IOSR Journal of Computer Engineering (IOSRJCE)
ISSN: 2278-0661, ISBN: 2278-8727Volume 8, Issue 1 (Nov. - Dec. 2012), PP 48-53
www.iosrjournals.org

A Two Stage Algorithm for Denoising of Speech Signal

[1]Shajeesh. K. U., [2]Sachin Kumar. S., [3]K. P. Soman.
[1,2,3](Centre for Excellence in Computational Engineering and Networking, Amrita Vishwa Vidyapeetham, Coimbatore, Tamil Nadu, India)

Abstract : *Spectral Subtraction based speech denoising suffers the major problem of introduction of new type of noise called musical noise. In this paper we introduce a two stage algorithm for denoising of noisy speech signal by minimizing the musical noise and additive white Gaussian noise. The algorithm works in two stages. First spectral subtraction method is applied on the noisy speech signal after applying pre-emphasis filter. Since the speech is highly corrupted by white Gaussian noise (3 dB SNR), the method introduces a musical noise. The enhanced speech is the processed using Savitzky–Golay smoothing filter to reduce the musical noise. The various objective as well as subjective measures are evaluated for the two stage method and compared against the individual methods. The two stage algorithm produces enhanced speech signal having high perceptual quality and intelligibility.*

Keywords: *Speech Enhancement, Spectral Subtraction, Savitzky Golay Filter, Quality Evaluation Metric, Speech Denoising.*

I. INTRODUCTION

The speech signal recorded in real time environments like in a class room or in an auditorium suffers background noise, reverberation and sometimes the speech from other speakers to result in degraded (noisy) speech [1]. The degraded speech is annoying to perceive and difficult for the listener to understand the message from it. Presence of background noise in speech signal reduces the performance of the speech processing tasks drastically [2]. Therefore, prior to any speech processing task, the degraded speech need to be processed for perceptual enhancement by removing the background noise. Speech denoising is the process of removing or reducing unwanted sounds from the speech signals other than the signal of interest.

Most of the speech processing systems rely on the quality and intelligibility of the speech signal. Quality of speech refers to attributes like the pleasantness, naturalness and speaker recognizability. Intelligibility refers to clarity of the message or information content of the speech signal [3]. The denoising method should take care of both intelligibility and quality of speech with at most importance.

For the last few decades several methods have been introduced for the enhancement of speech signals corrupted by background noise. The Spectral subtraction [4] is the simplest and widely used one such method. In spectral subtraction method, we process the signal in frequency domain. First we find the power spectrum of the noisy speech. We keep the phase intact without any change. Then we find the noise estimate and subtract this from the power spectrum and find the inverse Fourier transform with original phase to get the enhanced speech. The method is simple but introduces new noise types such as musical noise. Many methods have been introduced to reduce the musical noise in spectrally processed signals. As the noise intensity of noisy speech signal increases, musical noise also increases.

In this paper, we propose a two stage algorithm for enhancement of speech signal corrupted by background noise. In the first stage, noisy speech signal is processed with spectral subtraction algorithm. The enhanced signal from the first stage is fed as an input to second stage of algorithm, which is a recursive Savitzky – Golay filter. The result of proposed method is compared against the result of individual speech enhancement methods Spectral subtraction and S-G filter. The rest of the paper is organized as follows. Section 2, briefly describes the basic theory behind Spectral Subtraction and Savitzky–Golay filter. This section also describes various speech quality evaluation metrics used in this paper. Section 3 covers the experimental results and finally the conclusion is provided in section 4.

II. MATERIALS AND METHODS

2.1 Spectral Subtraction

Spectral subtraction process the signal in frequency domain. The method is based the assumption that the power spectrum of a noisy speech signal equal to the sum of the clean speech signal spectrum and the noise spectrum. That is the noise and the clean speech is uncorrelated and can be separated [4].

In this method, first we segment the signal in to small frames. Power spectra of the speech signal are found out for each windowed frame. The original phase, $\theta(\omega)$ of the speech signal is kept as such for reconstruction purpose. An estimate of noise power spectrum is estimated from the silence period of the speech

signal (minimum amplitude portion) and is subtracted from the power spectra of the noisy speech. The inverse Fourier transform of the resultant signal along with the original phase gives the enhance speech. The method can be represented mathematically as follows:

Let y(n) be the noisy speech signal. The y(n) is composed of the clean speech signal s(n) and noise signal w(n). The noisy speech signal can be represented as follows:

$$y(n) = s(n) + w(n) \tag{1}$$

For each of the windowed noisy speech signal we apply the Fourier transform.

$$Y(\omega) = \sum_{n=0}^{N-1} y(n) e^{-j\frac{2\pi kn}{N}} = |Y(\omega)| e^{j\theta(\omega)} \tag{2}$$

Where $\theta(\omega)$ is the original phase of the input noisy speech signal. Then we find the power spectra. It can be represented as

$$|Y(\omega)|^2 = |S(\omega)|^2 + |W(\omega)|^2 + S(\omega).W^*(\omega) + S^*(\omega).W(\omega) \tag{3}$$

Here $W^*(\omega)$ represents the complex conjugates of $W(\omega)$ and $S^*(\omega)$ represents the complex conjugate of $S(\omega)$. Also $|S(\omega)|^2$ and $|W(\omega)|^2$ are the short term power spectrum of speech and noise respectively. The $|W(\omega)|^2$ is known as the noise power estimate and can be found out by various noise estimation techniques. One such method is explained in [5].

Let $|Y(\omega)|^2$ and $|S(\omega)|^2$ can be replaced as $P_y(\omega)$ and $P_w(\omega)$ respectively, then the Enhanced speech spectrum $P_s^{'}(\omega)$ can be calculated as:

Let $D(\omega) = P_y(\omega) - P_w(\omega)$

$$P_s^{'}(\omega) = \begin{cases} D(\omega), \text{ if } D(\omega) > 0 \\ 0, \text{ othewise} \end{cases} \tag{4}$$

Enhanced speech is obtained by applying inverse Fourier transform on $P_s^{'}(\omega)$ using the original phase.

$$s^{'}(t) = F^{-1}\left\{ \sqrt{P_s^{'}(\omega)}\ e^{j\theta(\omega)} \right\} \tag{5}$$

This method has a disadvantage that it introduces new type of noise, called musical noise in the enhanced speech. Also some of the broadband noise is still present in the enhanced speech.

2.1.1 Musical Noise and the Broadband Noise

If we analyze the power spectrum of white noise, we can see some peaks and valleys in the spectrum. They vary randomly in frequency and amplitude in different speech frames [6]. If we subtract the noise estimate from the noisy speech spectrum, the spectral peaks shifted down and the valleys are set to zero. The spectral peaks will remain after the subtracting the noise estimate. Now there are wider peaks and narrower peaks in the spectrum. The remaining wider peaks are perceived as broadband noise whereas the narrower peaks are perceived as the musical noise since they have relatively large spectral excursions because of the deep valleys in it.

2.1.2 Modified Spectral Subtraction

In spectral subtraction method, musical noise can be reduced by introducing two new parameters the subtraction factor (α) and the spectral floor parameter (β). The modified algorithm is shown below [4].

$$\text{Let } D(\omega) = P_s(\omega) - \alpha P_w(\omega) \tag{6}$$

$$P^{'}{}_s(\omega) = \begin{cases} D(\omega), \text{ if } D(\omega) > \beta P_w(\omega) \\ \beta P_w(\omega), \text{ otherwise} \end{cases} \tag{7}$$

$$\text{with } \alpha \geq 1, \quad \textit{and} \quad 0 < \beta << 1$$

The subtraction factor, α reduces the broadband noise whereas the spectral floor parameter, β, reduces the musical noise by filling in the deep valleys surrounding the narrow peaks. By properly adjusting the value of α and β, we can reduce the musical noise and the remaining broadband noise.

But if the noise intensity is very high (in the order of 2-5 dB SNR), the method fails to remove the musical noise, which is highly intrusive to perceive. The remaining musical noise can be filtered by a polynomial filter, Savitzky-Golay Smoothing filter.

2.2 Savitzky-Golay Filter (SG)

Savitzky Golay filter is a digital, polynomial, smoothing filter. In polynomial filter, the signal samples are fitted using a polynomial function (of a certain degree) with the help of least square method. Through fitting an approximate coefficient value is found for the signal data sample. During filtering operation, the approximate value is found with the help of neighboring points of the windowed signal. In this way the data sample at the centre of the moving window is replaced [7]-[10]. There will be equal number of points to the left and right of the central point. After computing this, the window moves one sample to the right or the window is shifted, to find a polynomial fit to the next central point. This is repeated to all the data points. Considering a 2M+1 data sample window, centered at n=0, an approximate polynomial filter coefficient for the input data sample is calculated as,

$$p(i) = \sum_{k=-\frac{N-1}{2}}^{\frac{N-1}{2}} c_k i^k \ , \tag{8}$$

where ' $p(i)$ ' is the approximate value corresponding to the 'ith' data sample in the window, ' $-M \le i \le M$ ' where 'M' decides the number of data points (ie; 2M+1 data points), 'N' denotes the order of the polynomial, and ' c_k ' denotes the coefficient of the polynomial. For the input data window with 2M+1 samples, ' $x(-M)...x(M)$ ', a least square polynomial fit by polynomial vectors $p(-M)...p(M)$ with degree 'm' is found. To estimate the approximated coefficients for data points on both boundaries, zeros are padded at both ends. Consider a polynomial basis matrix 'A' is needed with basis as $t^0, t^1,\ldots,t^N$.

$$A = \begin{bmatrix} | & | & | & | & | & | \\ t^0 & t^1 & . & . & . & t^N \\ | & | & | & | & | & | \end{bmatrix} \quad \text{and} \quad A^t = \begin{bmatrix} - & t^0 & - \\ - & t^1 & - \\ - & t^2 & - \\ - & . & - \\ - & . & - \\ - & t^N & - \end{bmatrix}$$

For example, 'M=3' and 'm=5', the transposed polynomial basis matrix will be,

$$A^t = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -3 & -2 & -1 & 0 & 1 & 2 & 3 \\ 9 & 4 & 1 & 0 & 1 & 4 & 9 \\ -27 & -8 & -1 & 0 & 1 & 8 & 27 \\ 81 & 16 & 1 & 0 & 1 & 16 & 81 \\ -243 & -32 & -1 & 0 & 1 & 32 & 243 \end{bmatrix} \begin{matrix} \leftarrow t^0 \\ \leftarrow t^1 \\ \leftarrow t^2 \\ \leftarrow t^3 \\ \leftarrow t^4 \\ \leftarrow t^5 \end{matrix}$$

For a linear system of equation of the form Ax=b, with matrix 'A' having more rows than columns or more equations than unknowns, then 'b' will not lie in the column space of 'A'. In such situation the solution will be apporximate. The error vector 'e' is denoted as , e=b-Ax. When this error reduces, b_{new} the new solution will be $A\, x_{new} = b_{new}$. When 'e' becomes zero, an exact solution exists. Through least square projection method 'x_{new}' is obtained by solving

$$\left(A^t A\right) x = A^t b \tag{9}$$

$$\therefore x_{new} = \left(A^t A\right)^{-1} A^t b \tag{10}$$

This equation is obtained from the fact that the error vector is perpendicular to the column space of 'A'. Therefore the new estimated coefficients corresponding to the input data sample is

$$b_{new} = A(A^t A)^{-1} A^t b \tag{11}$$

2.3. Quality Evaluation Metrics

Quality of the enhanced speech evaluated based on some metrics. They are categorized as objective quality evaluation method and subjective quality evaluation methods.

2.2.1Subjective Quality Evaluations:

In subjective quality evaluation, a group of listeners (also known as test subjects) rate the enhanced speech signal based on three factors. They are:

- The speech signal part in the enhanced speech alone is rated based on signal distortion (SIG).
- The background noise alone is rated based on background disturbances (BAK).
- The overall quality of enhanced speech is rated as the mean of SIG and BAK Scale values (OVRL).

The rating is based on a five point scale [11] and is listed in the Table 1.

Table 1: Description of SIG and BAK Scale

Rating	SIG Scale	BAK Scale
5	Purely Natural, no degradation	Not perceptible
4	Fairly Natural, slight degradation	Somewhat noticeable
3	Somewhat natural, somewhat degraded	Noticeable but not intrusive
2	Fairly unnatural, fairly degraded	Fairly Noticeable, somewhat intrusive
1	Quite unnatural, Highly degraded	Quite Noticeable, Highly Intrusive

2.2.2 Objective Quality Evaluations:

Subjective Quality Measures are time consuming and costly. So we go for Objective Quality Measures. These are based on mathematical measures. Objective Quality Measures evaluate the speech quality by comparing clean speech and enhanced speech based on some mathematical measures. In this paper we have chosen four objective measures, Segmental SNR (SNRseg), Weighted Slope Spectral distance (WSS), Perceptual Evaluation of Speech Quality (PESQ) and Log Likelihood Ratio (LLR) [11][12]. Enhance speech with a lower value of WSS and higher value of SNRSeg indicates better quality. The LLR lies in the range between 0 and 2.

Conventional objective measures like SNRseg and LLR, are not correlated highly with speech/noise distortions and overall quality. In order to overcome this, a new measure called composite measures are introduced. It is formed as the linear combination of basic objective measures as well as subjective measures to form a new and more accurate measure. In this paper, we have chosen a composite measure for signal distortion (CSIG), a composite measure for noise distortion (CBAK), and a composite measure for overall speech quality (COVRL). These values are obtained by linearly combining the existing objective measures by the following relations [12]:

$$Csig = 3.093 – 1.029LLR + 0.603PESQ - 0.009WSS$$
$$Cbak = 1.634 + 0.478PESQ - 0.007WSS + 0.063segSNR$$
$$Covl = 1.594 + 0.805PESQ - 0.512LLR - 0.007WSS \quad (12)$$

III. RESULTS AND DISCUSSION

The proposed two stage algorithm is experimented with five speech signals from TIMIT database. The additive white Gaussian noise is added to signal at noise intensity level of 3 dB SNR using matlab.

The two stage method is applied on the speech files and the result is shown in figure 1. Figure 1 (a) is the clean speech signal, 1(b) is the noisy speech and 1(c) is the enhanced speech for two stage algorithm. Results show that although noise intensity is very high, almost all the noise is removed from the speech signal. The individual method is also applied on the speech signal and the results are shown in figure 2 and figure 3. Figure 2 (a) is the clean speech signal, 2(b) is the noisy speech and 2(c) is the enhanced speech for spectral subtraction method. Figure 3 (a) is the clean speech signal, 3(b) is the noisy speech and 3(c) is the enhanced speech for Savitzky – Golay filter.

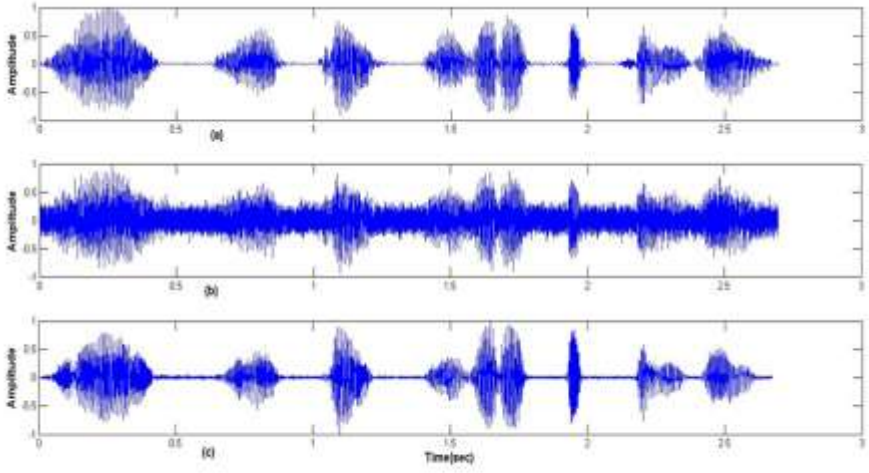

Figure 1: Two Stage Algorithm for Speech Enhancement

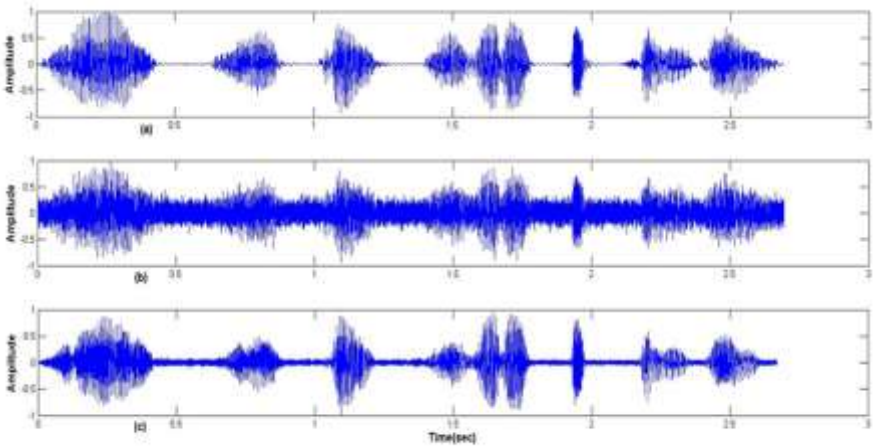

Figure 2: Spectral Subtraction Method for Speech Enhancement

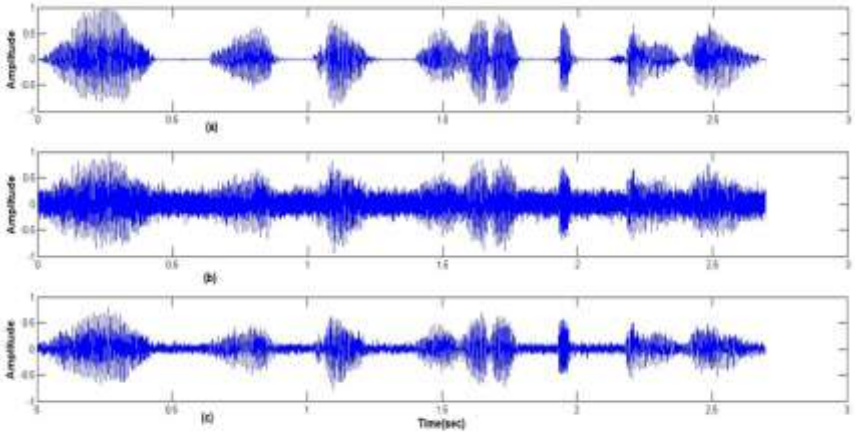

Figure 3: S-G filter based denoising of speech signal

Various objective quality measures are evaluated for the proposed method. The objective quality measures are evaluated in two levels [10]. In the first level, the objective measures are evaluated for the clean speech signal and noisy speech signal. This measure gives to what extend the clean speech is degraded by background noise. In second level, the clean speech and the enhanced speech signal is processed. This gives the measure of similarity between enhanced speech signal and clean speech signal. The results are shown in table 1. The objective measures for spectral subtraction alone is evaluated and shown in table 2. Comparing table 1 and 2, it is evident that two stage algorithm has considerable improvements over individual methods.

Table 1: Objective Quality Measures for Two Stage Algorithm for Denoising

Signal	B/w	CSig	CBak	COvrl	LLR	SNRSeg	WSS	PESQ
	O & N	-4.01681	1.388201	-1.44125	7.09642	-6.46757	68.04901	1.33473
SA1	O & E	-0.26196	1.292952	0.271181	3.080899	-3.30234	99.94096	1.185326
	O & N	-4.11492	1.121777	-1.69382	6.94519	-7.50781	69.78645	0.939905
SA2	O & E	-0.9989	0.792772	-0.31839	3.396593	-6.10852	124.3088	0.865628
	O & N	-3.64839	1.793758	-0.77923	7.307128	-7.38919	65.7773	2.271377
SA3	O & E	-0.36238	1.306368	0.353141	3.286602	-4.50567	107.6705	1.485186
	O & N	-4.49118	0.876743	-2.09742	7.035323	-7.83877	73.98252	0.53235
SA4	O & E	-0.71033	0.803679	-0.30271	3.039953	-4.92231	111.8974	0.550347
	O & N	-4.53935	1.140488	-1.90568	7.358441	-7.22027	69.52008	0.93725
SA5	O & E	-0.61527	0.980291	-0.03115	3.20032	-5.39417	113.1852	1.000878

. Table 2: Objective Quality Measures for Spectral Subtraction Method

Signal	B/w	CSig	CBak	COvrl	LLR	SNRSeg	WSS	PESQ
	O & N	-4.01681	1.388201	-1.44125	7.09642	-6.46757	68.04901	1.33473
SA1	O & E	-3.72709	1.287296	-1.47082	6.483102	-3.82152	90.41185	1.102373
SA2	O & N	-4.11492	1.121777	-1.69382	6.94519	-7.50781	69.78645	0.939905

	O & E	-3.41432	0.885732	-1.50964	5.857883	-6.07094	106.8422	0.799366
	O & N	-3.64839	1.793758	-0.77923	7.307128	-7.38919	65.7773	2.271377
SA3	O & E	-4.1447	0.995918	-1.87577	6.639081	-5.21918	94.03187	0.73002
	O & N	-4.49118	0.876743	-2.09742	7.035323	-7.83877	73.98252	0.53235
SA4	O & E	-3.97893	0.823571	-1.92157	6.299684	-5.65709	99.08739	0.501212
	O & N	-4.53935	1.140488	-1.90568	7.358441	-7.22027	69.52008	0.93725
SA5	O & E	-3.82584	1.227533	-1.42151	6.630464	-5.45856	101.2124	1.351274

The subjective quality measures are evaluated and tabulated in table 3. Subjective measures shows that two stage method gives perceptually better quality enhanced signal compared to the individual methods.

Table 3.4: Subjective Quality Measures for Total S-G filter based denoising

Test Subject	2S-SIG	2S-BAK	2S-OVRL	SS-SIG	SS-BAK	SS-OVRL	SG-SIG	SG-BAK	SG-OVRL
1	5	4	4	4	3	3	3	3	3
2	5	4	4	3	3	3	3	3	3
3	4	4	4	4	3	3	3	3	3
4	4	4	4	3	3	3	3	2	2
5	5	5	5	3	3	3	3	2	2
6	5	4	4	3	3	3	3	3	3
7	4	4	4	4	3	4	3	2	2
8	5	4	4	3	3	3	2	2	2
9	5	5	5	4	4	4	3	3	3
10	4	4	4	4	3	3	3	3	3

IV. CONCLUSION

This paper introduces a two stage algorithm for removal of additive white Gaussian noise from speech signal. First stage, spectral subtraction is applied to the signal. On the second stage, recursive Savitzky – Golay filter is applied. The method is proved to be efficient for removing white Gaussian noise with high noise intensity. Method also removes the musical noise from the speech signal. The method can be used as a pre-processing step in speech applications like speech recognition and speaker verification. The method is compared against the individual methods and results show that two- stage method is ideal for enhancement of speech signal corrupted by white noise.

REFERENCES

[1] P. Krishnamoorthy and S. R. M. Prasanna, "Enhancement of noisy speech by temporal and spectral processing", Speech Communication, Volume 53, Issue 2, February 2011.

[2] S. V. Vasighi and P. J. W. Rayner, "Detection and suppression of impulsive noise in speech communication systems,IEE Proc. of Communications, Speech and Vision, vol. 137, Pt. 1, no. 12, pp. 38-46, February 1990.

[3] J. R. Deller, J. H. Hansen, and J. G. Proakis, "Discrete Time Processing of Speech Signals", 1st edition, Upper Saddle River, NJ, USA: Prentice Hall PTR, 1993.

[4] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," Proc. IEEE Int. Conf. Acoustics, Speech,Signal Processing, pp. 208–211, March 1979.

[5] Rainer Martin, "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics," IEEE Transactions on Speech and Audio Processing, vol. 9, no. 5, July 2001.

[6] Mukul Bhatnagar, A Modified Spectral Subtraction Method Combined with Perceptual Weighting for Speech Enhancement, MS Thesis, The University of Texas at Dallas, August 2002.

[7] A. Savitzky and M. J. E. Golay, "Soothing and differentiation of data by simplified least squares procedures," Anal. chem., vol. 36, pp. 1627–1639, 1964.

[8] J. Riordon, E. Zubritsky, and A. Newman, "Top 10 articles," Anal. Chem., vol. 72, no. 9, pp. 24A– 329A, May 2000.

[9] Ronald W. Schafer, "What Is a Savitzky-Golay Filter?", IEEE Signal Processing Magazine, July 2011.

[10] Shajeesh K. U., Sachin Kumar S., Pravena D. and K. P. Soman, "Speech Enhancement based on Savitzky–Golay Smoothing Filter," International Journal of Computer Applications, Volume 57, No.21, November 2012.

[11] Hu, Y., Loizou, P. C., "Evaluation of Objective Quality Measures for Speech Enhancement". IEEE Transactions on audio, speech and language processing, Vol. 16, No. 1, pp. 229-238, January 2008.

[12] Krishnamoorthy P, An Overview of Subjective and Objective Quality Measures for Noisy Speech Enhancement Algorithms. IETE Tech Rev, Volume 28, Issue 4, p. 292-301, August 2011.