Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 8 July 2021: 5619 – 5628

Research Article

Speech Enhancement using Adaptive Filtering with Different Window Functions and Overlapping Sizes

SenthamizhSelvi R[1], Sathish Kumar P[2], Sri Krishna R[3], Surya Rao S[4]

ABSTRACT

Speech is the essential form of human communication. Speech processing is the research on speech signals and its processing methods. Noise is the unwanted sound in speech. Noise can cause communication issues, hearing problems, psychological health problems and many more. In most of themodern communication systems, speech enhancement plays a vital role. While transmitting the speech signal, the quality of that signal will degrade due to interference in the surrounding it is passing through. This paper is focused on performance analysis on enhancing the speech by using various windows, transformation techniques and overlapping percentage. The Kalman filter is used for filtering degraded speech signal. The windows used to perform analysis on this work are Hanning, Blackman, Hamming and Cosh window. The transformations used are Fast Fourier Transform(FFT) andDiscrete Cosine Transform(DCT). In this work, the noisy input signal is divided into multiple number of frames, each frames are sent through the window, after which the overlapping ofthose frames will be done. This overlapped signal will be applied to transformation technique and filtering process will be done. The output signal will be the enhanced and noise will be reduced to a certain extent. In this way, various combinations of windows, transformations and overlapping percentage, the amount of enhancement obtained is measured by taking the values of Signal to Noise Ratio(SNR) and performance analysis is done with those values. The highest value of SNR of 44.2409 dB is obtained by using the combination of Cosh window in FFT transform of overlapping percentage of 50%.

Keywords- Speech Enhancement, Windowing, Overlapping sizes, Fast Fourier Transform, Discrete Cosine Transform, Kalman filter, Signal to Noise Ratio.

I. INTRODUCTION

Speech enhancement system is used to improve the intelligibility and quality of speech, faded in the presence of degraded signal. Many sophisticated algorithms have been developed and the intense research is going on for the past five decades. When a listener and speaker are near to

[1,2,34] Electronics and Communication Engineering, Easwari Engineering College, Chennai, India
[1]senthamizhselvi.r@eec.srmrmp.edu.in, [2]sathish037358@gmail.com, [3]1616krishna@gmail.com,
[4]suryastunning.13@gmail.com

each other in a quiet surrounding, communication will be easy and accurate. However, in a noisy surrounding, it will be difficult to understand. Speech signals get distorted due to various types of background noises which results in listener fatigue. Hence, it is very essential to develop a system which can enhance the speech signal. Several algorithms were proposed to reduce the effect of background noise for improving the speech quality and intelligibility. Hidden Markov models in Mel-frequency domain is used for Speech Enhancement by Markov [1], enhancement of speech is achieved using Markov models in frequency domain in which efficiency is very low. Noise suppression based on an Analysis–synthesis approach is used to suppress the noise in degraded speech signal [2], helps to get a clear picture on reduction of noise. The noise in the audio signal will be reduced when its been passed through windows. In this work, the signal with noise is divided into several frames and passed through different windows and to reduce the noise, different transformation techniques are used and also Kalman filter is also used. It increases the spectral signal which has been disturbed by background noise, so as to improve the intelligibility and quality of speech. The speech enhancement system is evaluated using objective measures based on output SNR. This work performs analysis on different combinations of windows, overlapping sizes and transformation techniques. Enhancement of the speech by compressing the bandwidth of noisy signal [3], the noise is deducted also with that some of the clean speech also gets faded due to compressing bandwidth. So, here the signal is enhanced without reducing the bandwidth to achieve full information from the signal.

The remainder of this paper is organized as follows, Section II, provides the methodology on speech enhancement using various windows and overlapping sizes. Section III, tells about performance measures for speech enhancement and presents evaluation results. Section IV concludes the paper.

II. METHODOLOGY

Speech Enhancement system reduces the noise in the degraded signal and increases the intelligibility of the speech. To enhance the speech signal, multiple methods can be used. In this work, various windows and different overlapping sizes are implemented to enhance the speech signal. The following block diagram depicts the flow of this work,

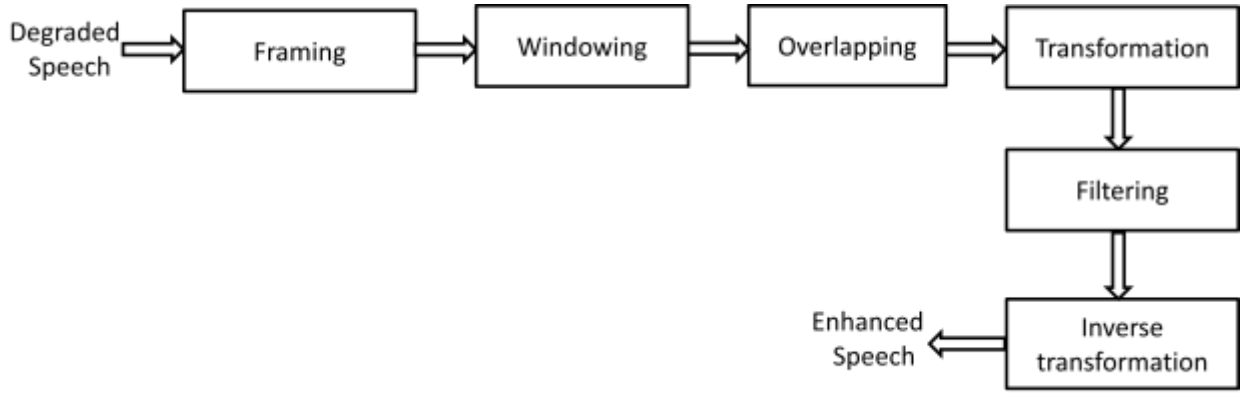

Fig.1. Block diagram of Speech Enhancement system

a.Framing:

The noisy signal is divided into multiple number of frames and each frame is passed through the window. This involved in preprocessing methods is the division of the speech signal into small pieces in which the frame with noise can be easily identified. The separation of frames will be modified with different sizes and the corresponding output for each modification is observed and tabulated.

b. Windowing:

In this section, each divided frame will be passed through window. Windowing methods act on raw data to reduce the effects of the leakage that occurs during an framing of the data. Four windows are used in this work. They are Blackman, Hanning, Hamming and Cosh window.

Blackman Window:

$$w(n) = 0.42 - 0.5\cos\left(\frac{2\pi n}{N-1}\right) + 0.08\left(\frac{4\pi n}{N-1}\right) \quad n = 0, \quad (1)$$

Cosh Window:

$$w[n] = \sum_{k=0}^{K}(-1)^{k} a_k \cos\left(\frac{2\pi kn}{N}\right), \quad 0 \le n \le N. \quad (2)$$

Hanning window:

$$\omega_0(x) \triangleq \begin{cases} \frac{1}{2}\left(1 + \cos\left(\frac{2\pi x}{L}\right)\right) = \cos^{2}\left(\frac{\pi x}{L}\right), & |x| \le L/2 \\ 0, & |x| > L/2 \end{cases} \quad (3)$$

HammingWindow:

$$h(n) = \alpha + (1.0 - \alpha)\cos\left[\left(\frac{2\pi}{N}\right)n\right] \quad (4)$$

c. Overlapping:

The frames passed through the window are collected and then combined into a single signal with different amount of overlapping percentages. This section is mainly used in times where there is the increasing number of frames in the overlap, which increases the speech error that missed even if one of the frames is not assigned in the correct form. The output values of these combinations are observed and tabulated.

d. Transformation:

The overlapped audio signal is allowed to perform transformation. It is used to transform the audio signal into a new form which is in certain aspects better than the original one Two transformations are used in this work. They are Fast Fourier Transform (FFT) and Discrete Cosine Transform (DCT).

Fast Fourier Transform:

$F(\omega) = \int_{-\infty}^{\infty} f(x) e^{-i\omega x}\, dx$ (5)

Discrete Cosine transform:

$$\mathrm{DCT(i,j)} = \frac{1}{\sqrt{2N}} \mathrm{C(i)C(j)} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \mathrm{pixel(x,y)} \cos\left[\frac{(2x+1)i\pi}{2N}\right] \cos\left[\frac{(2y+1)j\pi}{2N}\right] (6)$$

e. Filtering:

In this section, the signal in which transformation is applied is then passed through filter which is used to reduce the extra noise which is present in the signal after the transformation. In this work, Kalman filter is used which is a method that provides the value of some unknown variables given that the measurements observed over time. The mathematical expression of Kalman filter is given by,

$x(n) = -\sum_{i=1}^{p} \alpha_i x(n-i) + u(n)$(7)

f. Inverse Transformation:

In this section, the filtered audio signal is inversely transformed into its original state before transformation. In each combination, corresponding inverse transformation will take place according to the transformation used.

Inverse Fast Fourier Transform:

$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{i\omega x} d\omega$(8)

Inverse Discrete Cosine Transform:

$x[n] = w[n] \sum_{k=1}^{N} y[k] \cos\left(\frac{\pi(2k-1)(n-1)}{2N}\right) \ for\ 1 \le n \le N$ (9)

The audio signal which is obtained after the inverse transformation is the signal in which noise is reduced to a certain extent which is called as 'Enhanced Signal'. The SNR value of this enhanced signal calculated for each combination is tabulated and analysis is done based on the SNR value. The SNR value is inversely proportional to the noise present in the signal. Higher the SNR value, lower the noise.

III. RESULTS AND DISCUSSION

To examine the enhancement of speech by changing overlapping amount and windows with Kalman tracking technique, TIMIT database has been used. The clean audio speech data is taken from the TIMIT acoustic phonetic speech corpus. For evaluation, one audio of a male speaker is used from the database. The speech data are initially sampled at 16 kHz and quantized to 16 bits. The data is appropriately filtered and down sampled to 8 kHz to obtain the narrow band speech which are used in the analysis.

a. PERFORMANCE ANALYSIS:

In this, evaluation of performance of Speech Enhancement will be done. Two types of noise namely train noise, car noise are taken to analyze. Clean speech is manually changed at SNR level of 0, 5, 10 and 15 dB. The corrupted audio signal without performing enhancement is denoted as 'Noisy signal'. The clean speech signal is denoted as 'Clean signal'. The signal after enhancement process is denoted as 'Enhanced signal'. To appraise the modified speech with the joined effect of windowing, overlapping, transformations and filtering of all the stages of the system are studied.

The performance of enhancing the signal is measured through Spectrogram and waveform of the signal. Fig.2 shows the study of spectrogram for the combination of the signal with overlapping of 50%, cosh window and FFT transformation. Fig.2(a) shows the spectrogram of clean signals. Fig. 2(b), infers the spectrogram of noisy signal, car noise where the speech's harmonic part is not clearly visible because of the noise in the speech signal. Fig.2(c) illustrates the spectrogram of enhanced signal which shows the desired output. Fig.3 illustrates the waveform for the combination of the signal with overlapping of 50%, Cosh window and FFT transformation. Fig.3 (a) shows the waveform of clean signals. In Fig. 3(b), infers the waveform of noisy signal, car noise where the speech's harmonic part is not clearly visible because of the noise in the speech signal. Fig.3(c) represents the waveform of enhanced signal.

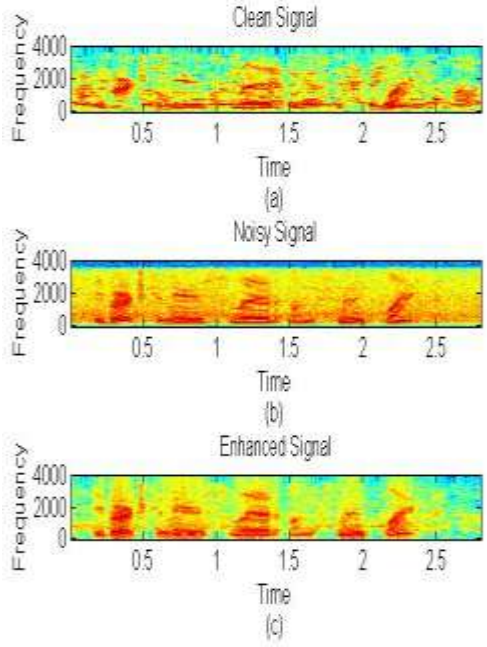

Fig.2. Spectrogram of (a) Clean signal (b) Noisy signal(c) Enhanced signal

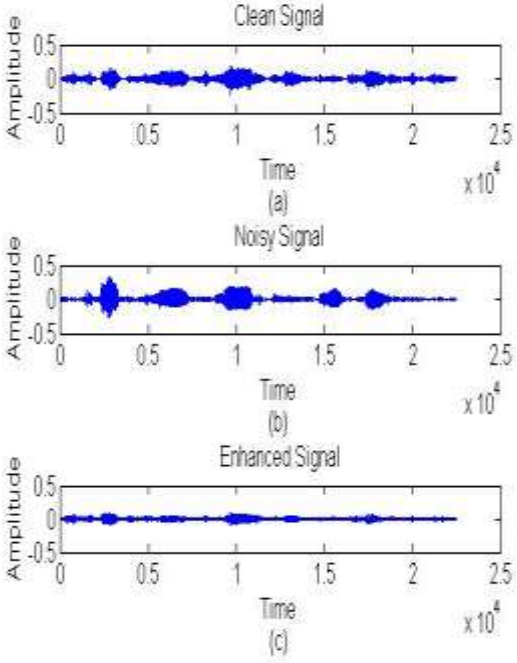

Fig.3. Waveform of (a) clean signal (b) Noisy signal (c) Enhanced signal

b. OBJECTIVE ANALYSIS:

The algorithm's performance is measured by the objective evaluation tool , Signal to Noise Ratio(SNR) which is most commonly used measure for quality of speech Experimental results of SNR values for overlapping amount of both 40% and 50% methods are represented in Tables I and II, respectively. The most commonly used measure for quality of speech is signal-to-noise ratio (SNR). An average SNR measure is defined as,

$$\text{SNR} = 10\log_{10}\left(\frac{P_S}{P_N}\right) db(12)$$

Where P_S and P_N are the power of signal and noise respectively.

Table 1 Output Snr Value Of 40% Overlapping With Different Windows And Transformations

Noise Type	**Input SNR(dB)**	**Output SNR(dB)**							
		Blackman window		**Hamming Window**		**Hanning window**		**Cosh window**	
		DCT	**FFT**	**DCT**	**FFT**	**DCT**	**FFT**	**DCT**	**FFT**
Car noise	**0**	5.0396	5.6797	2.9842	3.6206	3.6043	4.2224	25.8936	25.9753
	5	8.8512	9.1097	6.7064	6.9542	7.3494	7.5843	32.9107	32.9603
	10	12.8982	12.9650	10.7414	10.7912	11.3886	11.4392	38.7649	38.7875
	15	**17.6024**	**17.5942**	**15.4108**	**15.3866**	**16.0558**	**16.0312**	**44.1560**	**44.1798**
Train noise	**0**	3.3922	5.4506	3.4302	3.3558	3.0042	3.9442	25.6839	25.9671
	5	7.3745	8.9517	7.2274	6.8169	7.3376	7.4488	32.1993	30.4581
	10	11.3944	13.0051	11.4329	10.8278	11.6223	11.4779	37.4334	38.8394
	15	**16.0661**	**17.5573**	**16.2582**	**15.3957**	**16.2582**	**16.0413**	**42.0506**	**43.0981**

In Table 1, the output SNR values are compared with the different combinations of windows and transformations at different levels with overlapping of 40%. The experimental results shows that the combination of coshwindow with FFT transformation at 15dB provides high SNR value for 40% overlapping.

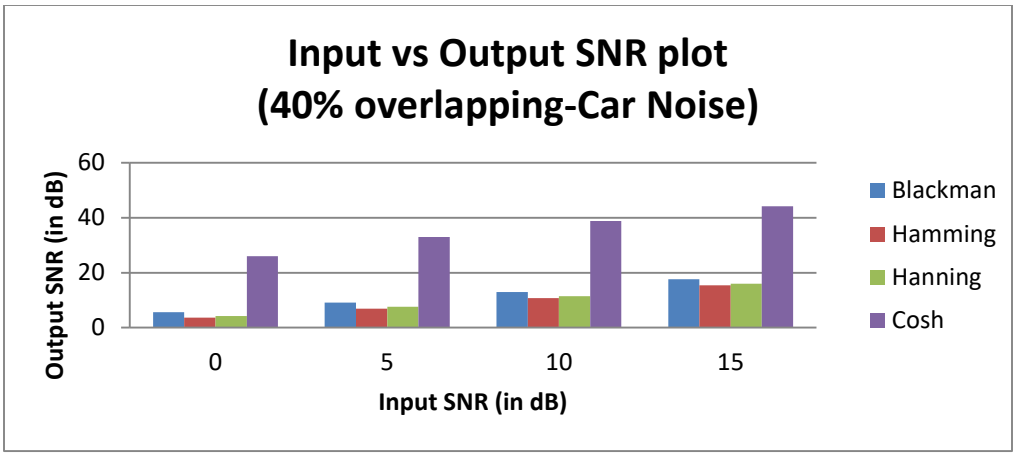

Fig.4. Input vs Output SNR plot

Fig.4 represents the Bar graph of output SNR values of different combinations of windows and transformations at different levels with overlapping of 40%.

Table 2 Output Snr Value Of 50% Overlapping With Different Windows And Transformations

Noise Type	Input SNR(dB)	Output SNR(dB)							
		Blackman window		Hamming Window		Hanning Window		Cosh Window	
		DCT	FFT	DCT	FFT	DCT	FFT	DCT	FFT
Car noise	0	6.2923	6.9702	4.5174	5.1717	5.0941	5.7411	25.9218	25.9753
	5	10.1683	10.4483	8.2754	8.5425	8.8782	9.1302	32.3221	32.9603
	10	14.2702	14.3746	12.3534	12.4376	12.9637	13.0399	38.7650	38.9253
	15	**18.9322**	**18.9499**	**17.0235**	**17.0339**	**18.6322**	**17.6362**	**44.0220**	**44.2409**
Train noise	0	5.8268	6.6606	4.8478	4.8311	5.2184	5.4100	25.6280	25.9581
	5	8.8101	10.3832	7.8322	8.4638	10.4328	9.0644	31.4592	32.9557
	10	12.9675	14.3538	12.9772	12.4086	12.5488	13.0162	37.6263	38.9385
	15	**17.6303**	**18.9421**	**17.6007**	**17.0167**	**17.6823**	**17.6224**	**43.5892**	**44.1658**

In Table 2, the output SNR values are compared with the different combinations of windows and transformations at different levels with overlapping of 50%. The experimental results show that the combination of Coshwindow with FFT transformation at 15dB provideshigh SNR value for 50% overlapping.

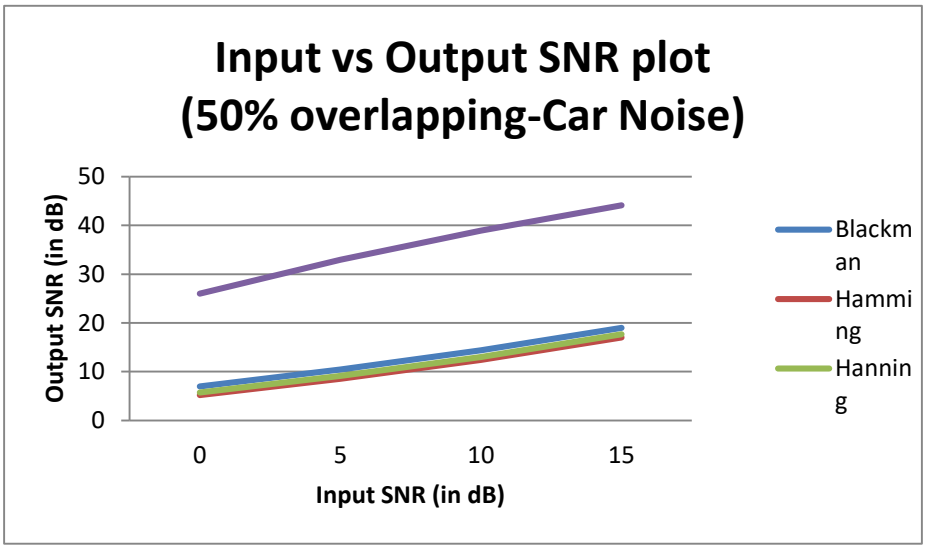

Fig.5.Input Vs. Output SNR plot

Fig.5 represents the Bar graph of output SNR values of different combinations of windows and transformations at different levels with overlapping of 40%.Comparing the highest SNR values of both 40% and 50% overlapping, the combination of Cosh window usingFFT transformation at 15dB with overlapping of 50% provides high SNR value of 44.2409dB for car noise and 44.1658dB for train noise.

IV. CONCLUSION

The proposed speech enhancement system provides the time to frequency characteristics of speech. Performance analysis of speech enhancement using different windows, overlapping sizes and transformation techniques gives idea of using the effective combination of methods used to reduce the noise and increase the intelligibility of the speech according to the user needs. Future development of obtaining noise free signal using other types of windows, transforms which will be more reliable than the windows and transforms used in this work. This analysis will be useful to most of the real life applications where the effective combination of this method can be implemented in the hearing aid devices and also in mobile phones to reduce noise. This method of enhancing speech is more efficient than the existing algorithms based on the analysis on SNR. The highest value of SNR of 44.2409 dB is obtained by using the combination of Cosh window in FFT transform of overlapping percentage of 50%.

REFERENCES

1. H. Veisi and H. Sameti, “Speech Enhancement using Hidden Markov models in Mel-frequency domain,” Speech Communication, vol. 55, no. 2, pp. 205–220, feb. 2013.
2. R.F.Chen, C.F.Chan, and H.C.So, “Noise suppression based on an analysis–synthesis approach,” in Proc. Eur. Signal Process. Conf.(EUSIPCO), Aug. 2010, pp. 1539–1543.1336 IEEE Transactions On Audio, Speech, And Language Processing, VOL. 20, NO. 4, MAY 2012.

3. J.S.Lim and V. O. Alan, "Enhancement and bandwidth compression of noisy speech," Proceedings of the IEEE, vol. 67, no. 12, pp. 1586–1604, 1979.
4. Akarsh K.A and Selvi R.S, "Speech enhancement using non negative matrix factorization and enhanced NMF" IEEE Conference Publications Circuit, Power and Computing Technologies (ICCPCT), Pages: 1 - 7, 2015.
5. J. S. Erkelens, R. C. Hendriks, R. Heusdens, and J. Jensen, "Minimum Mean-Square Error estimation of discrete Fourier coefficients with generalized Gamma priors," IEEE Trans. Audio, Speech, and LanguageProcess., vol. 15, no. 6, pp. 1741–1752, 2007.
6. D. Wang and J. Lim, "The unimportance of phase in speech enhancement," IEEE Trans. Acoust., Speech, Signal Process., vol. 30, no. 4, pp. 679–681, Aug. 1982.
7. K. W. Wilson, B. Raj, and P. Smaragdis, "Regularized non-negative matrix factorization with temporal dependencies for speech denoising," Interspeech, pp. 411–414, 2008.
8. P. C. Loizou, Speech Enhancement: Theory and Practice. Boca Raton: Taylor & Francis Group, 2007.
9. Qin Yan Saeed VaseghiEsfandiarZavarehei Ben Milner, "Kalman filter with linear predictor and harmonic noise models for noisy speech enhancement", 14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006.
10. C. F. Chan and E. W. M. Yu, "Improving pitch estimation for efficient multiband excitation coding of speech," Electron. Lett., vol. 32, no. 10, pp. 870–872, May 1996.
11. Y. Soon, S. N. Koh, and C. K. Yeo, "Improved noise suppression filter using self-adaptive estimator of probability of speech absence," Signal Processing, vol. 75, no. 2, pp. 151–159, 1999.
12. R. Senthamizh Selvi1 , G. R. Suresh, and S. Kanaga Suba Raja "A New Hybridized Speech Enhancement Technique for Stationary and Non-Stationary Noisy Environments Journal of Computational and Theoretical Nanoscience Vol. 14, 1–8, 2017
13. SenthamizhSelvi.R and G.R. Suresh, "'Hybridization of Spectral Filtering with Particle Swarm Optimization for Speech Signal Enhancement', International Journal of Speech Technology, DOI 10.1007/s10772-015-9317-1, 2015.
14. R.SenthamizhSelvi, R. Kishore, G.R. Suresh, S.Kanaga Suba Raja "Embedding data in audio signals using HSA-EMD Algorithm" ICONSTEM- COMPACTDISC-17
15. Akarsh K.A., SenthamizhSelviR.,and Suresh G.R."Real Time Speech Enhancement and Recognition using Enhanced Non Negative Matrix Factorization", 2015, in "Springer International Conference on Soft Computing Systems".
16. Akarsh K.A., SenthamizhSelvi R., and Suresh G.R. "Speech Enhancement using Non negative Matrix Factorization and Enhanced NMF", 978-1-4799-7074-2/15/$31.00 ©2015 IEEE
17. I.Sangeetha, R. SenthamizhSelvi and G.R. Suresh, "Speech Enhancement Based on Harmonic Noise Model with Kalman Tracking", Proceedings of IEEE sponsored Fourth international conference on Recent Trends in Information Technology(ICRTIT 2014), April 2014, MIT, Anna University, Chennai.
18. Ram prakash B., SenthamizhSelvi R. and Suresh G.R., "Parallel spectral and cepstral modeling based speech enhancement using Hidden Markov Model," *2014 International Conference on Communication and Signal Processing*, 2014, pp. 1467-1471, doi: 10.1109/ICCSP.2014.6950092.
19. Murugan, S., Jeyalaksshmi, S., Mahalakshmi, B., Suseendran, G., Jabeen, T. N., & Manikandan, R. (2020). Comparison of ACO and PSO algorithm using energy

consumption and load balancing in emerging MANET and VANET infrastructure. *Journal of Critical Reviews*, *7*(9), 2020.

20. Subramaniyan, Murugan, et al. "Deep Learning Approach Using 3D-ImpCNN Classification for Coronavirus Disease." Artificial Intelligence and Machine Learning for COVID-19 (2021): 141-152.
21. Alzubi, J. A. (2021). Bipolar fully recurrent deep structured neural learning based attack detection for securing industrial sensor networks. *Transactions on Emerging Telecommunications Technologies*, *32*(7), e4069.
22. Alzubi, J. A., Jain, R., Kathuria, A., Khandelwal, A., Saxena, A., & Singh, A. (2020). Paraphrase identification using collaborative adversarial networks. *Journal of Intelligent & Fuzzy Systems*, *39*(1), 1021-1032.