iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** VIII **Month of publication:** August 2017

DOI: http://doi.org/10.22214/ijraset.2017.8033

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue VIII, August 2017- Available at www.ijraset.com

Enhancement of Speech Using Kalman Filter With Phase and Magnitude Compensation

G. Sandeep[1], Mr CH. D. Uma shankar[2]
[1]*M. Tech Student Department of ECE Acharya Nagarjuna University, Guntur*
[2]*Assistant Professor Department of ECE Acharya Nagarjuna University, Guntur*

Abstract: During conversation, both hearing and speaking adapt to the background noise in a noisy environment. 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 the SNR and to allow for the listener to grasp the speech clearly, a speech enhancement method should be applied. In this work a novel speech enhancement method using phase and magnitude spectrum compensation using kalman filter is proposed. These method is tested on speech data from the NOIZEUS corpus, which have been corrupted with different types of noise (white, train and babble) of different signal to noise ratios 0dB, 5dB, 10dB and 15dB. The compensated signal by varying both magnitude and phase spectrum is given to kalman filter to overcome the limitations of wiener filter and to reduce the minimum mean square error (MMSE). The performance of the system is calculated based on mean PESQ scores and compared with MMSE, SSUB and PSC methods using Wiener filter.
Keywords: Speech enhancement, NOIZEUS, MMSE, PESQ, SSUB, Kalman filter.

I. INTRODUCTION

In speech communication, the speech signal is always accompanied by some noise. In most cases the background noise of the environment where the source of speech lies, is the main component of noise that adds to the
speech signal. Though the obvious effect of this noise addition is to make the listening task difficult for a direct listener, there are many more far reaching negative effects
when we process the degraded speech for some other applications. A related problem is processing degraded speech in preparation for coding by a bandwidth compression system. Hence speech enhancement not only involves processing speech signals for human listening but also for further processing prior to listening.

Main objective of speech enhancement is to improve the perceptual aspects of speech such as overall quality, intelligibility, or degree of listener fatigue. In this work, one of the speech enhancement technique to enhance the quality of speech is proposed. Speech can be corrupted with noise at any stage before it reaches the end listener called speech degradation. The different ways in which the speech can be degraded can be broadly categorized as below.

A. At the source of speech

When the source itself is in a noisy environment, the background noise gets added to the speech signal. The background noise may be noise-like such as aircraft cockpit, other moving vehicle or environmental sounds; or it may be speech-like, comprised of competing speakers. Besides this, noise effect can also arise due to room reverberation where the source of speech is present. Here the noise involved is multiplicative noise, where the noise gets convolved with the speech signal.

B. During the transmission of speech signal

Speech is transmitted over transmission channels like telephone lines to reach the listener at a distance.
During this transmission, due to the non-idealities of channel behavior, additional noise is added to the speech signal. Noise can also be added either during the data conversion done before the transmission or during the speech reproduction at the listeners end.

C. Noise at the listener end

Sometimes though the speech source may be in a quiet environment, the end listener may be in a highly noisy environment. Here also the listening fatigue arises as the quality of speech is apparently lowered. Hence speech enhancement is needed in this case also.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue VIII, August 2017- Available at www.ijraset.com

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 obtains the estimation of the clean signal magnitude spectrum, but the phase spectrum is unchanged [2].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 [3]. Wiener filter [4] was proposed by Wiener. Hansen and Jensen first presented the Wiener method in the single channel case enhancement [5]. Doclo and Moonen further extended the Wiener method in the mutil-channel case [6]. Ephraim and Van Trees proposed the linear predictive factors to estimate the pure speech signal [7]. 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[8].But recently, it is found that the phase spectrum may be useful in speech processing applications[9]. Kamil Wójcicki et.al proposed the speech enhancement method of phase spectrum compensation (PSC) in 2008[10] [11]. In this work, the role of phase processing for speech enhancement in the context of assisted listening and speech communication devices is reviewed and why recently phase processing is attracting increasing interest in the speech enhancement community. Furthermore, both early and recent methods for phase processing in speech enhancement are considered. The aim is to show that phase processing is an exciting field of research with the potential to make assisted listening and speech communication devices more robust in acoustically challenging environments.

The rest of the paper is structured as follows. Section II gives methodologies of speech enhancement. Section III proposed work. Finally results are presented in section IV.

II. METHODOLOGIES OF SPEECH ENHANCEMENT

Most of the existing speech enhancement algorithms only change the magnitude spectrum of the noisy speech by keeping phase constant. Some of the techniques for speech enhancement are Spectral subtraction, minimum mean square error using wiener filters.

A. Spectral Subtraction

The oldest and simplest method for getting rid of background noise is spectral subtraction. The amplitude spectra are employed to give the amplitude spectrum of the enhanced speech the phases will not be touched in this system. One must take care of not to generate negative amplitudes in the subtraction. In practice, some positive minimum value is given to limit the amplitudes. Another old method is to employ the power spectral densities, giving the spectrum of the enhanced speech.

B. Minimum mean square error

The realistic and optimal object power spectrum estimation without artefacts requires model based statistical methods. The stochastic estimation methods such as minimum mean square error (MMSE) and its variant MMSE log spectral amplitude (LSA) suggested by Ephrahim and Malah are commonly used estimation methods. They are based on modelling spectral components of speech and noise processes as independent Gaussian variables. Almost all literature mentions that the performance of Wiener filter and MMSE LSA is outstanding in terms of both subjective and objective evaluations. The stochastic estimation method called MAP (maximum a posteriori) is very close in performance with MMSE LSA with simpler computations.

C. Wiener filter

This filter is the precursor of Kalman filter. The goal of Wiener filter is to remove the noise from a corrupted signal. Normally, the filters are designed for a specific frequency, but in Wiener filters, first of all, one must have the knowledge about the spectral properties of the original 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.

D. Kalman filter

The filter is a mathematical procedure which operates through a prediction and correction mechanism. In essence, this algorithm predicts a new state from its previous estimation by adding a correction term proportional to the predicted error. In this way, this error is statistically minimized. This filter is the main algorithm to estimate dynamic systems specified in state-space form. A Kalman filter is simply an optimal recursive data processing algorithm. It processes all the measurements available, regardless the precision, to estimate the current value of the interest variables, using:

1) Knowledge of the system and the measurement devices.

2) Statistic description of the system noises, measurements of errors and the uncertainty of the dynamics models.
3) Any information available about the initials conditions of the variables under study.

III. PROPOSED WORK

The main idea in this work 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 compensated signal. After that, the idea is to use the Kalman filtering as a tool to estimate the minimum mean square error (MMSE) and to enhance the compensated signal.

A. Database

To develop this work 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.

B. Experimental steps

Once loaded with the original clean signal and the signal corrupted with noise, look for the amount of noise added. Consider an additive noise mode.

$$x(n) = s(n) + d(n)$$

Where $x(n), s(n)$ and $d(n)$ denote discrete-time signals of noisy speech, clean speech, and noise, respectively.

Converting the discrete time signal to frequency signal using discrete short-time Fourier transform (DSTFT) is done. The DSTFT of the corrupted speech signal is given by

$$X(n,k) = S(n,k) + D(n,k)$$

Where $X(n,k), S(n,k)$ and $D(n,k)$ are the DSTFTs of noisy speech, clean speech, and noise respectively. Each of them can be described in terms of the magnitude spectrum and the phase spectrum. Here magnitude and phase spectrum are driven for $X(n,k)$ and can be given in the polar form as

$$X(n,k) = |X(n,k)| e^{j\angle x(n,k)} = A_k e^{j a_k}$$

Where, $|X(n,k)| = A_k$ is the magnitude spectrum and $\angle X(n,k) = \alpha_k$ is the phase spectrum.

The magnitude of the signal can be estimated by magnitude estimator which is represented as,

$$\hat{A}_k = \frac{\sqrt{\pi}}{2} \frac{\sqrt{V_K}}{\gamma_K} \exp\left(-V_k/2\right)\left[(1+V_k) I_0\left(\frac{V_k}{2}\right) + V_k I_1\left(\frac{V_k}{2}\right)\right] Y_k$$

The phase spectrum compensation function can be obtained by

$$\wedge(n,k) = \lambda\varphi(k) \left| D(\hat{n},k) \right|$$

The next step is offsetting the complex spectrum of the noisy speech by the additive real-valued phase spectrum compensation function.

$$X_{\wedge}(n,k) = X(n,k) + \wedge(n,k)$$

From above we obtain the magnitude estimation and the compensated phase spectrum, then the modified complex spectrum is given as

$$\hat{S}(n,k) = \hat{A}_k \, e^{j\angle x_A(n,k)}$$

At last the IDSTFT is used to convert the frequency domain to the time-domain representation. The resulting time-domain frames may be complex, by employing the overlapped procedure, the compensated time-domain signal is obtained.
compensated signal is given to kalman filter, output of this filter is given by

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue VIII, August 2017- Available at www.ijraset.com

$$x(n,k)=\sum_{m=-\infty}^{\infty}x(m)w(n-m)e^{-j2\pi km/N}$$

Where k denotes the kth discrete-frequency of N uniformly spaced frequencies and w (n) is an analysis window function of short duration in kalman filter. Thus enhanced signal with minimum mean square error is obtained.

IV. EXPERIMENTAL RESULTS

Comparison of different speech enhancement methods is given in table 1.

SNR (dB)	METHODS					
	Clean	Noisy	SSUB	MMSE	PMSC& Wiener	Proposed
0 dB	5.5	1.59	1.76	1.96	2.08	2.82
5dB	5.5	1.83	2.22	2.3	2.43	3.19
10 dB	5.5	2.14	2.64	2.62	2.78	3.33
15 dB	5.5	2.47	3.06	2.92	3.13	3.89

Table 1. Mean PESQ scores.

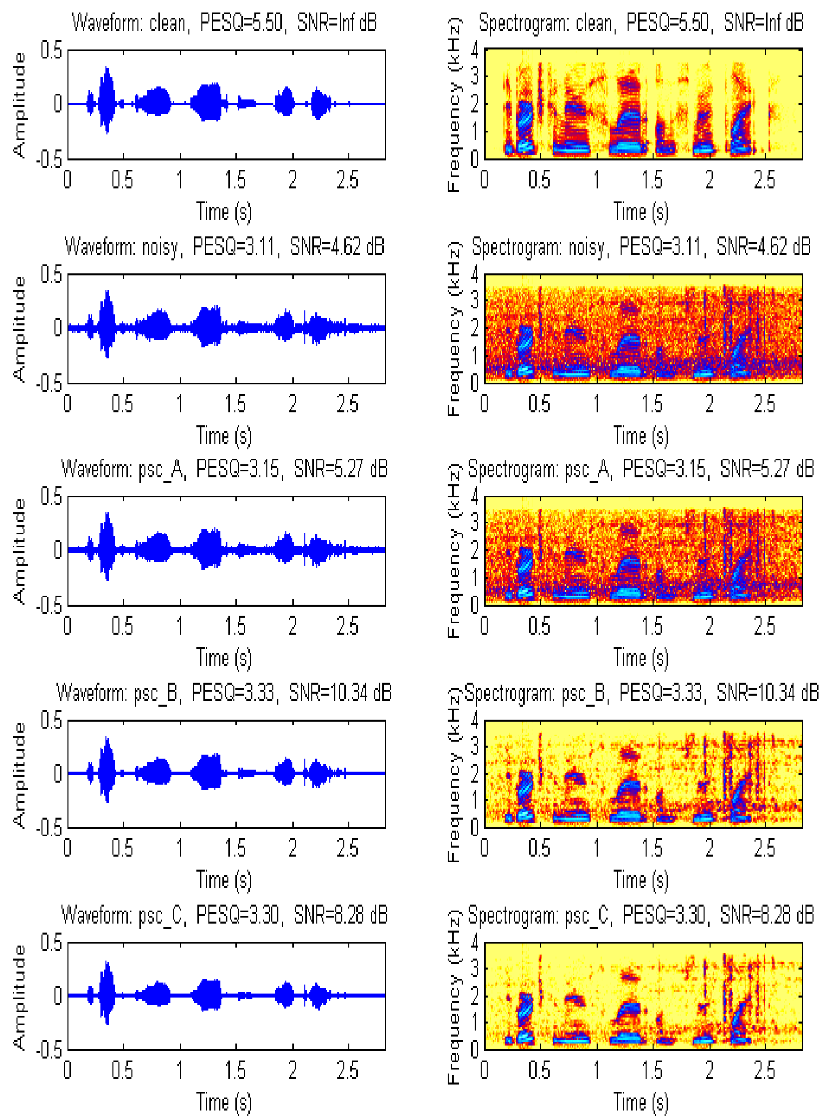

Figure.1 Spectrograms of sp01 NOIZEUS utterance at 10dB SNR.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue VIII, August 2017- Available at www.ijraset.com

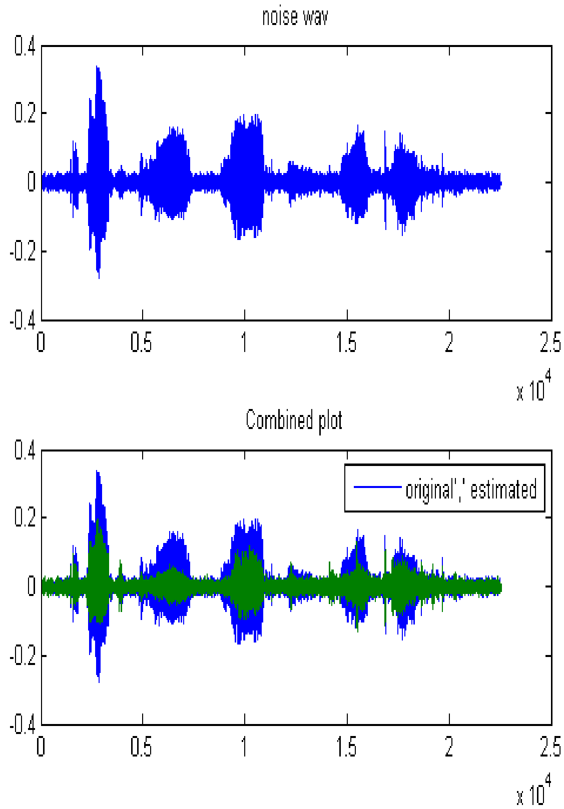

Figure. 2 Shows the combination of both noise wave and speech enhanced signal.

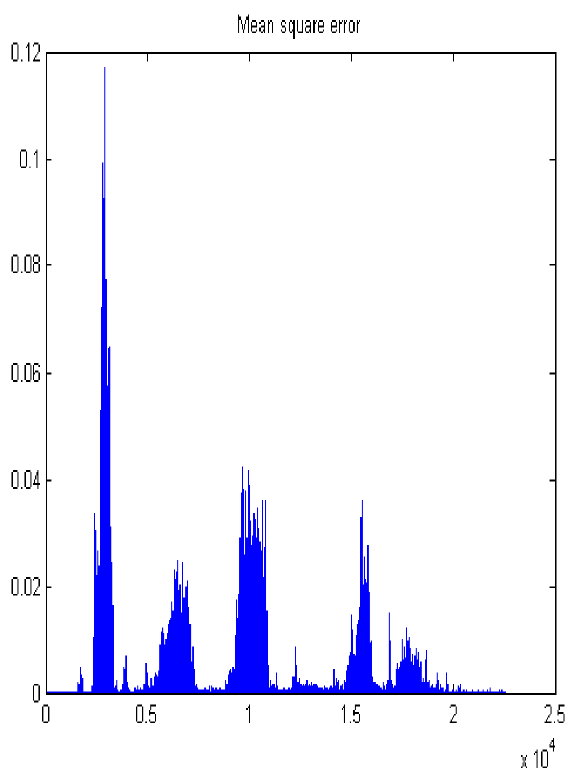

Figure. 3 Shows the plot of mean square error using kalman filter.

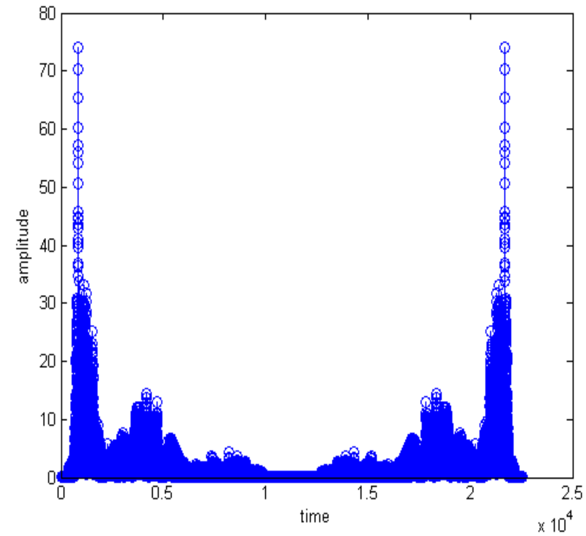

Figure 4 Shows that the magnitude change before and after the speech enhancement

V. CONCLUSION

In this paper a novel speech enhancement method using phase and magnitude spectrum compensation using kalman filter is proposed. These method is tested on speech data from the NOIZEUS corpus, which have been corrupted with different types of noise (white, train and babble) of different signal to noise ratios 0dB, 5dB, 10dB and 15dB.The compensated signal by varying both magnitude and phase spectrum is given to kalman filter to overcome the limitation of wiener filter and to reduce the minimum mean square error (MMSE). The performance of the system is calculated based on mean PESQ scores. The performance of our proposed algorithm is compared with MMSE, SSUB and PSC methods using Wiener filter. Proposed algorithm achieved the best results in terms of mean PESQ scores at any input speech SNRs and can improve speech intelligibility for low SNR levels.

REFERENCES

[1] P. Loizou, Speech Enhancement: Theory and Practice. Boca Raton,FL: CRC, 2007.
[2] BOLL S F. Suppression of acoustic noise in speech using spectral subtraction[J]. IEEE Trans. Acoustics, Speech, Signal Processing, 1979, 27(2):113-120.
[3] 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): 1109-112
[4] N. Wiener, The Extrapolation, Interpolation, and Smoothing of Stationary Time Series With Engineering Applications. New York: Wiley, 1949.
[5] 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.
[6] 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.
[7] . Ephraim and H. V. Trees, “A signal subspace approach for speech enhancement,” IEEE Trans. Speech Audio Process., vol. 3, no. 4, pp.251–266, Jul. 1995.
[8] 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
[9] K. Paliwal, L Alsteris, “Usefulness of phase in speech processing”,Proc. IPSJ Spoken Language Processing Workshop, Gifu, Japan, pp. 1-6, 2003
[10] Kamil Wójcicki, MitarMilacic, Anthony Stark, James Lyons, KuldipPaliwal. Exploiting Conjugate Symmetry of the Short-Time Fourier Spectrum for Speech Enhancement [A].IEEE Signal Process[C].Lett, 2008,15:461-464.
[11] Stark, A., Wójcicki, K.K., Lyons, J.G. and K. Paliwal, “Noise driven short time phase spectrum compensation procedure for speech enhancement”, Proceedings of the 10th International Conference on Spoken Language Processing (INTERSPEECH-ICSLP), Brisbane.