Speech Enhancement based on Savitzky–Golay Smoothing Filter

Shajeesh K. U., Sachin Kumar S., Pravena D., K. P. Soman
Centre for Excellence in Computational Engineering and Networking,
Amrita Vishwa Vidyapeetham,
Coimbatore, Tamil Nadu, India.

ABSTRACT

Speech denoising is the process of removing unwanted sounds from the speech signal. In the presence of noise, it is difficult for the listener to understand the message of the speech signal. Also, the presence of noise in speech signal will degrade the performance of various signal processing tasks like speech recognition, speaker recognition, speaker verification etc. Many methods have been widely used to eliminate noise from speech signal like linear and nonlinear filtering methods, total variation denoising, wavelet based denoising etc. This paper addresses the problem of reducing additive white Gaussian noise from speech signal while preserving the intelligibility and quality of the speech signal. The method is based on Savitzky-Golay smoothing filter, which is basically a low pass filter that performs a polynomial regression on the signal values. The results of S-G filter based denoising method are compared against two widely used enhancement methods, Spectral subtraction method and Total variation denoising. Objective and subjective quality evaluation are performed for the three speech enhancement schemes. The results show that S-G based method is ideal for the removal of additive white Gaussian noise from the speech signals.

General Terms

Speech pre-processing, Speech Enhancement

Keywords

Speech Enhancement, Savitzky–Golay filter, Noise removal, Speech Signal Denoising.

1. INTRODUCTION

Speech enhancement is the process of improving the quality of the speech signal by reducing the background noise and other unwanted sounds from the speech signal. Speech signal quality is often weighed by its clarity, intelligibility and pleasantness [1]. Speech enhancement is a preliminary procedure in the speech processing area, including speech recognition, speech synthesis, speech analysis and speech coding.

Speech signals recorded in a real time environment may contain unwanted sound such as barking of dogs, playing loud announcement by people, sound of fan, AC etc. These are considered under the category of noise. To listeners, these interferences are highly unpleasant and should be reduced in order to enhance the quality and intelligibility of speech signal. Also the speech-signal processing algorithms are based on the assumption that the speech signal is free from background noise. Presence of background noise in speech signals will lower the performance of the speech processing system significantly [2]. So we go for noise removal as a pre-processing step in all the speech processing tasks.

Many methods have been introduced for the enhancement of speech signal corrupted by background noise. The prior information about the type of noise present in the signal and noise characteristics is necessary before processing of speech enhancement algorithm. Here in this paper we mainly focus on the enhancement of speech signal corrupted by fan noise as well as AC noise. Fan noise is additive in nature and has most of its energy concentrated in the lower frequency spectrum. Also there is less overlap with the speech spectrum.

In [3], B. Yegnanarayana et al. presented a novel method enhancement of speech signal using linear prediction residual. The algorithm assumes the noise is additive in nature and follows Gaussian distribution. High SNR regions in the noisy speech are processed in both the spectral and time domains in three levels.Speech and noise regions are identified in time domain, followed by low and high SNR portion, then find the short time spectrum. LP analysis is applied on the blocks and based on the LPC residual signal block is classified as noise and speech region. In the speech regions the inverse spectral flatness is significantly higher than in the noisy regions. As the noise intensity increases, the accuracy and intelligibility of enhanced speech decreases.

In [4], M. Berouti et al. presented a method for enhancing noisy speech signal corrupted by broadband white Gaussian noise based on the modified spectral noise subtraction method. In classical method, an estimate of the noise power spectrum is subtracting from the speech power spectrum, setting negative differences to zero.This method reduces broadband noise, but introduces new type noise "musical noise". The modified spectral subtraction method eliminates the musical noise by introducing another parameter, spectral floor. This parameter prevents the resultant spectral components from going below a pre-set minimum level. The method gives good results if the noise intensity is less. As the intensity of noise increases, enhanced speech quality is also decreases. Various subjective tests were performed to determine the quality and intelligibility of speech enhanced by spectral subtraction method. The scores show that the method is ideal for fan noise removal.

In [5], Dalei Wu et al. presented a novel method for speech enhancement based on compressive sensing. Compressive sensing is a novel technique recently proposed and studied in image and signal processing as an alternative to Shannon's sampling theorem. The method works based on the principle, speech and audio signals are k-sparse while noises are not sparse. Therefore noise from speech signal can be theoretically separable by using compressive sensing. The sparsity can be brought in to the signal by applying the wavelet transform. The noise reduction problem can be formulated as an L1 norm minimization problem with constraint term by using a random partial Fourier transform operator. The optimization problem solved using a gradient descend line search (GDLS) algorithm effectively. The method gives good results when the speech signals corrupted with white Gaussian noise.

In [6], Saeed Gazor and Wei Zhang proposed an efficient speech enhancement algorithm that uses Laplacian – Gaussian mixture. Here, degraded speech signal is first de-correlated and the clean speech components are estimated from the de-correlated speech. The distribution of clean speech is assumed to be Laplacian and the noise signals to be Gaussian. Maximum likelihood (ML) or minimum-mean-square-error (MMSE) estimators are used to estimate the clean speech components. The parameters required for these estimators are adaptively extracted by the ML approach during the active speech or silence intervals, respectively. Also, a voice activity detector (VAD) is used to detect the speech and non- speech regions. The experiment results show that the method performs better than Wiener filtering and computational complexity is very low.

In [12, 13] Savitzky and Golay proposed a method that will fit the data using different polynomials with degree ‘m’, which is equivalent to performing a discrete convolution on data sample with filter coefficient. The filter tries to fit the data sample at the center of the window in time-domain using a lower order polynomial to find a smooth value for all the data point. In other words, it performs a low pass filtering (smoothing) on the data points. In [14] Ronald W. Shafer’s lecture note examines the Savitzky-Golay filter and its properties from frequency-domain.

In this paper, we propose a robust speech enhancement method for speech signal corrupted by fan noise as well as AC noise which is under the category of additive white Gaussian noise. The method is based on Savitzky–Golay (S-G) filter. The S-G filter performs a linear polynomial regression to each sample in the noisy speech.The method is compared against the well-known speech enhancement methods, Spectral subtraction and Total variation denoising. Section 2, briefly describes the basic theory behind Savitzky–Golay filter and the various quality evaluation metrics used to evaluate the performance of the method. Section 3 covers the experimental results and finally the conclusion is provided in section 4.

2. MATERIALS AND METHODS

2.1 Savitzky-Golay Filter

The basic idea in S-G filter is to fit a polynomial of a certain degree to the data points. During digital filtering operation, the data point is replaced by an unweighted average value computed from its neighboring points In SG filter, an approximated value for the data sample at the centre of the moving window is found [12, 13, 14]. 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,

$$p(i) = \sum_{k=0}^{N} c_k i^k \ , \qquad (1)$$

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. In SG filter, the approximated coefficient is estimated for the data sample at the centre of the window. When the window moves by ‘k’ steps, the same process is repeated. To estimate the approximated coefficients for data points on both boundaries, an adjustment at the boundaries will serve the purpose (in this experiment zeros are padded at both ends). In SG filter, the new data coefficients are estimated through least square based polynomial fitting. For this a polynomial basis matrix ‘A’ is needed with basis as t^0, t^1,…,t^N. . For example, ‘M=3’ and ‘m=5’, the transpose d polynomial basis matrix will be,

$$A = \begin{bmatrix} | & | & | & | & | & | \\ t^0 & t^1 & . & . & . & t^N \\ | & | & | & | & | & | \end{bmatrix}$$

And

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

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

For a linear system of equation of the form Ax=b, in which matrix ‘A’ contains more rows than columns or more equations than unknowns, then ‘b’ will not lie in the column space of ‘A’. In such situation an approximate solution is found. The error vector ‘e’ will be, e=b-Ax. When error length reduces, b$_{new}$ will be the new solution (A x$_{new}$ = b$_{new}$). When ‘e’ becomes zero, an exact solution exists. Through least square projection method ‘x$_{new}$’ is obtained by solving

$$\left(A^t A\right) x = A^t b \qquad (2)$$

$$\therefore x_{new} = \left(A^t A\right)^{-1} A^t b \qquad (3)$$

This equation is obtained from the fact that the error vector will be perpendicular to the column space of ‘A’. Therefore the new estimated coefficients corresponding to the data point will be

$$b_{new} = A(A^t A)^{-1} A^t b \qquad (4)$$

2.2 Quality Evaluation Metrics

The quality of enhanced speech is often measured using objective quality evaluation method and subjective quality evaluation methods.

2.2.1 Subjective Quality Evaluations:

Subjective quality evaluations are based on the opinion of a group of listeners also known as test subjects. Listeners rate the enhanced speech signal based on three factors. They are:

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

The enhanced speech quality is expressed using a specific unit, called Mean Opinion Score (MOS). The SIG and BAK scale [7] are 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 evaluation method is time consuming and costly. So we go for objective quality measures. They are evaluated based on the mathematical measures. For the evaluation purpose, we have used four objective measures such as Segmental SNR (SNRseg), Weighted Slope Spectral distance (WSS), Perceptual Evaluation of Speech Quality (PESQ) and Log Likelihood Ratio (LLR) [8]. A good quality enhanced speech may have a higher SNRSeg value and a lower WSS value compared to the noisy signal. LLR may vary in the range between 0 and 2.

Composite objective measures are the linear combination of two or more basic objective measures and subjective measures to form a new and more accurate measure. These measures highly correlate with speech/noise distortions and overall quality. 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 [8]:

$$Csig = 3.093 - 1.029LLR + 0.603PESQ - 0.009WSS$$

$$Cbak = 1.634 + 0.478PESQ - 0.007WSS + 0.063segSNR$$

$$Covl = 1.594 + 0.805PESQ - 0.512LLR - 0.007WSS \tag{5}$$

3. RESULTS AND DISCUSSIONS

S-G Filter based speech enhancement method presented in section 2.1 is used for experimenting with various speech signals. Testing speech signals are recorded with 16 KHz sampling rate 16 bit resolution and then the speech was stored as uncompressed .wav format. The white Gaussian noise is simulated using matlab in various noise intensities. Some test speech signals are recorded in the presence of fan noise.

The SG filter is applied on speech signal corrupted by white Gaussian noise with intensity varies from 30 db to 10 db. The parameters such as frame length and degree are adjusted such a way that the clarity and intelligibility of enhanced speech is maximum. The optimal value for frame length and degree are found to be 9 and 3 respectively. The method gives extremely good results for the noisy speech with noise level 30 db. As the noise intensity increases, performance of the system also reduces. Figure 1 shows the out of S-G filter based denoising applied on a signal at noise level 10 db. Figure 1(a) is clean speech signal, 1(b) is noisy speech signal, 1(c) is enhanced speech signal and 1(d) is the error signal. Figure 2 shows the more detailed representation of processed signal at noise level 10 db. Figure 2(a) represents the clean speech signal, 2(b) is noisy speech signal and 2(c) is enhanced speech signal.

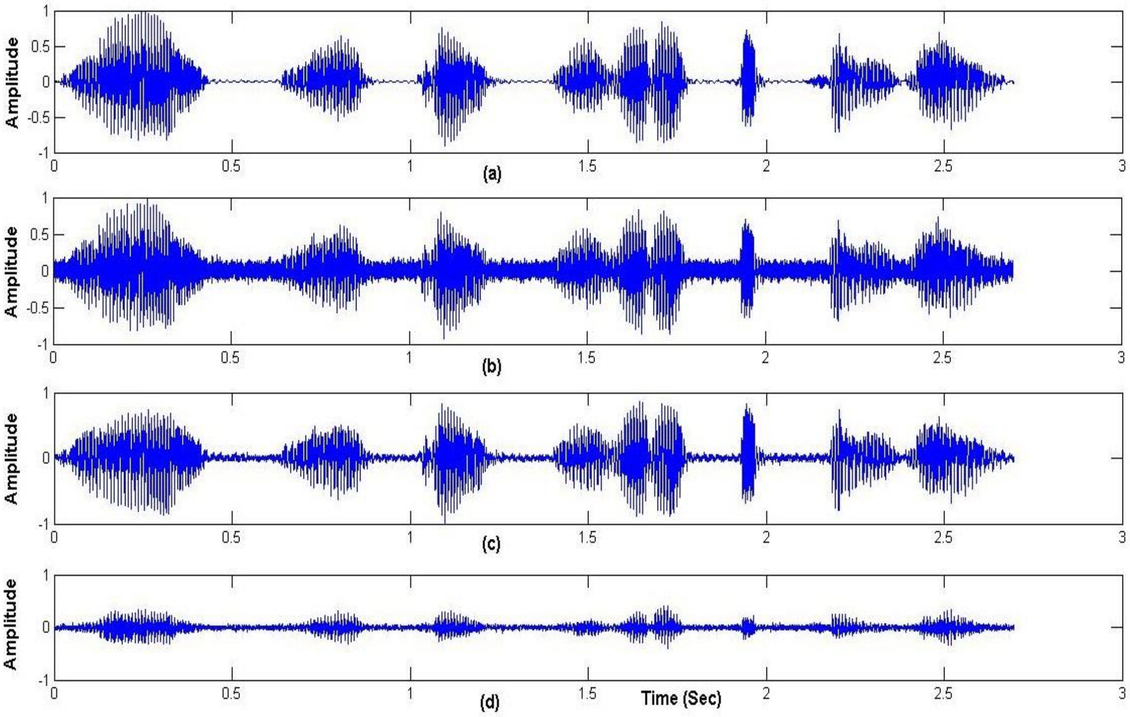

Figure 1: S-G filter based denoising

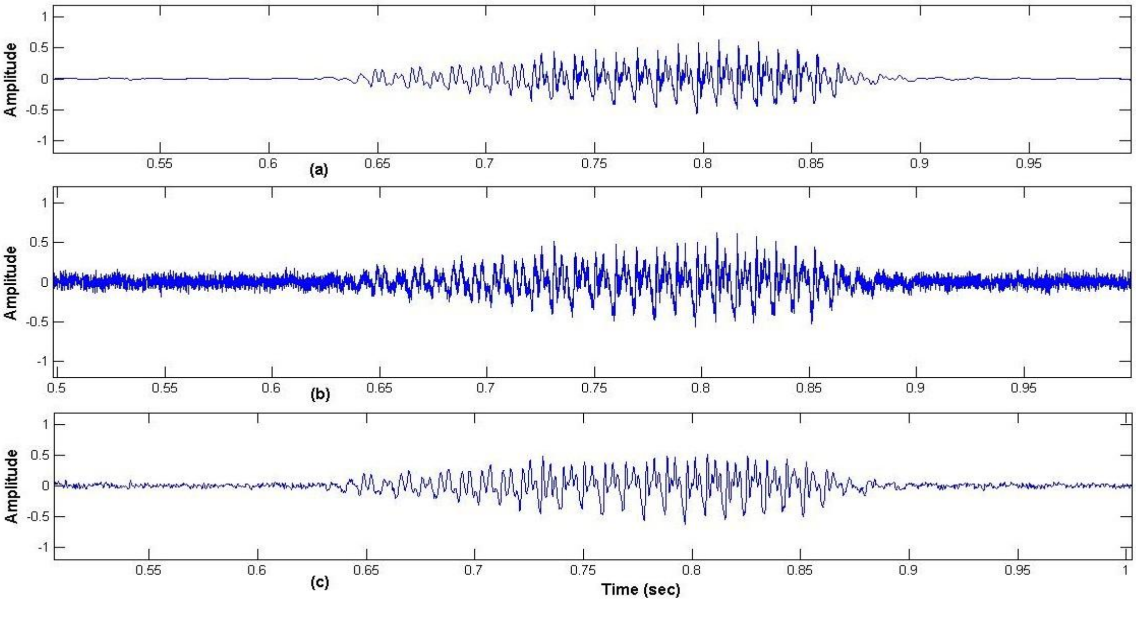

Figure 2: S-G filter based denoising of speech signal with 1 sec duration

The objective quality measures are evaluated in three levels. 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. In the last level, the noisy speech and enhanced speech is processed for the evaluation. This measure indicates the improvement of enhanced speech over the noisy speech. Various objective and subjective quality evaluations are performed and the results are tabulated in table 2. The high PESQ score and composite measures show that S-G filter based method is ideal for fan noise removal.

Table 2: Objective Quality Measures for S-G filter based denoising

Signal	CSig	CBak	COvrl	LLR	SNRSeg	WSS	PESQ	At SNR
O & N	1.796339	2.438033	2.34936	2.833192	-7.62407	8.975553	2.818363	30
O & D	1.536421	2.480801	2.369452	3.185949	-8.2689	21.55967	3.177113	
N & D	1.829166	4.329643	2.988043	3.484515	13.54116	15.67744	4.084294	
O & N	0.256813	2.111603	1.299268	3.963507	-8.08226	15.18954	2.286845	20
O & D	1.043226	2.164323	1.861148	3.301776	-8.60776	31.58927	2.706562	
N & D	0.837889	4.083229	2.456295	4.356266	10.7985	23.7346	4.048255	
O & N	-0.62969	1.926877	0.687968	4.601375	-8.12825	19.68063	1.97222	15
O & D	0.84715	1.974108	1.607118	3.267547	-8.73254	39.01215	2.433773	
N & D	0.346409	3.887646	2.164565	4.743607	8.849983	29.83993	3.985306	
O & N	-1.55922	1.726527	0.042244	5.264427	-8.20758	24.43159	1.633108	10
O & D	0.688636	1.768093	1.354003	3.178437	-8.86936	46.04836	2.123852	
N & D	-0.04153	3.656829	1.899982	5.001959	6.709416	35.95878	3.874157	

The S-G filter based method is compared against the two well-known speech enhancement techniques, spectral subtraction [8][9], and Total variation denoising [10][11]. Various parameter values in these four algorithms such as lambda, tolerance, spectral floor and subtraction factor are adjusted to reduce maximum of noise keeping the quality of the speech signal. The subjective as well as objective values are evaluated and tabulated in table 3 and 4. From the objective measures, it is evident that SG based denoising outperforms the other two methods. Table 5 shows various subjective measures.

Subjective quality measures are evaluated with the help of ten test subjects for all the three methods. Results are shown in Table 3.4. S-G filter based method gives better subjective measures than the other methods.

Table 3: Objective Quality Measures for Spectral Subtraction based denoising

Signal	CSig	CBak	COvrl	LLR	SNRSeg	WSS	PESQ	At SNR
O & N	1.796339	2.438033	2.34936	2.833192	-7.62407	8.975553	2.818363	30
O & D	2.628281	2.901091	3.022266	2.288269	-3.70008	15.32347	3.362887	
N & D	4.954808	4.243325	4.354838	0.267177	14.51304	8.995466	3.677764	
O & N	0.256813	2.111603	1.299268	3.963507	-8.08226	15.18954	2.286845	20
O & D	1.23121	2.563487	2.044346	3.289521	-4.45415	20.18355	2.827162	
N & D	4.688299	3.839752	4.001388	0.266932	11.46258	9.503492	3.242958	
O & N	-1.55922	1.726527	0.042244	5.264427	-8.20758	24.43159	1.633108	15
O & D	-0.3792	2.086528	0.900615	4.381408	-5.92677	33.19742	2.21401	
N & D	4.313615	3.277953	3.573611	0.311371	6.669737	16.30913	2.799002	
O & N	-1.55922	1.726527	0.042244	5.264427	-8.20758	24.43159	1.633108	10
O & D	-0.3792	2.086528	0.900615	4.381408	-5.92677	33.19742	2.21401	
N & D	4.313615	3.277953	3.573611	0.311371	6.669737	16.30913	2.799002	

Table 4: Objective Quality Measures for Total Variation based denoising

Signal	CSig	CBak	COvrl	LLR	SNRSeg	WSS	PESQ	At SNR
O & N	1.796339	2.438033	2.34936	2.833192	-7.62407	8.975553	2.818363	30
O & D	1.21206	2.557072	2.311857	3.672472	-9.25759	12.82233	3.33903	
N & D	2.123844	4.099635	3.13191	3.243733	9.317823	7.294263	4.036971	
O & N	0.256813	2.111603	1.299268	3.963507	-8.08226	15.18954	2.286845	20
O & D	0.63258	2.273575	1.754016	3.88943	-9.28748	18.56645	2.833999	
N & D	1.383586	3.970505	2.745833	3.931278	7.676395	9.228213	4.011484	

O & N	-0.62969	1.926877	0.687968	4.601375	-8.12825	19.68063	1.97222	15
O & D	0.26719	2.122985	1.436645	4.062205	-9.3264	22.86158	2.586987	
N & D	0.887586	3.860721	2.47114	4.373997	6.440128	10.41839	3.96218	
O & N	-1.55922	1.726527	0.042244	5.264427	-8.20758	24.43159	1.633108	10
O & D	-0.14235	1.952523	1.078562	4.254865	-9.38314	27.37542	2.303952	
N & D	0.316242	3.712859	2.132697	4.85917	4.997684	11.65879	3.861122	

Table 5: Subjective Quality Measures for Total S-G filter based denoising

Test Subject	**CS-SIG**	**CS-BAK**	**CS-OVRL**	**SS-SIG**	**SS-BAK**	**SS-OVRL**	**TV-SIG**	**TV-BAK**	**TV-OVRL**
1	5	4	4	3	4	4	3	4	4
2	4	4	4	3	3	3	3	4	4
3	5	5	5	3	4	3	3	3	3
4	4	4	4	4	3	3	4	3	3
5	4	4	4	3	3	3	4	4	4
6	4	3	4	3	3	3	3	4	3
7	4	4	4	4	3	4	3	3	3
8	4	4	4	3	4	4	3	3	3
9	5	5	5	4	3	3	4	4	4
10	4	4	4	3	3	3	3	3	3

4. CONCLUSION

This paper introduces a simple and efficient method for enhancement of speech signals corrupted by additive white Gaussian noise. The method is based on S-G smoothing filter, which is basically a low pass smoothing filter. The method is simple but gives better results compared to well-known methods like spectral subtraction and total variation denoising. The proposed method is evaluated using different objective and subjective tests like WSS, LLR, SNRSeg, PESQ etc. From the results, it is evident that SG filter based method is ideal for fan noise removal.

5. REFERENCES

[1] S.ChinaVenkateswarlu, K.Satya Prasad and SubbaRami Reddy, "Improve Speech Enhancement Using Weiner Filtering", Global Journal of Computer Science and Technology, Vol. 11, Iss. 7, Ver 1.0, May 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] B. Yegnanarayana, Carlos Avendano, HynekHermansky and P. Satyanarayana Murthy, "Speech enhancement using linear prediction residual", Elsevier Speech Communication, Vol. 28, No. 1, pp. 25–42, May 1999.

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

[5] Dalei Wu, Wei-Ping Zhu, and M N S Swamy, "A compressive sensing method for noise reduction of speech and audio signals", Proceedings of IEEE International Midwest Symposium on Circuits and Systems, Vol. 7, No. 10, pp.1-4, August 2011.

[6] SaeedGazor and Wei Zhang, "Speech Enhancement Employing Laplacian–Gaussian Mixture", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 13, No. 5, September 2005.

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

[8] Upadhyay, Navneet, Karmakar and Abhijit, "The spectral subtractive-type algorithms for enhancing speech in noisy environments," 2012, 1st International Conference on Recent Advances in Information Technology (RAIT) , pp.841-847, 15-17, March 2012.

[9] Miyazaki. R, Saruwatari. H, Inoue. T, Takahashi Y, Shikano K and Kondo. K, "Musical-Noise-Free Speech Enhancement Based on Optimized Iterative Spectral Subtraction," IEEE Transactions onAudio, Speech, and Language Processing, 2012.

[10] Ivan W. Selesnick and IlkerBayram, "Total Variation Filtering," February 4, 2010.

[11] G. R Vogel and M. E. Oman, "Iterative methods for total variation denoising", SIAM J. Sci. Computing.

[12] 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.

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

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