TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 11283 - 11288

Objective Evaluation Measures for Speech Enhancement for Non Stationary Noise Environments

Dr. G. Amjad Khan[1], M. Devendra[2]Dr. S. M. Shamsheer Daula[3]G. Pavan Kumar [4]

[1,4]*Assistant Professor ,G.Pulla Reddy Engineering College (Autonomous),,Kurnool, Andhra Pradesh, India*
[2,3]*Associate Professor,G.Pulla Reddy Engineering College (Autonomous),,Kurnool, Andhra Pradesh, India*

Article Info
Volume 82
Page Number: 11283 - 11288
Publication Issue:
January-February 2020

Article History
Article Received*: 18 May 2019*
Revised*: 14 July 2019*
Accepted*: 22 December 2019*
Publication*: 21 February 2020*

Abstract
A new speech enhancement algorithm was once proposed in this paper with an intention of decreasing the non-stationary noises introduced on the easy speech signals. Suppression of non-stationary noise in a imperative issue. Because, the traits of noises are one of a kind shape noise to noise as nicely as varies with admire to the environment. To solve this issue, a novel non-stationary noise suppression mechanism is proposed in this paper based on the Sub band Adaptive Filtering. The main advantage with the accomplishment of SAF is faster convergence and also better quality achievement. Along with SAF, a noise classification mechanism additionally proposed in this paper to limit the extra computational complexity in the noise identification. Extensive simulations are carried out over the proposed mechanism thru distinctive speech indicators with different noise at exclusive SNRs. The performance evaluation is carried out through the performance metrics, PESQ and WSS and reveals the outstanding performance of proposed mechanism.

Keywords: *Speech enhancement, Non-stationary noise, Sub band adaptive filtering, PESQ, SegSNR and WSS.*

I. INTRODUCTION

In recent years, a drastic boom in the speech oriented applications like present day mobile communications, Automatic Speech recognition and Computer-Human Interaction systems thru voice communications; speech enhancement through suppressing the unnecessary noises will become an critical research component to enlarge the high-quality and intelligibility of speech [1]. Several techniques are proposed in earlier to suppress the noises form a noisy speech signal. One of the most popular noise suppression techniques is spectral subtraction which was developed with an aim of reducing the effects of noise in the speech signal. In this method, the spectral The basic requirement of this method is the determination of noise spectrum form non-speech segments which results in the loss of intelligibility and results in a residual noise in the enhanced speech. The subband based speech enhancement is another typical method developed with an aim of reducing the noise in speech signals [3]. In this paper, an adaptive speech enhancement algorithm is proposed based on Sub Band Adaptive Filtering (SAF). Unlike the conventional approaches which was not efficient under non-stationary noisy environments, this paper addressed the problem of finding a best fit noise part though the accomplishment of Normalized Least Mean Square (NLMS) algorithm. Along with this, the proposed approach also developed a novel subband adaptive filtering technique for noise suppression after finding a best fit noise part form the noisy speech through NLMS.

January - February 2020
ISSN: 0193 - 4120 Page No. 11283 - 11288

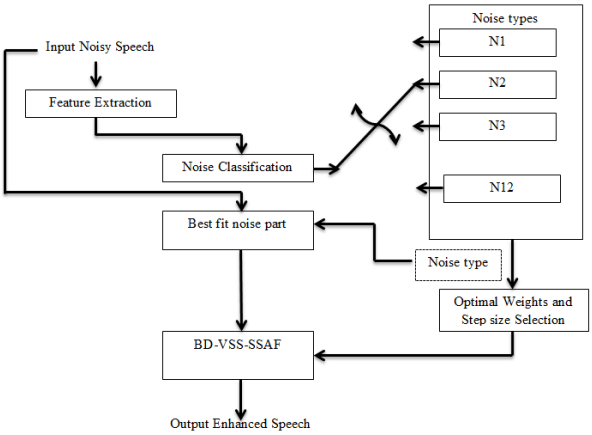
II. LITERATURE SURVEY

Various authors developed various approaches to enhance the speech signal under different noisy circumstances. Mainly there are two major classes of speech enhancement approaches such as time domain speech enhancement technique, for example sub space methods [4-6] and frequency domain approaches such as spectral subtraction methods [7-10].

By modifying the gain function of Log-spectral amplitudes (LSA), based on two hypotheses, an optimally modified LSA (OM-LSA) is proposed in [7], had shown a substantial enhancement in the speech enhancement. In [8], a single channel speech enhancement technique is proposed for reduction of combined non-stationary noises. The proposed method is based on wavelet packet and ideal binary mask thresholding function for speech enhancementVarying step size based noise suppression method is developed in [12] and found to be having a good convergence rate and also less MSE compared to the LMS. In order to improve the quality of speech signal and adaptability of cochlear implant under strong noise background, an improved method was proposed in [13] based on the combination of spectral subtraction and variable-step Least Mean Square error (LMS) [15]. Concerning the problem of slow convergence rate and big steady-state error, the squared term of output error was used to adjust the step size of variable-step LMS adaptive filtering algorithm; besides, the combination of fixed and changed values of step was also considered, thus improved the adaptability and quality of speech signal. Due to the simplicity, these algorithms are were used but the convergence rate is observed to be low which results in a prolonged time to achieve the desired signal. Various authors developed various approaches based on the sub band adaptive filtering. [16-19]

III. PROPOSED APPROACH

To solve the problems with conventional approaches, this paper proposes a novel band dependent variable size sign subband adaptive filtering (BD-VSS-SSAF) technique. The proposed method is an extension to the most popular sub band adaptive filtering [21]. The overall working procedure is represented in figure.1.



As shown in figure.1, initially the input noisy speech signal is subjected to the noise classification through the Support vector machine algorithm. Here totally 12 types of noises are trained to the system though SVM and in the testing phase, the input noisy speech is processed for noise type detection. Once the noise is detected, the optimal weights and step sizes are chosen. Further the input noisy speech is correlated with standard noise signal to extract the best fit noise part. This is to find the noise dominant regions in the noisy speech. Since processing a complete signal results an increased complexity, the proposed best fit noise evaluation part extracts only the noise dominant regions and only these regions are processed for proposed noise suppression mechanism, i.e., band dependent variable step size SSAF. The details are explained in the further subsections.

A. Band Dependent Variable Step Size, Sign Sub bad Adaptive filtering (BD-VSS-SSAF)

Consider a desired signal $d(n)$ derived from the unknown noise suppression system is given by

$$d(n) = n(n) + \boldsymbol{u}^T(n)\boldsymbol{w}_0 \quad (1)$$

Where $\boldsymbol{w}_0$ denotes a weight vector which we needs to be estimated through the adaptive filter, $\boldsymbol{u}(n)$ is an input signal vector, T denotes the transpose and $v(n)$ denotes the additive noise.

January - February 2020
ISSN: 0193 - 4120 Page No. 11283 - 11288

In this noise suppression system, the desired signal $d(n)$ and input signal $u(n)$ are partitioned into N sub band signals, $d_i(n)$and $u_i(n)$, $i = 0,1,2,\dots,N-1$, through the analysis filter bank $\{H_i(z), i = 0,1,2,\dots,N-1\}$. Then the sub band signals $y_i(n)$ and $d_i(n)$ are decimated such that the decimated sub band signals are represented as $y_{i,D}(k)$ and $d_{i,D}(k)$ where n denotes the original signal and k represents the decimated signal.

Further sub band error signal $e_{i,D}(k)$ is evaluated as,

$$e_{i,D}(k) = d_{i,D}(k) - \boldsymbol{u}_i^T(k)\boldsymbol{w}(k) \quad (2)$$

For $i = 0,1,2,\dots,N-1$.

Where $\boldsymbol{w}(k) = [w_1(k), w_2(k), \dots, w_M(k)]$ denotes the weight vector, $d_{i,D}(k) = d_i(kN)$and

$$\boldsymbol{u}_i(k) = [\boldsymbol{u}_i(kN), \boldsymbol{u}_i(kN-1), \dots, \boldsymbol{u}_i(kN-M-1].$$

Representing the expression (2) in vector form as,

$$\boldsymbol{e}_D(k) = \boldsymbol{d}_D(k) - U^T\boldsymbol{w}(k) \quad (3)$$

Where
$\boldsymbol{d}_D(k) = [\boldsymbol{d}_{0,D}(k), \boldsymbol{d}_{1,D}(k), \dots, \boldsymbol{d}_{N-1,\ D}(k)]$and
$\boldsymbol{U}(k) = [\boldsymbol{u}_0(k), \boldsymbol{u}_1(k), \dots, \boldsymbol{u}_{N-1}(k)]$.

According to the SSAF reported in [16], the weight vector of the original SSAF is obtained as

$$w(k+1) = w(k) + \mu \frac{\boldsymbol{U}(k)sgn(\boldsymbol{e}_D(k))}{\sqrt{\sum_{i=0}^{N-1}\boldsymbol{u}_i^T(k)\boldsymbol{u}_i(k)+\epsilon}} \quad (4)$$

Where μ is step size and ϵ is a small arbitrary constant to avoid the division by zero and sgn(.) denotes the sign function of every element in $\boldsymbol{e}_D(k)$.

The expression shown in equation (4) is obtained by minimizing the weight cost function as

$$J(k) = \sum_{i=0}^{N-1}\lambda_i\left|e_{i,D}(k)\right| \quad (5)$$

Where λ is considered as weight factor of ith sub band. According to the gradient descent theory,

$$w(k+1) = w(k) + \mu \frac{\boldsymbol{u}_i(k)\ sgn(\boldsymbol{e}_{i,D}(k))}{\sqrt{\sum_{i=0}^{N-1}\boldsymbol{u}_i^T(k)\boldsymbol{u}_i(k)+\epsilon}} \quad (6)$$

In the case of equal weight factors,

$$\lambda_1 = \lambda_2 =, \dots, \lambda_{N-1} = \frac{1}{\sqrt{\sum_{i=0}^{N-1}\boldsymbol{u}_i^T(k)\boldsymbol{u}_i(k)+\epsilon}} \quad (7)$$

B. Best fit noise part

The evaluation of bets fit noise part tries to extract the best noise part form the noisy speech by comparing it with a standard reference noise signal. For this purpose, the standard reference noise signal is processed frame by frame. Initially the noise reference signal is partitioned into N number of frames and every frame is compared with noisy speech signal to find a best fit noise part. Once the noise signal is split into the frames, the frame to frame comparison is accomplished through normalized LMS algorithm. According to the NLMS [22], the error updates and the weight vector update are evaluated as,

$$e_q(k) = d_q(k) - x_i^T(k)w_q(k) \quad (8)$$

$$w_q(k+1) = w_q(k) + \frac{\mu_q e_q(k)x_q(k)}{x_i^T(k)x_q(k)} \quad (9)$$

Where $q \in [1\ M]$ and M denotes the number of frames obtained after splitting the noise reference signal into frames of length L. since the pure LMS is more sensitive to scaling of input signals, it leads to the misguidance of weight vectors. This makes it very tough to choose a step size that provides a guarantee about the stability of algorithm. Hence this method accomplishes NLMS to find a best fit noise part. The frame having correlation components are shown in figure.2(a) and the frame with no correlation is represented in the figure.2(b).

January - February 2020
ISSN: 0193 - 4120 Page No. 11283 - 11288

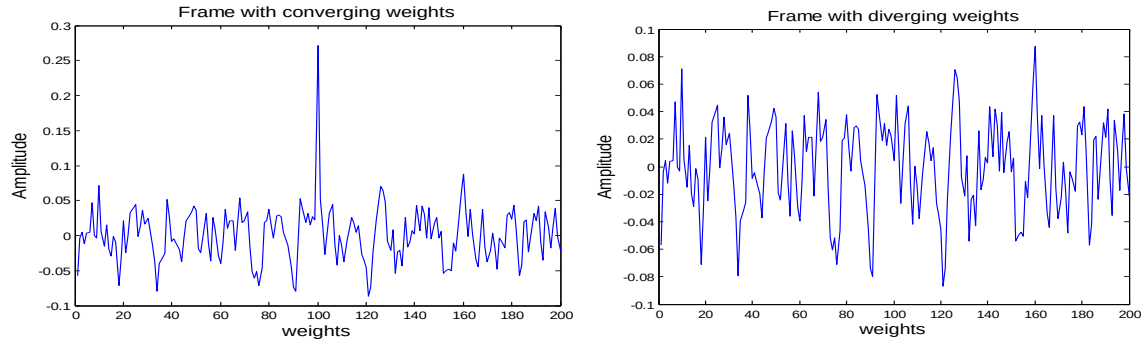


Figure. 2 Weight Convergence

V. SIMULATION RESULTS

This section describes the performance evaluation of the proposed approach. To illustrate the efficacy of the proposed speech enhancement algorithm, mock-ups were carried out with 10 male and 10 female utterances, indiscriminately selected from the TIMIT database. Totally 12 types of noises are considered form the NOISEX-92 database, i.e. These noises are mixed with the clean speech signals received form the TIMIT database at various SNRs such as 0dB, 5dB and 10 dB. The assimilated all clean speeches measured here is elapsed a time of 2sec. Totally 20 clean speech samples are deliberated here for performance evaluation and are composed of 23000 samples on an average After this, the proposed BD-VSS-SSAF is accomplished to perform noise suppression and finally the obtained noise free speech is processed for performance evaluation.

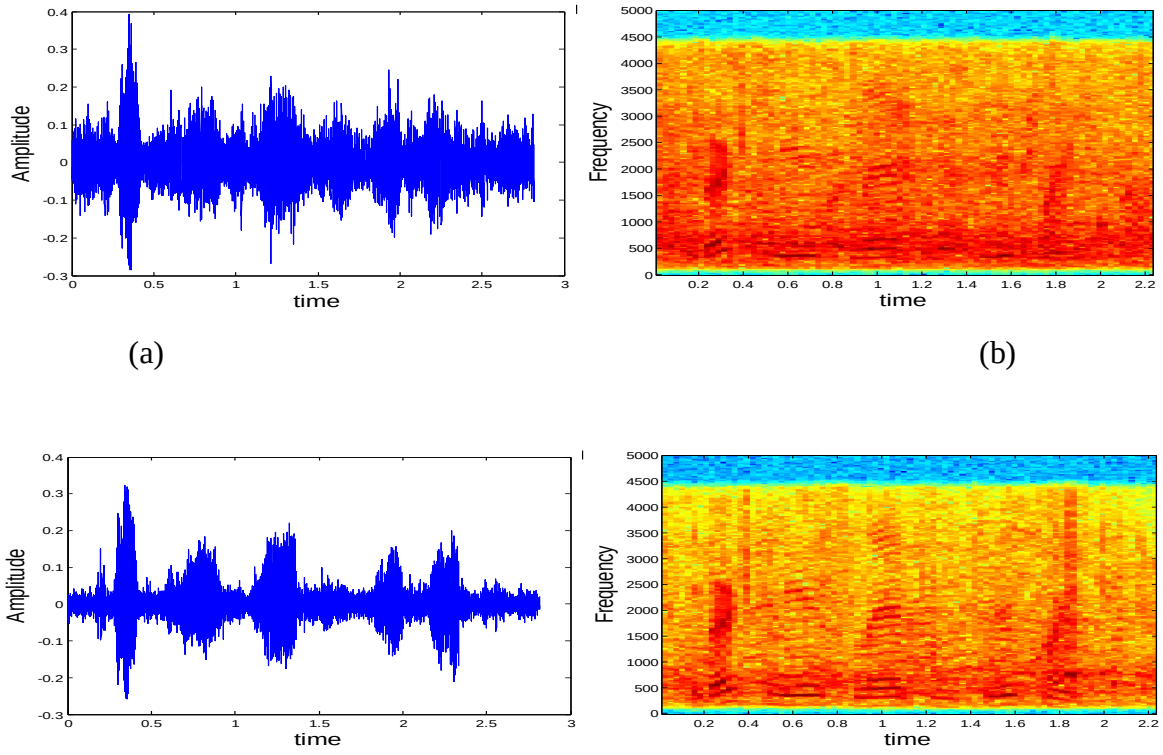


Figure.3 Enhanced speech signals, (a) Input Noisy Speech (Factory 2 noise at SNR = 0 dB), (b) Spectrogram of (a), (c) Output Enhanced Speech, (d) Spectrogram of (c)

January - February 2020
ISSN: 0193 - 4120 Page No. 11283 - 11288

Table.1 Results of composite measure for signal distortionC_{sig}

Noise	SNR	OM-LSA	MMSE-BC	IMCRAvg	Proposed
Military vehicle	0	3.92	3.72	3.95	4.11
	5	4.34	4.19	4.37	4.52
	10	4.80	4.65	4.83	4.89
Factory 2	0	3.23	3.04	3.34	3.47
	5	3.76	3.52	3.82	3.98
	10	4.27	4.07	4.31	4.44
Jet 2	0	2.18	1.96	2.25	2.57
	5	2.86	2.61	2.96	3.16
	10	3.43	3.28	3.52	3.69

Table.2 Results of composite measure for Background Intrusiveness C_{bak}

Noise	SNR	OM-LSA	MMSE-BC	IMCRAvg	Proposed
Military vehicle	0	2.94	2.88	2.96	3.17
	5	3.30	3.28	3.30	3.56
	10	3.75	3.71	3.77	3.89
Factory 2	0	2.52	2.42	2.57	2.67
	5	2.99	2.85	3.01	3.11
	10	3.44	3.36	3.46	3.59
Jet 2	0	2.12	1.95	2.15	2.28
	5	2.58	2.42	2.62	2.79
	10	3.02	2.94	3.08	3.19

Table.3 Results of composite measure for Overall Quality C_{ovl}

Noise	SNR	OM-LSA	MMSE-BC	IMCRAvg	Proposed
Military vehicle	0	3.35	3.26	3.39	3.61
	5	3.76	3.69	3.78	3.87

January - February 2020
ISSN: 0193 - 4120 Page No. 11283 - 11288

	10	4.17	4.11	4.20	4.41
Factory 2	0	2.74	2.63	2.84	2.99
	5	3.24	3.08	3.29	3.45
	10	3.70	3.59	3.73	3.86
Jet 2	0	1.96	1.83	2.03	2.25
	5	2.55	2.40	2.64	2.79

VI. CONCLUSION

In this paper, a new speech enhancement approach is proposed by combining the sub band adaptive filtering with machine algorithm. The innovative idea of this approach is the identification of best fit noise part form a noisy speech signal through the Normalized LMS algorithm. Further the proposed approach also accomplished to find the type of noise present in the input noisy speech signal which makes the system very fast by reducing the number of samples needs to be processed for suppression. The simulation results reveal the performance effectiveness of proposed approach through the performance metrics such as C_{sig}, C_{bak} and C_{ovl}. The obtained performance metrics revealed the performance enhancement in the suppression of non-stationary noises and also succeeded in the preservation of speech quality and intelligibility.

REFERENCES

[1] A. Sankaar, S. F. Beaufaysa, and V. Digalakise, "Training data clustering for improved speech recognition", In: *Proc. of European Conference on Speech Communication and Technology (EUROSPEECH)*, Madrid, Spain, pp.1-4, 1995.

[2] Y. Ephraime and D. Malahe, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 32, No. 6, pp. 1109–1121, 1984.

[3]. A. Saadouna, A. Amrouches, S., Perceptual subspace speech enhancement using variance of the reconstruction error, 2014, Digital Signal Processing: A Review Journal, vol. 24, no. 1, pp.187-196.

[4]. J. R. Jensenz, J. Benestx, M. G. Christensen, , A class of optimal rectangular filtering matrices for single-channel signal enhancement in the time domain, 2013, IEEE Transactions on Audio, Speech, and Language Processing, vol. 21, no. 12, pp.2595-2606.

[5].Cohens I, Berduge B. Speech enhancement for non-stationary noiseenvironments. Signal Process 2001;81(11):2403–18.

[6]. S. Singh, M. Tripathe, R. S. Anands, Single channel speech enhancement for mixed non-stationary noise environments, 2014, Advances in Signal Processing and Intelligent Recognition Systems, vol. 264, pp. 545-555.

[7] Shambhu Shankar Bhartis, Manish sharma, and SuneetaAgarwal, "A NewSpectral Subtraction Method for Speech Enhancement using AdaptiveNoise Estimation", 2016 3rd International Conference on Recent Advancesin Information Technology (RAIT),pp. 128-132, 2016.

[8] Ching-Ta Lu, Kun-Fu Tsen, Yung-Yue Chen, Ling-Ling Wang, and Chung-Lin Leita, "Speech Enhancement Using Spectral SubtractionAlgorithm with Over-Subtraction and Reservation Factors Adapted byHarmonic Properties", 2016 International Conference on Applied SystemInnovation (ICASI),pp. 1-5, 2016.

[9] I. Cohens and B Bergud "Noise Estimation by Minima Controlled Recursive Averaging for Robust Speech Enhancement", *IEEE Signal Processing Letters*, Vol.9, No.1, pp.12-15, 2002.