Advances in Computer Science and Information Technology (ACSIT)
p-ISSN: 2393-9907; e-ISSN: 2393-9915; Volume 4, Issue 2; January-March, 2017, pp. 111-114

http://www.krishisanskriti.org/Publication.html

Implementation of Magnitude and Phase Spectrum Compensation to Achieve an Enhanced Speech Signal

Sakil Ansari

(B.TECH) Department Of CSE, College Of Engineering, JNTUH,
Kukatpally, Hyderabad, Telangana, (500085), INDIA
E-mail: sakilansari4@gmail.com

Abstract—*Speech enhancement method is used to improve the speech quality by suppressing the background noise or estimating the background noise. Removal of background noise is one of the crucial task because the speech signal should not be harmed by noise signal. Typically many speech enhancement algorithms have been proposed by concentrating only on short-time magnitude spectrum or else by compensating the phase spectrum by keeping the magnitude spectrum unchanged. In this paper, we propose an innovative technique by changing both magnitude and phase spectra to improve the quality of a speech signal by producing an modified complex spectrum. In simulation results, both objective speech quality measure PESQ and spectrogram analysis has shown the performance of an enhanced speech signal.*

Keywords - *speech enhancement, magnitude spectrum, phase spectrum compensation.*

1. INTRODUCTION

Speech enhancement is nothing but noise suppression technology. It perceives most important significance to improve the quality and intelligibility of the speech signal in the noise environments. Natural background noises are sounds better when compared to unwanted twisted noises such as traffic sounds, music sounds, some other person is speaking in the background etc, which may degrades the perceptual quality and intelligibility of the desired speech signal.

Enhancement of speech signal is based on magnitude and phase spectrum compensation to restore the original signal from several background noises.

Let the noisy speech be defined as

$$x(n) = s(n) + d(n) \qquad (1)$$

where $x(n)$ represents discrete-time signals of noisy speech,

$s(n)$ represents original speech

$d(n)$ represents noise component.

The discrete short-time Fourier transform (DSTFT) is a set of real and complex numbers of $x(n)$to produce a periodic function of frequency variable. Therefore, the DSTFT of the corrupted speech signal is defined as

$$X(n,k) = \sum_{m=-\infty}^{\infty} x(m)\,\omega(n-m)e^{-j2\pi km/N} \qquad (2)$$

where k defines as the kth discrete frequency of uniformly spaced N frequencies and $\omega(n)$ is an analysis window function. Therefore by using DSTF transform Eq.(1) can be represented as

$$X(n,k) = S(n,k) + D(n,k) \qquad (3)$$

where $X(n,k)$, $S(n,k)$ and $D(n,k)$ represents DSTF transform's of the total noise speech, clean speech, and noise, respectively. The above equation is determined to describe each part of the speech signal in terms of the DSTFT magnitude spectrum and the DSTFT phase spectrum. Here, the clean speech signal $S(n,k)$ can be represented in terms of polar form as

$$s(n,k) = |s(n,k)|^{e^{j<s(n,k)}} = A_k e^{j\alpha_k} \qquad (4)$$

where $|S(n,k)|$ is the magnitude spectrum, and $\angle S(n,k)$ is the phase spectrum.

Many of the existing speech enhancement algorithms to remove the noisy speech as well as to look at the effects of intelligibility of speech signal to improve the quality of the speech signal. Actually, the speech signal is converted into polar form to generate magnitude and phase spectrum of the speech signal with respect to frequency. Here, many of the speech models has been proposed for speech enhancement, which is based only on the statistical model of magnitude to produce a modified complex spectrum. Thus, the modified complex spectrum is known as the estimated clean speech spectrum.

The following are some of the existing speech enhancement algorithm which is used to change only the magnitude spectrum of the noisy speech signal and then combines with the unchanged phase spectrum to form a modified complex spectrum. Boll proposed Spectral subtraction (SSUB) in 1979. The basic idea of this method is that the noisy speech signal is the combination of both clean speech and noise as shown in Eqn.(1). Now, the magnitude spectrum of noise is get subtracted by the magnitude spectrum of the noisy speech. Therefore, the resultant magnitude spectrum of clean speech is combines with the unchanged phase spectrum[2] to obtain modified complex spectrum. Wiener filter[4]was proposed by Wiener. Hansen and Jensen. The basic principle of the weiner filter is to obtain an estimate of clean signal from the signal corrupted by the noise component. The estimate clean speech is obtained by minimizing the mean square error(MSE) between the desired signal $s(n)$ and the estimated signal $\hat{s}(n)$. The MMSE estimator[3], which is presented by Ephraim and Malah in 1984. The MMSE method is analyzed by using short time spectral amplitude (STSA) estimator to enhance the desired speech signal quality. This method is also based on statistical analysis. 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].

Conventional speech enhancement method enhances the magnitude spectrum and usue the corrupted speech phase spectrum for signal recovery. Mainly, all the existing are used to change the magnitude spectrum of the speech signal because the phase spectrum provides less perceptual effect at high signal to noise ratio(SNR) level[8].Later, it has been found that phase spectrum is also helpful for speech enhancement applications[9].

This paper proposed a novel method by focusing on both components such as magnitude spectrum and phase spectrum of a speech signal to improve the quality and intelligibility of a speech signal. The quality of the sppech signal is measured by PESQ (Perceptual Estimation of Speech Quality) score compared with conventional methods.

2. LITERATURE REVIEW

i. Spectral subtractive algorithms:

In AMS (analysis–modification–synthesis) framework, the modulation domain processing was added by Paliwaland Schwerin[10]. To know the importance of the modulation stage for speech intelligibility, they had done three experiments, they are

1. Investigate the relative contributions to intelligibility of the modulation magnitude

2. Modulation phase, and

3. Acoustic phase spectra.

In these 3 experiments they found that that the intelligibility of stimuli constructed from only y the modulation magnitude or phase spectra is significantly lower than the intelligibility of the acoustic magnitude spectrum. So here we can recognize that there is an effect of the modulation frame duration on intelligibility for both, the modulation magnitude and phase spectrum as the speech reconstructed from only the short-time modulation phase spectrum has highest intelligibility when long modulation frame durations (>256 ms) are used.

ii. Statistical-model-based algorithms

The effects of window shape and its length on the quality of phase-only and also magnitude-only reconstructed speech was investigated by Loveimi, Ahadi (Loveimi, Ahadi, 2010)[11]. The Speech signal is reconstructed through Least Square Error Estimation (LSEE) and Overlap Add (OLA) methods from its magnitude-only and phase-only spectra. If we want to reconstruct the speech from its magnitude spectrum they select sequence of random uniformly distributed numbers in the range of $(-\pi\ , \pi, \varphi)$, as the phase sequence, or substitute phase spectrum with zero.

iii. Subspace algorithms

The Combined TSP method gives a relatively better performance compared to temporal or spectral processing alone was invented by Krishnamoorthy & Prasanna[11]. They propose a noisy speech enhancement method by combining linear prediction (LP) residual weighting in the time domain and spectral processing in the frequency domain is used to provide better noise suppression as well as better enhancement in the speech regions. By estimating the sum of the peaks in the discrete Fourier transform (DFT) spectrum, the gross level features are computed. Using the knowledge of the instants of significant excitation the fine level features are identified. This system is performed only on the high SNR regions of the spectrally processed speech.

3. PROPOSED METHOD

In our proposed method a noisy speech signal is transformed into magnitude and phase spectrum to produce a modified complex spectrum to obtain better intelligibility and high speech quality.

The magnitude spectrum of clean speech is defined as

$$\hat{A}_k = \frac{\sqrt{\pi}}{2}\frac{\sqrt{V_k}}{\gamma_k} exp\left(-\frac{V_k}{2}\right)\left[(1+V_k)I_0\left(\frac{V_k}{2}\right) + V_k I_1\left(\frac{V_k}{2}\right)\right]Y_k \qquad (5)$$

where $I_0(\cdot)$ and $I_1(\cdot)$ represents the modified Bessel functions of zero and first order, respectively. V_k can be defined as

$$V_k = \frac{\xi_k}{1+\xi_k}\gamma_k \qquad (6)$$

where ξ_k and γ_k can de defined as

$$\xi_k = \frac{\lambda_x(k)}{\lambda_d(k)} \gamma_k = \frac{Y_k^2}{\lambda_d(k)} \qquad (7)$$

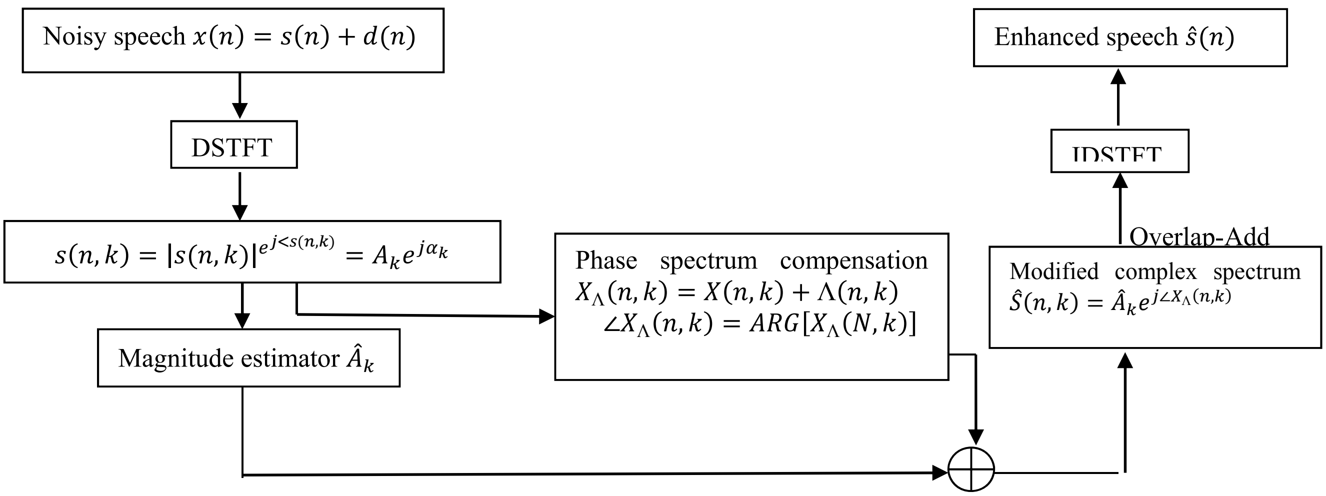

Fig. 1. Block Diagram of Proposed Speech Enhancement Method

Now the phase spectrum compensation function can be defined as

$$\Lambda(n,k) = \lambda \psi(k) \left|\hat{D}(n,k)\right| \qquad (8)$$

Where λ represents real-valued empirically determined constant, let the constant value of λ be 3.74. $\psi(k)$ denotes as antisymmetry function, and it is given by

$$\psi(k) = \begin{cases} 1\,, if\ 0 < k/N < 0.5 \\ -1, if\ 0.5 < k/N < 1 \\ 0, \quad otherwise \end{cases} \qquad (9)$$

Here, the speech signal consists of both real and conjugate vectors. Zero weighting is applied to the values of non conjugate vectors of DSTF transform (i.e, $k = 0$ and $k = N/2$ for $N = even$).The next step is to generate a phase spectrum compensation to reduce the noise of the speech signal and to generate a complex spectrum.

$$X_\Lambda(n,k) = X(n,k) + \Lambda(n,k) \qquad (10)$$

Now the phase spectrum compensation function is obtained by

$$\angle X_\Lambda(n,k) = ARG[X_\Lambda(n,k)] \qquad (11)$$

ARG=complex angle function

The above equation is giving the information about magnitude estimation and compensated phase spectrum. After this we can get the remold equation as complex spectrum is given below.

$$\hat{S}(n,k) = \hat{A}_k e^{j\angle X_\Lambda(n,k)} \qquad (12)$$

To convert the frequency domain representation to the time domain representation, we are using the IDSTFT of $\hat{S}(n,k)$. The output of this may be complex because it is in time representation. So in PSC method the imaginary part is removed. And the final result, i.e enhanced time domain signal is obtained by applying overlap-Add method procedure,$\hat{s}(n)$.

4. DISCUSSION

1. A clean speech is corrupted by the noise at the receiver section. To improve the quality of a speech signal as well as the intelligibility proposed method is implemented.

2. Many of the previous methods are used to operate only on short timemagnitude spectrum by keeping short time phase spectrum to enhance the quality of the speech signal.

3. In the proposed method both the magnitude and phase spectrum compensation are implemented to denoise the signal and to avoid the complexities usually occurred in speech processing.

4. Proposed method provides better noise cancellation when compared to the existing methods.

5. The results were analyzed mainly by mean PESQ scores and spectrograms were used to present the performance on objective quality and speech intelligibility.

5. SIMULATION RESULTS

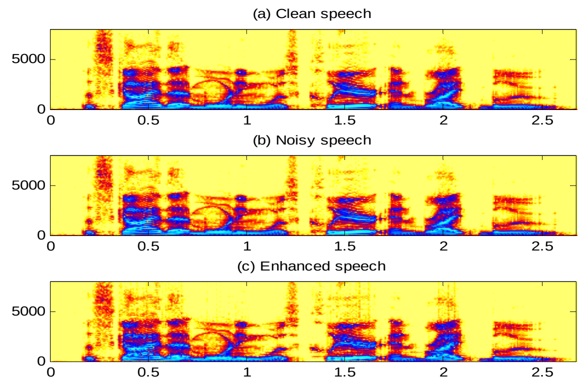

Fig.2. Spectrogram analysis of (a) Clean speech, (b) Noisy speech and (c) Enhanced speech.

Analysis: Spectrogram analysis of an original speech signal (a) is transmitted to a channel. At the receiver section, the clean or the original speech signal is corrupted by noise. The noisy speech signal spectrogram is shown in fig (b). The enhanced spectrogram of a speech signal of the proposed method is shown in fig (c).

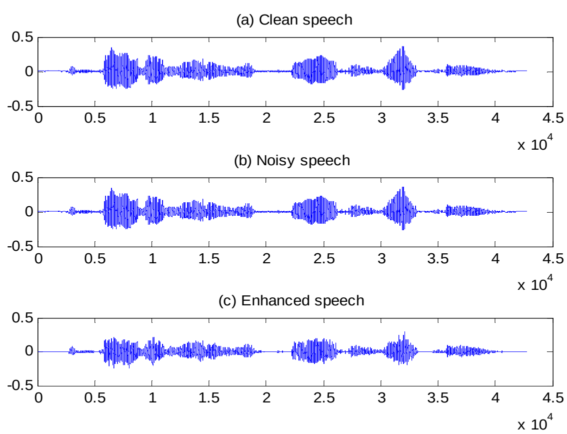

Fig.3. Spectrogram analysis of (a) Clean speech, (b) Noisy speech and (c) Enhanced speech.

Analysis 2: spectrum analysis of a clean speech signal is plotted in fig (a) and the noisy speech signal is shown in fig (b) and the enhanced speech signal of the proposed method is shown in fig (c).

6. CONCLUSION

In this paper, we propose an innovative technique by changing both magnitude and phase spectra to improve the quality of a speech signal by producing an modified complex spectrum.The signal is analysed in both of the spectrums and then added to produce an enhanced noise free speech signal. In simulation results, both objective speech quality measure PESQ and spectrogram analysis has shown the performance of an enhanced speech signal.

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): 1109-1121

[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] Y. 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] K. Paliwal, B. Schwerin, and K. Wójcicki, “Role of modulation magnitude and phase spectrum towards speech intelligibility,” Speech Commun., vol. 53, no. 3, pp. 327–339, Mar. 2011.

[9] E. Loveimi, S. Member, S. M. Ahadi, and S. Member, “Objective Evaluation of Magnitude and Phase Only Spectrum-,” no. March, pp. 3–5, 2010.

[10] P. Krishnamoorthy and S. R. M. Prasanna, “Enhancement of noisy speech by temporal and spectral processing,” Speech Commun., vol. 53, no. 2, pp. 154–174, Feb. 2011.