International Journal of Engineering & Technology, 7 (2.8) (2018) 456-460

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research Paper

Speech intelligibility improvement based on adaptive exponential smoothing factor

S Janardhanarao[1*], J Krishna chaitanya[2], B A David[3]

[1,2]Vardhaman college of Engineering
[3]Krest Embedded Technologies Pvt Ltd
**Corresponding Author Email: jsyamalapalli@gmail.com*

Abstract

This paper aims to increase the intelligibility of the speech content in the noisy speech signal. The approach presented in this paper tries to improve the intelligibility by adaptively re distributing the speech energy with exponential smoothing factor. Experiments were conducted on the different types of noises in noisy signal and found that there is a significant increase in intelligibility while preserving the quality of it.

Index terms*: Speech Enhancement, Intelligibility Improvement.*

1. Introduction

Speech enhancement or noise reduction is treated as one of the attractive research area in speech processing from past few decades. An extraneous component when merged with the original signal leads to the noisy speech signal with poor quality and intelligibility. These distorted components may be additive or multiplicative, however in practice the multiplicative noise should be considered due to reverberation. The main intention of this work is to increase the quality and informative regions of the distorted speech signal such that it can be clearly understand at the receiver.

For the near end listener the speech intelligibility is very poor due to the insertion of the noise from the background environment and also from the far end reflections. This can be overcome with pre processing before play back in order to become more intelligible w.r.t background noise.

In order to increase the informative and intelligibility parts of speech the level of speech can be increased but due to the audibility limitation of the loudspeakers the play back level cannot be increased after a certain point and also creates an unpleasant environment. One of the alternate solutions for this is to increase the distribution of energy within the speech signal in time or frequency transform domains.

One of the best solutions for near end problem and for improving the speech intelligibility is to boost high frequency components at a cost of low frequencies. Many solutions were proposed like, Griffiths et.al, Hall et. al in [2],[3] proposed to whiten the speech spectrum which is independent of noise type.

Sauert et.al and Zorila et. al in [4],[5] proposed to restructure the spectral components SNR values to be equal in magnitude over a band o frequencies. In [6] Taal et.al proposed a strategy to increase the intelligibility and improve speech intelligibility index (SII) with the use of linear filter.

In this paper an improvisation to the work done by Taal et.al [1] is presented by implementing the exponentially adaptive smoothing factor. The main motivation of this work is that the energy distributed over a time interval is flexible and this makes it possible to use this strategy in the applications where high delay is tolerated.

This paper is organized as follows, section 1 presents the need and necessity of the work with a brief introduction regarding the earlier works, section 2 presents the proposed algorithm and its mechanism to increase the intelligibility factor. Section 3 presents the experimental results that were obtained with the proposed approach

2. Proposed Processing Algorithm

Figure 1 shows the basic structure of the distortion measurement, let x denote the speech signal, ε is a far end background noise and $x + \varepsilon$ is noisy version of it. The distortion measure is the factor interprets the audibility of the given speech signal, if the measure is high then more the audible noise and vice versa.

The distortion measure factor is evaluated for three important factors firstly, this approach considers spectro-auditory model for the measurement and also includes the temporal envelope within a short interval of time period around 20-40 milli seconds. Secondly, this approach of distortion measurement obeys mathematical properties which makes it's more evitable to implement it for optimization. Lastly, this measurement correlates to the intelligibility of speech content in the signal [1], [4].

As shown in the figure1, the speech is decomposed into time frequency components and is further segmented into short intervals of 32ms each over squared Hanning window. These segmented short intervals of speech samples are passed into auditory filter bank which is further passed to a low pass filter. These steps of filtering process are included to extract the temporal envelope from the speech segments. Non-linear systems using controllers[11-20].

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

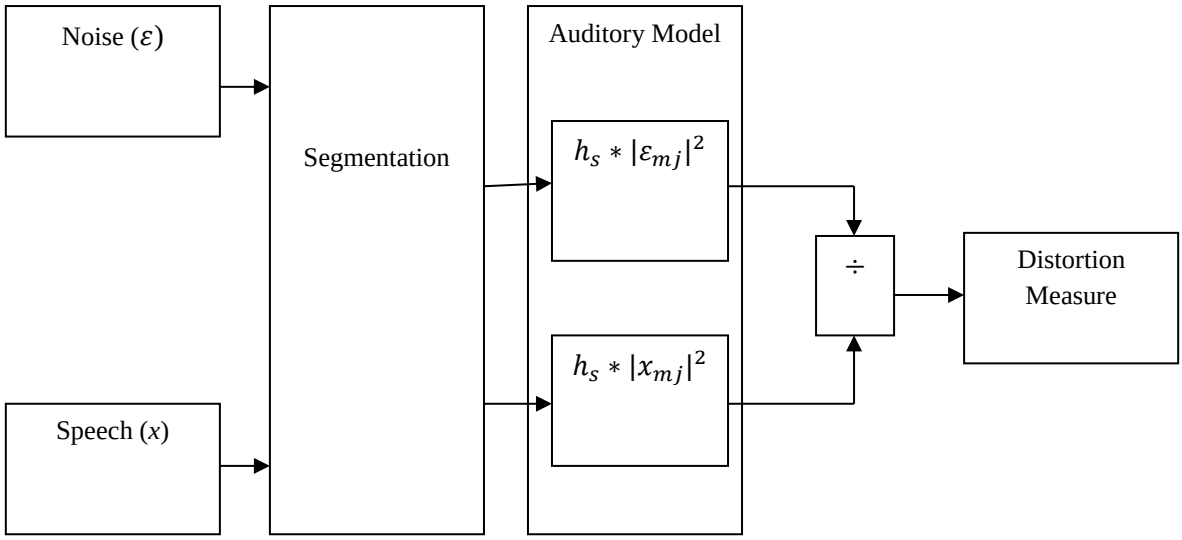


Fig. 1: Block diagram of distortion measurement

Let h_s denote the smoothing low pass filter response , then the distortion measure for the signal is given as

$$d(x_{mi,}\varepsilon_{m,i}) = \sum_n \frac{(h_s * |\varepsilon_{mj}|^2)(n)}{(h_s * |x_{mj}|^2)(n)} \quad (1)$$

Here 'n' denotes the time index within one short time frame. In order to increase the speech intelligibility equation (1) has to be minimized. This can be achieved by applying time/ frequency dependent gain function α.. Every individual speech frame's energy is redistributed by scaling.

$$\alpha_{m,i} = \underset{argmin}{} \sum_{\{m,i\}\in L} E[d(\alpha_{m,i}\, x_{m,i}, \varepsilon_{m,i})] \quad (2)$$

In the above equation the second term relates to the power measured at the output of auditory filters. The term 'E' represents the notation for expectation, in [1] Taal has stated 2 reasons for speech energy. Firstly, different algorithms are compared against each other under listening text and this is conducted under fixed SNR and optimal energy. Secondly, in contrast to loudness the approach is to be mathematically tractable despite of complex computations.
Let 'λ' denote a lagrangian multiplier and the cost function can be represented as

$$J = \sum_{\{m,i\}\in L} E[d(\alpha_{m,i}\, x_{m,i}, \varepsilon_{m,i})] + \lambda\left(\sum_{\{m,i\}\varepsilon L} \left|\left|\alpha_{m,i} x_{m,i}\right|\right|^2 - \left|\left|x_{m,i}\right|\right|^2\right) \quad (3)$$

The solution is given as

$$\alpha_{m,i}^2 = \frac{\left|\left|x_{m,i}\right|\right|^2 \beta_{m,i}^2}{\sum_{\{m,i\}\in L} \beta_{m',i'}^2 \left|\left|x_{m,i}\right|\right|^2} \quad (4)$$

Where $\beta_{m,i} = \left(\frac{E(x_{m,i}, \varepsilon_{m,i})}{\left|\left|x_{m,i}\right|\right|^2}\right)^{1/4}$

An exponential function is applied to $\alpha_{m,i}$ for smoothening; this results in reduction of sudden variation that may affect the quality of the speech.

$$\alpha_{m,\iota} = (1 - \widehat{\gamma)\alpha}_{m,\iota} + \gamma\widehat{\alpha_{m-1,\iota}} \quad (5)$$

It was stated that better results obtained when $\gamma = 0.9$ but this will not work for every signal. So in this paper a new approach is defined to calculate this gamma function given as

$$\gamma = \frac{1}{1+\exp(-a(SNR-T))}$$

Where $a = \frac{2}{1-e^{-2\mu SNR}} - 1$

Here is the step size taken as 0.25 and $3 \leq T < 5$

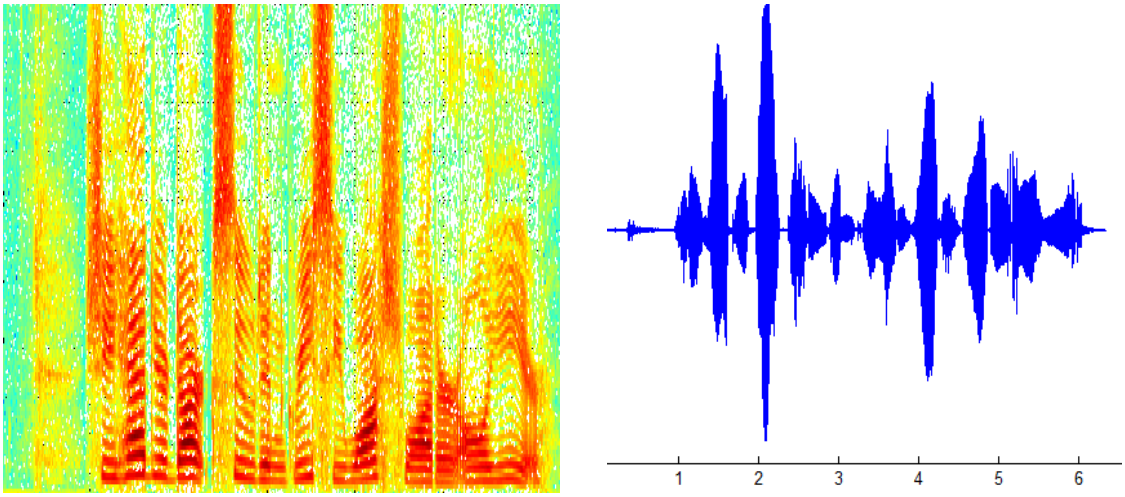
Estimation of noise statistics

The real valued frequency components in the frequency transformed domain are multiplied by the low pass filter coefficients. At reconstruction, the signal is reconstructed by adding all the tie frequency components under a squared Hanning window. Three important points are considered in estimating the properties of noise

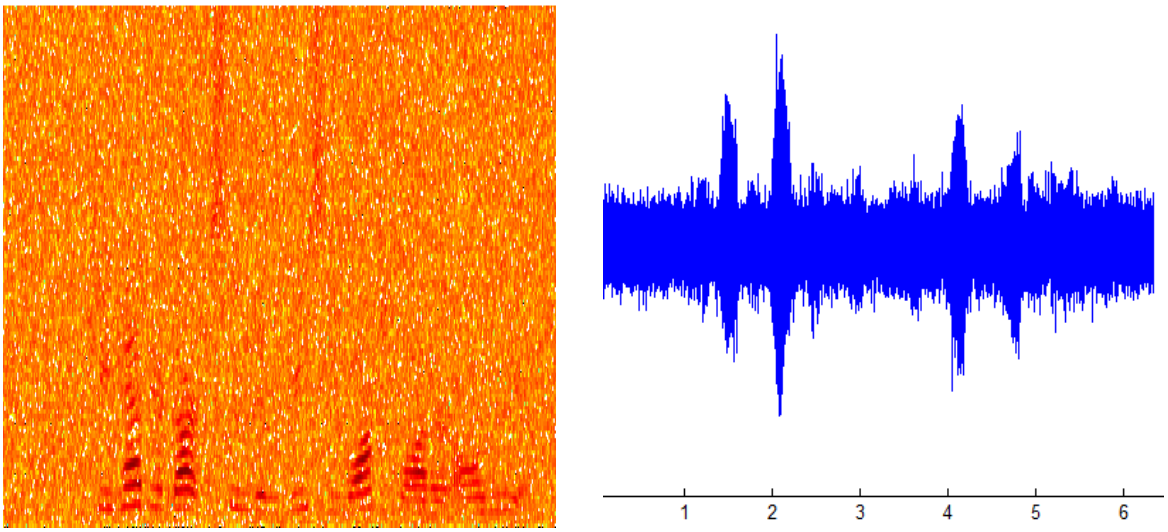
(i) The power spectral density is estimated from the previous time frames which are further used to calculate the intelligibility.
(ii) For clean unprocessed speech signals the PSD tracker is not applied, however, they are applied to processed speech signals.
(iii) In order to deal with delays and coloration of the signals the transfer function from the microphone to the loud speaker has to be known.

3. Experimental Results

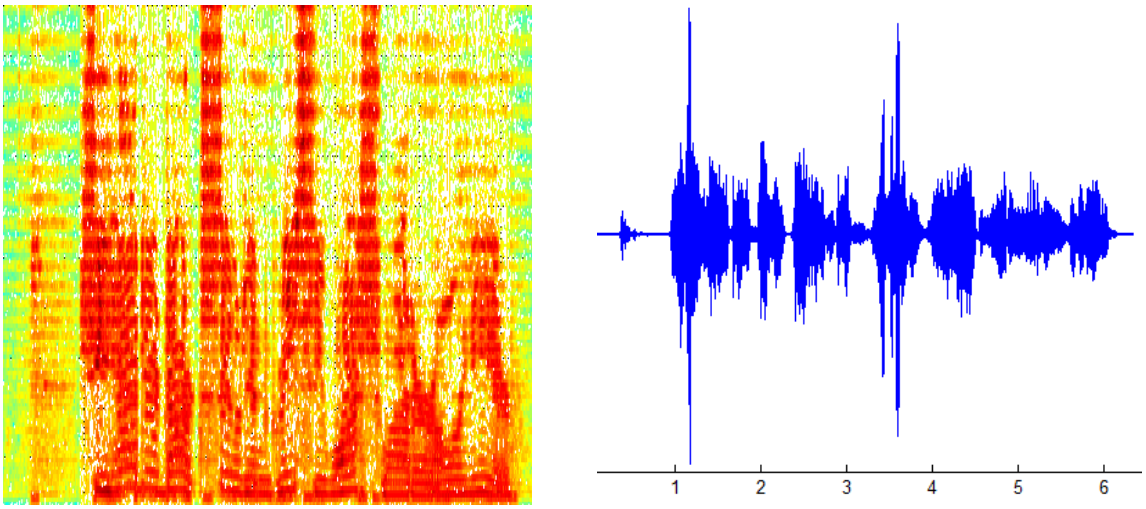
The proposed approach is tested with different noise speeches and compared against the Taal [1] method. Both qualitative and subjective analysis is been made, Perceptual evaluation of speech quality (PESQ) [8] and short time objective intelligibility (STOI) [9] were calculated for comparison.



(a) Original Clean speech & its spectrum

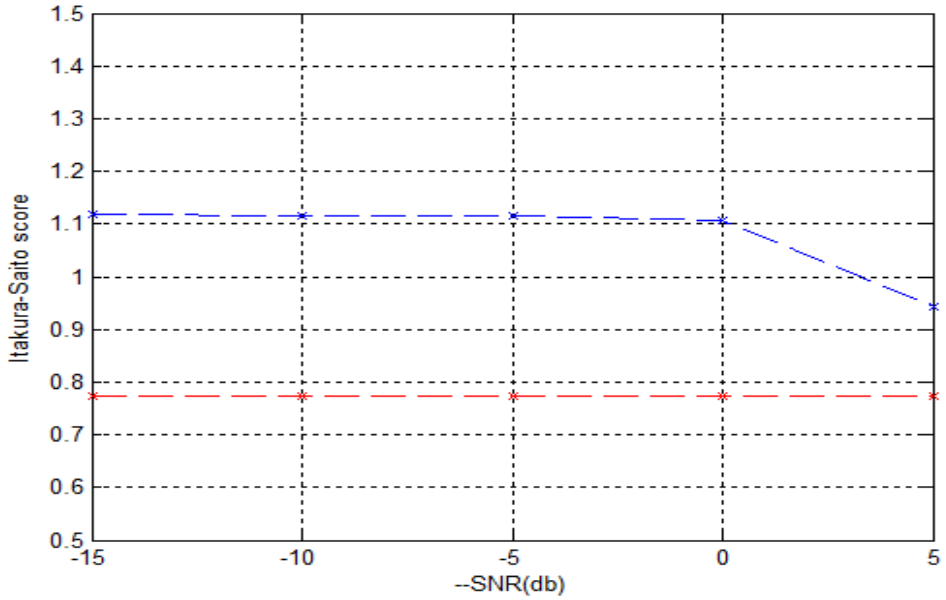


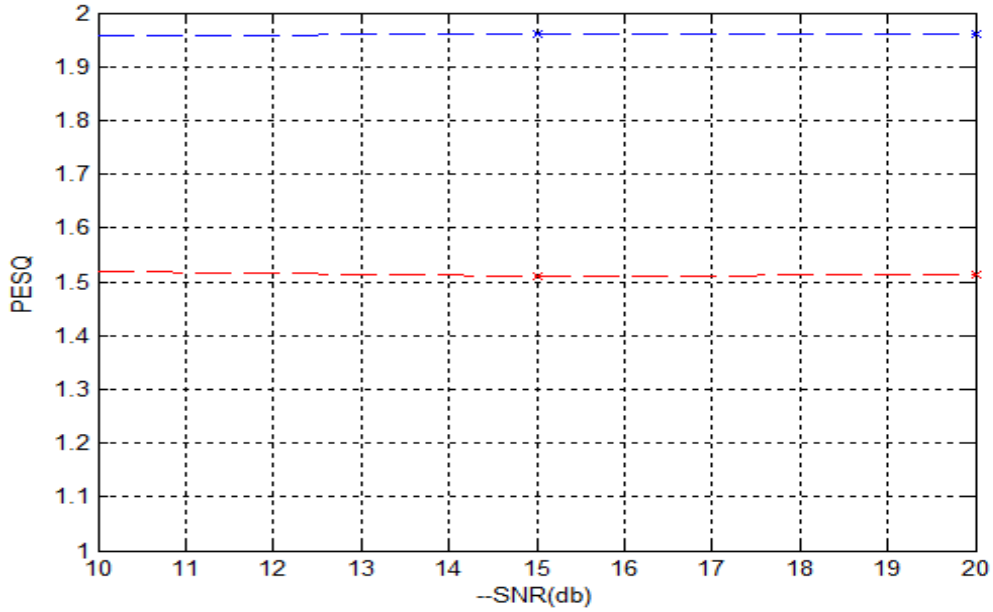
(b) Corrupted Signal with Noise & its Spectrum

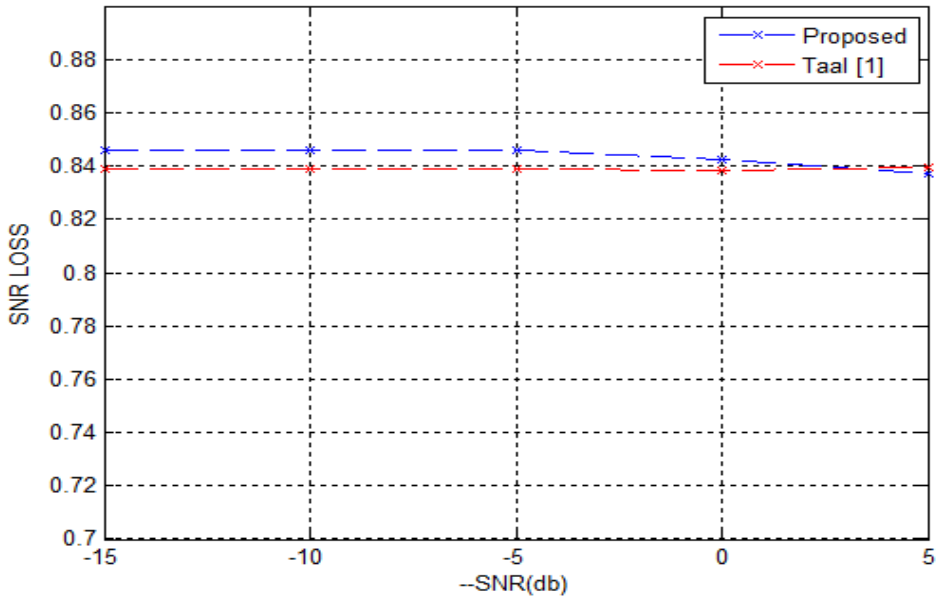


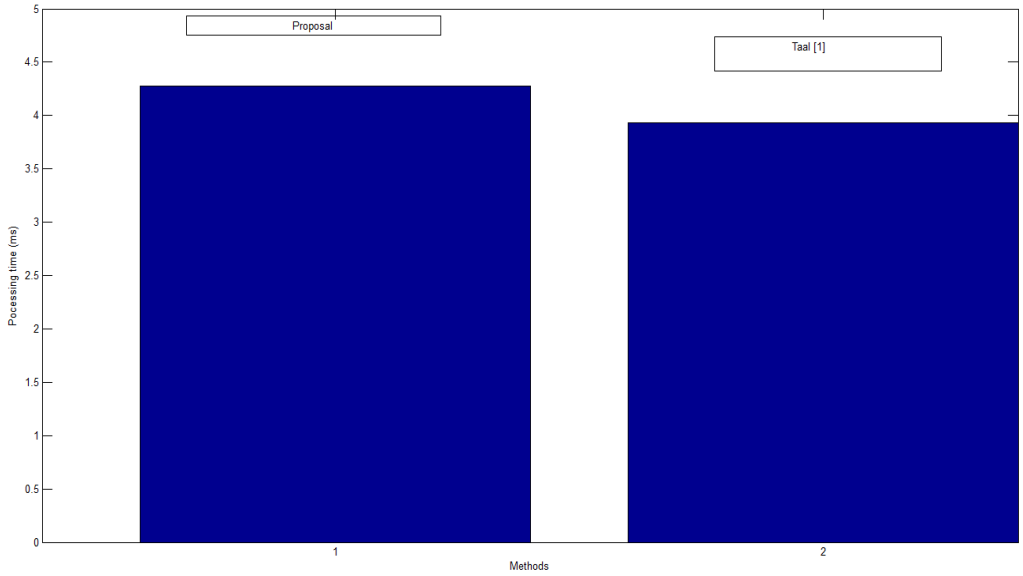
(c) Enhanced Signal & Its Spectrum

Fig. 2: Experimental results obtained for a speech signal


Itakura-Saito score
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
-15
-10
-5
0
5
--SNR(db)


PESQ
2
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1
10
11
12
13
14
15
16
17
18
19
20
--SNR(db)


Proposed
Taal [1]
SNR LOSS
0.88
0.86
0.84
0.82
0.8
0.78
0.76
0.74
0.72
0.7
-15
-10
-5
0
5
--SNR(db)



4. Conclusion

The perceptual speech intelligibility measurement factor is proposed in this paper, the approach is stable and can able to attain considerable results when compared traditional Taal approach. The PESQ value is 1.52 for Taal approach on the other hand its 1.95 for proposed approach. However, the proposed approach consumes more CPU time than the earlier one.

References

[1] Taal, C.H., et al., Speech energy redistribution for intelligibility improvement in noise based on a perceptual distortion measure. Comput. Speech Lang. (2013), http://dx.doi.org/10.1016/j.csl.2013.11.003

[2] J. D. Griffiths, "Optimum linear filter for speech transmission," J. Acoust. Soc. Am., vol. 43, no. 1, pp. 81–86, 1968.

[3] J. L. Hall and J. L. Flanagan, "Intelligibility and listener preference of telephone speech in the presence of babble noise," J. Acoust. Soc. Am., vol. 127, no. 1, pp. 280–285, 2010.

[4] B. Sauert, G. Enzner, and P. Vary, "Near end listening enhancement with strict loudspeaker output power constraining," in Proceedings of International Workshop on Acoustic Echo and Noise Control (IWAENC), 2006.

[5] T. Zorila, V. Kandia, and Y. Stylianou, "Speech-in-noise intelligibility improvement based on power recovery and dynamic range compression," in Proc. EUSIPCO, 2012, pp. 2075–2079.

[6] Taal, C.H., Jensen, J., Leijon, A., 2013. On optimal linear filtering of speech for near-end listening enhancement. IEEE Signal Processing Letters20 (3), 225–228

[7] Taal, C.H., Hendriks, R.C., Heusdens, R., 2012. A low-complexity spectro temporal distortion measure for audio processing applications. IEEE Transactions on Audio Speech and Language Processing 20 (5), 1553–1564

[8] A. Rix, J. Beerends, M. Hollier, and A. Hekstra, "Perceptual evaluation of speech quality (pesq) - a new method for speech quality assessment of telephone networks and codecs," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, 2001, pp. 749–752

[9] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "A short time objective intelligibility measure for time-frequency weighted noisy speech," in Proc. ICASSP, 2010, pp. 4214–4217.

[10] Tang, Yan, and Martin Cooke. "Subjective and Objective Evaluation of Speech Intelligibility Enhancement under Constant Energy and Duration Constraints."Inter speech. 2011.

[11] R. Kalaivani, K. Ramash Kumar, S. Jeevananthan, "Implementation of VSBSMC plus PDIC for Fundamental Positive Output Super Lift-Luo Converter," Journal of Electrical Engineering, Vol. 16, Edition: 4, 2016, pp. 243-258.

[12] K. Ramash Kumar,"Implementation of Sliding Mode Controller plus Proportional Integral Controller for Negative Output Elementary Boost Converter," Alexandria Engineering Journal (Elsevier), 2016, Vol. 55, No. 2, pp. 1429-1445.

[13] P. Sivakumar, V. Rajasekaran, K. Ramash Kumar, "Investigation of Intelligent Controllers for Varibale Speeed PFC Buck-Boost Rectifier Fed BLDC Motor Drive," Journal of Electrical Engineering (Romania), Vol.17, No.4, 2017, pp. 459-471.

[14] K. Ramash Kumar, D.Kalyankumar, DR.V.Kirbakaran" An Hybrid Multi level Inverter Based DSTATCOM Control, Majlesi Journal of Electrical Engineering, Vol. 5. No. 2, pp. 17-22, June 2011, ISSN: 0000-0388.

[15] K. Ramash Kumar, S. Jeevananthan, "A Sliding Mode Control for Positive Output Elementary Luo Converter," Journal of Electrical Engineering, Volume 10/4, December 2010, pp. 115-127.

[16] K. Ramash Kumar, Dr.S. Jeevananthan," Design of a Hybrid Posicast Control for a DC-DC Boost Converter Operated in Continuous Conduction Mode" (IEEE-conference PROCEEDINGS OF ICETECT 2011), pp-240-248, 978-1-4244-7925-2/11.

[17] K. Ramash Kumar, Dr. S. Jeevananthan," Design of Sliding Mode Control for Negative Output Elementary Super Lift Luo Converter Operated in Continuous Conduction Mode", (IEEE conference Proceeding of ICCCCT-2010), pp. 138-148, 978-1-4244-7768-5/10.

[18] K. Ramash Kumar, S. Jeevananthan, S. Ramamurthy" Improved Performance of the Positive Output Elementary Split Inductor-Type Boost Converter using Sliding Mode Controller plus Fuzzy Logic Controller, WSEAS TRANSACTIONS on SYSTEMS and CONTROL, Volume 9, 2014, pp. 215-228.

[19] N. Arunkumar, T.S. Sivakumaran, K. Ramash Kumar, S. Saranya, "Reduced Order Linear Quadratic Regulator plus Proportional Double Integral Based Controller for a Positive Output Elementary Super Lift Luo-Converter," JOURNAL OF THEORETICAL AND APPLIED INFORMATION TECHNOLOGY, July 2014. Vol. 65 No.3, pp. 890-901.

[20] Arunkumar, T.S. Sivakumaran, K. Ramash Kumar, "Improved Performance of Linear Quadratic Regulator plus Fuzzy Logic Controller for Positive Output Super Lift Luo-Converter," Journal of Electrical Engineering, Vol. 16, Edition:3, 2016, pp. 397-408.

[21] S.V.Manikanthan and V.Rama"Optimal Performance of Key Predistribution Protocol In Wireless Sensor Networks" International Innovative Research Journal of Engineering and Technology ,ISSN NO: 2456-1983,Vol-2,Issue –Special –March 2017.

[22] T. Padmapriya, V.Saminadan, "Performance Improvement in long term Evolution-advanced network using multiple imput multiple output technique", Journal of Advanced Research in Dynamical and Control Systems, Vol. 9, Sp-6, pp: 990-1010, 2017.