SPEECH ENHANCEMENT USING SPECTRAL SUBTRACTION TECHNIQUE WITH MINIMIZED CROSS SPECTRAL COMPONENTS

Soumya Jolad[1], Shridhar[2]

[1]*Department of Electronics and Communication Engineering Basaveshwar Engineering CollegeBagalkot, India*
soumyajolad4@gmail.com
[2] *Department of Electronics and Communication Engineering Basaveshwar Engineering College Bagalkot, India*
shridhar.ece@gmail.com

Abstract

The aim of speech enhancement is to get significant reduction of noise and enhanced speech from noisy speech. There are several approaches for speech enhancement .earlier approaches didn't consider cross spectral terms into account. Cross spectral terms become prominent when processing window size becomes small i.e. 20ms-30ms. In this paper, an enhancement method is proposed for significant reduction of noise, and improvement in the quality and perceptibility of speech degraded by correlated additive background noise. The proposed method is based on the spectral subtraction technique. The simple spectral subtraction technique results in poor reduction of noise. One of the main reasons for this is neglecting the cross spectral terms of speech and noise, based on the appropriation that clean speech and noise signals are completely uncorrelated to each other, which is not true on short time basis. In this paper an improvement in reduction of the noise is achieved as compared to the earlier methods. This fact is mainly attributed to the cross spectral terms between speech and noise. This algorithm can be implemented and used in hearing aids for the benefit of hearing impaired people. Objective speech quality measures, spectrogram analyses and subjective listening tests conforms the proposed method is more effective in comparison with earlier speech enhancement techniques.

Keywords: *Spectral Subtaction,Cross Spectral Components*

--***--

1. INTRODUCTION

Improve the intelligibility and overall perceptual quality of noisy speech is main objective of the speech enhancement. To achieve this objective it uses many noise suppression algorithms. Today study on speech enhancement has more importance because of its real time applications such as mobile phones, teleconferencing systems, speech recognition, hearing aids and speech coding.

There are number of speech enhancement algorithms aimed to achieve the quality and perceptibility of speech. Filtering techniques, spectral restoration and speech model based techniques are mainly three categories in the speech enhancement techniques [1], [2]. Designing a linear filter or transformation is the basic principle of filtering techniques so that noise component will be attenuated when the noisy speech is passed through such a filter the. Spectral subtraction method [3], [8], wiener filtering, signals subspace approaches come under filtering techniques. In spectral restoration the clean speech spectrum is estimated from the noisy speech spectrum. Minimum mean square error short time spectral amplitude estimation (MMSE STSAE) [4] comes under spectral restoration techniques. In speech model based technique a discrete speech model is generated, using this model noise reduction problem can be converted into parameter estimation problem in a much lower dimension space [2]. Harmonic model based noise reduction [5], [6], linear prediction based noise reduction come under speech model based techniques [2].

The spectral subtraction algorithm is one of the simple, single channel speech enhancement algorithms. There are two classes in speech enhancement in that first one uses subtractive type algorithm and second one applies spectral subtraction filter to noisy speech [7]. Our approach belongs to first class of speech enhancement. There are mainly two problems in enhanced speech by spectral subtraction. One is a musical noise present in enhanced speech and other is cross spectral terms present in enhanced speech. Musical noise is mainly because of no optimized estimation of noise spectrum. And presence of cross spectral terms is because of neglect of product of Noise and clean speech spectral terms (cross spectral terms) while doing the spectral subtraction. These results an inefficient reduction of noise in enhanced speech. This paper mainly deals with second problem which is cross spectral terms reduction. Here noise is estimated during absence of speech. This paper is arranged as follows. In section II We proposed a new method of spectral subtraction with considering the cross spectral terms. We have showed the spectrogram plots and evaluated it in section III.

2. PROPOSED WORK

In current environments, the degradation of speech is mainly because of additive noises, so we can model noisy speech as a sum of the clean speech and noise signal. Initially noisy speech is divided in to frames and Hamming or Hanning window is applied to each frame. A windowed noisy speech can be represented as

$$x(n) = c(n) + r(n) \tag{1}$$

n is the discrete-time index, *x* (*n*), *c (n)* and *r (n)* are the windowed noisy speech, clean speech and random noise, reactively. To convert from time domain to frequency domain, we apply FFT on each frame so Eq. (1) can be expressed in frequency domain as

$$X(f) = C(f) + R(f) \tag{2}$$

f is the frequency variable. *X (f), C (f)* and *R (f)* denotes noisy speech, clean speech and noise in a frequency domain, respectively. The *X (f)* can be written in polar form a

$$X(f) = |X(f)|\, e^{j\theta x} \tag{3}$$

In Eq.(3), $|Y(f)|$ represents the magnitude spectrum and θ_x denotes the phase spectrum. To obtain power spectrum, apply square on equation (2)

$$|X(f)|^2 = |C(f)|^2 + |R(f)|^2 + C(f).R(f)^* + C(f)^*.R(f) \tag{4}$$

The last 2 terms in LHS side, which is the product of magnitude spectrum of clean speech and magnitude spectrum of noise is called cross spectral terms .In conventional spectral subtraction method this cross spectral terms are neglected or made it equal to 0 by assuming clean and noise signals are uncorrelated to each other.

For statistically large data this assumption is valid, but in spectral subtraction method we deal with short time intervals where this assumption does not holds good [9]. This assumption not holds for correlated or colour noise [11]. As the effect of neglecting the cross terms Significant degradations in performance of speech recognition were noted at SNR levels near 0 dB [14].Some meaningful attempts have been made to take into account of cross spectral terms in spectral subtraction [9], [10], [11].In this paper we have derived for cross spectral terms using below equations. In below equation we have substituted *C (f)+R (f)* for *X(f)*

$$X^*(f)\, R(f) + X(f)R^*(f) = (C^*(f) + R^*(f))R(f) + (C(f) + R(f))R^*(f)$$

$$= C(f)R^*(f) + C^*(f)R(f) + 2|R(f)|^2 \tag{5}$$

$$C(f)R^*(f) + C^*(f)R(f) = X^*(f)R(f) + X(f)R^*(f) - 2|R(f)|^2 \tag{6}$$

Substitute (6) in (4) we get

$$|C(f)|^2 = |X(f)^2| - |R(f)|^2 - X^*(f)\, R(f) - X(f)\, R^*(f) + 2|R(f)|^2$$

$$|C(f)|^2 = |X(f)|^2 + |R(f)|^2 - 2|X(f)R(f)| \tag{7}$$

$$|C(f)| = \sqrt{\ } |X(f)^2| + |R(f)|^2 - 2|X(f)R(f)| \tag{8}$$

By applying the IFFT and using available noisy speech phase samples with above equation, got clean speech in time domain such as

$$c(n) = \text{IFFT}(C(f)e^{-j\Theta x}) \tag{9}$$

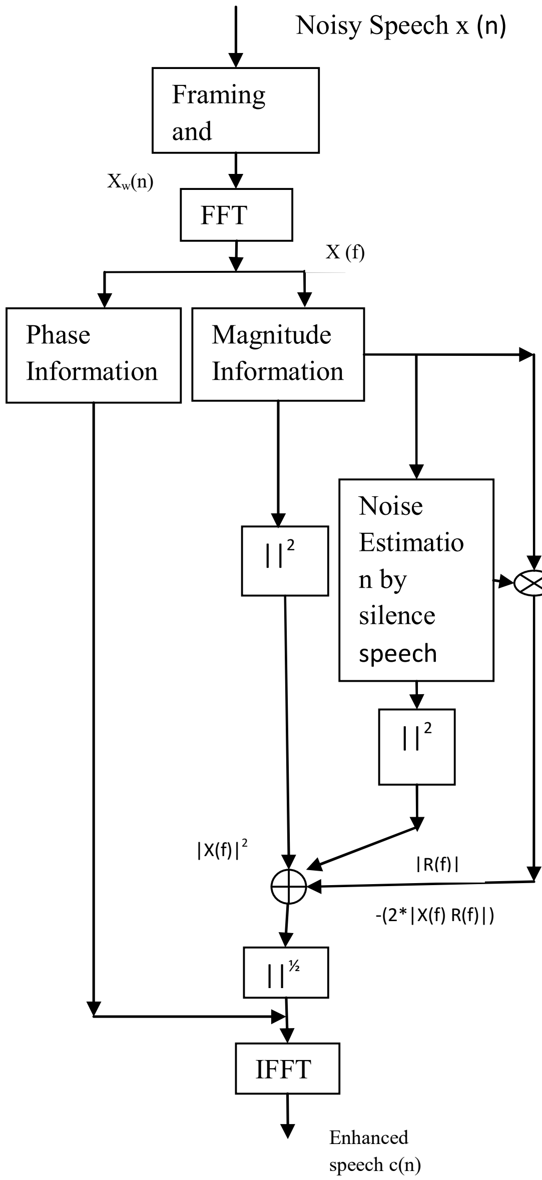


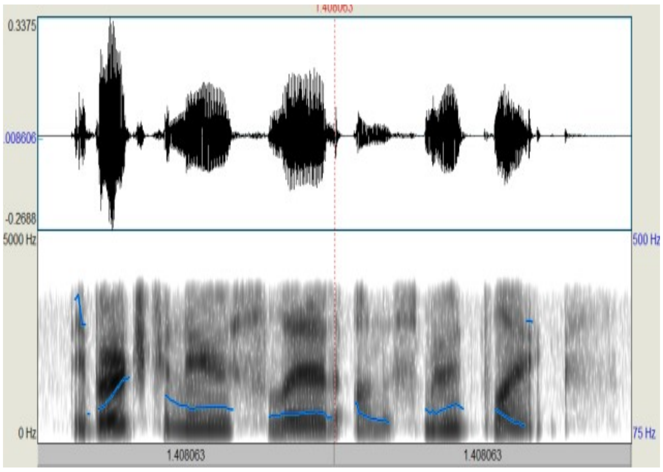
Fig.1-Block diagram for proposed method

Above block diagram explains different steps used in proposed method, what we have explained above in detail.

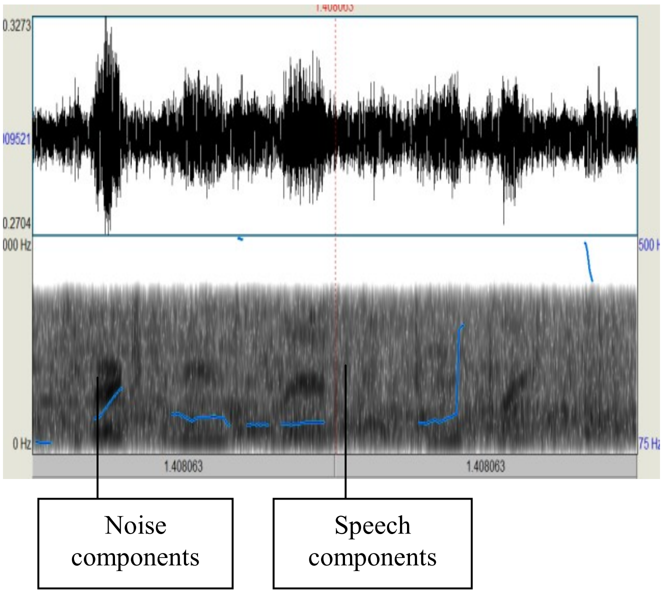
3. RESULTS

In this paper the assessment of speech quality is done using Spectrogram analysis. By using this test Performance of enhanced speech using new method is compared with performance of enhanced speech using simple spectral subtraction.

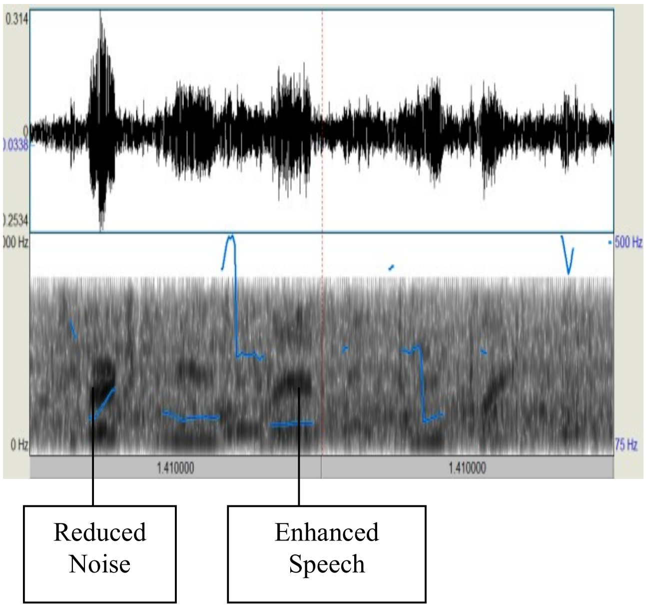
A spectrogram is a plot that represents how the frequency content of a signal changes with time. Hence spectrogram is plot of time and frequency along the x-axis and y-axis respectively and level of grey in spectrogram represents the amount of energy in the signal at any given time and frequency. Whitish colour in spectrogram is because of silence speech, and frequency regions where there is little energy.



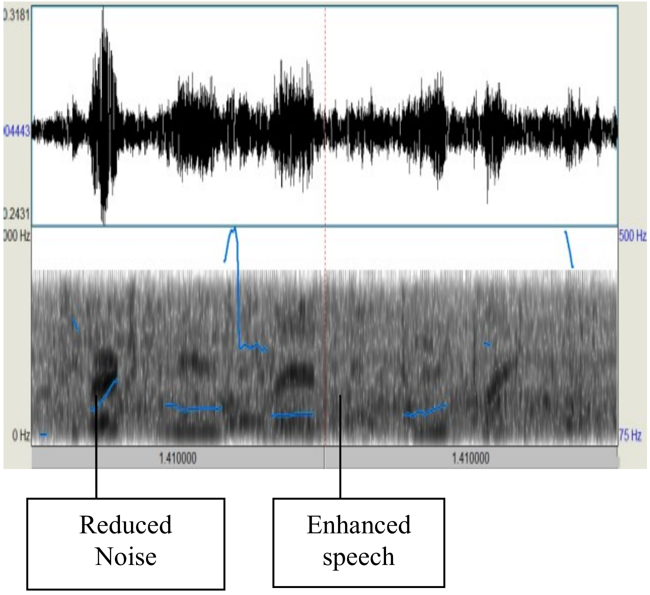
(a)



(b)



(c)



(d)

Fig.2. Time plot and Spectrogram plot of (a)Clean speech, (b)Noisy speech (c)Enhanced speech with simple spectral subtraction,(d)Enhanced speech with proposed Method for airport noise at SNR of 0dB

From fig.2 it is observed that dark gray colour at place of noise components in Fig.2 (b) is converted into light gray colour in Fig.2(c) and very light gray colour in Fig.2(d) which indicates the reduction of or absence of noise components. Along with this, dark gray colour and very dark gray colour is observed at place of speech components in Fig.2(c) and (d) respectively, compared to in Fig.2 (b) which indicates enhancement of clean speech. Hence form the Fig.2 it is proved that better enhancement is done with new method compared to simple spectral subtraction method.

CONCLUSION

The problem associated with simple spectral subtraction is considered in this paper. And this problem is solved using proposed approach in which the crosses spectral term is taken into account. In proposed method we derived the equation which is the product of noisy speech and noise to obtain cross spectral terms. All results show that the proposed method provides consistently better results in the sense of better clarity in spectrogram of proposed method than that of simple spectral subtraction method hence providing a more enhanced speech. From the results it is also concluded that proposed method performs well in such situations where clean speech is corrupted by correlated noise and at low SNR.

ACKNOWLEDGMENT

The authors wish to thank Dr. Shridhar S.K for fruitful and helpful discussions in the course of this work. Author also wishes to thank the authorities of Basaveshwara Engineering College Bagalkot for providing the required facilities in order to complete this portion of work.

REFERENCES

[1]. Sukhvinder Kaur1 , Anil Garg , Pankaj Bactor "An Overview Of Speech Enhancement Techniques And Evaluation" Vol. 1, Issue VI, July 2013 ISSN 2320-6802

[2]. Benesty, sondi,Houng "Springes Hand Book on Speech Processing"

[3]. Paruav Goel and Anil Garg "Review of spectral subtraction Technique for speech enhancement" IJECT Vol 2 December 2011.

[4]. R. Martin, ―Speech enhancement using a minimum mean-square error short-time spectral amplitude estimation,‖ in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, pp. 504—512, 2002

[5]. M. Seltzer, J. Droppo, and A. Acero, "A harmonicmodel-based front end for robust speech recognition," Eurospeech, 2003, To appear.

[6]. C. Eunjoon, J. Smith, and B. Widrow, "Exploiting the harmonic structure for speech enhancement," in IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 4569–4572, 2012.

[7]. N. Upadhyay and A. Karmakar, "The Spectral Subtractive-Type Algorithms for Enhancement of Noisy Speech: A Review," International Journal of Research and Reviews in Signal Acquisition and Processing, Vol. 1, No. 3, 2011, pp. 43-49.]

[8]. Kaladharan N "Speech Enhancement by Spectral Subtraction Method" *International Journal of Computer Applications (0975 – 8887) Volume 96– No.13, June 2014*

[9]. Md. T. Islam, C. Shahnaz and S.A. Fattah, "Speech Enhancement Based on a Modified Spectral Subtraction Method", *IEEE 2014*, Pg No 1085 – 1088

[10]. Yang Lu, Philipos C. Loizou, "A geometric approach to spectral subtraction" , *Speech Communication* 50 (2008) 453–466

[11]. Yi Hu , Mukul Bhatnagar and Philip C. Loizou, "A Cross-Correlation Technique for Enhancing Speech Corrupted with Correlated Noise", *2001 IEEE*, No 673-676

[12]. Sukhvinder Kaur, Anil Garg "An Effective Evaluation Study of Objective Measures Using Spectral Subtractive Enhanced Signal" International Journal of Research in Engineering and Technology Volume: 04 Issue: 03 | Mar-2015

[13]. Hu, Y., Loizou, P., 2008. Evaluation of objective quality measures for speech enhancement. IEEE Trans. Audio Speech Language Processing 16 (1), 229–238.

[14]. Evans, N., Mason, J., Liu, W., Fauve, B., 2006. An assessment on the fundamental limitations of spectral subtraction. In: Proc. IEEE Internat. Conf. on Acoustics, Speech, Signal Processing, Vol. I. pp. 145–148.