ISSN 2250-3153

GUI Based Performance Analysis of Speech Enhancement Techniques

Shishir Banchhor*, Jimish Dodia, Darshana Gowda*****

[*] Electronics & Telecommunication, Mumbai University
[**] Electronics & Telecommunication, Mumbai University
[***] Electronics & Telecommunication, Mumbai University

Abstract- Speech, being a fundamental way of communication, has been embedded in various applications. In many unavoidable situations, we are rendered helpless trying to deduce the intelligibility of the speech and this is where Speech enhancing technique i.e. removal of unwanted background noise, comes into picture. In this paper, an attempt has been made towards studying Speech Enhancement techniques such as Spectral Subtraction, Minimum Mean Square Error (MMSE), Kalman and Wiener filter. Based on our observations and analysis of various performance parameters, we conclude which of the methods is most suitable for speech enhancement. The implementation of the code for various filters is done using Graphic User Interface on MATLAB.

Index Terms- Kalman filter, MMSE, SNR, Spectral subtraction, Speech enhancement, Wiener filter.

I. Introduction

Speech is a fundamental and common medium, hence important for us, to communicate. Advancements in technology have made way for many more speech oriented applications like cellular voice calls, VoIP, teleconferencing systems, speech recognition, and hearing aids, etc. In many cases, these systems work well in nearly noise-free conditions, but their performance deteriorates rapidly in noisy conditions. In general, there exists a need to increase the reliability of these systems in noisy environments. Therefore, improvement in existing pre-processing algorithms or introducing entire new class of algorithm for speech enhancement is the basic objective of research community. In speech enhancement, the goal is to improve the quality of degraded speech. Speech enhancement algorithms are noise suppression techniques, using the knowledge from the field of hearing science, that mitigate the effect of the corrupting background noise, and hence improve the perceived speech quality and speech intelligibility.

The problem of improving performance of speech communication systems in noisy environments has been a challenging area for research for more than three decades now. For making speech coders robust to noise, speech enhancement focuses on improving the quality and intelligibility of speech damaged by noise. But if speech is over processed, intelligibility of noisy speech may decrease rapidly hindering effective improvement in enhancing the performance of other speech applications such as speech coding/compression and speech recognition [1-4]. This project presents an overview of different speech enhancement. Using tables and graphs we compare and review the techniques.

II. TYPES OF SPEECH ENHANCEMENT

Speech Enhancement methods can be classified in many ways. A standard algorithm alone is insufficient for all the types of noise present in the surrounding. Therefore, speech enhancement algorithms are generated based on the applications. The assumptions and other constraints are determined completely according to the application and the environment. The performance of speech enhancement algorithm is limited by factors like number of noise sources available, limitations in time variations and available a priori data. Model based constraints like restriction of algorithm to uncorrelated noise also play an important role. The Speech Enhancement systems can be classified based on number of channels used i.e. single or multiple, domain of processing i.e. time or frequency and the type of algorithms.

Speech processing strategies can be broadly divided into single and multichannel enhancement techniques [5-6].

A. Single channel speech enhancement

In applications like hearing aids and mobile phones, where an alternate channel is unavailable, single channel enhancement is used. Single channel enhancement techniques are very easy to build and are less expensive when compared to its multichannel counterpart. A disadvantage of this method is that the signal cannot be pre-processed since there is no noise reference signal present. Hence, single channel enhancement systems make use of a different statistical approach to the unwanted noise [7]. Most of the single channel enhancement algorithms work on the assumption that noise is stationary in nature.

B. Multi-channel speech enhancement

This method of speech enhancement is better than the previously mentioned method due to presence of reference channel. This channel helps us eliminate noise in an effective way. Phase alignment is performed in one of the channels to reject the undesired noise components. The limitations of single channel enhancement method are overcome and non-stationary noise sources can be addressed in this method of speech enhancement. Multi-channel systems are more complex. This method is implemented in adaptive noise cancellation devices where a reference of contaminating noise is present in an auxiliary channel. This reference input is then filtered. Noise is hence subtracted from noisy speech input by following an adaptive algorithm. The main drawbacks are fabrication cost and complexity of the algorithm. This technique is currently in application in mobile phones for Noise Reduction using Dual Mic.

III. BASIC BLOCK DIAGRAM

The basic block diagram for speech enhancement is as shown below in Fig. 1.

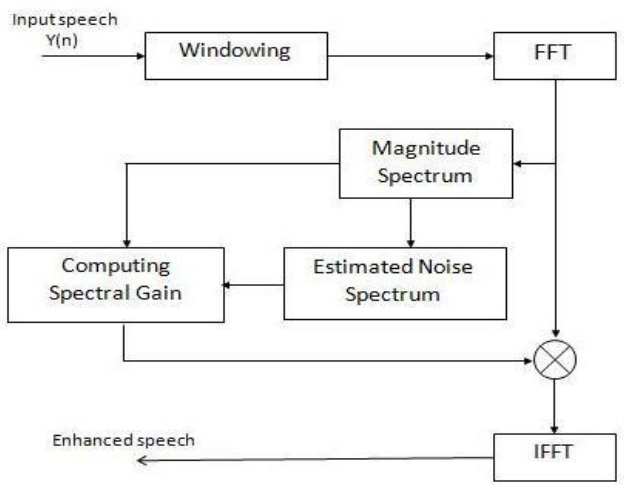

Figure 1: Basic block diagram

The noisy input signal is sent through the analysis window. Here, a few samples of the signal are selected at a time as the signal is continuous and big and cannot be processed in one go. Fast Fourier Transform is applied to convert the signal from time domain to frequency domain. The magnitude of noise and noisy speech are compared and noise is subtracted from the affected speech. The enhanced speech received is in frequency domain and hence requires to be converted back to frequency domain. This is done by taking Inverse Fourier Transform. Overlap and add method is applied to the recovered enhanced signal so as to compensate for the windowing method applied in the beginning.

IV. TYPES OF FILTERS

A. Spectral Subtraction method

The Spectral Subtraction method is one of the most widely used methods of speech enhancement. This is because of the simplicity of implementation and lower computational load. The power spectrum or magnitude Spectrum of the speech signal can be easily restored using approach. The estimated noise spectrum is subtracted from the noisy speech input in order to obtain clean speech. It reduces the effect of background noise based on the STSA estimation technique [8-9].

The effectiveness in enhancing the speech signal, degraded by noise, makes this technique a popular one. The basic principle includes subtraction of magnitude spectrum of noise from that of noisy speech. The noise spectrum is estimated during the periods when only noise is present. The noise is assumed to be uncorrelated and additive in nature. Since phase distortion is not perceived by human ear, it is kept unchanged.

The simple subtraction process perhaps comes with a price. The subtraction process needs to be done carefully in order to avoid any distortion. If too much noise is subtracted, then some speech information might be lost. On the other hand if too little is subtracted then the distortion will remain in the speech.

The block diagram given in [10] is as shown in Fig. 2.

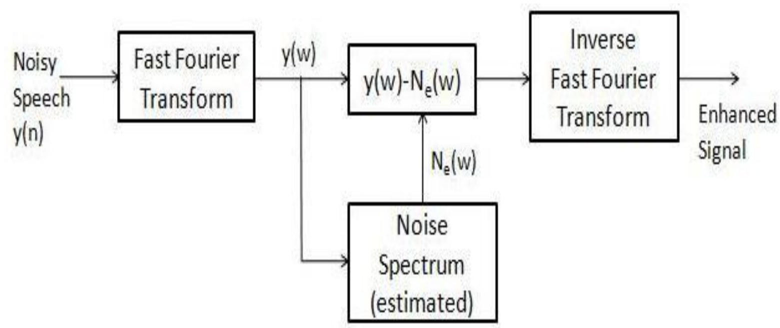

Figure 2:Spectral Subtraction

Noisy Speech, y(n), is given as an input to this filter. Windowing is done in order to take fixed number samples of the signal which is continuous in nature. In this method only the magnitude is considered. The phase part is not taken into consideration as it increases the complexity and calculations. Fourier transform is applied to the signal in order to convert the signal from time domain to frequency domain. This helps us to obtain magnitude and phase as separate values. The magnitude of estimated noise, $N_e(w)$, is subtracted from the magnitude of noisy signal and an enhanced Speech is obtained at the output of spectral modification block. Inverse Fourier transform of the enhanced speech is taken so as to obtain the signal in its time domain form. Phase of signal, in its original form, is added to the magnitude at this stage. Thus we obtain an enhanced version of the noisy Speech signal at the output end.

B. Weiner Filter

For more than two decades, speech processing has been a growing and dynamic field. This clearly indicated further growth and development. To optimize the filter we need to minimize the mean square error value of the filter. This is done by calculating the difference between the desired response and the actual filter output. Weiner filter is described as a class of optimum linear filters [11-12]. These filters involve linear estimation of a desired signal sequence from another related sequence. This technique is widely used in the field of signal processing. Weiner filter is a common as well as an adaptive filtering technique and is the solution for stationary input signals.

The filter first originated in Kalman's document (1960). In the document [13], it is described as a recursive solution for the linear filtering problems. Estimation of values through the recursive least squares was based on state-space models. Based on different application requirements, a Weiner filter is designed to enhance the signal for that very desired frequency response. But before the actual processing, the spectral properties of the original signal and noise should be known.

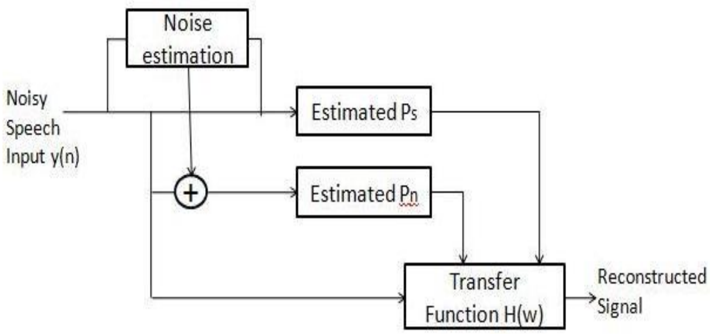

Figure 3: Weiner filter

Shown above in Fig. 3, is the block diagram of Weiner filter. In this process, mean the noisy speech is represented as $P_s(w)$ and mean of noise power is represented as $P_n(w)$, which is calculated for recovering the original signal.

C. Minimum Mean Square Error

Minimum mean square error technique can be implemented when the input SNR is known. This algorithm minimizes the estimated error due to both noise and clutter. It is an implementation of Weiner Filter. The STSA estimation problem is to estimate the modulus of each complex Fourier expansion coefficient of the speech signal from the noisy speech in that analysis frame [14]. Fourier expansion series are samples of its Fourier Transform and have a close relation [15-17]. The relation between Fourier series expansion and discrete Fourier transform enables use of FFT for efficient implementation of the algorithm.

For implementation of the MMSE STSA estimator, the a-priori-probability distribution of the Speech and noise Fourier expansion coefficients should be known. Since in real-time they are unknown, each probability distribution can be measured or a reasonable statistical model be assumed. In this paper, the Speech and the noise are neither stationary nor Ergodic processes. Thus probability distribution can be obtained by examining independent sample functions belonging to the ensemble of each process. However, due to non-stationary nature of process, probability distributions are time varying, their measurement and characterization by the above way is complicated, and the procedure seems to be impracticable.

The only disadvantage of the MMSE processor [18] is additional complexity in determining the linear estimator. Further, for large problems, the matrix inverse operation required to implement the MMSE estimator is very problematic and can slow down processing speed, like in the field of radar signalling. Another implementation of the MMSE algorithm can be developed and the data vector can be split into smaller segments to reduce processing time.

The basic block diagram of MMSE filter is shown in Fig. 4

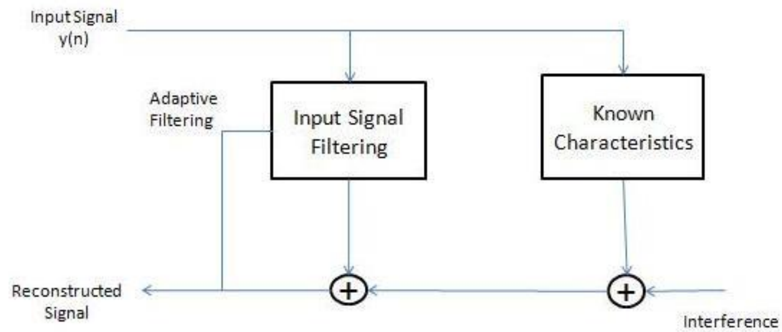

Figure 4: MMSE Filter

D. Kalman Filter

Kalman filter (KF) algorithm is based on continual implementation of the MMSE estimator [19]. It is a set of mathematical equations providing an efficient computational means to estimate the state of a process. In this way, it recursively minimizes the mean of the squared error. It uses state space model to deliver an admirable way of extracting a signal from noise. These state space models are described by an Auto-Regressive (AR) process. Hence, this fast process is able to produce practical solutions to problems which posed a significant hurdle in the Weiner filter.

The Kalman is a digital filter that automatically adjusts itself (when the input signal and interfering noise change) to give you the best possible idea of what the current state of the system is. The Kalman filter keeps updating and improving itself by trying to predict the newest value and then comparing this to the observed /estimated value. It does not directly refer to estimated iteration, but in fact it makes the best possible use of the entire history of the data to figure out what next estimation, despite the interfering noise. The covariance formed from the prediction described in [20], with a new estimate every step is repeated with every sample of data can be shown as:

$$X_k = K_k . Z_k + (1 - K_k). X_k \quad (1)$$

Where,

X_k is the current estimation,
X_{k-1} is the previous estimation,
K_k is Kalman Gain and
Z_k is measured value.

It is an optimal estimator i.e. it infers parameters of interest from ambiguous, unreliable and indistinct observations. If all noise is White Gaussian, the Kalman filter minimizes the mean square error of the estimated parameters. Given only the mean and standard deviation of noise, the Kalman filter is the best *linear* estimator. Based on a state space approach it models signal generation and the noisy and distorted observation signal is modelled by an observation equation [21-22]. The use of Kalman filter for Speech Enhancement was first introduced by Paliwal (1987).

The advantages of Kalman Filtering Technique are good results in practice due to optimality and structure. The filter is distinguished by its skill to predict the state of a model in the past, present and future, although the exact nature of the modelled system is unknown. The dynamic modelling of a system is one of the key features which distinguish the Kalman method. Because of its apparently simple and easily programmed algorithm, the Kalman filter will continue to play a very important role in GPS-based navigation systems.

The disadvantages of Kalman Filtering Technique are that it is necessary to know the initial conditions of the mean and variance state vector to start the recursive algorithm. There is no general consent over the way of determining the initial conditions.

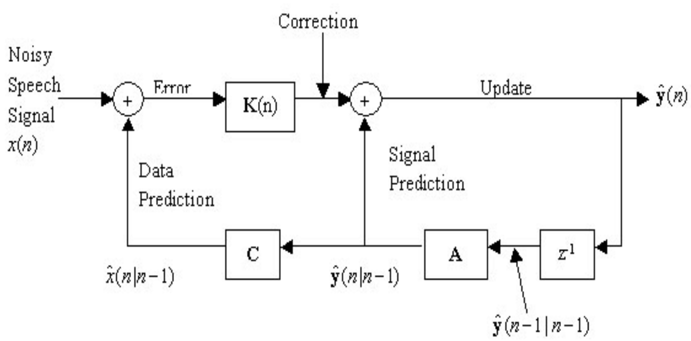

Figure 5: Kalman Filter

As shown in Fig. 5, the input Speech signal is taken and distortion of noise in the signal is found.The current output is based on the past output and current input which is solved using Yule's equation. All the parameters are represented in the form of state space matrix because it makes calculations easier. Next filter gain is calculated and noise is then removed from the noisy Speech input to get enhanced Speech signal.

V. MEASURES OF PERFORMANCE PARAMETERS

A. Signal-to-Noise Ratio (SNR)

Signal-to-noise ratio is a measure used in science and engineering that compares the level of a desired signal to the level of background noise as given in [3].

Signal-to-noise ratio is sometimes used to refer to the ratio of useful information to irrelevant data in a conversation or exchange. Signal-to- noise ratio is defined as the power ratio between a signal (meaningful information) and the background noise (unwanted signal). It is measured in dB and is represented as SNR or S/N. The Signal to Noise ratio can be represented as:

$$\mathrm{SNR} = 10 \times \log_{10} \frac{\mathrm{mean}\,(\mathrm{Input}^2)}{\mathrm{mean}\,(\mathrm{Input}^2 - \mathrm{Enhanced}^2)}$$

B. Mean Square Error (MSE)

The Mean Squared Error (MSE) measured in dB is one of the ways to determine the difference between values implied by an estimator and the true values of the quantity being estimated. MSE corresponds to the expected value of the squared error loss. MSE measures the average of the squares of the "errors". The difference between

the values implied by estimator and the quantity which is to be estimated gives us the error value of the signal. The randomness of signal or the inability of the system to produce accurate values at the output gives rise to error in the signal. The Mean Square Error can be represented as:

$$\mathrm{MSE} = \frac{1}{\mathrm{length}\,(\mathrm{Input})} \times \sum (\mathrm{Enhanced} - \mathrm{Input})^2$$

C. Normalized Root Mean Square Error (NRMSE)

The Root Mean Square Error (RMSE) also known as Root-Mean-Square Deviation (RMSD) is a frequently used measure to find the normalised value of difference between the values predicted by an estimator and the values that are actually observed. The RMSE provides us with an average of the magnitude of errors calculated over a period of time. RMSE is a good measure of accuracy of a system. It can be calculated by using the generalized equation given below:

$$\mathrm{NRMSE} = \frac{\sqrt{\mathrm{mean}\,[(\mathrm{Input} - \mathrm{Enhanced})^2]}}{\sqrt{\mathrm{mean}\,\{[\mathrm{Input} - \mathrm{mean}\,(\mathrm{Input})]^2\}}}$$

D. Peak Signal-to-Noise Ratio(PSNR)

Peak Signal-to-Noise Ratio, often abbreviated PSNR, is used to represent the ratio of maximum possible value of a signal to the power of the undesired noise signal. Because many signals have a very wide dynamic range, PSNR is usually expressed in terms of the logarithmic decibel scale. PSNR is generally used to measure the quality of reconstruction of lossy compression codes.

International Journal of Scientific and Research Publications, Volume 3, Issue 9, September 2013
ISSN 2250-3153

In case of image compression, for example, the signal is the original image and the noise is the error introduced due to compression. PSNR is basically an approximate value to the human perception of the quality of reconstruction of speech signal. A higher PSNR indicates that the reconstructed signal has a better quality but the reverse may be true in case of some applications. PSNR can be calculated by:

$$PSNR = 10 \times \log_{10} \frac{length \times \max[Input^2]}{Input^2 - Enhanced^2}$$

E. Distortion (AAD)

Distortion is basically the alteration or warping of the original shape or other characteristics of an object, image, sound or waveform. Distortion is usually unwanted, and often efforts are made to reduce it as much as possible. The addition of noise or other outside signals (hum, interference) may lead to distortion of signal. In this project we use the parameter AAD to measure the distortion in the given Speech signal. AAD is generally equated as shown below:

$$AAD = \frac{1}{length(Input)} \times \sum(Enhanced - Input)$$

VI. EXPERIMENTAL RESULTS

The above mentioned techniques of speech enhancement were applied to the noisy speech input and the performance parameters were evaluated as below.

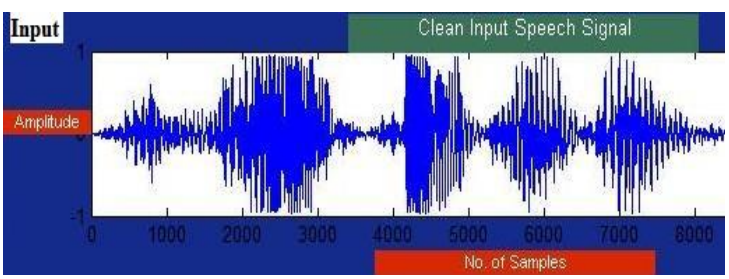

Figure 6: Clean Speech

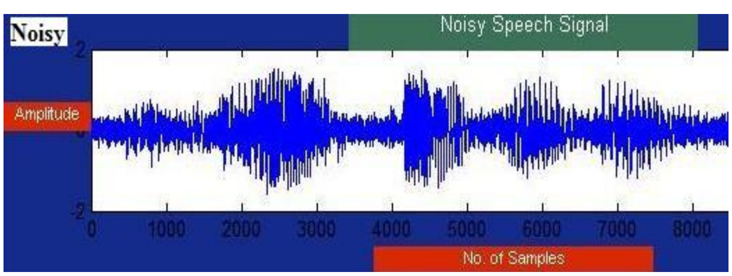

Figure 7: Noisy Speech

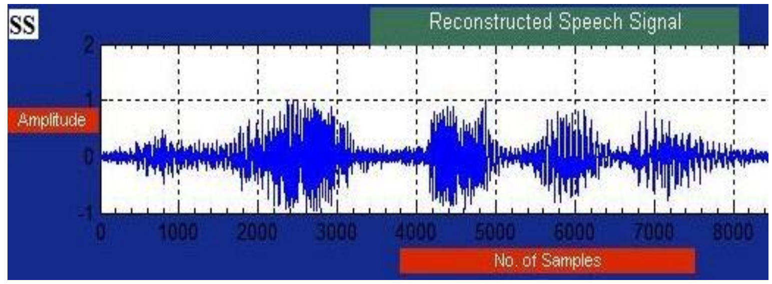

Figure 8: Spectral Subtraction Output

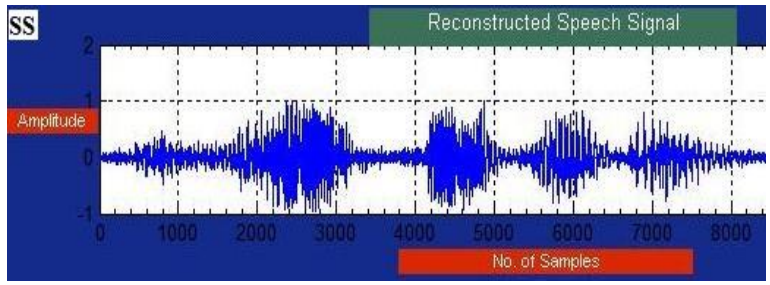

Figure 9: Weiner Filter Output

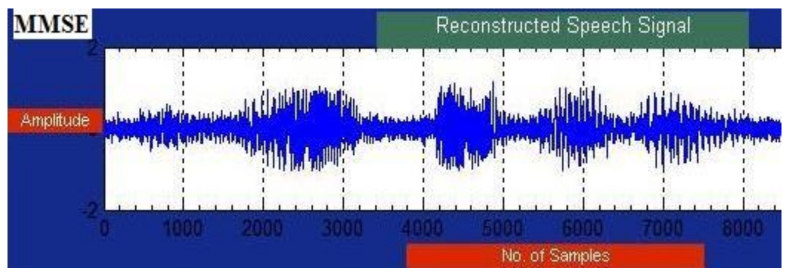

Figure 10: MMSE Filter Output

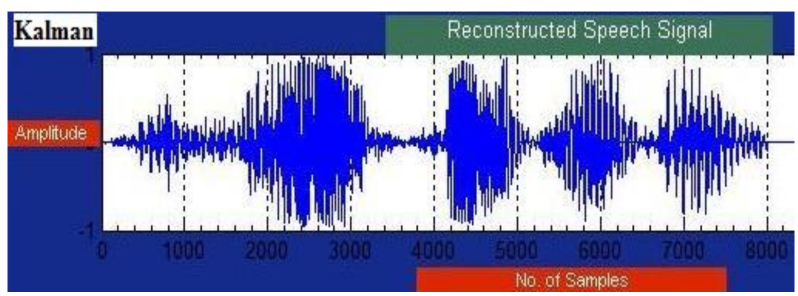

Figure 11: Kalman Filter Output

VII. CONCLUSION

The technique most suitable for speech enhancement is the one which provides robustness to environmental noise contributing factors and robustness to acoustical inputs.

Table 1: Parameters for input SNR of 2dB

PARAMETERS	Weiner	SS	MMSE	Kalman
Output SNR	2	2	4	3
PSNR	25	22	20	23
MSE	5.661e-007	4.405e-006	2.998e-006	5.300e-006
NRMSE	3.306e-001	5.246e-001	6.477e-001	4.621e-001
AAD	8.165e-006	2.277e005	1.879e-005	2.498e-005

Table 2: Parameters for input SNR of 5dB

PARAMETERS	Weiner	SS	MMSE	Kalman
Output SNR	4	3	5	4
PSNR	29	24	20	24
MSE	1.306e-007	2.168e-006	1.574e-006	2.853e-006
NRMSE	2.173e-001	3.829e-001	5.594e-001	3.657e-001
AAD	3.921e-006	1.598e-005	1.361e-005	1.833e-005

Table 3: Parameters for input SNR of 10dB

PARAMETERS	Weiner	SS	MMSE	Kalman
Output SNR	6	4	7	4
PSNR	32	24	20	25
MSE	9.301e-011	2.676e-008	5.544e-008	3.590s-007
NRMSE	1.186e-001	2.863e-001	4.560e-001	2.662e-001
AAD	2.391e-007	1.363e-006	1.761e-005	2.407e-006

Table 4: Parameters for input SNR of 15dB

PARAMETERS	Weiner	SS	MMSE	Kalman
Output SNR	8	4	8	5
PSNR	35	25	21	26
MSE	1.348e-010	1.281e-007	7.085e-007	1.435e-008
NRMSE	7.726e-002	2.365e-001	3.948e-001	2.144e-001
AAD	1.260e-007	3.884e-006	1.663e-005	7.458e-006

Table 5: Parameters for input SNR of 20dB

PARAMETERS	Weiner	SS	MMSE	Kalman
Output SNR	9	4	9	5
PSNR	36	26	21	26
MSE	1.321e-012	3.942e-009	3.696e-007	7.397e-011
NRMSE	6.500e-002	2.186e-001	3.663e-001	1.965e-001
AAD	1.247e-008	6.814e-007	6.597e-006	9.333e-008

In this paper, we have reviewed the methodologies and principles of various techniques and presented the analysis in GUI MATLAB. Based on the performance parameters the following points have been concluded:

- Wiener Filter follows statistical approach and could be tuned to provide optimal performance.
- Kalman has the ability to estimate accurately by using autoregressive (AR) modeling and is suitable for real-time applications.
- Spectral Subtraction is a real time filter which is relatively easy to implement for stationary noise.
- MMSE provides best values for the most parameters under given conditions and hence is most suitable technique for speech enhancement.

A graphical representation for comparison of the above mentioned techniques based on SNR and PSNR values is as below:

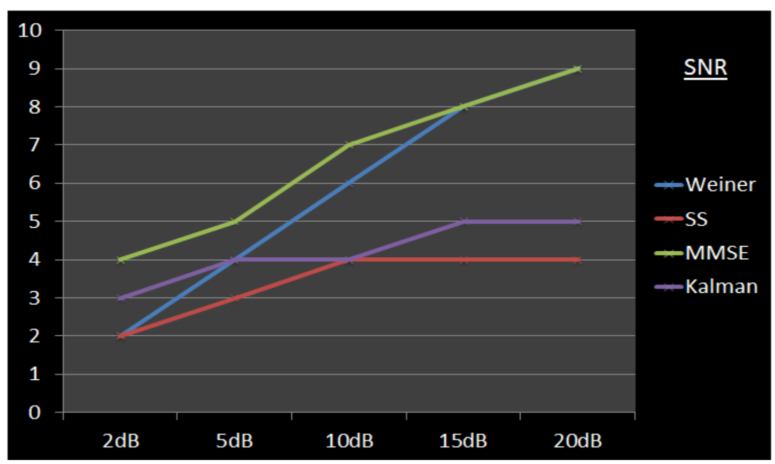

Figure 12: SNR comparison

The above graph provides a comparison between input SNR for each technique and their respective output SNR. The signal to noise ratio for MMSE is more than all filters for any value of input SNR whereas that of Spectral subtraction is the least for all values of inputs SNR.

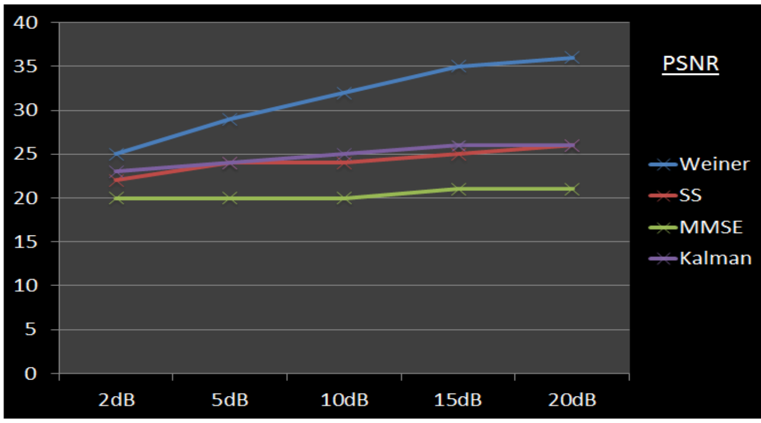

Figure 13: PSNR comparison

The graph given shows the value of peak Signal-to-noise ratio for all speech enhancement techniques. The value of PSNR is greatest for Wiener filter for all values of input SNR. Whereas MMSE filter has the least Peak Signal to Noise Ratio.

VIII. APPLICATIONS AND FUTURE SCOPE

- Cell phone call quality enhancement
- Pay phones in a noisy environment
- Air-ground communication systems
- Teleconferencing systems
- Hearing aids

IX. CONCLUSION

From the graphs we can clearly see and eventually conclude that MMSE filter provides the best SNR ratio which a major requirement as we cannot accept a filter degrading the original speech. It also filters out the noise and renders us with clean speech in a more flat line Peak SNR value in a varying loudness of speech. We tried out playing the reconstructed signal after processing the noisy speech using the four filters and MMSE definitely sounded better than the rest.

REFERENCES

[1] Yariv Ephraim and Israel Cohen, “Recent Advancements in Speech Enhancement”, *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol.32, no.5, pp. 1109- 1121, March 9, 2004.
[2] T. Lalith Kumar and Soudara Rajan, “Speech Enhancement Using Adaptive Filters”, *VSRD-IJEECE*, Vol. 2 (2), 2012, 92-99, 1997.
[3] http://en.wikipedia.org/wiki
[4] http://www.mathworks.com/support
[5] Javier Ortega-Garca and Joaqun GonzlezRodrguez, “Overview of Speech Enhancement Techniques for Automatic Speaker Recognition”, CICYT under Project TIC94-0030., 1999.
[6] Prof. MilindNemade, Prof. Satish K. Shah, “Performance Comparison of Single Channel Speech Enhancement Techniques for Personal Communication”, for *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 1, Issue 1, March 2013.
[7] Yao Ren and Michael T. Johnson, “Improved SNR Estimator for Speech Enhancement”, *ICASSP-IEEE*, 2008.
[8] Saeed V. Vaseghi, “Spectral Subtraction”, *Advanced Digital Signal Processing and Noise Reduction, Second Edition*, 1996.
[9] MukulBhatnagar, “A modified spectral subtraction method combined with perceptual weighting for speech enhancement”, *Thesis presented to The University of Texas*, August 2002.
[10] Boll S. F., “Suppression of acoustic noise in speech using Spectral Subtraction”, *IEEE transaction on Acoustic Speech, Signal Processing*, Vol. ASSP-27, pp. 113-120, 1979.
[11] M. A. Abd E-Fattah, M. I. Dessouky, S. M. Diab and F. E. Abd El-Samie, “Speech Enhancement using Adaptive Weiner Filtering Approach”, *Progress in Electromagnetic research M*, Vol. 4, pp. 167-184, 2008.
[12] Bram Cornelis, Marc Moonen and Jan Wouters, “Performance Analysis of Multichannel Wiener & Filter-Based Noise Reduction in Hearing Aids Under Second Order Statistics Estimation Errors”, *IEEE Transactions on Audio, Speech and Language Processing*, VOL. 19, No. 5, July 2011, 2011.
[13] Kalman, R. E. 1960. “A New Approach to Linear Filtering and Prediction Problems,” Transaction of the ASME—*Journal of Basic Engineering*, pp. 35-45, March 1960.
[14] Dr. Bin Chen, “Speech Enhancement using a Laplacian based MMSE Estimator of the magnitude spectrum”, 2005.
[15] Ephraim, Y., Malah D., “Speech enhancement using a Minimum mean square error short-time spectral amplitude estimator,” *Acoustics, Speech and Signal Processing, IEEE Transactions on* , vol.32, no.6, pp. 1109-1121, Dec 1984.
[16] Bin Chen, Philipos C. Loizou, “A Laplacian based MMSE estimator for speech enhancement”, December 20, 2006.
[17] Peter Kabal, “Minimum Mean-Square Error Filtering: Autocorrelation/Covariance, General Delays, and Multirate Systems”, April 2011.
[18] Deepa, “Dual channel speech enhancement using Hadamard LMS algorithm with DCT pre-processing technique”, *International Journal of Engineering Science and Technology*, Vol. 2(9), 2010.
[19] Ning Ma, Martin Bouchard and Rafik A. Goubran, “Perceptual Kalman filtering for speech enhancement in colored noise”, 2004
[20] Marcel Gabrea, “Adaptive Kalman filtering-based speech enhancement algorithm*”, Ecole de TechnologieSuperieure*, Vol. 3, Issue 7, 2004.
[21] Sandeep and K. Samprasad, “A Robust Speech Enhancement by using Adaptive Kalman Filtering Algorithm*”, International Journal of Engineering Research and Applications (IJERA)* Vol. 2, Issue 4, August 2012.

AUTHORS

First Author – Shishir Banchhor, K.J. Somaiya (MU), Mumbai. Email – shishir_skb@yahoo.co.in

Second Author – Jimish Dodia, K.J. Somaiya (MU), Mumbai. Email – jimishdodia@gmail.com

Third Author – Darshana Gowda, K.J. Somaiya (MU), Mumbai. Email – gowda.darshana@gmail.com