

Deepa Srinivasa, Completed her B.E. from BNMIT, Bengaluru, Karnataka, India and. M.Tech in VLSI and Embedded System from the Dept. of ECE Engg., BNMIT, Bengaluru, Karnataka, India. This paper is based on the Project work carried out under the guidance of Dr. P. A.Vijaya.

Dr. P.A.Vijaya, Completed her B.E. from MCE, Hassan, Karnataka, India and M.Tech and Ph.D. from IISC, Bengaluru, India. She worked in MCE, Hassan, Karnataka, India for about 27 years. Presently she is professor and Head, in the Dept. of ECE Engg. BNMIT, Bengaluru, India from 2013. Three students have obtained Ph.D. under her guidance and four more are doing Ph.D. Her research areas are pattern recognition, Image Processing, VLSI Design, Embedded System and RTOS.