Scientific Journal of Impact Factor(SJIF): 3.134

e-ISSN(O): 2348-4470
p-ISSN(P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 2,Issue 5, May -2015

Speech Enhancement Using Conjugate Symmetric Property of Short Time Fourier Transform

Mahiboob, Dr. Shridhar S. K, S. R. Bharamagoudar, Somashekhar S. Pujar

Department of Electronics and Communication Engineering,Basaveshwara Engineering College Bagalkot, India

Abstract: *The main objective of speech enhancement is to improve quality and intelligibility. At the same time listener fatigue should also be reduced. Short Time Fourier Transform (STFT) is employed in the present work. STFT magnitude and phase spectrum of noisy speech are computed. The magnitude spectrum is kept as it is, whereas phase spectrum is modified empirically. The modified phase is combined with the unmodified magnitude to get the complex spectrum. The complex spectrum is converted into time domain signal through inverse STFT. The resulting time domain signal quality and intelligibility are superior compared to the noisy speech signal. The proposed algorithm is evaluated through subjective listening tests, objective measures and spectrogram analysis. The results are superior as compared to earlier speech enhancement methods.*

Key word: *Speech enhancement, Magnitude spectrum, Phase spectrum, Phase modification.*

I. INTRODUCTION

Speech enhancement aims at improving noisy speech quality, intelligibility and listener fatigue. Earlier speech enhancement algorithms were based on Spectral Subtraction [1], Wiener filtering [2], MMSE estimation [3], kalman filtering [4] and Signal Subspace method [5]. Most of the earlier methods used STFT (Short Time Fourier Transform), DCT (Discrete Cosine Transform), and DWT (Discrete Wavelet Transform) [6, 7] to transform signal from time domain to frequency domain.

In real time speech signal is corrupted by acoustic interferences such as additive noise. Additive noise is typically the background and is uncorrelated with the clean speech. The noisy speech signal can be modeled as a sum of clean speech and the random noise as

$$y(n) = c(n) + r(n)\ ;\ n = 0,1,.....N-1 \quad (1)$$

Where *n* is *the* discrete time index, *y(n)* , *c(n)* and *r(n)* are the n^{th} sample of the discrete-time signal of noisy speech, clean speech and noise respectively [6]. Speech is non- stationary in nature, rather it is quasi stationary. It is analysed frame-wise using its magnitude spectrum and retaining its phase spectrum as it is in the Analysis Modification Synthesis (AMS) framework through the short-time Fourier analysis. The discrete short-time Fourier transform (DSTFT) of the noisy speech signal $y(n)$ is given by

$$Y(n,k) = \sum_{n=0}^{N-1} y(n)w(n)e^{-i\frac{2\pi kn}{N}} \quad (2)$$

Where k denotes the k^{th} discrete-frequency of uniformly spaced frequencies and $w(n)$ is an analysis window (Hamming) function [6]. In the proposed method the length of the window is 20 msec (160 samples) and with an overlap of 50%. Frame length of 20 msec gives almost stationary speech frames upon which Fourier transformation can be employed [7]. The DSTFT of the time windowed noisy speech, assuming noise is additive, can be written as

$$Y(n,k) = C(n,k) + R(n,k) \quad (3)$$

Where $Y(n,k)$, $C(n,k)$ and $R(n,k)$ are the DSTFT of noisy speech, clean speech and noise respectively. Each of these can be expressed in terms of the DSTFT magnitude spectrum and the DSTFT phase spectrum. For instance, the DSTFT of the noisy speech signal can be written in polar form as

$$Y(n,k) = |Y(n,k)|e^{i\angle Y(n,k)} \quad (4)$$

Where $|Y(n,k)|$ is the magnitude spectrum and $\angle Y(n,k)$ is the phase spectrum of noisy speech signal.

The AMS-based speech enhancement methods modify only the magnitude spectrum while keeping the phase spectrum unchanged for synthesis. We propose a novel approach to speech enhancement, in which the noisy magnitude

spectrum is recombined with a modified phase spectrum to produce a modified complex spectrum. During synthesis, low energy components of the modified complex spectrum cancel out more than the high energy components, resulting in background noise reduction. The values of phase modification factor K are based on listening tests. The proposed algorithm is evaluated through subjective listening tests, objective measures and spectrogram analysis. The results are better than the earlier conventional spectral subtraction based speech enhancement methods.

II. PROPOSED METHOD

The proposed method is based on Analysis Modification Synthesis (AMS) framework. The AMS framework consists of three stages they are (1) Analysis (2) Modification and (3) Synthesis. The block diagram of the proposed method is as shown in figure 1.

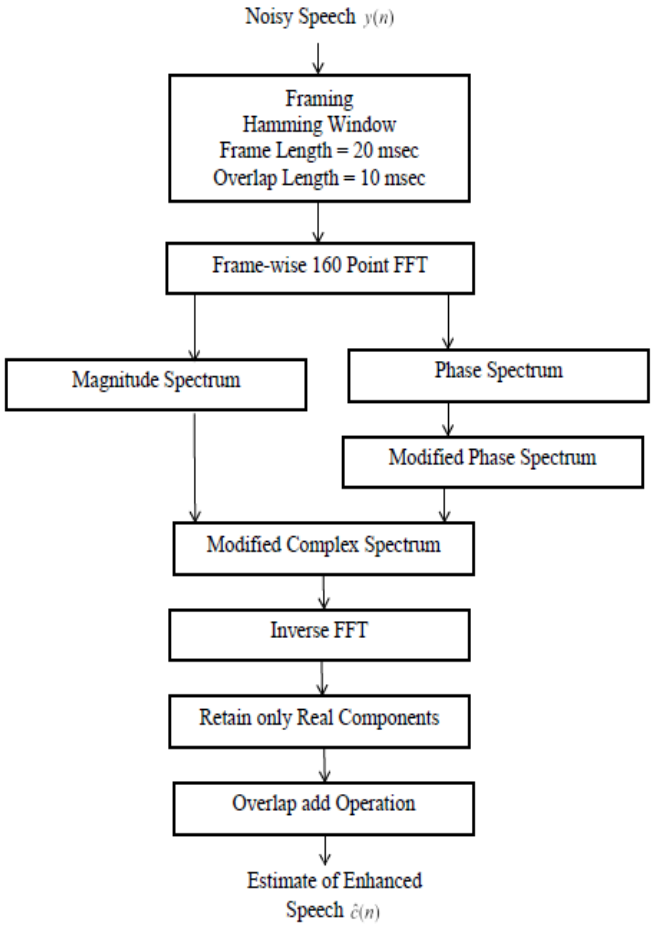

Figure1. Block diagram of the proposed method.

The noisy speech signal is a real valued signal. Hence its Fourier transform obeys conjugate symmetry, given by

$$Y(n,k) = Y^*(n, N-k) \qquad (5)$$

The degree of reinforcement of these complex conjugates can be controlled by modifying their phase. For this purpose an anti-symmetric function is used. Mathematically it is given by

$$Y'(n,k) = \begin{cases} Y(n,k) + k_1; & 0 \le k < \dfrac{N}{2} \\ Y(n,k) - k_1; & \dfrac{N}{2} \le k \ge N-1 \end{cases} \qquad (6)$$

In our approach Phase modification factor $k_1$is determined empherically for a given value of SNR of noisy speech signal. The magnitude spectrum of noisy speech is kept as it is, while its modified phase spectrum is computed from equation (6). Modified complex spectrum is obtained by combining the unmodified magnitude with modified phase given by

$$Y''(n,k) = |Y(n,k)| e^{j\angle Y'(n,k)} \qquad (7)$$

Here N is assumed to be even and k1 is real valued constant and anti-symmetric with respect to N/2 point.
The modified complex spectrum is converted back into time domain to have real samples. But conjugate symmetry is disturbed by k_1 hence samples obtained will be complex. Further, only real part of the complex samples is retained. Finally overlap-add operation is performed on the real samples to get enhanced speech.

III. EXPERIMENTATION AND EVALUATION

A. *Speech Database*

NOIZEUS speech database is used in the experimentation of the proposed algorithm. This database is composed of non-stationary noisy speech samples at different SNRs. It also contains phonetically balanced clean speech samples of both male and female. The sampling frequency being used is 8 kHz [8].

B. *Experimentation Procedure*

Five normal hearing subjects and two impaired hearing loss subjects of age group 20-25 years are chosen for the listening test. Noisy speech stimuli of different noise sources such as babble, restaurant, car, exhibition, AWGN and airport at SNR 0, 5, 10, and 15 dB are used for experimentation. Each subject is presented with noisy speech stimuli and enhanced speech stimuli and asked to give the opinion rating of both noisy and enhanced speech samples on 1-5 scale without wearing any hearing aid. The rating criteria are based on the quality of speech. Further mean opinion score (MOS) of all the subjects is computed.

C. *Other Evaluation methods*

Other than subjective listening tests described above, objective measures such as increment in segmental SNR, Itakura-Saito Distance (ISD) measure, Weighted Likelihood Ratio (WLR) and Log-likelihood Ratio (LLR) are used to evaluate the algorithms. These are used to validate the subjective listening tests. Finally spectrogram analysis of enhanced speech samples is also carried out. Spectrogram is a time-varying spectral representation that shows how the spectral density of a signal varies with time. Also known as spectral waterfalls, sonograms, voiceprints, or voice grams,

IV. RESULTS AND DISCUSSION

The proposed method performs best in the case of AWGN noise as compared to Babble, car, restaurant, exhibition and airport noise environments. This fact is evident from subjective listening tests and objective measures. In case of AWGN it is simple to estimate the noise power as compared to other class of competing noise sources.
The results of spectrogram analysis are shown in figure 7, 8, 9, 10 and 11. The enhanced signal for the white noise case does not exhibit speech distortion, while the background noise has been attenuated. In the other noises cases, though the background noise is suppressed, a small amount of speech signal distortion exists because, the factor *k1* being constant for all values of frequencies within the frame.
The mean opinion score of normal hearing subjective listening tests are presented in table I. The objective measures: Increment in segmental SNR, Itakura Saito distance, weighted likelihood ratio and log likelihood ratio to validate the subjective listening tests is presented in table II, III, IV and V respectively. All these objective measures are highly correlated with subjective listening tests.

Table I. Mean opinion score of normal hearing subjects.

Noise Type/SNR	MEAN OPINION SCORE							
	0 dB		5 dB		10 dB		15 dB	
	NS	ES	NS	ES	NS	ES	NS	ES
Babble	1.5	3.51	2.1	3.7	3.4	4.3	4.05	4.51
Car	1.5	2.8	2.4	3.3	3.1	3.6	3.9	4.3
Restaurant	1.1	3.2	2.1	3.3	2.9	4	3.8	4.3
Exhibition	2.1	3.5	2.7	3.8	3.4	4.2	4.3	4.72
AWGN	2.5	4	3.1	4.1	3.6	4.3	4	4.5
Airport	1.4	2.9	2.1	3.5	3.4	4.2	4	4.5

Table II. Mean opinion score of two impaired hearing loss subjects

Noise Type/SNR	MEAN OPINION SCORE							
	0 dB		5 dB		10 dB		15 dB	
	NS	ES	NS	ES	NS	ES	NS	ES
Babble	1.12	1.81	1.4	2.52	2.4	3.4	2.9	3.94
Car	1.2	1.8	1.8	2.88	2.6	3.6	3.2	4.1
AWGN	1.14	2.24	1.5	2.66	2.1	3.6	2.8	3.8
Airport	1.1	2.9	1.9	3.1	2.8	3.9	3.5	4.3

Table III. Increment in segmental SNR of enhanced speech with respect to clean speech

Noise type/SNR	0 dB		5 dB		10 dB		15 dB	
	ES	ISNR	ES	ISNR	ES	ISNR	ES	ISNR
Babble	9.89	9.89	16.62	11.62	18.08	8.08	18.2	3.2
Restaurant	13.9	13.9	14.30	9.30	16.56	6.56	18.6	3.66
Car	9.67	9.67	6.03	1.03	15.53	5.53	22.8	7.81
Exhibition	3.62	3.62	21.24	16.24	13.5	3.58	18.4	3.41

ES= Enhanced Speech, ISNR= Increment in Segmental SNR

Table IV. Itakura Saito Distance (ISD) of enhanced speech with respect to clean speech

Noise type/SNR	0 dB	5 dB	10 dB	15 dB
Babble	37.18	41.45	52.94	52.04
Car	36.17	49.61	48.47	51.46
Restaurant	34.12	50.19	46.48	49.81
Exhibition	29.61	44.63	45.89	49.20
AWGN	37.35	46.06	50.34	50.94
Airport	42.35	46.09	50.78	51.32

Table V. Weighted Likelihood Ratio (WLR) of enhanced speech with respect to clean speech

Noise type/SNR	0 dB	5 dB	10 dB	15 dB
Babble	108.9	110.7	113.3	113.3
Car	111.4	113.7	112.4	113.3
Restaurant	104.5	111.7	113.1	113.2
Exhibition	108.1	111.8	112.7	113.0
AWGN	107.6	111.5	112.6	113.0
Airport	109.6	112.0	112.8	113.2

Table VI. Log Likelihood Ratio (LLR) of enhanced speech with respect to clean speech

Noise type/SNR	0 dB	5 dB	10 dB	15 dB
Babble	1.93	1.96	2.01	2.01
Car	1.98	2.01	1.99	2.01
Restaurant	1.85	1.98	2.01	2.01
Exhibition	1.92	1.98	2.01	2.01
AWGN	1.91	1.97	2.0	2.01
Airport	1.94	1.99	2.0	2.01

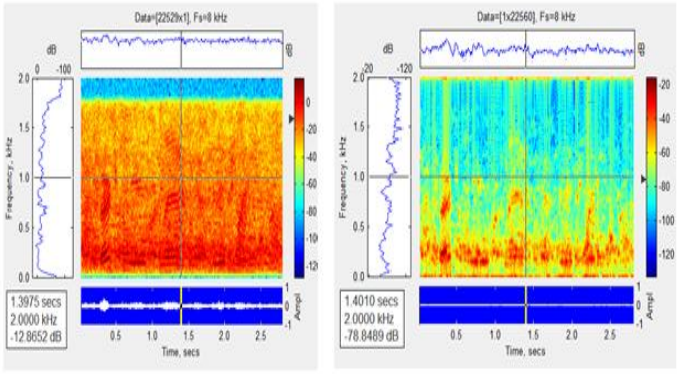

Figure2. Spectrogram of noisy speech and enhanced speech sample (Babble Noise) of SNR 0dB

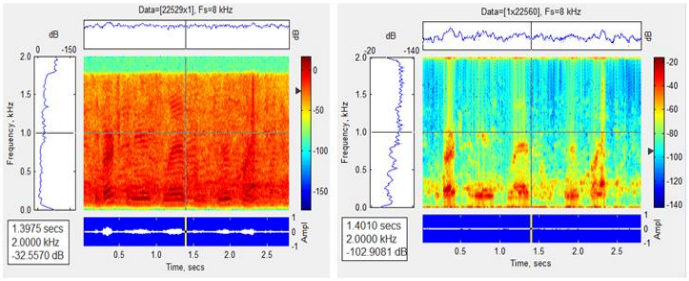

Figure2. Spectrogram of noisy speech and enhanced speech sample (Babble Noise) of SNR 5dB

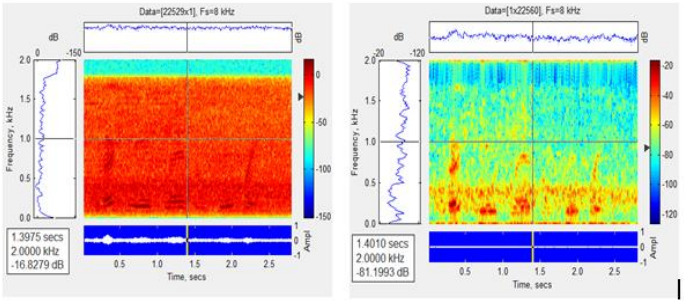

Figure3. Spectrogram of noisy speech and enhanced speech sample (Car Noise) of SNR 0dB

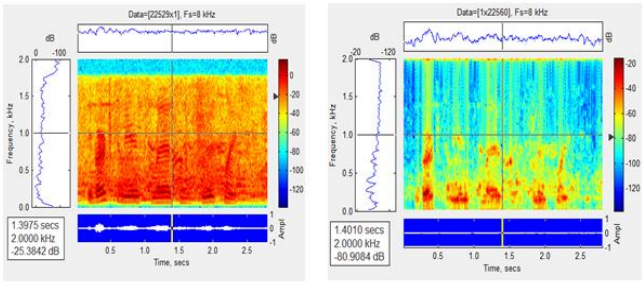

Figure4. Spectrogram of noisy speech and enhanced speech sample (Restaurant Noise) of SNR 5dB

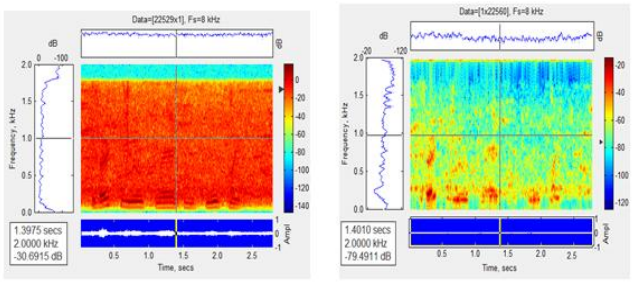

Figure5. Spectrogram of noisy speech and enhanced speech sample (Exhibition Noise) of SNR 5dB

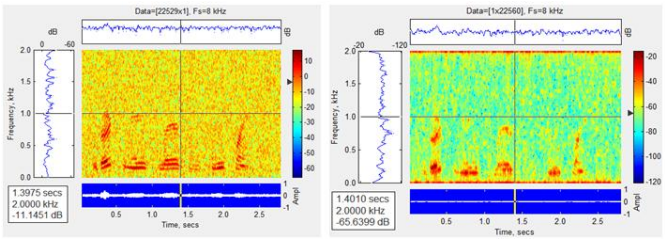

Figure6. Spectrogram of noisy speech and enhanced speech sample (AWGN Noise) of SNR 0dB

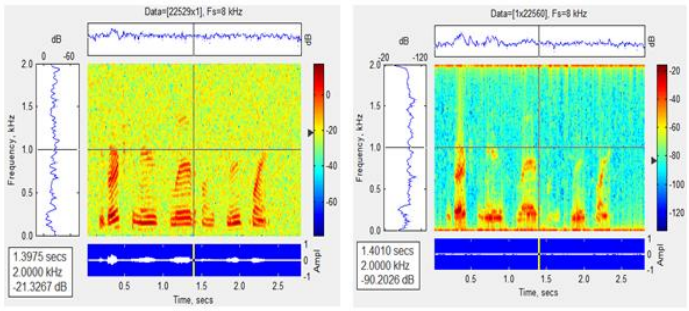

Figure6. Spectrogram of noisy speech and enhanced speech sample (AWGN Noise) of SNR 10dB

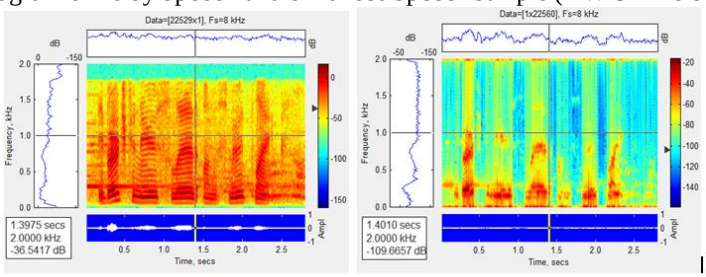

Figure7. Spectrogram of noisy speech and enhanced speech sample (Airport Noise) of SNR 0dB

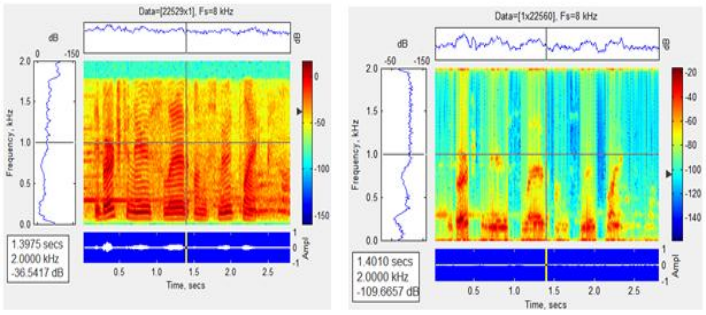

Figure7. Spectrogram of noisy speech and enhanced speech sample (Airport Noise) of SNR 15dB

CONCLUSION

In this paper a novel speech enhancement algorithm is proposed. Its performance is better than the conventional spectral subtraction based class of algorithms. Its novelty is its simplicity and control over the quality of speech. Control over the quality is through *k1*. This approach of speech enhancement is unique because here phase of Fourier transform is processed. In this method low energy components (noise) cancels out more as compared to high-energy components (speech). When magnitudes of the spectral components are greater than the value of k1, the spectral components remain as it is. On the other hand when magnitudes of spectral components are smaller than the value of k1, the spectral components gets suppressed further i.e. small spectral components means noise components. This is as per the basic assumption where speech and noise are additive.

ACKNOWLEDGMENT

This paper is made possible through the help and support from the guide Prof S.R.Bharamagoudar. The first author would like to gratefully and sincerely thank Dr. K. Sridhar for his guidance, understanding, and patience during my M.Tech course at B.E.C Bagalkot and during the all phases of this paper.

REFERENCES

[1] Navneet Upadhyay, and Abhijit Karmakar "Single Channel Speech Enhancement Utilizing Iterative Processing of Multi-Band Spectral Subtraction Algorithm" Proc. 2nd International Conference on Power, Control and Embedded Systems, 2012.
[2] Y P. Fardkhaleghi and M.H. Savoji "New Approaches to Speech Enhancement Using Phase Correction in Wiener Filtering" proc. 5th International Symposium on Telecommunications (IST'2010).
[3] Y.Ephraim and D. Malah, "Speech enhancement using a minimum mean–square error log-spectral amplitude estimator, "IEEE Trans. Acoustics, Speech Signal Processing, vol.ASSP-33, no.2, pp.443-445, Apr.1985.
[4] K.Paliwal and A.Basu, "A Speech enhancement method based on kalman filtering" , in proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'87), Apr.1987, vol.12, pp.297 – 300.
[5] Y. Ephraim and H. V. Trees, "A Signal Subspace approach for speech enhancement" IEEE Trans. Speech Audio Processing.,vol.3,no.4. pp. 251-266, July.1995.
[6] Kamil Wojcicki ,Mitar Milacic,Anthony Stark, James Lyons, and Kuldip Paliwal, "Exploiting conjugate symmetry of the short – time Fourier spectrum for speech enhancement, "IEEE Signal processing letters , Vol.15.2008. pp. 461-464
[7] Ing Yann Soon, Soo Ngee Koh , Chai Kiat Yeo "Noisy speech enhancement using discrete cosine transform" Speech Communication 24 (1998) 249±257
[8] Sharon Gannot, David Burshtein, and Ehud Weinstein "Signal Enhancement Using Beamforming and Nonstationarity with Applications to Speech" IEEE Transactions on signal processing, vol. 49, no. 8, August 2001
[9] Bin Jiang and Jun Yang "Preferred Frame Length for the Short-Time Magnitude Spectrum on Speech Intelligibility and Speech Quality" IEEE 2011
[10] Yu Gwang Jin, Jong Won Shin, and Nam Soo Kim "Spectro-Temporal Filtering for Multichannel Speech Enhancement in Short-Time Fourier Transform Domain" IEEE signal processing letters, vol. 21, no. 3, March 2014
[11] Martin Krawczyk and Timo Gerkmann "STFT Phase Reconstruction in Voiced Speech for an Improved Single Channel Speech Enhancement" IEEE/ACM Transactions on audio, speech, and language processing, vol. 22, no. 12, December 2014
[12] Ing Yann Soon, and Soo Ngee Koh "Speech Enhancement Using 2-D Fourier Transform" IEEE Transactions on speech and audio processing, vol. 11, no. 6, November 2003
[13] Sanjay P. Patil, John N. Gowdy "Exploiting The Baseband Phase Structure Of The Voiced Speech For Speech Enhancement" 2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)

