# Denoising of Speech Signal using Empirical Mode Decomposition and Kalman Filter

**Nandhini A, Bharath K P, Mahalti Mohammed Sohail, Rajesh Kumar M**

***Abstract*: Speech denoising is the process of removing the noise from the noise corrupted speech. The applications of speech denoising are used in speech enhancement, speech recognition and many more. In this work, a new approach is proposed to de-noise the speech which is corrupted from different noises, Empirical mode decomposition and the Kalman filter (EMD-KF) is used for speech denoising in the proposed work. The clean speech is corrupted by the noise with the different SNR's, and further Empirical mode decomposition (EMD) is applied to the noise corrupted speech later the obtained resultant speech is passed through the Kalman filter (KF) which gives the denoised speech. The result shows that the mean squared error (MSE) values of EMD-KF are extremely less when compared to other methods like discrete wavelet transform (wavelet families like Daubechies and Symlet), empirical mode decomposition (EMD) and moving average filter followed by empirical mode decomposition (MA-EMD). As an application the proposed algorithm is used in the feature extraction for speech recognition. Mel frequency cepstral coefficient (MFCC) is performed on both the original speech and the denoised speech and found majority of the denoised speech features are similar to the original speech features and few denoised speech features are nearby to the original speech features.***

***Keywords*: Empirical mode decomposition (EMD), Kalman filter (KF), Mel-frequency cepstral coefficient (MFCC), Speech denoising.***

## I. INTRODUCTION

Speech is a subset of audio signal that refers to the sound made by a human speaker. The process of removing the noise from the corrupted speech is known as speech denoising. Initially the original speech is corrupted by the noise then the denoising method will be applied to the noise corrupted speech in order to recover the original speech. Yash Varadhan Varshney et al. [1] separated the speech from the noise corrupted speech by performing wavelet decomposition before sparse non-negative matrix factorization (SNMF).

**Revised Manuscript Received on June 30, 2020.**
***** Correspondence Author

**Nandhini A**, School of Electronics Engineering, VIT University, Vellore, India. Email: 18nandhiniakilan@gmail.com

**Bharath K P**, School of Electronics Engineering, VIT University, Vellore, India. Email: bharathkp25@gmail.com

**Mahalti Mohammed Sohail**, School of Electronics Engineering, VIT University, Vellore, India. Email: msohail17925@gmail.com

**Rajesh Kumar M***, School of Electronics Engineering, VIT University, Vellore, India. Email: mrajeshkumar@vit.ac.in

Hongqing Liu et al. [2] suppressed impulsive noise by optimization and gaussian noise by wiener filter in different transform domains like short time fourier transform (STFT), wavelet transform (WT) and wavelet synchrosqueezed transform (WSST). Peng Xiong et al. [3] used to filter most of the noise found in the electrocardiogram signals by wavelet transform with scale adaptive thresholding method and the residual noise is removed by a deep neural network based on improved denoising autoencoder. Shubhratha S et al. [4] analyzed the performance of two denoising methods (discrete wavelet transform and empirical mode decomposition). Tassadaq Hussain et al. [5] used extreme learning machine to remove the noise from the speech by random selection of hidden units. Yu-Cheng Su et al. [6] used generalized maximum a posteriori spectral amplitude (GMAPA) algorithm which takes up a smaller scale to avoid amends during high SNR and uses a larger scale to remove noise components during low SNR.

Christoph F. Stallmann et al. [7] employed various artificial neural networks to identify and remove the noise in musical sound waves. Chengli Sun et al. [8] transformed the noise corrupted speech into time-frequency domain where the noise and the speech are considered as a low-rank and sparse component respectively. Szu-Wei Fu et al. [9] employed fully convolutional network to preserve the local temporal structures particularly the high frequency components of the speech. Nikolaos Dionelis et al. [10] minimized the noise by modulation-domain kalman filtering which estimates the posterior distribution for inter frame speech in log-magnitude spectrum.

Ying-Hui Lai et al. [11] restored the original speech from the noise corrupted speech by deep denoising autoencoder (DDAE) and got better results when compared to other methods like wiener, Karhunen-Loève theorem (KLT) and log minimum mean square error (log MMSE). Haifa Touati et al. [12] removed the noise from the noise corrupted speech by empirical mode decomposition followed by adaptive least mean square (LMS) filter. Tomohiro Nakatani et al. [13] performed both denoising and dereverberation at the same time by weighted power minimization distortionless response (WPD) beamformer.

In this paper, the noise is reduced by passing the noise corrupted speech to EMD and then to the Kalman filter. The MSE values are calculated and compared with other denoising techniques. The original speech features and the denoised speech features are analyzed.

*Retrieval Number: H6313069820/2020©BEIESP*
*DOI: 10.35940/ijitee.H6313.069820*
*Journal Website: www.ijitee.org*

*Published By:*
*Blue Eyes Intelligence Engineering*
*& Sciences Publication*

## II. METHODOLOGY

In this section, the denoising methods and the proposed work are discussed as follows:

### A. Discrete Wavelet Transform (DWT)

DWT has two functions namely scaling functions (1) and wavelet functions (2). The wavelet function and the scaling function are associated with the high pass filter and the low pass filter respectively. Detail coefficients are obtained when the speech is passed through the high pass filter. Approximation coefficients are obtained when the speech is passed through the low pass filter. Detail coefficients are considered as the DWT coefficients for first level and the approximation coefficients are again passed to high and low pass filters and the process continues. The speech is reconstructed by adding the coefficients which are obtained from last level to first level. Similarly, detail and approximation coefficients are passed through high and low pass filters respectively.

The forward DWT is given by,

$$S_\alpha(i_0,l)=\frac{1}{\sqrt{N}}\sum_{m=0}^{N-1}x(m)\alpha_{i_0,l}(m) \tag{1}$$

$$W_\beta(i,l)=\frac{1}{\sqrt{N}}\sum_{m=0}^{N-1}x(m)\beta_{i,l}(m) \tag{2}$$

The inverse DWT is given by (3),

$$x(m)=\frac{1}{\sqrt{N}}\sum_{l}S_\alpha(i_0,l)\alpha_{i_0,l}(m)+\sum_{i=i_0}^{I-1}\sum_{l}W_\beta(i,l)\beta_{i,l}(m) \tag{3}$$

where N=$2^I$ and $i_0$=0.

$x(m)$ is speech, $S_\alpha$ is scaling coefficient, $W_\beta$ is wavelet coefficient and $N$ is total samples.

### B. Empirical Mode Decomposition (EMD)

EMD breaks down the speech into components called intrinsic mode functions (IMF). This process of converting the speech into IMF is known as sifting process. The mean is calculated between upper and lower envelopes of the speech by cubic-spline interpolation. The first component is obtained by taking difference between the original speech and mean and it is given by (4),

$$I_1=y(t)-M_1 \tag{4}$$

where $y(t)$ is speech, $I_1$ is first component and $M_1$ is mean.

The first component is considered as data and mean is computed by taking its upper and lower envelopes during the second iteration and it is given by (5),

$$I_{11}=I_1-M_{11} \tag{5}$$

The sifting process continues till an IMF is obtained and it is given by (6),

$$I_{1n}=I_{1(n-1)}-M_{1n} \tag{6}$$

where $n$ is number of iterations.

The residual is obtained by subtracting the IMF from the original speech and it is given by (7),

$$R_1=y(t)-I_{1n} \tag{7}$$

where $R_1$ is first residual and $I_{1n}$ is first IMF.

### C. Proposed method (EMD-KF)

In our work, the original speech is corrupted by the noise. The noise corrupted speech is given as input to empirical mode decomposition whose output speech is passed through the Kalman filter. Hence the denoised speech is obtained as the output of the filter. The block diagram of the proposed work is as shown in Fig. 1.

The speech used in this work is ‘sp01.wav’ which is taken from [14]. It is having time period of 20ms, sampling frequency of 8 kHz and total samples are 22529. The noise added in this work is additive white gaussian noise (AWGN) whose signal-to-noise ratio (SNR) is varying from 0 decibel to 20 decibel with increase of 5 decibel. The maximum number of iterations used for the sifting process is 30 and the maximum number of IMF obtained after EMD is 9.

Kalman filter predict the speech from the resultant speech obtained after applying EMD by following equations (8), (9), (10), (11) and (12).

$$\tilde{Y}(m|m-1)=\psi\tilde{Y}(m-1|m-1) \tag{8}$$

where $\tilde{Y}(m|m-1)$ is priori estimate of present state vector $Y(m)$ and $\psi$ is state transition matrix.

$$Q(m|m-1)=\psi Q(m-1|m-1)\psi^T+HPH^T \tag{9}$$

where $Q(m|m-1)$ is priori estimate error covariance matrix, $H$ is input matrix and $P$ is matrix of process noise covariance.

$$k(m)=Q(m|m-1)G^T(GQ(m|m-1)G^T+S)^{-1} \tag{10}$$

where $k(m)$ is Kalman gain for the $m$ th instant, $G$ is observation matrix and $S$ is matrix of measurement noise covariance.

$$\tilde{Y}(m|m)=\tilde{Y}(m|m-1)+k(m)(x(m)-G\tilde{Y}(m|m-1)) \tag{11}$$

where $Y\%(m|m)$ is posteriori estimate and $x(m)$ is speech corrupted by noise.

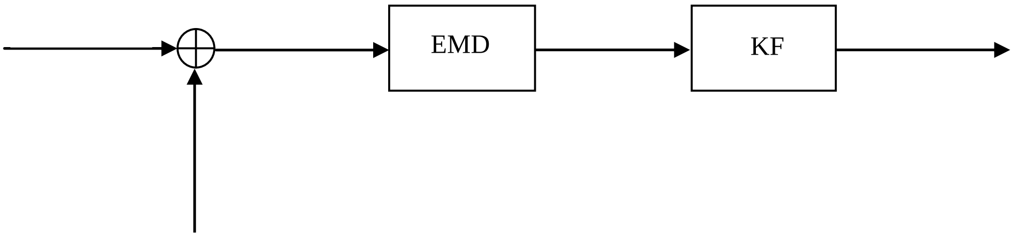

**Fig. 1.Block diagram of speech denoising using EMD-KF.**

$$Q(m|m)=(I-k(m)G)Q(m|m-1) \qquad (12)$$

where $Q(m|m)$ is posteriori estimate error covariance matrix and $I$ is identity matrix.

## III. APPLICATIONS

### A. Feature Extraction for Speech Recognition

Mel Frequency Cepstral Coefficient (MFCC) is the process of extracting features from the speech. MFCC is used in recognition of speech. The various processes in MFCC are pre-emphasis, framing, windowing, fast fourier transform (FFT), mel filter bank and discrete cosine transform (DCT) [15] and [16]. In pre-emphasis, the speech is passed through a filter which supports the high frequency.

In framing, the speech samples are divided into frames which contain the original characteristics of the signal. The discontinuities in the speech after framing are avoided by Hamming windowing technique. The Hamming windowing function is given by (13),

$$\mathrm{S[m]} = 0.54 - 0.46\cos\left[\frac{2\pi m}{K-1}\right] \qquad (13)$$

where K is the number of samples in one frame. FFT converts frame to frequency domain from time domain. Mel filter bank is the band pass filter which applies mel frequency scaling on the resultant signal. MFCC coefficients are obtained after applying DCT.

## IV. RESULTS

### A. Metric

Mean squared error (MSE) is the average squared difference between the actual speech and the noise corrupted speech computed by (14),

$$MSE=\frac{1}{m}\sum_{j=1}^{m}(X_j-\hat{X}_j)^2 \qquad (14)$$

where $X$ is actual speech, $\hat{X}$ is noise corrupted speech and $m$ is total samples.

### B. Experimental Results

The speech used in this work is ‘sp01.wav’ which is taken from [14]. It is having time period of 20ms, sampling frequency of 8 kHz and total samples are 22529. The original speech is corrupted by the noise. The noise added in our work is additive white gaussian noise (AWGN) whose signal-to-noise ratio (SNR) is 0dB. EMD is applied to the noise corrupted speech. The maximum number of iterations and the maximum number of intrinsic mode functions are 30 and 9 respectively. The resultant speech obtained after applying EMD to the noise corrupted speech is passed through the Kalman filter which gives the denoised speech.

Fig. 2 shows the intrinsic mode functions and residual obtained after applying EMD to the noise corrupted speech. The denoised speech is obtained after applying Kalman filter to the resultant speech as shown in Fig. 3.

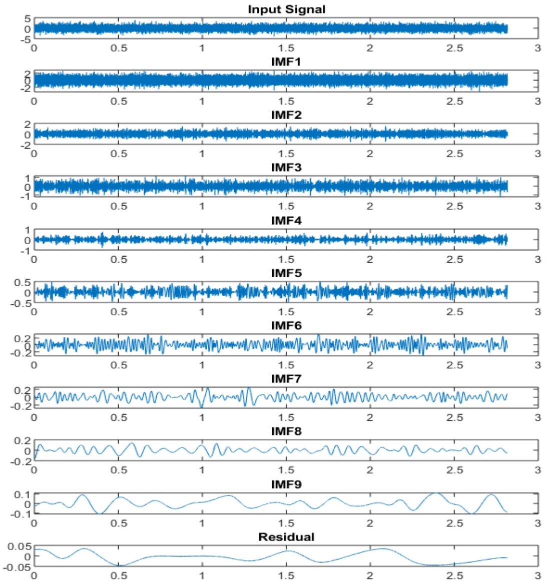

**Fig. 2.Intrinsic mode functions (IMF) and residual**

**obtained after performing EMD to the noise corrupted speech.**

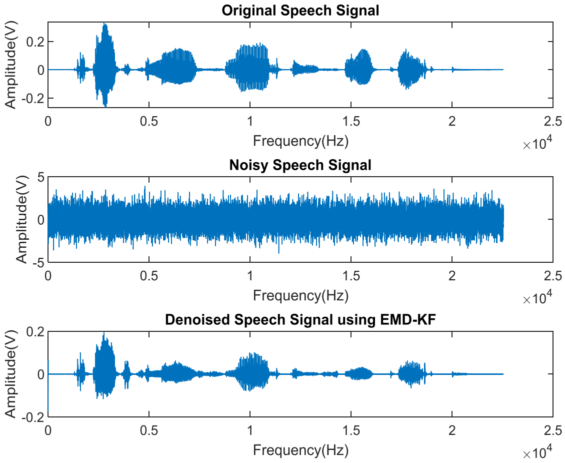

**Fig. 3.Speech signals obtained after applying EMD-KF.**

Table- I illustrates the mean squared error (MSE) values obtained for various denoising methods. The denoising methods compared are discrete wavelet transform (wavelet families – Daubechies 3, Daubechies 5, Symlet 4), empirical mode decomposition and moving average filter is applied before empirical mode decomposition.

The MSE values are obtained for SNR ranging from 0 decibel to 20 decibel with increase of 5 decibel. The MSE values of the proposed work are compared with the above mentioned denoising techniques and found that the MSE values obtained for EMD-KF are extremely less when compared with the MSE values obtained for other denoising methods. The lower value of MSE shows that the noise has been minimized in the noise corrupted speech. Hence the proposed (EMD-KF) denoising method gave better results when compared to other denoising techniques.

The MSE vs SNR is plotted for various denoising methods is shown in Fig. 4 and Fig. 5 shows that the bar graph is being plotted for the MSE vs SNR values for various denoising methods.

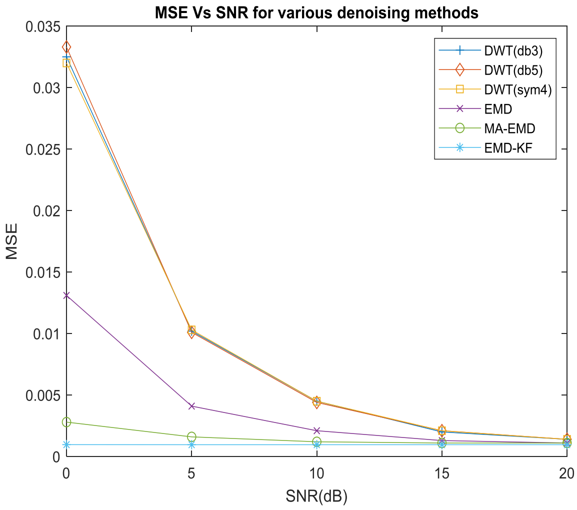

**Fig. 4.Mean squared error (MSE) vs signal-to-noise ratio (SNR) for different denoising methods – Plot.**

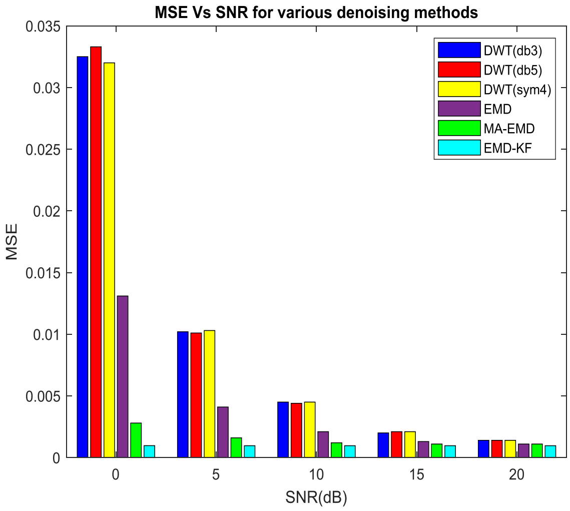

**Fig. 5.Mean squared error (MSE) vs signal-to-noise ratio (SNR) for different denoising methods – Bar graph.**

Fig. 6 and Fig. 7 show the original speech and the denoised speech. As an application the proposed work is used in the feature extraction for speech recognition. MFCC is applied for the original speech and the denoised speech.

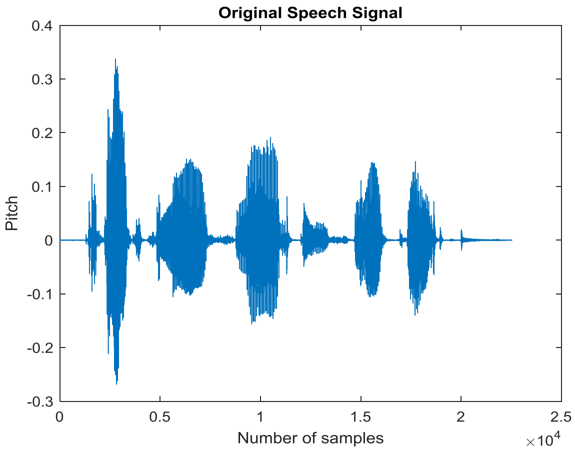

**Fig. 6.Original speech.**

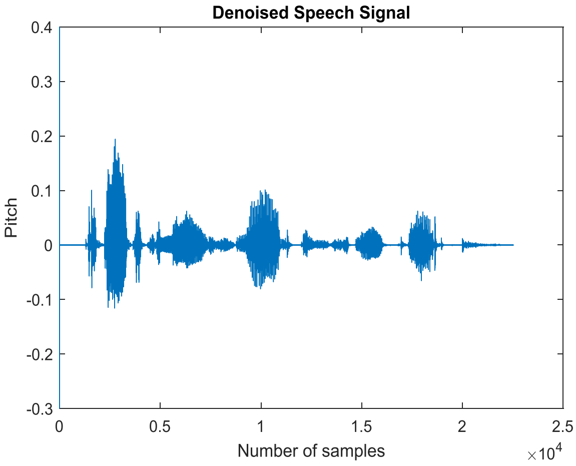

**Fig. 7.Denoised speech.**

*Retrieval Number: H6313069820/2020©BEIESP*
*DOI: 10.35940/ijitee.H6313.069820*
*Journal Website: www.ijitee.org*

*Published By:*
*Blue Eyes Intelligence Engineering*
*& Sciences Publication*

**Table- I: MSE values obtained for various denoising methods**

| SNR (dB) | MSE (original speech, noisy speech) | DWT | | | EMD | | |
|---|---|---|---|---|---|---|---|
| | | *MSE (original speech, db3)* | *MSE (original speech, db5)* | *MSE (original speech, sym4)* | *MSE (original speech, EMD)* | *MSE (original speech, MA-EMD)* | *MSE (original speech, EMD-KF)* |
| 0 | 0.9890 | 0.0325 | 0.0333 | 0.0320 | 0.0131 | 0.0028 | **0.00096792** |
| 5 | 0.3116 | 0.0102 | 0.0101 | 0.0103 | 0.0041 | 0.0016 | **0.00096283** |
| 10 | 0.1009 | 0.0045 | 0.0044 | 0.0045 | 0.0021 | 0.0012 | **0.00096293** |
| 15 | 0.0311 | 0.0020 | 0.0021 | 0.0021 | 0.0013 | 0.0011 | **0.00096281** |
| 20 | 0.0102 | 0.0014 | 0.0014 | 0.0014 | 0.0011 | 0.0011 | **0.00096269** |

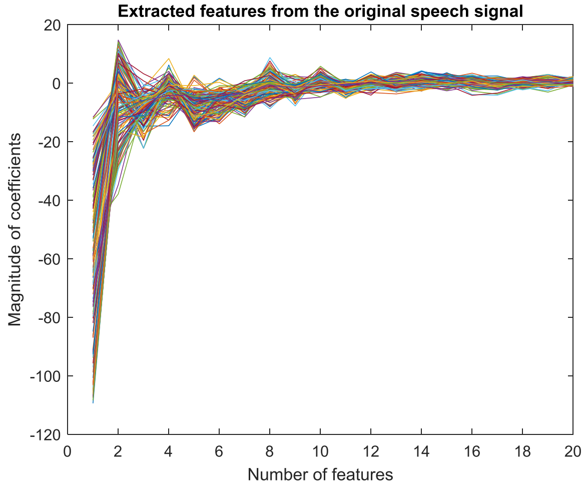

**Fig. 8.Features obtained after applying MFCC to the original speech.**

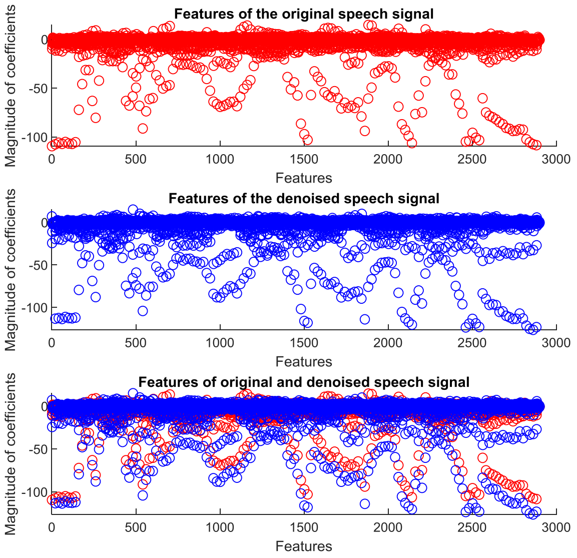

**Fig. 10. Features of original and denoised speech signals.**

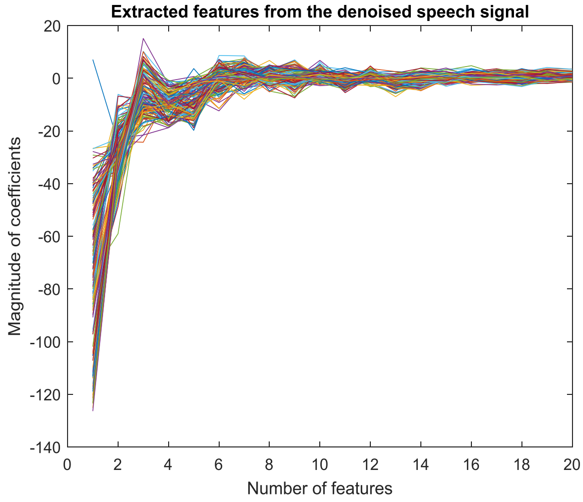

**Fig. 9.Features obtained after applying MFCC to the denoised speech.**

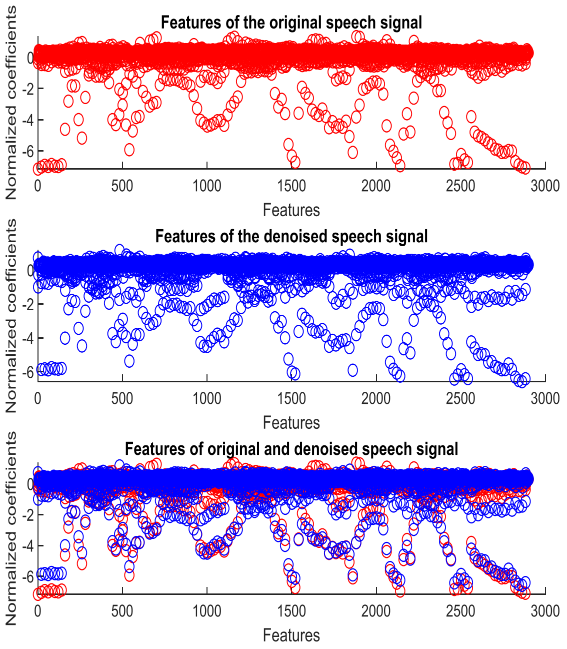

**Fig. 11. Features of original and denoised speech signals after normalization.**

Fig. 8 and Fig. 9 are the original speech features and the denoised speech features which are obtained after applying MFCC to the original speech and the denoised speech respectively.

The original speech features and the denoised speech features are plotted by scatter plot is shown in Fig. 10. The magnitude of coefficients values obtained for the original speech and the denoised speech are normalized and plotted by scatter plot is shown in Fig. 11 and found that the most of the denoised speech features are similar to the original speech features and few denoised speech features are nearby to the original speech features.

## V. CONCLUSION

In our work, the original speech is corrupted by the noise. EMD is applied to the noise corrupted speech and the Kalman filter is applied to the resultant speech obtained after EMD. The MSE is calculated for the proposed work (EMD-KF) and also for other denoising methods like DWT, EMD and MA-EMD. The MSE values obtained for noisy speech signal, DWT (Daubechies3), DWT (Daubechies5), DWT (Symlet4), EMD, MA-EMD and EMD-KF for AWGN having 0dB as SNR are 0.9890, 0.0325, 0.0333, 0.0320, 0.0131, 0.0028 and 0.00096792 respectively. It is found that the MSE values of EMD-KF are much reduced when compared to other denoising techniques. The lower value of MSE shows that the noise has been minimized. As an application the proposed approach is used in the feature extraction for speech recognition. MFCC is applied to both the original speech and the denoised speech and their features are obtained and found that the most of the denoised speech features are similar to the original speech features and few denoised speech features are nearby to the original speech features.

## REFERENCES

1. Yash Vardhan Varshney, Z. A. Abbasi, M. R. Abidi, and Omar Farooq, "SNMF based speech denoising with wavelet decomposed signal selection," *International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 2017.
2. Hongqing Liu, Ruibo Zhang, Yi Zhou, Xiaorong Jing, and Trieu-Kien Truong, "Speech denoising using transform domains in the presence of impulsive and gaussian noises," *IEEE Access*, vol. 5, pp. 21193 - 21203, 2017.
3. Peng Xiong, Hongrui Wang, Ming Liu, Suiping Zhou, Zengguang Hou, and Xiuling Liu, "ECG signal enhancement based on improved denoising auto-encoder," *Engineering Applications of Artificial Intelligence*, vol. 52, pp. 194 – 202, 2016.
4. Shubhratha S, D. K. Kumuda, "Performance of empirical mode decomposition and wavelet transform in denoising of audio signal," *International Journal of Industrial Electronics and Electrical Engineering*, vol. 4, 2016.
5. Tassadaq Hussain, Sabato Marco Siniscalchi, Chi-Chun Lee, Syu-Siang Wang, Yu Tsao, and Wen-Hung Liao, "Experimental study on extreme learning machine applications for speech enhancement," *IEEE Access*, vol. 5, pp. 25542–25554, 2017.
6. Yu-Cheng Su, Yu Tsao, Jung-En Wu, and Fu-Rong Jean, "Speech enhancement using generalized maximum a posteriori spectral amplitude estimator," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7467–7471, 2013.
7. Christoph F. Stallmann, and Andries P. Engelbrecht, "Gramophone noise detection and reconstruction using time delay artificial neural networks," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 6, pp. 893–905, 2017.
8. Chengli Sun, Qin Zhang, Jian Wang, and Jianxiao Xie, "Noise reduction based on robust principal component analysis," *Journal of Computational Information Systems*, vol. 10, no. 10, pp. 4403–4410, 2014.
9. Szu-Wei Fu, Yu Tsao, Xugang Lu, and Hisashi Kawai, "Raw waveform-based speech enhancement by fully convolutional networks ," *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2017.
10. Nikolaos Dionelis, and Mike Brookes, "Modulation-Domain kalman filtering for monaural blind speech denoising and dereverberation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 4, pp. 799-814, April 2019.
11. Ying-Hui Lai, Fei Chen, Syu-Siang Wang, Xugang Lu, Yu Tsao, and Chin-Hui Lee, "A Deep denoising autoecoder approach to improving the intelligibility of vocoded speech in cochlear implant simulation," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1568-1578, July 2017.
12. Haifa Touati, and Kais Khaldi, "Speech denoising by adaptive filter LMS in the EMD framework," *International Multi-Conference on Systems, Signals & Devices (SSD)*, 2018.
13. Tomohiro Nakatani, and Keisuke Kinoshita, "A Unified convolutional beamformer for simultaneous denoising and dereverberation," *IEEE Signal Processing Letters*, vol. 26, no. 6, pp. 903-907, June 2019.
14. https://ecs.utdallas.edu/loizou/speech/noizeus/
15. M. Mallikarjunan, P. Karmali Radha, K. P. Bharath, and Rajesh Kumar Muthu, "Text-Independent speaker recognition in clean and noisy backgrounds using modified VQ-LBG algorithm," *Circuits, Systems, and Signal Processing*, vol. 38, no. 6, pp. 2810-2828, June 2019.
16. Risanuri Hidayat, Agus Bejo, Sujoko Sumaryono, and Anggun Winursito, "Denoising speech for MFCC feature extraction using wavelet transformation in speech recognition system," *International Conference on Information Technology and Electrical Engineering (ICITEE)*, 2018.

## AUTHORS PROFILE

**Nandhini A** received B.E degree in Electronics and Communication Engineering from Kingston Engineering College, Vellore, India in 2018 and currently pursuing M.Tech degree in Communication Engineering at VIT University, Vellore, India. Her areas of interest are signal processing and image processing.

**Bharath K P** received B.E degree in Electronics and Communication Engineering from Amruta Institute of Engineering and Management Sciences, Bangalore, India in 2014 and M.Tech degree in Digital Communication Engineering from Siddaganga Institute of Technology, Tumkur, India in 2016. He is currently pursuing Ph.D. in the School of Electronics Engineering, VIT University, Vellore, India; alongside he is CSIR-Senior Research Fellow, Govt India (From April-2019 to present). He is currently working on speaker identification and speaker verification with various spoofing attacks. His areas of interest are speech processing, signal processing and image processing.

**Mahalti Mohammed Sohail** currently pursuing B.Tech degree in Electronics and Communication Engineering at VIT University, Vellore, India. His areas of interest are signal processing and digital logic design.

**Rajesh Kumar M** (SM'16) received B.E degree in Electronics and Communication Engineering from Madras University, Chennai, India in 2000, M.E degree in Electrical Drives and Embedded Control from the College of Engineering, Anna University, Chennai, India in 2005, and M.B.A degree in Systems from Alagappa University, Karaikudi, India in 2008. He was the recipient of the Research Scholarships award for his Ph.D from Northumbria University, Newcastle upon Tyne, U.K. in 2010. He is a Chartered Engineer (with a membership of the IET, U.K.). Since 2000, he has been in teaching with the Department of Electronics and Communication Engineering (India) and Faculty of Engineering and Environment (U.K.). He is currently an Associate Professor with the School of Electronics Engineering, VIT University, Vellore, India. His research interests include image engineering and security, biometrics, speech signal processing, computer vision, perceptual hashing, pattern recognition and optimization technique.