International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-5, June 2019

Improved Spectral Subtraction with Time Recursive Noise Estimation

P.Sunitha, K.Satya Prasad

Abstract*: This Paper proposes a method for single channel speech enhancement with noise estimation algorithm in presence of additive back ground noise which overcomes the drawbacks of basic spectral subtraction. Since speech is non-Stationary signal, noise distribution is non-uniform i.e few frequency components are affected severely than others. So speech enhancementalgorithms requires an accurate noise estimation to remove the noise effectively. The proposed method uses noise estimation by First order recursive averaging algorithm based on a posteriori Signal -to-Noise-Ratio and its performance evaluated in terms of objective performance measures such as segmental SNR, Cepstrum distance, Log Likelihood Ratio and Perceptual Evaluation of Speech quality under eight different real-world noises at three ranges of input SNR. The performance of Proposed method was compared with Basic Spectral Subtraction method ,Spectral subtraction with Various noise estimation algorithms .The results shows that ,proposed method shows superior performance in all the cases considered.

Index Terms*: speech enhancement, Voice Activity Detection, Noise Power Estimation, Signal -to –Noise, Ratio, Objective Performance Measures.

I. INTRODUCTION

In speech enhancement techniques noise power estimation plays a key role to enhance the speech quality in presence of highly non stationary noise. Most of the speech enhancement techniques use Voice Activity Detection (VAD) to discriminate speech segments from non-speech segments.Generally Voice Activity Detection algorithms update the noise spectrum in non-speech segments and remains unaltered in speech presence segments. VAD algorithms extract features from noisy speech during speech absence periods and compared against a threshold value .If the measured value is greater than the threshold value then it is considered as a speech presence segment otherwise it is treated as a speech absence segment. Several VAD algorithms are available in literature .VAD algorithm based on statistical model given in[1],detects the speech presence and absence segments based on likelihood criterion. Statistical based algorithms are easy to implement and effective, but it fails when average of the likelihood ratio is greater than the threshold as it uses fixed threshold.One more drawback i.e it severely effectsthe performance of noise estimation due to large delay.This can be bypassed using an adaptive threshold [2],by this we can detect many silence regions those cannot be detected when fixed threshold was used. Improvements can be found when VAD implemented with multiple observation likelihood ratio test[3].

Manuscript published on 30 June 2019.
* Correspondence Author (s)
P.Sunitha, Research Scholar, Dept. of ECE, JNTUK,India,
K.Satya Prasad, Rector, VFSTR,Guntur, India.

Noise estimation with the aid of recursive averaging was proposed in [4] Grapf implemented VAD with various features and comparative analysis was done based on features. Recursive averaging with adaptive parameter was used to implement VAD for non-stationary noise [5].In [6,7] implemented speech enhancement with a priori SNR estimation and noise estimation based on entropy and improvements can be found interms of segmental SNR. These algorithms work well under stationary noise but fails in non-stationary noise. In speech enhancement applications requires accurate noise estimation at all times during speech presence segments also .In such cases VAD might not be sufficient to get precise noise power spectrum. When the noise power spectrum estimate is more accurate then enhanced signal is more nearer to the clean signal. To improve the speech quality and intelligibility in presence of highly non stationary noise,speech enhancementalgorithms requires noise estimation algorithms which update the noise spectrum continuously.Most of the speech enhancement applications in non-stationary scenarios use noise estimation methods algorithms which track the noise spectrum continuously [8]. Now, researchers focus their attention to improve the speech quality and intelligibility using efficient noise estimation algorithms. Minimum statistics is one among them ,which updates the noise on a frame by frame basis by considering the minimum over a periodogram of input noisy speech signal over a small window[8].The main issue in this method is tracking of minimum leads to the over estimation of noise spectrum ,because of its inability to respond fast changes of the noise spectrum.Minima Controlled Recursive Algorithm (MCRA) proposed by Cohen estimates the noise power spectrum by taking the average of past spectral values of speech against a time – frequency dependent smoothing factor.The key problem with this method is it requires twice the number of frames [9]. To overcome this Cohen proposed, Improved Minima Controlled Recursive Algorithm (IMCRA) by continuously monitoring the minimum of the noise estimate from noisy speech based on speech presence probability[10].Mostof the minimum statistics algorithms introduce too much residual noise because of inaccurate estimate of noise power. Time recursive averaging algorithms exploits the fact that, the noise power spectrum has non-uniform effect on the speech spectrum,some regions will effect more adversely than others.For any type of noise power spectrum estimate it is not possible to obtain an estimation and updating of noise in each frequency bin of the entire noise spectrum. This observation makes the use of recursive type algorithms, whosenoise spectrum is updated by considering theweighted average of past and present noisy speech spectrum estimates.

The weight changes according to the speech presence probability[11]. This paper implements spectral subtraction with first order time recursive noise estimation algorithm and its performance is evaluated in terms of performance measures.

The structure of the paper as follows, section II gives Basic Spectral Subtraction methodfor Speech enhancement , Section III presents Noise Estimation by First order time recursive averaging Algorithm and Section IV gives objective quality measures finally results and conclusion are presented in section V.

II. BASIC SPECTRAL SUBTRACTION METHOD:

Spectral subtraction method is one of the basic method suggested for improving the speech quality in presence of additive background noise. In this method estimate of clean speech spectrum is obtained by taking the difference between noisy speech power spectrum and noise power spectrum. Estimation and updating of noise spectrum can be done by using Voice Activity Detection in speech absence segments. This method is more popular because of its straight forward implementation as it involves forward and inverse transform only. The noisy speech signal y(n),consists of both clean speech c(n) and additive noise d(n).Mathematically in time it can be written as

$$y(n)= c(n) + d(n), \quad (1)$$

Since speech is non-stationary one, to make it as stationary framing can be used by using a suitable window with a duration of 10-30 ms. In most of the speech processing applications Hamming window is preferable because of less spectral leakage in side lobes. Analysis of noisy speech can be doneby computing the noise power spectrum of the each frame using Fast Fourier Transform (FFT). Now noise power spectrum of noisy speech is given by

$$|Y(K)|^2 = |C(K)|^2+|D(K)|^2 \quad (2)$$

FFT of y(n) is given by Y(K) ,can be calculated as

$$Y(K) = \sum_{n=0}^{N-1} y(n)e^{\frac{-j2\pi nK}{N}} = |Y(K)|e^{j\phi(k)} \quad (3)$$

Equation (3) represents Y(K) is represented in terms of both phase and magnitude power spectrums of noisy Speech. In speech enhancement process phase information is kept aside, because it doesn't perceived by the human ear. An estimate of clean speech power spectrum obtained after the subtraction process is given by

$$\left|\hat{C}(K)\right|^2 = |Y(K)|^2 - \left|\hat{D}(K)\right|^2 (4)$$

To reconstruct the estimate of clean speech in time domain ,the estimate of clean power spectrum is combined with the phase information by using any one of the synthesis techniques like overlap-add (OLA), Filter Bank Summation (FBS) or Least Square Error Synthesis(LSE)[12].In this paper overlap add technique is used to reconstruct time – domain signal .

$$\hat{c}(n) = IFFT\{|\hat{C}(K)|.\exp(j\phi(k))\} (5)$$

In spectral subtraction process an estimate of noise power spectrum is subtracted from the noisy power spectrum which may results in negative values. To supress this one can use Half Wave Rectification [17] which in turn creates a bi-product named as residual noise in the estimated speech. To avoid this Full-Wave Rectification can be used, but it shows poor performance in attenuation of noise.

III. NOISE ESTIMATION BY FIRST ORDER TIME RECURSIVE AVERAGING ALGORITHM:

Fundamental requirement of any speech enhancement algorithm is an accurate estimate of noise spectrum. Most of the Spectral subtractive type speech enhancement algorithms uses Voice Activity Detection (VAD) to separate the voiced and unvoiced regions which works well in case of stationary noise but fails in highly non-stationary noise. To overcome this , the proposed method uses first order recursive averaging noise estimation algorithm, to improve the performance of Spectral subtraction method for speech enhancement with a smoothing parameter which is controlled by a-posteriori SNR. In recursive type averaging algorithms noise estimate is obtained by considering the average weight of the past and present noise estimation of the noise speech spectrum [11].Depending on the speech presence probability smoothening parameter will be varied and updated on frame by frame basis as a sigmoid function changes according to the a-posteriori SNR. Depending on the SNR of each frame weights changes adaptively by smoothening parameter. Estimate of the noise can be evaluated using first order recursive algorithm as

$$\hat{\sigma}_d^2(\omega,k) = \lambda(\omega,k)\hat{\sigma}_d^2(\omega,k-1) + \big(1-\lambda(\omega,k)\big)|Y(\omega,k)|^2 (6)$$

Where$|Y(\omega,k)|^2$gives the estimate of noise PSD with respect to the frame K and frequency ω, $\hat{\sigma}_d^2(\omega,k)$ is the noisy speech spectrum and the time –frequency dependent smoothing factor is given by $\lambda(\omega,k)$depending on the presence or absence of speech at frequency bin k[12][23].

$$\lambda(\omega,k) = 1/(1+\exp(-\beta(\mu_k(\alpha-1.5)))) \quad (7)$$

a-posteriori SNR is given by

$$\mu_k(\omega) = 10log_{10}\left(\frac{|Y(\omega,k)|^2}{\frac{1}{m}\sigma_d^2\sum_{p=1}^{m}(\omega,k-p)}\right) \quad (8)$$

Denominator part in equation (13) gives the average of the estimated noise power spectrum in the past ' m' frames.Larger values of $\mu_k(\omega)$ yields $\lambda(\omega,k)$=1 then equation (11) $changes\ as \hat{\sigma}_d^2(\omega,k) = \lambda(\omega,k)\hat{\sigma}_d^2(\omega,k-1)$,then noise update will cease and the noise estimate will remain same as the previous frame's estimate, indicates that speech is present.Similarly smaller values of a posteriori SNR $\mu_k(\omega)$ yields $\lambda(\omega,k)$=0 .As a result $\hat{\sigma}_d^2(\omega,k) = |Y(\omega,k)|^2$, then noise estimate follow the noise power spectrum ,indicates the speech absence. The parameter 'β'should be chosen carefully. In this paper 'β' value is selected as 0.6 yields good tracking of the noise spectrum.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

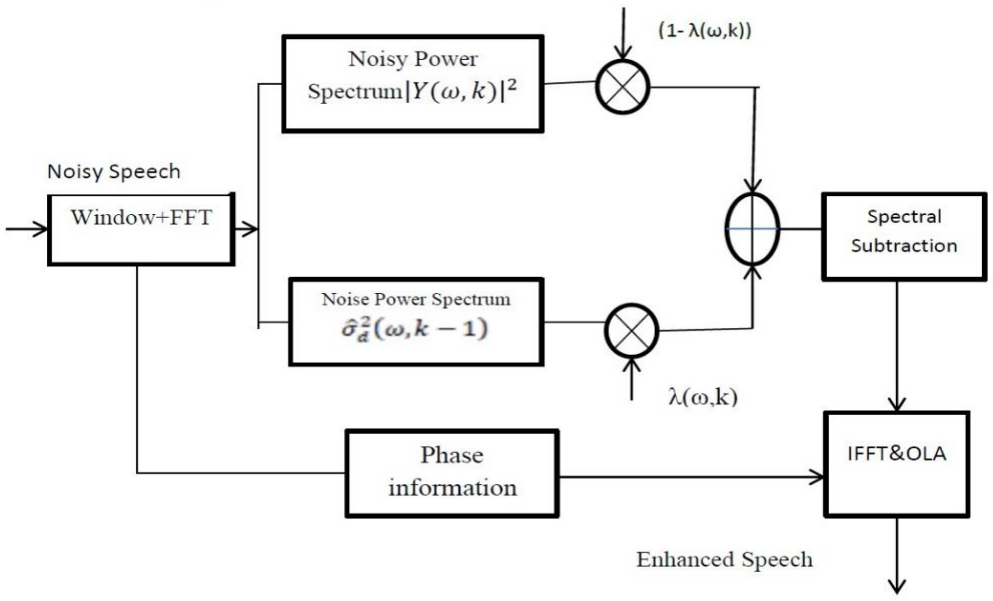

Fig.1,Block Diagram Of Proposed Method

IV. OBJECTIVE QUALITY MEASURES FOR PERFORMANCE EVALUATION:

Performance of speech enhancement techniques can be evaluatedin terms of either subjective listening test or objective measures. Comparison of original clean speech and reconstructed (enhanced) speech signals by a group of listeners on a predetermined scale is known as subjective quality evaluation[19].One can use objective evaluation to quantify the quality between the original clean and reconstructed speech signals[18].This paper presents the performance evaluation based on three different quality measures which are ,Cepstrum Distance, Segmental SNR,Log Likelihood Ratioand Perceptual Evaluation of Speech Quality.

i)CepstrumDistancee (d_{CEP}): The spectral distance between the clean and estimated speech spectrums on a logarithmic scale IS DEFINED AS Cepstrum distance [13] and it can be evaluated using the following equation

$$d_{CEP}(\vec{c}_c,\vec{c}_{\hat{c}}) = \frac{10}{log10}\sqrt{2\sum_{k=1}^{p}(c_c(k)-c_{\hat{c}}(k))^2} \quad (8)$$

Cepstrum Coefficients of the clean and enhanced signals are given by $c_c(k)$and $c_{\hat{c}}(k)$.

ii)Segmental Signal-to-Noise Ratio:

The segmental Signal-to-Noise Ratio (SNR seg) in the time domain can be expressed as

$$SNR_{seg} = \frac{10}{M}\sum_{M=0}^{M-1} log_{10}\frac{\sum_{n=NM}^{Nm+N-1}c^2(n)}{\sum_{n=NM}^{Nm+N-1}(c(n)-\hat{c}(n))^2} \quad (9)$$

Here c(n) represents the original clean speech signal.$\hat{c}$(n) is the reconstructed signal ,frame length is given by N and the number of frames is given by M. The geometric mean of all frames of the speech signal is SNR seg [15].

*iii).*Log Likelihood Ratio (LLR):

$$LLR(\overrightarrow{a_x},\vec{a}_{\hat{x}}) = log\left(\frac{\vec{a}_{\hat{x}}R_x\vec{a}_{\hat{x}}^T}{\vec{a}_xR_x\vec{a}_x^T}\right) \quad (10)$$

$\vec{a}_x$, $\vec{a}_{\hat{x}}^T$are the LPC coefficients of the Clean and enhanced signals. R_x is the autocorrelation matrix of the Clean signal .In LLR denominator term is always lower than numerator therefore LLR is always positive [19]and the LLR values are in the range of (0-2).

iv). Perceptual Evaluation Of Speech Quality (PESQ): One among the objective quality measures which provides an accurate speech quality recommended by ITU_T[16] which involves more complexity in computation. A linear combination of average asymmetrical disturbance A_{ind} and average disturbance D_{ind} is given by PESQ .

$$PESQ=4.754-0.186D_{ind}-0.008\ A_{ind} \quad (17)$$

V. Results and conclusion:

To evaluate the quality of the proposed method objective quality measures likeCepstrumdistance,Segmental SNR, Log Likelihood Ratio and PESQwere used further the results are compared with Power Spectral Subtraction method. Simulations are done in the MATLAB environment. The English noisy speech sample has taken from a NOIZEUS [12]database which consists of 30 sentences belonging to six speakers, degraded by eight different types of colorednoises (airport, exhibition, car,babble, restaurant, street,station and train)at different ranges of SNRs(0dB,5dB,10dB). Originally the sentences were sampled at 25 KHz and down sampled to 8KHz. In this algorithm speech sample is taken from a male speaker, English sentence was "**we can find joy in the simplest things"**. In simulation, the speech signal is divided into frames of 30 ms duration using Hamming window with 50% overlapping and the smoothing factor' β' is taken as 0.6.

Table.1: Cepstrum Distance Measure for Basic Spectral subtraction, Multi Band Spectral Subtraction and Improved spectral Subtraction methods

Type of noise	Input SNR	BasicSpectral subtraction	Multi Band Spectral Subtraction	Improved spectral Subtraction
Airport	0dB	6.43	6.25	6.02
	5dB	5.12	5.01	4.77
	10dB	4.87	4.75	4.10
Babble	0dB	6.73	6.47	6.15
	5dB	5.96	5.77	5.65
	10dB	4.94	4.59	4.38
car	0dB	6.83	6.45	6.04
	5dB	5.91	5.76	5.14
	10dB	4.98	4.89	4.19
Exhibition	0dB	6.85	6.49	5.98
	5dB	5.85	5.73	5.57
	10dB	5.03	4.98	5.01
Restaurant	0dB	6.06	5.82	5.72
	5dB	6.91	6.05	5.73
	10dB	4.95	4.64	4.12
Station	0dB	6.54	6.39	6.20
	5dB	5.95	5.81	5.14
	10dB	5.94	5.17	5.70
Street	0dB	6.90	6.52	6.48
	5dB	5.70	5.69	5.36
	10dB	4.08	4.01	3.90
Train	0dB	6.42	6.32	3.72
	5dB	5.95	5.66	5.50
	10dB	5.10	4.90	4.67

Table.2: Segmental SNR Measure forBasicSpectral subtraction, Multi Band Spectral Subtraction and Improved spectral Subtraction methods

Type of noise	Input SNR	Basic Spectral subtraction	Multi Band Spectral Subtraction	Improved spectral Subtraction
Airport	0dB	-3.65	-2.18	-1.93
	5dB	-1.34	-1.12	0.63
	10dB	2.11	2.18	3.72
Babble	0dB	-3.65	-2.60	-1.98
	5dB	-0.77	-0.54	0.15
	10dB	1.82	2.58	3.18
car	0dB	-2.66	-2.45	-1.33
	5dB	-0.11	-0.04	1.48
	10dB	2.46	2.57	4.21
Exhibition	0dB	-3.27	-2.16	-1.57
	5dB	-3.27	-0.28	1.01
	10dB	2.30	2.93	4.03
Restaurant	0dB	-3.94	-3.59	-2.52
	5dB	-1.97	-1.18	-0.12
	10dB	1.89	2.12	3.05
Station	0dB	-3.80	-3.71	-1.14
	5dB	-0.94	-0.54	0.62
	10dB	2.26	3.00	3.50
Street	0dB	-3.32	-2.75	-2.03
	5dB	-1.97	-1.95	-0.01
	10dB	2.99	3.05	4.31
Train	0dB	-2.66	-1.96	-1.51
	5dB	0.11	0.01	0.95
	10dB	2.68	2.76	3.81

Table.3: Log Likelihood Ratio Measure forBasic Spectral subtraction, Multi Band Spectral Subtraction and Improved spectral Subtraction methods

Type of noise	Input SNR	Basic Spectral subtraction	Multi Band Spectral Subtraction	Improved spectral Subtraction
Airport	0dB	1.78	1.52	1.15
	5dB	1.24	1.02	0.74
	10dB	0.98	0.87	0.69
Babble	0dB	1.93	1.75	1.18
	5dB	1.57	1.44	0.98
	10dB	0.95	0.85	0.66
car	0dB	1.98	1.73	1.21
	5dB	1.63	1.25	1.02
	10dB	1.09	0.97	0.76
Exhibition	0dB	2.01	1.89	1.28
	5dB	1.96	1.75	1.01
	10dB	1.12	1.01	0.78
Restaurant	0dB	1.15	1.05	0.95
	5dB	0.99	0.85	0.93
	10dB	0.79	0.71	0.63
Station	0dB	1.96	1.67	1.12
	5dB	1.87	1.58	1.04
	10dB	1.21	1.04	0.78
Street	0dB	2.10	1.92	1.22
	5dB	1.13	1.04	0.93
	10dB	0.68	0.62	0.59
Train	0dB	1.96	1.67	1.17
	5dB	1.34	1.12	1.09
	10dB	1.24	1.05	0.72

Table.4: PESQ Measure forBasicSpectral subtraction, Multi Band Spectral Subtraction and Improved spectral Subtraction methods

Type of noise	Input SNR	Basic Spectral subtraction	Multi Band Spectral Subtraction	Improved spectral Subtraction
Airport	0dB	1.60	1.85	1.86
	5dB	2.28	2.35	2.44
	10dB	2.30	2.35	2.55
Babble	0dB	1.24	1.71	1.82
	5dB	1.94	2.09	2.23
	10dB	2.45	2.58	2.71
car	0dB	1.55	1.75	1.95
	5dB	2.24	2.16	2.30
	10dB	2.63	2.56	2.62
Exhibition	0dB	1.48	1.74	1.85
	5dB	1.95	2.06	2.21
	10dB	2.15	2.37	2.48
Restaurant	0dB	1.22	1.95	2.01
	5dB	2.07	2.07	2.18
	10dB	2.47	2.52	2.65
Station	0dB	1.47	1.69	1.75
	5dB	2.17	2.14	2.19
	10dB	2.47	2.48	2.59
Street	0dB	1.39	1.69	1.76
	5dB	2.06	2.03	2.14
	10dB	2.57	2.49	2.59
Train	0dB	1.02	1.76	2.09
	5dB	1.98	2.16	2.13
	10dB	2.23	2.31	2.51

The evaluation of the subjective quality of the proposed method reported in this section .The proposed method is compared in terms of Cepstrum Distance, , segmental SNR ,Log Likelihood Ratio and Perceptual Evaluation of Speech Quality, by taking average over all eight different types of noises over three SNRs. Figure2 to figure 5 shows the comparisons of the objective quality measures for Spectral Subtraction alone, Multi Band Spectral Subtraction and

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

Improved Spectral Subtraction method with Noise estimation by time recursivealgorithm . The results were shown in table1 to table4the same can be represented in figures 2 to 5.

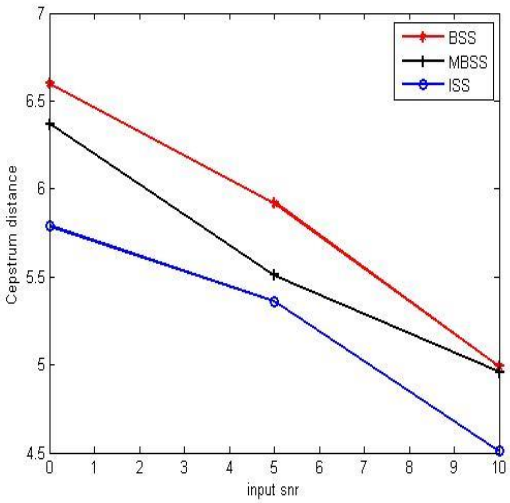

Fig.2:Cepstrum Distance

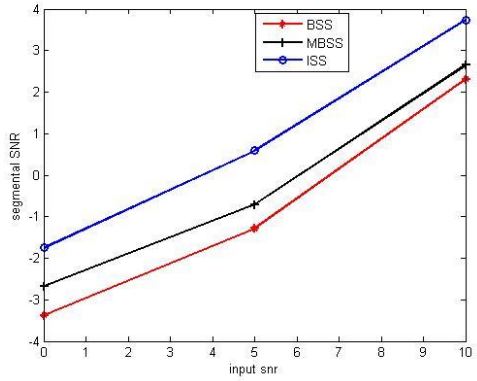

Fig.3:Segmental SNR

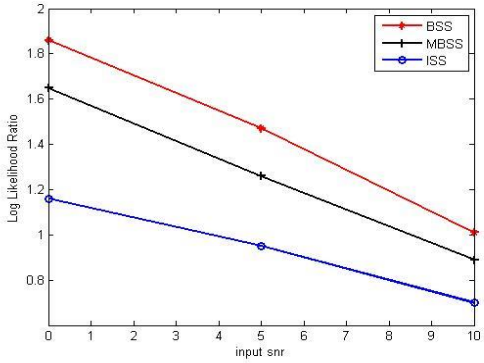

Fig.4:Log Likelihood Ratio

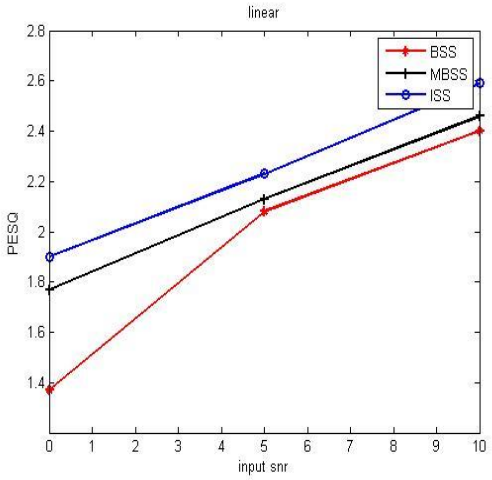

Fig.5:Perceptual Evaluation of Speech Quality

Basic Spectral Subtraction (BSS)

Multi Band Spectral Subtraction (MBSS)

Improved Spectral Subtraction (ISS)

From figures 2&4 it can be concluded that Cepstrum Distance and Log Likelihood Ratio measures are minimum for the proposed method i.e Improved Spectral Subtraction with time recursive noise estimation when compared to others. Higher the values of segmental SNR and PESQ shows the superior performance of proposed method as shown in figures 3&5. Therefore it should be noted that proposed spectral subtraction method with time recursive noise estimation algorithm results in higher values of Segmental SNR,PESQ and lower values of Cepstrum distance and Log Likelihood Ratio.

REFERENCES

1. Sohn,NS .Kim, and W.Sung,"A Statistical model-based Voice Activity Detection", IEEE Signal Processing Letters,6(1),Pg.No:1-3,1999.
2. YD.Cho,A.Kondoz,"Analysis and improvement of a Statistical model-based Voice Activity Detector", IEEE Signal Processing Letters,8(10), Pg.No:276-278,2001
3. J.Ramirez,JC.Segura,C.Benitez,"Statistical Voice Activity detectionusing a multiple observation likelihood ratio test", IEEE Signal Processing Letters,12(10), Pg.No :689-692,2005.
4. K.Nakayama,S.Higashi , A.Hirano,"A Noise spectral estimation approach based on VAD and recursive averaging using new adaptive parameters for non-stationary Noise environments", Proceedings of international symposium on intelligent signal processing and communication systems,1-4,2008.
5. S.Graf,T.Hrrbig,M.Buck,"Features for Voice Activity Detection:AComparitive Analysis", EURASIP Journal on Advances in Signal Processing,2015(1),1-15,2015.
6. Y.Ma,A. Nishihara, "Efficient Voice detection algorithm using long term spectral flatness measure",EURASIP Journal on Advances in Signal Processing,2013(1),1-18,2013.
7. R.Yao,Zeng ,Zhu, "A priori SNR estimation and Noise estimation for speech enhancement",2016(1),2016.
8. R.Martin, "Noise Power Spectral Density Estimation based on Optimal Smoothing and Minimum statistics", IEEE Transactions on Audio, Speech Processing 9 (5), pp.504–512, 2001.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

9. I.Cohen, “Noise Estimation by Minima controlled recursive averaging for robust speech enhancement ”, IEEE Signal Processing. Letter 9 (1), pp.12–15,2002
10. Loizou, R.Sundarajan,Y. Hu,”Noise estimation Algorithm with rapid Adaption for highly non-stationary environments “,Proceedings on IEEE International Conference on Acoustic Speech Signal Processing,2004.
11. Loizou, R.Sundarajan,.”A Noise estimation Algorithm for highly non-stationary Environments”. Speech Communication ,48 ,Science Direct ,Pp.220-231,2006.
12. A Noisy Speech Corpus for Assessment of Speech Enhancement Algorithms. https: //ecs.utdallas.edu/ Loizou /speech/noizeous
13. Yi.Hu ,P.C .Loizou.,"Evaluation of objective Quality Measures for Speech Enhancement " ,IEEE Transactions on Audio, Speech and Language Processing ,vol.16,no.1,pp.229-238,Jan.2008.
14. ITU_T Rec ,”Perceptual evaluation of speech quality(PESQ), An objective method for end to end speech quality assessment of narrowband telephone networks and speech codecs”.,International Telecommunications Union ,Geneva Switzerland, February 2001.
15. Boll,S.F,”Suppression of acoustic noise in speech using spectral subtraction”. IEEE Transactions on Acoustics Speech and Signal Processing, 1979,27(2), 113–120.