INTERNATIONAL JOURNAL OF RESEARCH
In Advanced Engineering Technologies

Volume 6, Issue 2 NOV 2017

A NOVEL APPROACH OF SPEECH ENHANCEMENT USING MODIFIED COMPLEX SPECTRUM

SHABISTA NUDRAT (PG SCHOLAR) [1]

M. TEJASWI (MTECH EMDEDD SYSTEM)[2]

DR.B.R.VIKRAM(M.E,PH.D,LMISTE,MIEEE)[3]

[1,2,3]Vijay Rural Engineering College, Nizamabad, Telangana, 503003, INDIA

shabista.nud@gmail.com[1] thejaswi11@gmail.com[2] vikramom2007@gmail.com[3]

Abstract

In speech processing, enhancement of speech signal can be defend by multiple enhancement techniques. Background noise is one of the common problem in speech processing, due to this the quality and accuracy of speech reduces automatically. Previously, many of the speech enhancement algorithms works only on short- time magnitude spectrum, while keeping short time magnitude spectrum remains unchanged or else operates only on short- time phase spectrum by keeping short- time magnitude spectrum remains the same. There is no such technique was implemented to work on both the spectrums to enhance the quality of the speech signal. In this paper, a novel speech enhancement technique is proposed to change the characteristics of both the magnitude and phase spectrums to produce a modified complex spectrum with improved speech quality. The test of an objective speech quality measure PESQ, and spectrogram analysis had showed that the proposed method can obtain better enhancement performance.

Keywords: Speech, Speech enhancement, Noise measurement, Signal to noise ratio, Spectrogram, Estimation, phase spectrum compensation, speech enhancement, magnitude spectrum

1. INTRODUCTION

1.1. Background

In the field of speech enhancement we are interested in enhancing the quality of speech corrupted by additive noise distortion. Depending on the number of audio channels available, speech enhancement methods can be grouped into single-channel and multi-channel approaches. Various single-channel speech enhancement approaches have been proposed in the literature. These can be grouped into spectral subtraction, minimum mean-square error (MMSE) estimation, Wiener filtering (linear MMSE), Kalman filtering and subspace methods. Several of these methods employ the short-time Fourier analysis modification-synthesis (AMS) framework. We focus here on the AMS-based approach to speech enhancement.

1.2. AMS framework based speech enhancement

The AMS framework consists of three stages:

1. The analysis stage, where the input speech is processed using the short-time Fourier transform (STFT) analysis;

2. The modification stage, where the noisy spectrum undergoes some kind of modification; and

3. The synthesis stage, where the inverse STFT is followed by the overlap-add synthesis to construct the output signal.

Let us consider an additive noise model

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

where x(n), s(n) and d(n) denote discrete-time signals of noisy speech, clean speech and noise, respectively. Since speech can be assumed to be quasi-stationary,

it is analysed frame-wise using the short-time Fourier analysis. The STFT of the corrupted speech signal x(n) is given by

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

where k refers to the index of the discrete frequency, L is the length of frequency analysis, and w(n) is an analysis window function.2 In speech processing, the Hamming window with 20–40 ms duration is typically employed. Using STFT analysis we can represent Eq. (1) as

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

where $X(n,k)$, $S(n,k)$ and $D(n,k)$ are the STFTs of noisy speech, clean speech and noise, respectively. Each of these can be expressed in terms of the STFT magnitude spectrum and the STFT phase spectrum. For instance, the STFT of the noisy speech signal can be written in polar form as

$$S(n,k) = |S(n,k)|e^{j\angle S(n,k)} = A_k e^{j\alpha_k} \quad (4)$$

1.3. Earlier studies on the usefulness of the short-time phase spectrum in speech processing

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 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 multi-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].

1.4. Aims of the paper and its organization

In this paper, a new approach to speech enhancement is developed, where not only the short time magnitude is noise compensated but also the short time phase spectrum is altered to handle the noise causing unwanted distortion in the enhanced speech. Based on the fact that noisy speech spectrum in low frequency region is equivalent to the noisy spectrum in that region, a noise estimation approach is introduced with the conditional spectral subtraction method in order to track the time variation of non-stationary noise. Unlike the conventional speech enhancement methods that help the short time phase spectrum unchanged, we proposed to incorporate the estimate noise spectrum in a procedure of noise compensation in the phase spectrum. the new complex spectrum obtained by exploiting the modified magnitude and phase spectra is found effective is producing enhanced speech with improved quality with minimal distortion as compared to some of the existing speech enhancement methods.

2. LITERATURE REVIEW

There are several techniques used for the speech enhancement.

Ching-Ta Lu et. al. [12] proposed a single channel speech enhancement method with the use of perceptual decision directed (TSDD) approach. The two-step decision-directed approach is used to improve the accuracy of approximated speech

spectra. This method can also be used to enhance the performance of TSDD approach. Experimental results show that this method enhance the capability of perceptual method in removing the residual noise and also improve the speech quality.

V. Ramakrishnan et. al. [13] introduced a two-stage method to solve the speech enhancement problem in real noisy world. This method comprises of general spectral subtraction method followed by a series of perceptually motivated postprocessing algorithms. Subtraction step removes the additive noise but adds some spectral artifacts which are removed by post-processing step. Test results show that performance is effective at SNR greater than 0 db.

A. Narayanan et. al. [14] introduced a SNR estimation system which is based on computational auditory scene analysis (CASA). It is a binary masking scheme. This method cannot be used for short-time SNR estimation. This method involves autocorrelation computation and envelope extraction at each T-F unit. Results of different experiments show that the proposed method works better than other long-term SNR estimation algorithms.

N. Yousefian et. al. [15] proposed a coherence-based dual microphone method for estimation of SNR. This technique can be used for hearing aids and cochlear implant devices. Different experiments have been conducted in different conditions. The results show that the proposed method gives significant performance in anechoic and mildly reverberant conditions.

N. Madhu et. al. [16] attempted to define a so called binary mask as the objective of binary mask estimation. Here, it is shown that methods using binary masks are able to improve the intelligibility at low SNR values. For relevant results, a low spectral resolution, modeled using the Bark-spectrum scale is to be used. The performance of IBM and IWF has compared. Intelligibility test shows the higher intelligibility values of IWF than IBM.

J. B. Crespo et. al. [17] presented a method for speech reinforcement in a case where there are many play back regions. In such a case, signals from one region go to other resulting in degradation of speech intelligibility. A smooth distortion is used to improve the quality or intelligibility. Results show the advantages of multizone processing over the iterated application of single zone algorithm.

J. Jensen et. al. [18] proposed a method based on mutual information for estimation of average intelligibility of noisy and processed speech signal. This method estimates the mutual information by comparing the critical-band amplitude envelopes of noisy or processed speech signal because mmse can be considered as an indicator for the intelligibility of noisy speech. Simulation results show that the proposed method can predict the intelligibility of speech distorted by both stationary and non-stationary noises.

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

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

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 remould equation as complex spectrum which 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 mthod the imaginary part is removed. And the final result i.e enhanced time domain signal is obtained by applying overlap-Add method procedure,$\hat{s}(n)$.

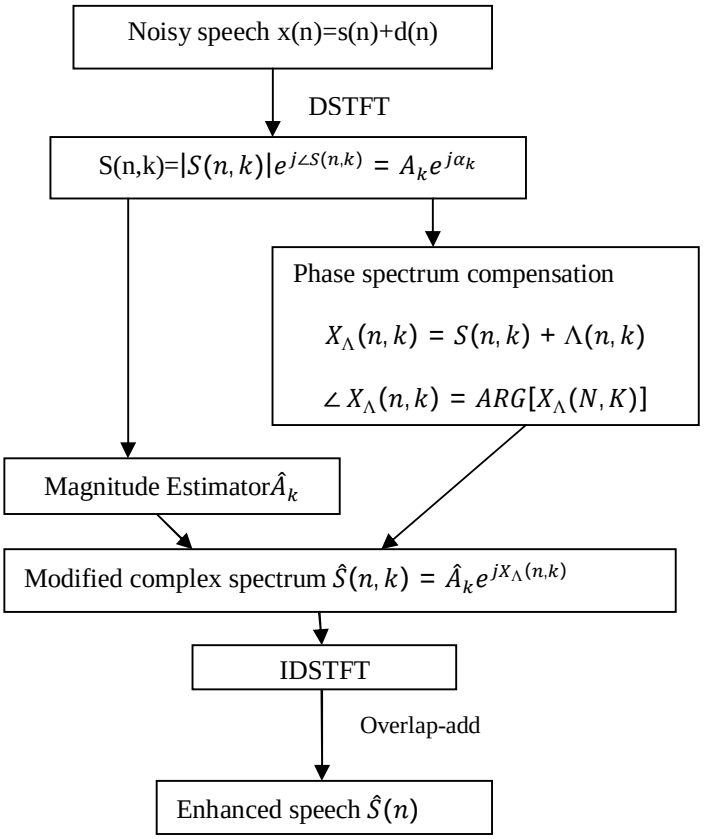

Fig. 1. Block Diagram of Proposed Speech Enhancement Method

4. SIMULATION RESULTS

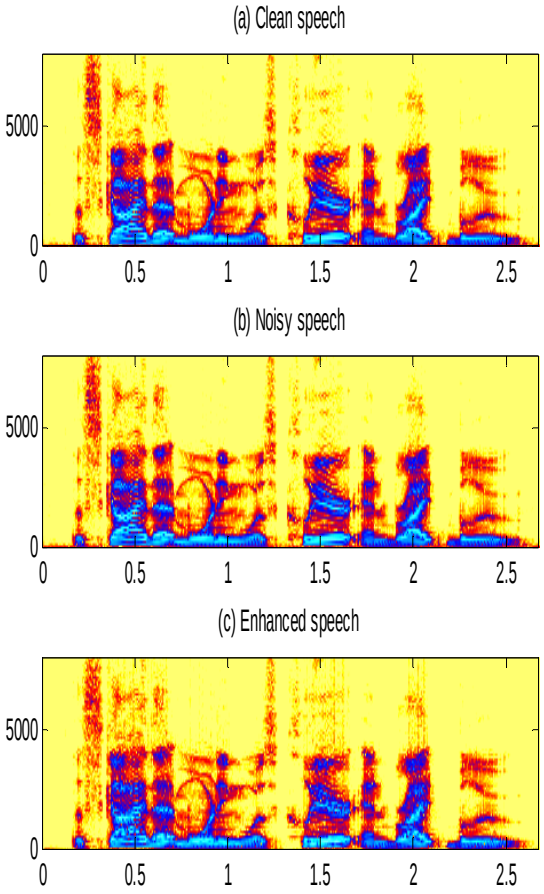

Fig.2. Spectrogram analysis of (a) Clean speech, (b) Noisy speech and (c) Enhanced speech.

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).

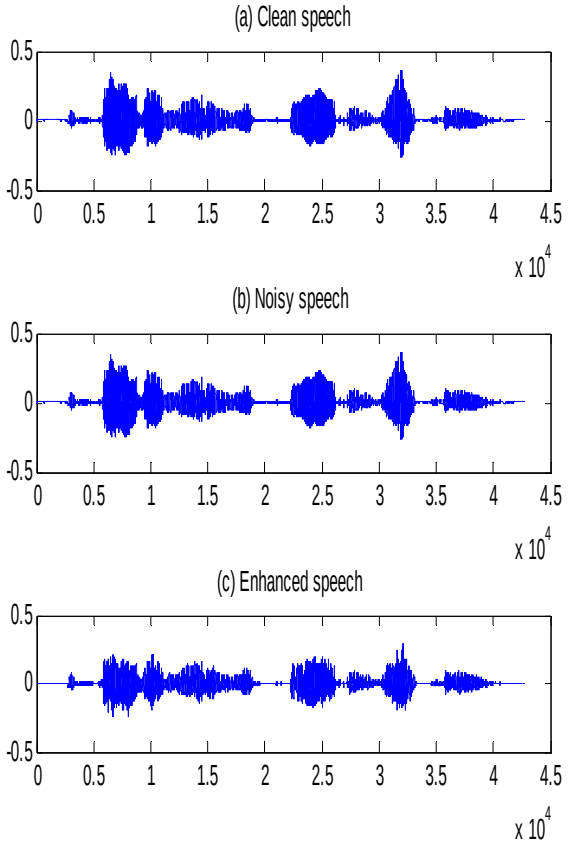

Fig.3. Spectrogram analysis of (a) Clean speech, (b) Noisy speech and (c) Enhanced speech.

Analysis: 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).

5. CONCLUSION

In this paper, a novel speech enhancement technique is proposed to change the characteristics of both the magnitude and phase spectrums to produce a modified complex spectrum with improved speech quality. The test of an objective speech quality measure PESQ and spectrogram analysis had showed that the proposed method can obtain better enhancement performance.

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.

[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 ,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.

[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] 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.