International Journal of Emerging Engineering Research and Technology
Volume 7, Issue 8, 2019, PP 1-7
ISSN 2349-4395 (Print) & ISSN 2349-4409 (Online)

Enhancement of Speech using Kalman Filter with Phase and Magnitude Spectrum Compensation

Ramy Abdul Mawjood Mohammed[1], B. Bhaskara Rao[2*], Ch. D Uma Sankar[3]

[1]*M.Tech Student, Dept. of ECE, University College of Engineering and Technology, Acharya Nagarjuna University, Guntur, AP, India*

[2,3]*Assistant Professor, Dept. of ECE, University College of Engineering and Technology, Acharya Nagarjuna University, Guntur, AP, India*

***Corresponding Author:** B. Bhaskara Rao, Assistant Professor, Dept. of ECE, University College of Engineering and Technology, Acharya Nagarjuna University, Guntur, AP, India, Email: b.bhaskararao.ece@gmail.com*

ABSTRACT

During conversation, both hearing and speaking adapt to the background noise in a noisy environment. It is therefore possible to have a conversation in quite disturbing background noise environments. However, when the conversation takes place over the telephone disturbances are more annoying. The occurrence of disturbance is a problem since the brain will not get the extra visual and other background information when interpreting the speech. The speech signal transmitted to the other party is picked up by a microphone connected to the telephone. The microphone signal contains both speech and noise at some ratio (Speech to Noise Ratio, SNR) depending on. To increase the SNR and to allow for the listener to grasp the speech clearly, a speech enhancement method should be applied. In this thesis enhancement of speech using kalman filter with phase and magnitude spectrum compensation is proposed. This method can be used in the systems which need to cancel the background noises such as speech recognition, speech communication, etc., and it can further improve the speech quality and intelligibility. This gives better PSEQ scores and SNR than other algorithms.

INTRODUCTION

Communication via speech is one of the essential functions of human beings. Humans possess varied ways to retrieve information from the outside world or to communicate with each other and the three most important sources of information are speech, images and written text. For many purposes, speech stands out as the most efficient and convenient one. Speech not only conveys linguistic contents, but also communicates other useful information like the mood of the speaker. When speaker and listener are near to each other in a quiet environment, communication is generally easy and accurate. However, at a distance or in a noisy background, the listener's ability to understand suffers (Douglas 2005). In many speech communication systems the quality and intelligibility of speech is of greatest importance for ease and accuracy of information exchange. The speech processing systems used to communicate or store speech is usually designed for a noise free environment but in a real-world environment, the presence of background interference in the form of additive background and channel noise drastically degrades the performance of these systems, causing inaccurate information exchange and listener fatigue. Over the year, researchers have developed a number of methods to enhance speech from the degraded speech. Yet, due to complexities of the speech signal, restoring the desired speech signal from the mixture of speech and background noise still poses a considerable challenge in speech processing and communication system research. Speech enhancement algorithms attempt to improve the performance of communication systems when their input or output signals are corrupted by noise. The presence of background noise causes the quality and intelligibility of speech to degrade. Here, the quality of speech refers how a speaker conveys an utterance and includes such attributes like naturalness and speaker recognisability. Intelligibility is concerned with what the speaker had said, that is, the meaning or information content behind the words (Hu and Loizou 2007). Therefore, a noisy environment reduces the speaker and listeners ability to communicate. To reduce the impact of this problem speech enhancement can be performed. It is usually difficult to reduce noise without distorting speech and thus, the performance of speech enhancement systems is limited by the trade-off between speech distortion and noise reduction (Boll 1979). Efforts to achieve higher quality and/or intelligibility of noisy speech may effectively end up improving performance of other speech applications such as speech coding/compression and speech recognition, hearing aids, voice communication systems and so on. The goal of speech enhancement varies according to specific applications, such as to reduce listener fatigue, to boost the overall speech quality, to increase intelligibility and to improve the performance of the voice communication device. Hence speech enhancement is necessary to avoid the degradation of speech quality and to overcome the limitations of human auditory systems.

LITERATURE REVIEW

Most of the existing speech enhancement algorithms only change the magnitude spectrum of the noisy speech. The modified magnitude then recombined with the unchanged phase spectrum to produce a modified complex spectrum, which is the estimated clean speech spectrum. These algorithms are called magnitude spectrum based methods. Boll proposed the method of spectral subtraction (SSUB) in 1979. Its basic principle is to subtract the magnitude spectrum of the noise from the noisy speech magnitude spectrum, and obtain the estimate of the clean signal magnitude spectrum, but 4 the phase spectrum is unchanged [1]. The MMSE estimator, which is presented by Ephraim and Malah in 1984. Its main idea is to minimize the mean-squared error (MSE) between the clean and estimated (magnitude or power) spectra [2]. The reason for ignoring the phase impact is that the phase spectrum has been found to have less perceptual effect at significantly higher signal to noise ratio (SNR) levels[3].But recently, it is found that the phase spectrum may be useful in speech processing applications[4]. Kamil Wójcicki et al. Proposed the speech enhancement method of phase spectrum compensation (PSC) in 2008[5]. Wiener filtering in the time domain or the frequency domain is a commonly used noise reduction technique for single-channel and multichannel signals [6], e.g., in speech enhancement applications. The standard Wiener filter minimizes the mean-square error between the filter output signal and the speech component in one of the microphone signals. Hence, the error signal typically consists of a term related to noise reduction and a term related to speech distortion. Whereas the standard Wiener filter assigns equal importance to both terms, a generalized version, the so-called speech-distortion weighted Wiener filter, provides a trade-off between noise reduction and speech distortion [7]. In, it has been proved that the output signal-to-noise ratio (SNR) after noise reduction with the single-channel Wiener filter is always larger than or equal to the input SNR, for any filter length and for all possible speech and noise correlation matrices. However, the proof in [8] is quite involved, requiring the generalized eigenvalue decomposition of the speech and the noise correlation matrices and using inductive reasoning. In the field of speech enhancement, we are interested in the reduction of noise from noise-corrupted speech in order to improve its intelligibility and quality. Various methods have been investigated in the literature for performing speech enhancement. These can be grouped into spectral subtraction [9], MMSE estimation [10], Wiener filtering (linear MMSE), Kalman filtering, and subspace methods. Several of these methods employ the analysis-modification-synthesis (AMS) framework. In the present work, a procedure that employs a noise estimate to compensate the phase and magnitude spectrum for additive noise distortion using an objective speech quality measure and spectrogram analysis are formulated and the proposed method compares favourably to other popular speech enhancement techniques.

ENHANCEMENT TECHNIQUES

Broadly the enhancement techniques can be classified as single channel and dual channel or multi-channel enhancement techniques. Single channel enhancement techniques apply to situations in which only one acquisition channel is available. In dual channel enhancement techniques, a reference signal for the noise is available and hence adaptive noise cancellation technique can be applied. Multi-channel enhancement techniques employ microphone arrays and take advantage of availability of multiple signal inputs to our system, to make possible the use of phase alignment to reject the undesired noise components.

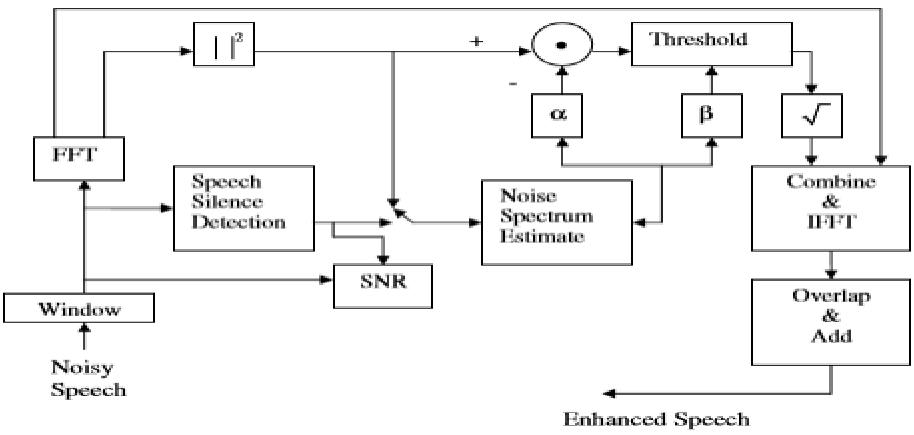
When only a single acquisition channel is available, single channel enhancement techniques are used. This may be imposed by the system used (as telephone based applications) or by the availability of the desired signal (as prerecorded applications). They are especially interesting due to the simplicity in microphone installation but the major constraint of single channel methods is that there is no reference signal for the noise available. Therefore the power spectral density of the noise has to be estimated based on the available noisy speech signal only and this is what makes it a challenging task. In all single channel enhancement techniques, we assume the available speech signal model as d(n) +s(n) = y(n). Where s(n) represents the pure speech signal, which is assumed to be a stationary signal whenever processing is done on a short time basis, d(n) is the uncorrelated additive noise and y(n) represents the degraded speech signal. There is broad set of applications for the single channel enhancement techniques.

SPECTRAL SUBTRACTION

Among all the methods spectral subtraction method is the most popular choice when one has to eliminate the background stationary noise. The greatest asset of spectral subtraction lies in its simplicity, and the fact that all that is required is an estimate of the mean noise power and that the algorithm doesn't need any signal assumptions. At the same time the latter is its great weakness. Within the framework occasional negative estimates of the power spectrum can occur. To make the estimates consistent some artificial flooring is required, which yields a very characteristic musical noise, caused by the remaining isolated patches of energy in the time-frequency representation. But with suitable additional steps taken we can minimize the musical noise. Spectral subtraction technique cannot only be used for enhancement of noisy speech for a direct listener, but also as a pre-processor noise reduction technique for digital voice processors used for speech compression, recognition and authentication. A typical example is spectral enhancement pre-processing for an LPC speech analysis-synthesis vocoder. In this section we will first discuss the basic spectral subtraction method and then a modified spectral subtraction method is studied in detail which has been suggested by Berouti. The algorithm’s strengths and limitations are identified.

The basic assumption of the method is treating the noise as uncorrelated additive noise, which is true in case of background noise. This allows us to treat the power spectrum of the degraded speech as equal to sum of the signal power spectrum and noise power spectrum, by considering the speech signal model degraded signal would represented as d(n) +s(n) = y(n). Where s(n) represents the actual speech signal, d(n) is the uncorrelated additive noise and y(n) represents the degraded speech signal. Another assumption that we make is that of assuming s(n) and d(n) to be stationary signals as processing is done on a short time basis. Above all, this method exploits the assumption that, for human perception the short time spectral amplitude is more important than the phase for intelligibility and quality.

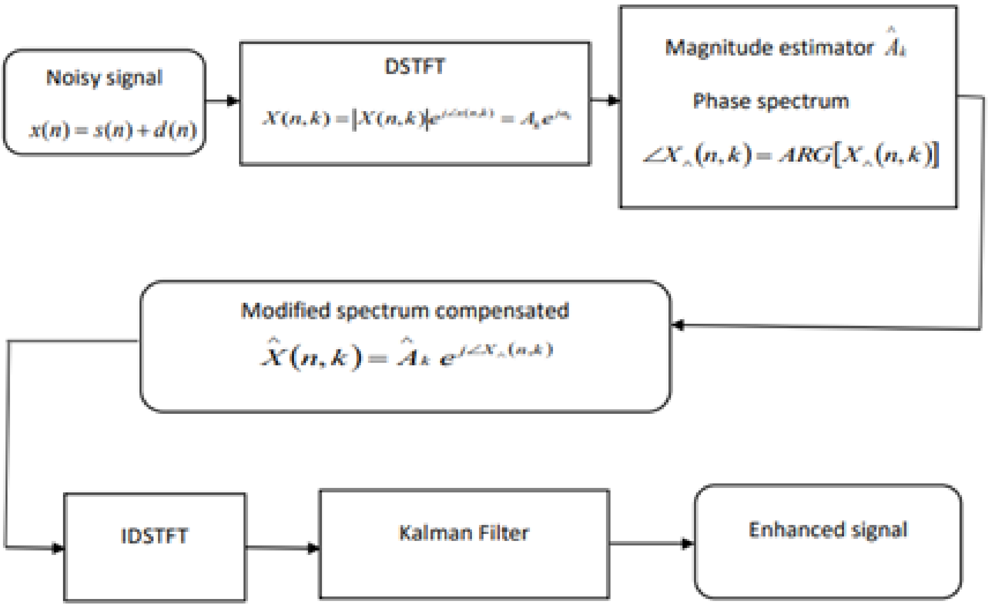
This assumption has been shown to be true, by many works, one such work is of Lim and Wang, wherein they observed that using the actual phase rather than the degraded speech phase does not improve the quality of the enhanced speech. Considering the signal model d(n) +s(n) =y(n). Taking Fourier transforms, we get D(w) +S(w)=Y(w) But as processing is carried out on a short time basis, denoting the corresponding windowed signals by respectively we have y (n) s (n) d (n) w = w+w and its short-time Fourier transform is given by



Block diagram of modified spectral subtraction method

PROPOSED METHOD

The main idea in this thesis is to recover the clean speech signal from a sample corrupted with background noise through a telephone conversation. In this speech spectrograms are estimated using Phase and Magnitude spectrum compensation to achieve the enhanced signal. After enhancing the signal using proposed method, the idea is to use the Kalman filtering as a tool to estimate the minimum mean square error (MMSE). Finally comparison is made between SSUB, MMSE using wiener filter.



DATABASE & SIMULATION RESULTS

To develop this research the recordings corrupted with noise are required. Here noisy the core test set of the NOIZEUS speech corpus which was composed of 30 phonetically balanced sentences belonging to six speakers (three males and three females) are used. There were 8 kinds of non-stationary noises at different SNRs except noisy speech with white noises. During the evaluation process, we generated a noisy speech set corrupted by additive white noise (taken from the NOISEX-92 noise database) at four SNR levels: 0 dB, 5 dB, 10 dB, and 15dB noises at different SNRs noise, babble, car, exhibition hall, restaurant, street, airport and train station noise. The IEEE database contains phonetically-balanced sentences with relatively low word context predictability. The sentences selected from this database for NOIZEUS include all phonemes in the American English language.

The sentences were originally sampled at 25 kHz and down sampled to 8 kHz. In this chapter, the results of Phase and magnitude spectrum compensation using Kalman filter algorithm are discussed, when applied to enhance a noise corrupted speech from the NOIZEUS database. To quantitatively measure the quality of the enhanced speech, and to compare it to the original clean speech, some evaluation metrics are needed. In this thesis the common objective measure SNR and another more commonly used subjective evaluator of speech is the PESQ (Perceptual Evaluation of Speech Quality) are calculated. SNR gives an indication of the amount of noise reduction, whereas PESQ gives an idea about the perceptual quality of enhanced speech. A very high segmental SNR can be rarely misleading when caused by a significant removal of spectral components of speech along with noise. In that case, the enhanced speech will have a low PESQ indicating that the high SNR was due to loss of intelligibility. Hence, both parameters complement each other, and are used together to evaluate speech enhancement algorithms.

From the above comparison table its shows that proposed method have better performance when compared to the other algorithms. And by using phase and magnitude spectrum compensation speech signal can be estimated more accurately.

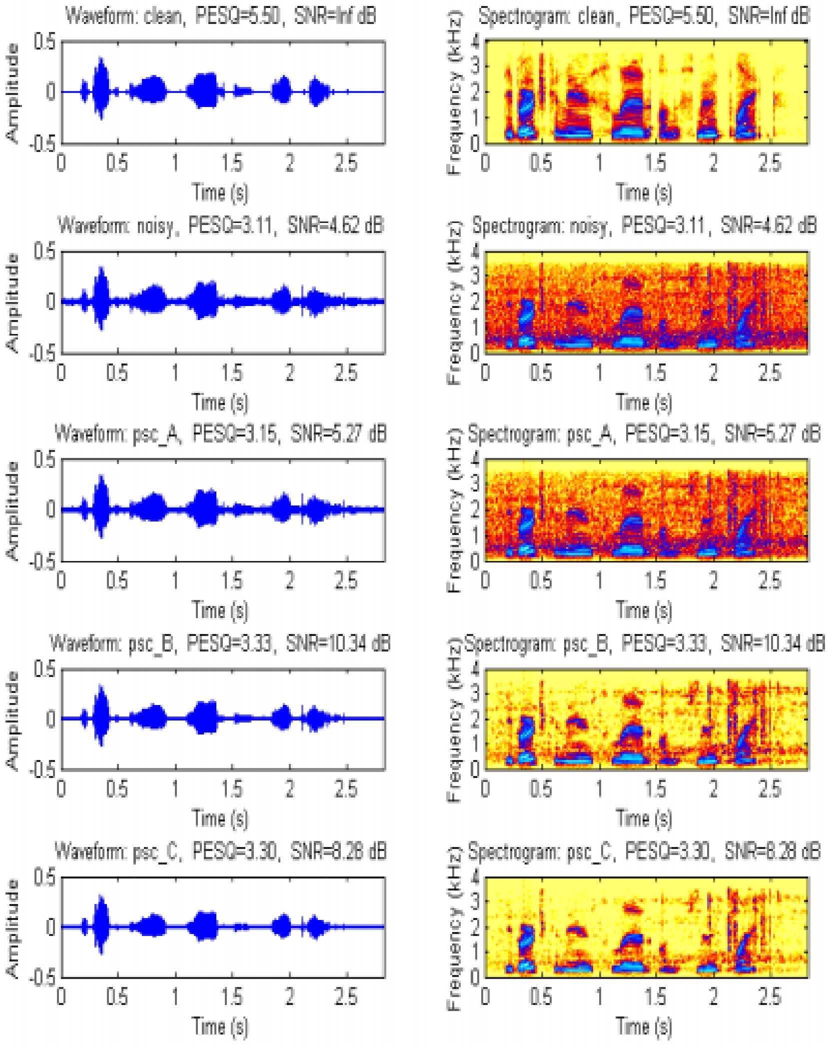
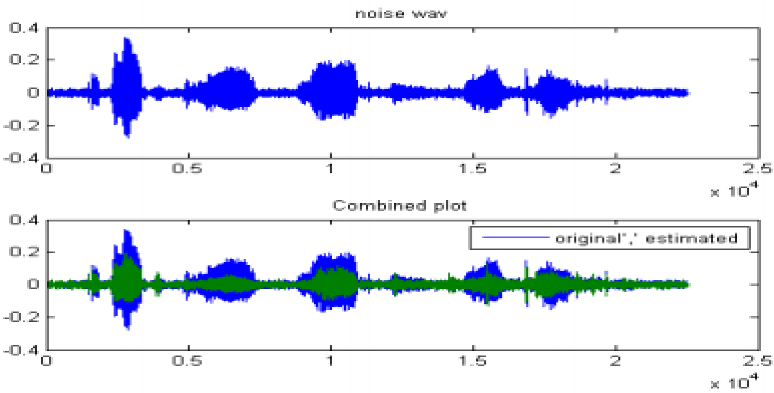


Figure. *Spectrograms of Sp01 NOIZEUS utterance at 10dB SNR.*



pe is to increase the model order. 59

CONCLUSION

Speech enhancement is a technique that improves quality of speech signal. The speech signal gets degraded because of various types of noise. The clean speech signal is necessary for applications such as speech or speaker recognition, hearing aids, mobile communication. Speech enhancement techniques are used to enhance the corrupted signal by reducing noise. It is assumed that the noise is additive. It is assumed that the noise characteristics change very slowly as compared to the signal. This is the underlying assumption in speech enhancement methods. However, it is very difficult to estimate both magnitude and phase, and using the noisy phase is an acceptable trade off in algorithms based on short-term magnitude estimation using wiener filter. In this thesis enhancement of speech using kalman filter with phase and magnitude spectrum compensation is proposed. Experiment results of the objective speech quality measure PESQ, and spectrogram analysis had showed that the proposed method achieve better speech quality than the conventional speech enhancement methods. The method can be used in the systems which need to cancel the background noises such as speech recognition, speech communication, etc., and it can further improve the speech quality and intelligibility. Which gives better PSEQ scores and SNR than other algorithms.

REFERENCES

[1] BOLL S F. Suppression of acoustic noise in speech using spectral subtraction [J]. IEEE Trans. Acoustics, Speech, Signal Processing, 1979, 27(2):113-120.

[2] Ephraim Y, Malah D. Speech enhancement using a minimum mean square error short time spectral amplitude estimator. IEEE Transactions on Acoustics, Speech, Signal Processing, 1984, 32(6): 1109-1121

[3] D.L. Wang and J.S. Lim, “The unimportance of phase in speech enhancements”, IEEE Trans. Acoust., Speech and Signal Process., Vol.30, pp. 679-681, Aug. 1982

[4] K. Paliwal, L Alsteris, “Usefulness of phase in speech processing”, Proc. IPSJ Spoken Language Processing Workshop, Gifu, Japan, pp. 1-6, 2003.

[5] Kamil Wójcicki, Mitar Milacic, Anthony Stark, James Lyons, Kuldip Paliwal. Exploiting Conjugate Symmetry of the Short-Time Fourier Spectrum for Speech Enhancement [A].IEEE Signal Process[C]. Lett, 2008, 15:461-464.

[6] N. Wiener, The Extrapolation, Interpolation, and Smoothing of Stationary Time Series With Engineering Applications. New York: Wiley, 1949.

[7] P. C. Hansen and S. H. Jensen, “FIR filter representations of Reduced rank noise reduction,” IEEE Trans. Signal Process., vol. 46, no.6, pp.1737--1741, Jun. 1998.

[8] S. Doclo and M. Moonen, “On the output SNR of the speech-distortion weighted multichannel Wiener filter,” IEEE Signal Process. Lett., vol.12, no. 12, pp. 809-- 811, Dec. 2005.

[9] Y. Hu and P. Loizou. Evaluation of objective quality measures for speech enhancement. IEEE Transactions on Speech and Audio Processing, 16:229–238, 2008.

[10] Kuldip Paliwal, Belinda Schwerin, ―Single-channel speech enhancement using spectral subtraction in the short-time modulation domain‖ in ELSEVIER 9 February 2010.

[11] P. Loizou, Speech Enhancement: Theory and Practice. New York: Taylor & Francis, 2007. 60

[12] Kotta Manohar, Preeti Rao, “Speech enhancement in nonstationary noise environments using noise properties”, Journal on Speech Communication, Elsevier, Vol. 48, Pages 96-109, 15 September 2005.

[13] Navneet Upadhyay, Abhijit Karmakar, “An Improved Multi-Band Spectral Subtraction Algorithm for Enhancing Speech in Various Noise Environments”, International Conference On Design and Manufacturing, IConDM 2013, Elsevier, Vol. 64, Pages 312-321, 13 November 2013.

[14] Yi Zhang, Yunxin Zhao, “Real and imaginary modulation spectral subtraction for speech enhancement”, Journal on Speech Communication, Elsevier, Vol. 55, Pages 509-522, 6 November 2012.

[15] Quanshen Mai, Dongzhi He, Yibin Hou, Zhangqin Huang, “Speech Enhancement Using Adapative Kalman Filterining Algorithm”, pp. 327-332, August 2011

[16] ZHANG Xiu-zhen, FU Xu-hui, WANG Xia, Improved kalman filter method for speech Improved enhancement. Computer Applications, Vol.28, enhancement. Computer pp.363-365, Dec 2008.

[17] Naritanabe, Toshiniro Furukawa, Shigeo Ts, ujji. Fast noise Suppression algorithm with Kalman filter Theory. Second International Symposium on Universal Communication, pp. 411-415, 2008.

[18] Nri Tanabe, Toshiniro Furukawa, Hideaki Matsue and Shigeo Tsujji. Kalman Filter for Robust Noise Suppression in White and Colored Noise. IEEE International Symposiumon Circiuts and Systems 2008. 1172-1175.

[19] Zhixin Chen ― Simulation of Spectral Subtraction Based Noise Reduction Method‖ International Journal of Advanced Computer Science and Applications, Vol. 2, No.8, 2011.

[20] Javier Onega-Garcia and Joaquin Gonzalez ―Overview of Speech Enhancement Techniques for Automatic Speaker Recognition in Proc. of IEEE Int. Joint Conference, 1997

[21] S. F. Boll, ―Suppression of acoustic noise in speech using spectral subtraction, in IEEE Trans. Acoust Speech and Signal Processing, vol.-2, ASSP-27, no.2, pp113-120, 1979

[22] SU Wan-Xin, HUANG Chun-mei LIU Peiwei, MA Ming-long. Application of adaptive Kalman filter in intial alignment of inertial navigation system. Journal of Chinese Inertial Technology Vol.18, No.1, pp. 44-47, Feb.2010. 61

[23] GAO Yu, ZHANG Jain-qui. Kalman filter with Wavelet-Based Unknown Measurement Noise Estimation and its Application for information Fusion. Acta Electronica Sinica, Vol.35, No.1, pp.108-111, Jan.2007.

[24] XIE Hua. Adaptive Speech Enhancement Base on Discrere Cosine Transformation in High Noise Environment. Harbin Engineering University, 2006.

Citation: *Ramy Abdul Mawjood Mohammed, B. Bhaskara Rao, Ch. D Uma Sankar, "Enhancement of Speech using Kalman Filter with Phase and Magnitude Spectrum Compensation", International Journal of Emerging Engineering Research and Technology, 7(8), 2019, pp. 1-7.*