Performance Comparison of STSA Based Speech Enhancement Methods

Boriwal Poojakumari Ramprasad[1], Naveen Jain[2], Vijendra Maurya[3]
[1]Dept of Electronic and Communication
[2, 3]Assistant Professor, Dept of Electronic and Communication
[1, 2, 3] GITS, Udaipur(Raj.) India

Abstract- *Speech enhancement is the term used to describe algorithms or devices whose purpose is to improve some perceptual aspects (quality) and intelligibility of speech for the human listener or to improve the speech signal. Speech enhancement algorithms have been applied to problems as diverse as background noise removal, cancellation of reverberation and multi-speech separation (speaker separation) in modern speech communication systems. In this paper, several short time spectral amplitude (STSA) based speech enhancement methods are discussed. This paper evaluates the performance of STSA based algorithms for different types of noises. The experimental and simulation results are presented in this paper. These methods are very effective in suppressing additive white noise but spectral subtraction based STSA methods generates another synthetic noise known as musical noise. The modifications required are suggested to overcome the reported problems.*

Keywords- Array processing, Matlab, Reverberation, White Noise, VAD

I. INTRODUCTION

Speech is used by the human being to communicate messages. Historically, just after the invention of telephone by Alexander Graham Bell in the year 1850, the task of advanced speech signal enhancement in the field of communication engineering was required to begin.

Lately, the use of wireless communication in cellular and mobile phones with or without 'hands free' system is increased in voice messaging service (voice mail), call service centers, voice over internet protocol (VOIP) phones, cord less hearing aids etc. require efficient real time speech enhancement strategies to combat with additive noise and convolutive distortion (e.g., reverberation and echo) that occur in any communication system [6]. The speech enhancement techniques can be divided into two basic categories: (i) Single channel and (ii) Multiple channels (array processing) based on speech received from single microphone or multiple microphone sources respectively [3]. Also the proposed method have other characteristics like real-time implementation, reasonable computational complexity while processing, low level of speech distortion, operation with low level SNR, separation as cleaned speech signal, adaptation to background noise, controlled level of noise suppression in speech. [1]
Page|1

The introduction about proposed research

The proposed method is statistical based method it has named MMSE-LSA_modified or MMSE 85_modifed algorithm. It has same output speech equation as MMSE-LSA or minimum mean square error logarithmic spectral amplitude. In proposed research, we have made some parameter more adaptive from other speech enhancement methods so there is trade-off between musical noise and transient noise. So we have solved the problem of non stationary and colored noise speech signal. In proposed paper we analysed the Spectrographic results using MATLAB. Results shows that MMSE-LSA modified is best approach to overcome the reported problems.

II. STSA BASED METHODS

The STSA based methods are most conventional transform doamain methods. They are single channel enhancement methods work on only extracting the spectral amplitude not phase from noisy signal at output. So they are called Short Time Spectral Amplitude methods[2].

Table 1 list of symbols

Symbol	Meaning
$y(n)$	Degraded Speech signal
$x(n)$	Clean speech signal
$d(n)$	Additive noise
α	Over Substraction Factor
β	Spectral floar Parameter
η	Smoothing Constant
K	Discrete frequency bin
δ	Tweaking Factor
$\xi(K)$	A priori SNR at frequency bin K $= \frac{\lvert\hat{X}(K)\rvert^2}{\lvert\hat{D}(K)\rvert^2}$
$\gamma(K)$	*A posteriori* SNR at frequency bin K $= \frac{\lvert Y(K)\rvert^2}{\lvert\hat{D}(K)\rvert^2}$
i	Frequency band
$\phi_y(K)$	Phase of signal y(n) at frequency bin K

1. Spectral Substraction Methods

A. Magnitude Spectral Subtraction (MSS)

S.F.Boll first proposed Spectral subtraction method[2]. taking only magnitude of spectrum we can write

$$|\hat{X}(K)| = \begin{cases} |Y(K)| - |\hat{D}(K)| \; if \; |Y(K)| > |\hat{D}(K)| \\ 0 \quad else \end{cases} \quad (1)$$

Hence, original speech estimate is given by,

$$X(K) = [|Y(K)| - |\hat{D}(K)|]e^{j\phi_y(K)} \quad (2)$$

B. Power Spectral Subtraction (PSS)

The preceding discussion of magnitude spectrum subtraction can be extended t power spectrum domain as[2]

$$|\hat{X}(K)|^2 = \begin{cases} |Y(K)|^2 - |\hat{D}(K)|^2 \; if \; |Y(K)|^2 > |\hat{D}(K)|^2 \\ 0 \quad else \end{cases} \quad (3)$$

C. Berouti Spectral Subtraction (BSS)

In Spectral subsraction method, one limitation is that algorithms introduce another artifact called musical noise. Berouti *et al*[3]. proposed an important variation of the original method, which improves the noise reduction compare to the basic spectral subtraction. It introduces an over subtraction factor $(\alpha \geq 1)$ and spectral floor parameter $(0 < \beta < 1)$; and it is defined as

$$|\hat{X}(K)|^2 = \begin{cases} |Y(K)|^2 - \alpha|\hat{D}(K)|^2 \; if \; |Y(K)|^2 > (\alpha+\beta)|\hat{D}(K)|^2 \\ \beta|\hat{D}(K)|^2 \quad else \end{cases} \quad (4)$$

The parameter β controls the amount of remaining residual noise and the amount of perceived musical noise.

D. Multiband Spectral Subtraction (MBSS)

This method is proposed by S.D.Kamath [4] performs spectral subtraction with different over subtraction factor in different non-overlapped frequency bands. It is based on the fact that speech signal is not affected uniformly over the whole spectrum. Some frequencies will be affected more adversely than others depending on the spectral characteristics of the noise. This can address the problem of colored noise reduction. The spectral subtraction rule in i^{th} frequency band is given by

$$|\hat{X}_i(K)|^2 = \begin{cases} |\bar{Y}_i(K)|^2 - \alpha_i\delta_i|\hat{D}_i(K)|^2 \; if \; |Y_i(K)|^2 > \alpha_i\delta_i|\hat{D}_i(K)|^2 \\ \beta|\bar{Y}_i(K)|^2 \quad else \qquad for \; b_i \leq K \leq e_i \end{cases} \quad (5)$$

where the spectral floor parameter β is set to 0.002. The over subtraction parameter in i^{th} band is specified as

$$\alpha_i = \begin{cases} 4.75 & SNR_i < -5\,dB \\ 4 - \frac{3}{20}(SNR_i) & -5\,dB \leq SNR_i \leq 20\,dB \\ 1 & SNR_i > 20\,dB \end{cases} \quad (6)$$

The additional over subtraction factor δ_i; called tweaking factor provides additional degree of control in each frequency band.

$$\delta_i = \begin{cases} 1 & f_i < 1\,kHz \\ 2.5 & 1\,kHz \leq f_i \leq \frac{F_s}{2} - 2\,kHz \\ 1.5 & f_i > \frac{F_s}{2} - 2\,kHz \end{cases} \quad (7)$$

2. Wiener Filtering methods

The Wiener filter is an optimal filter that minimizes the mean square error of a desired signal in time domain. The Wiener filter is given by[5]

$$|\hat{X}(K)| = \frac{\xi(K)}{1+\xi(K)}|Y(K)| \quad (8)$$

This filter is a function of *a priori* SNR.

A. Decision direct (DD) approach

As a solution, Ephraim and Malah [6] proposed the decision directed rule to estimate this ratio and it is used by Scalart *et al.* [9] with Wiener filter. The decision direct rule for frame *t* is given by

$$\xi^{(t)}(K) = \eta\frac{|\hat{X}^{(t-1)}(K)|^2}{|\hat{D}^{(t)}(K)|^2} + (1-\eta)max(\gamma^{(t)}(K)-1,0) \quad (9)$$

Where $0 \leq \eta \leq 1$ is smoothing constant and normally it is set to 0.98.

3. Statistical model based

An alternate approach is to use nonlinear estimators of the magnitude spectrum only using various statistical model and optimization criteria.

A. Maximum likelihood (ML) approach

First McAulay and Malpass proposed and applied the ML approach [6]. The pdf of noise Fourier transform coefficients is assumed to be zero-mean complex Gaussian. Based on this the ML estimation is given by

$$|\hat{X}(K)| = \frac{1}{2}\left(|Y(K)| + \sqrt{|Y(K)|^2 - |\hat{D}(K)|^2}\right) \quad (10)$$

B. Minimum mean square error (MMSE) approach

This method takes MMSE estimate of spectral amplitude rather than complex spectrum as in Wiener filter. The MMSE-SA optimization suggested by Ephrahim and Malah [6] is given by the equation

$$|\hat{X}(K)| = \frac{\sqrt{\pi}}{2}\frac{\sqrt{v(K)}}{\gamma(K)}e^{-\frac{v(K)}{2}}\left[(1+v(K))I_0\left(\frac{v(K)}{2}\right) + v(K)I_1\left(\frac{v(K)}{2}\right)\right]|Y(K)|;$$
$$v(K) = \frac{\xi(K)}{1+\xi(K)}\gamma(K) \quad (11)$$

Here $I_0(.)$ and $I_1(.)$ denote the modified Bessel functions of zero and first order.

C. MMSE log spectral amplitude (MMSE-LSA)

As a variant, Ephrahim and Malah [5] proposed MMSE log spectral amplitude (MMSE-LSA) estimator based on the fact that a distortion measure with the log spectral amplitudes is more suitable for speech processing. It minimizes the mean square error of the log amplitude spectra and the estimate of the clean speech is given by the equation

$$|\hat{X}(K)| = \frac{\xi(K)}{1+\xi(K)}\exp\left(-\frac{1}{2}\int_{v(K)}^{\infty}\frac{e^{-t}}{t}dt\right)|Y(K)| \quad (12)$$

This method reduces the residual noise considerably without introducing much speech distortion.

D.Proposed method (mmse-lsa modified/mmse85 modified)

The proposed method has same optimization clean speech equation as MMSE-LSA but we have made some parameter more adaptive than MMSE-LSA. The decision direct rule is introduced as given in wiener filtering methods. clean speech is given by the equation

The decision direct rule for frame *t* is given by equation (9)

Proposed modification of *a* priori snr

In the expression of given by MMSESTSA-LSA, the choice of η is critical.In general, η is given a value very close to 1. It has been shown that the closer the value of is to 1, the lesser is the *musical noise*, but there is more "transient distortion" to the resulting signal. Balancing these two effects, reported resultsin the literature usually set a constant value in the range 0.95–0.99 with a few exceptions. But using a constant has certain drawbacks.

Therefore, it will be logical to use adaptive rule for η

$$\eta(K)^t = \frac{1}{1+\left(\frac{\xi(K)^t - \bar{\xi}(K)^{t-1}}{\xi(K)^t+1}\right)^2}$$

In proposed method we η make adaptive for different frequency characteristics. so there is always trade off between musical noise and transient distortion.

III. SIMULATION OF SPECTROGRAPHIC RESULTS

The simulation is done using MATLAB software. We have analysed the result using Spectrograms. NOIZEUS is a noisy speech corpus recorded among research groups of IEEE. The noisy database contains 30 IEEE sentences [4]

produced were corrupted by eight different real-world noises (train, babble, car, exhibition hall, restaurant, street, airport and train-station noise)at different SNRs and recorded in sound proofing booth using Tucker DavisTechnology(TDT). In our experiment the AWGN of desired SNR (usually in the range 0-15dB) is added to the clean speech signal.

Spectrogram of both original speech and noisy speech and enhanced output using MMSE-LSA_modified algorithm are shown in Fig. 1. Remaining algorithms Spectrogram are shown in Fig. 3 and Fig. 4. As it can be seen there is a constant background noise present in the noisy speech.Spectrograms of remaining eight methods are shown in figure 3 and figure 4.The spectrographic analysis shows that if we compare the results with original spectrogram ,In statistical modeling method the ML method is worst while the MMSE-STSA85 (MMSE-LSA) modified gives best result and in spectral substraction , performance of MBSS is good and The Wiener filter method also gives less random dots but slightly more distortion in spectrogram (results in more residual noise or speech distortion) compared to MBSS and MMSESTSA85 modified methods. The MMSE-LSA_modified algorithm is found the best from these two from listening point of view. A more useful judgement is obtained using objective measures described in the following section.

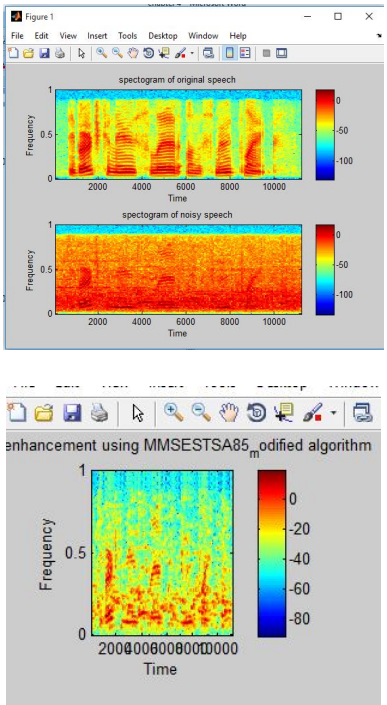

Fig. 1

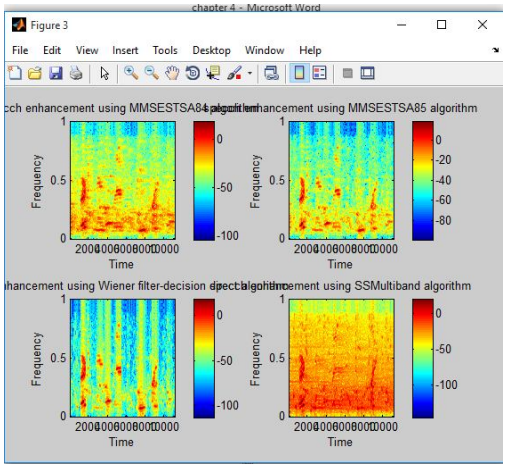

Fig. 2

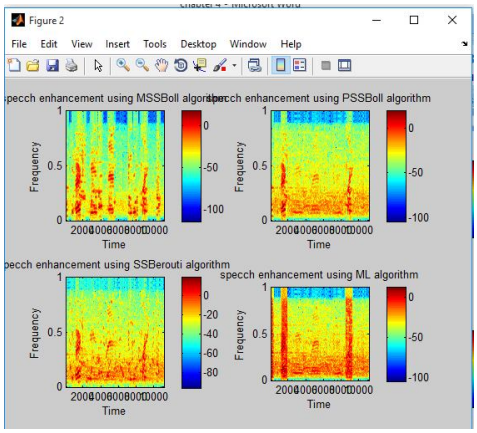

Fig 3

IV. ACKNOWLEDGMENT

I extend our greatest thanks to almighty GOD , who so ever give me such energy to complete this task. I would like to express my thanks to my guide Mr. Naveen Jain and Vijendra Maurya, Assistant Professor, Department of EC engineering, GITS, Udaipur(Raj.), India for their support during my work for this research. Finally, I would like to express my special gratitude to my family, specially my father Mr. Ramprasad Boriwal.

REFERENCES

[1] Thomas F. Quatieri, Discrete-time Speech Signal Processing, 1st Indian reprint, Pearson education signal processing series, Delhi, 2004.

[2] S. F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-27, pp.-113-120, April 1979.

[3] M. Berouti, R. Schwartz, J.Makhoul, "Enhancement of speech corrupted by acoustic noise," ICASSP'79, Vol.4, pp. 208-211 , April 1979.
[4] S..Kamath, P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, 2002.
[5] Ephrahim, D.Malah, "Speech enhancement using minimum mean square error short time spectral amplitude estimator," IEEE Trans. on Acoustics, Speech and Signal Processing,
Vol.ASSP-32,no. 6, pp. 1109-1121, December 1984.
[6] P.Scalart, J.V. Filho, "Speech enhancement based on a priori signal to noise ratio estimation,"in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 96, pp. 629-632, May 1996.