International Conference on Machine Learning and Computational Intelligence-2017
International Journal of Scientific Research in Computer Science, Engineering and Information Technology
 | Volume 2 | Issue 7 | ISSN : 2456-3307

# Fractional-DCT ADALINE method for Speech Enhancement

**R. Ram, M. N. Mohanty***
Department of Electronics and Communication Engineering Siksha 'O' Anusandhan University, Bhubaneswar, Odisha, India

## ABSTRACT

Enhancement of speech is an essential task in the most of the field along with social management. The quality of the speechisdegraded mostly in the noisy environment.Also the same can be obtained from the physical disable people. To improve the quality of the speech, different enhancement algorithms can be applied. In this paper an attempt has been taken by the help of adaptive neural network based model. The fractional DCT (FrDCT) has been utilized for the input to the model. Earlier to it the discrete cosine transform (DCT) coefficients are employed to the model for the sake of verifications. That follows the coefficients of FrDCT and the results are compared. The deteriorated speech considers in this are in vehicular environment as well as summer environment with the fan.The results obtained for different noise environment that the FrDCT-ADALINE method outperforms better than the other methods.
**Keywords:** Adaptive Linear Neuron, Fractional Discrete Cosine Transform, Filtering, Segmental Signal-to-Noise Ratio, Mean Opinion Score, Speech Enhancement.

## I. INTRODUCTION

Speech enhancement has been one of the major challenges in the speech community since a decade due to its practical applications in mobile telephony, speech recognition, hearing aids etc. The performance of these systems degraded owing to the presence of background noise, babble noise, cockpit noise, impulsive noise inducing distorted information exchange. To reduce the impact of these disturbances, many algorithms have been introduced to enhance the perceptual quality of the speech signals. It is generally a difficult task to regenerate the desired signal without affecting the speech signal and the performance is restricted between speech distortion and noise reduction (Loizou, P., 2007, Haykin,S., 2009).

Boll's spectral subtraction is one of the admired algorithm for speech enhnacement. On the other hand the presence of musical noise and the half wave rectification are the foremost drawbacks of it. Further, lots of modifications have been made in this method. Non-linear spectral subtraction, Multiband spectral subtraction, spectral over subtraction, spectral subtraction consisting of perceptual properties are some of them (Boll,S.F., 1979, Upadhyay, N., Karmakar, A., 2015). Different adaptive algorithms are also designed for this problem. Wiener filter, Least mean squares, Recursive Least Squares, State Space Recursive Least Squares are some of them (Ram, R Mohanty, M.N.,2016, Vihari,S., Murthy,A.S., Soni, P., Naik,D.C.,2016).

Neural networks can be applied to cancel and remove noise from noisy speech signal (Fah,L.B., Hussain, A., Samad,S.A., 2000). The covolutional neural network is used as a convolutional denoising autoencoder. This type of architecture reflects strong correlations between speech in time and features. The speech specttra and the ideal ratio mask are estimated and the performance are compared. Different language speech signals are trained in this network for measuring the objective quality of the speech (Kounovsky, T., Malek,J.,2017).

Youshen Xia and Jun Wang (2015) have proposed a neural network based Kalman filter for speech enhancement. The recurrent neural network is based on the noise constrained least squares estimate and it provides better performance against non gaussian noise. The neuaral network is also designed for cochlear implant users. A speaker dependent algorithm and a speaker independent algorithm are compared for speech enhancement. The intelligibilty of the speech is improved with a low computational complexity (Goehring,T., Bolner,F., Monaghan,J.J.M., Dijk, B., Zarowski,A., Bleeck,S., 2017).

The choice of fractional order is an important issue in different applications (Ozaktas, H.M., Ankan,O., Kutay,M.A., Bozdaki, G., 1996). For filtering the noisein Fractional domain, the Wigner distribution should be rotated with an appropriate angle.The filtering process is performed iteratively (Kutay, M.A., Ozaktas, H.M., Arikan, O., Onural, L., 1997). The fractional specral subtraction method is proposed by Wang Zhenli and Zhang Xiongwei (2005) for speech enhancement. To filter out the noise signal, an estimated fractional noise spectrum is subtracted from the fractional noise speech spectrum.The fractional fourier transform based adaptive filter can also be designed based on fractional Fourier transform (FrFT). Using different window and fractional order the adaptive filters are implemented (Ram, R., Mohanty, M.N., 2017). Different noisy signals are tested and FrDCT filter proves better than the FrFT filter. But DCT is much better than the DFT for removing the noise components from the speech signals (Cariolaro, G., Erseghe, T., Kraniauskas, P., 2002).

DCT based filtering is suitable on the signal dependent noise. The important issue is signal dependent and multiplicative noise present in speech signal. The DCT based filtering is found competant. Also in this case, block based processing is preferred (Jeeva, M.P.A., Nagarajan, T., Vijayalakshmi, P., 2016, Ram, R., Mohanty, M.N., 2017).

The paper is organized as follows : Section 1 provides the Introduction of the work. Section 2 deals with design of the model. The model is based on ADALINE. Section 3 explains the utilization of FrDCT in ADALINE model and the model is converted to FrDCT-ADALINE model. Section 4 discusses the result for different deterioted speech signal.Finally Section 5 concludes the piece of the work.

## II. ADALINE ADAPTIVE FILTER FOR SPEECH ENHANCEMENT

ADALINE is one of the most commonly used neural networks for noise cancellation. Based on this fact, this neural network approach is widely used in the field of signal processing applications. The weights and bias of this ADALINE are adapted the LMS learning rule based on the Widrow-Hoff (Daqrouq, K., Abu-Isbeih, I.N., Alfauori, M., 2009). The weights are adjusted to minimize the error. The structure of an ADALINE is shown in Figure 1.

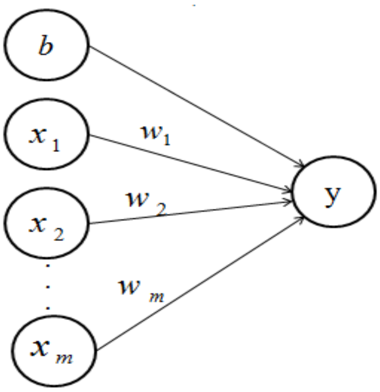


**Figure 1**. Structure of the ADALINE

The ADALINE has one output which receives the input from many neurons. To compute the output of each time sequence, the individual set of weight and bias are considered. The input layers $x_1, x_2,\dots x_m$ are interconnected to output y by weights $w_1$, $w_2,\dots w_m$ and bias *b*. Figure 2 presents the block diagram of the speech enhancement method using ADALINE. To enhance the noisy speech signal, the clean speech signal is considered as the target signal.

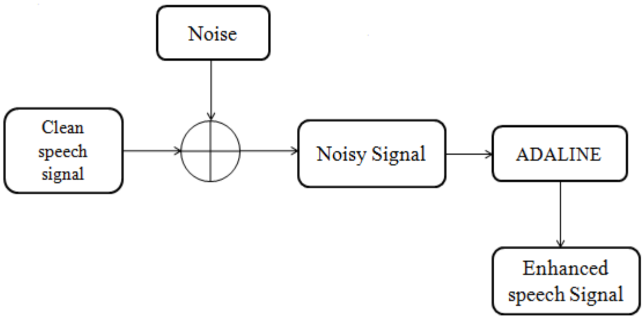


**Figure 2**. Speech enhancement using ADALINE

The following algorithm steps present the speech enhancement system.

1. The learning rate parameter (*l*) is set at 0.25. (experimentally)
2. The biases *(b(i))* and weights *(w(i))* are set at 0.95 and 0.35 respectively. (experimentally)
3. The clean signal is considered as target signal (t).
4. Set the noisy signal as the input signal *(x)*.
5. For each time index, the output *(y)* and the error *(e)* are calculated as

   $$y_i = w_i * l_i + b_i$$

   $$e_i = t_i - y_i$$

6. The weights and the biases are adjusted as

   $$w_i(new) = w_i(old) + 0.01(e_i * x_i)$$

   $$b_i(new) = b_i(old) + 0.01 * e_i$$

An additional factor of 0.01 is considered for fine tuning. All values are set experimentally for adjusting the weights and biases. Steps 5 and 6 are repeated for the first few samples of the speech signal. The enhanced signal is obtained as the error signal yielded by the adaptive network.

## III. FRACTIONAL DCT-ADALINE METHOD FOR SPEECH ENHANCEMENT

Due to the real coefficients of DCT, the spectral resolution is better for the equal size of Discrete Fourier transform. In FrDCT, both the amplitude and the phase of the noisy speech signal is enhanced and the upper bound on the maximum improvement of SNR is possible. Figure 3 presents the block diagram of the FrDCT-ADALINE enhancement method. The fractional order is selected arbitrarily to get maximum SNR.

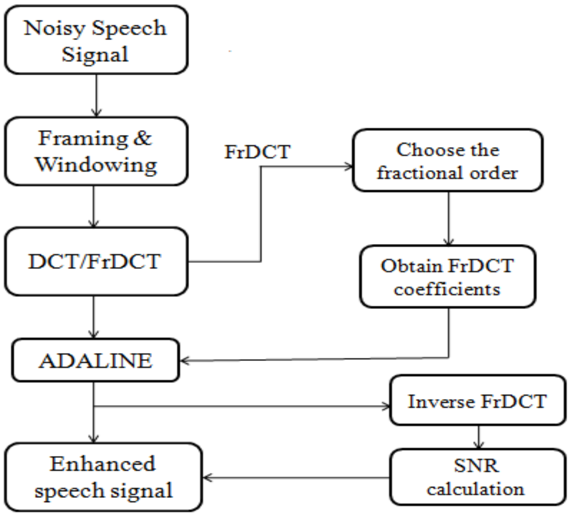


**Figure 3**. Block diagram of the Proposed Speech enhancement method

The forward DCT and inverse DCT of the discretesequence $x_m$ are defined as

$$X_k = \frac{1}{\sqrt{N}} c_k \sum_{m=0}^{N-1} x_n \cos\left(2\pi \frac{(2m+1)k}{4N}\right)$$

and

$$x_m = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} c_k X_k \cos\left(2\pi \frac{(2m+1)k}{4N}\right)$$

Where k=0, 1,……, N-1, and m=0,1, ….., N-1. N =length of the signal.

$c_0 = 1$ and $c_k = \sqrt{2}$ for *k>0*. The DCT matrix (P) of size $N \times N$ for column sequence $x_m$ and $X_k$ are expressed as

$$P = \left\| \frac{1}{\sqrt{N}} c_k \cos\left(2\pi \frac{(2m+1)k}{4N}\right) \right\|$$

In matrix form the signals are represented as

$X = Px$ and $x = P^{-1}X$ .

The FrDCT is produced from the real powers of *P*. The FRDCT can be defined as $P_\alpha : X_\alpha = P_\alpha x$ where $P_\alpha = A\Lambda^\alpha A^*$. To obtain the real values of the DCT matrix, the eigen values $\lambda_m$ of the matrix is replaced by $\lambda_m{}^a$ . Where a is the order of the fractional transform. Using the additive and orthogonality property, the inverse FrDCT is obtained as

$X_\alpha = P_\alpha s$ and $s = P_{-\alpha} X_\alpha$

Where $s$ is the enhanced signal obtained from the FrDCT which is real and orthogonal. The same steps are followed for ADALINE by considering the fractional DCT coefficients. For reconstruction of original signal, the output of the neural network will be back to the Inverse FrDCT. In nonstationary noisy cases, fractional DCT transform is better than the standard DCT. It can circumstantially care for the low frequency cotents also. Therefore the information will not be contaminated and the error estimation is less. This method is verified by using discrete FrDCT with ADALINE and the enhancement results are exhibited and compared in the result section.

## IV. RESULTS AND DISCUSSION

The objective of this proposed method is to obtain the clean signal from the noisy signal. To verify, 'have a nice day' is recorded by a female speaker in a class room at a sampling rate 8 KHz as shown in Figure 4. For enhancement, different noise is added to the speech signal such as bus noise, street noise, train noise, fan noise and babble noise. The corrupted bus noise speech signal is shown in Figure 5. The noisy signal is then applied as the input signal and the clean signal is the target signal to the ADALINE. According to the LMS rule, the linear network adapts to cancel the noise from the noisy signal. The learning rate is set at 0.25 which is determined experimentally. The bias and the weights of the network are set at 0.95 and 0.35 respectively. Figure 6 presents the result of enhancement method using ADALINE.

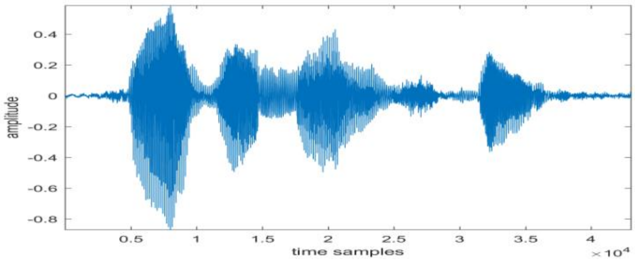


**Figure 4**. Clean speech signal 'Have a nice day'

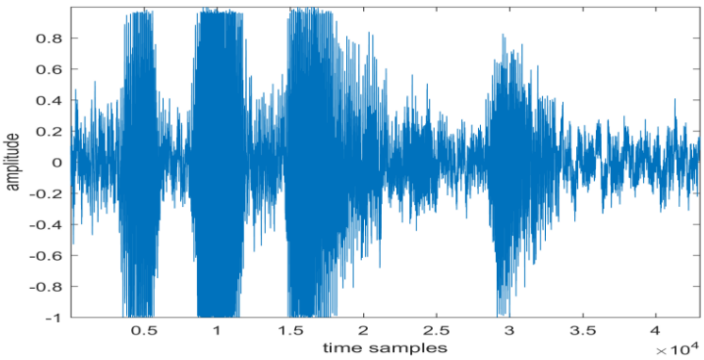


**Figure 5**. Noisy signal (bus noise)

To obtain better enhancement, the DCT and the FrDCT coefficients are acquired and employed to the ADALINE. Due to nonstationary nature of speech, the signal is first splitted into frames. A hamming window of length 512 is multiplied to each frame to avoid spectral leakage. The fractional order is selected arbitrarily and adaptively to achieve maximum SNR. The inverse transform is aimed to recover the enhanced signal. The enhanced signal of DCT-ADALINE method is shown in Figure 7 and Figure 8 presents the enhanced output of FrDCT_ADALINE method. The order 1.8 of FrDCT provides better enhanced signal.

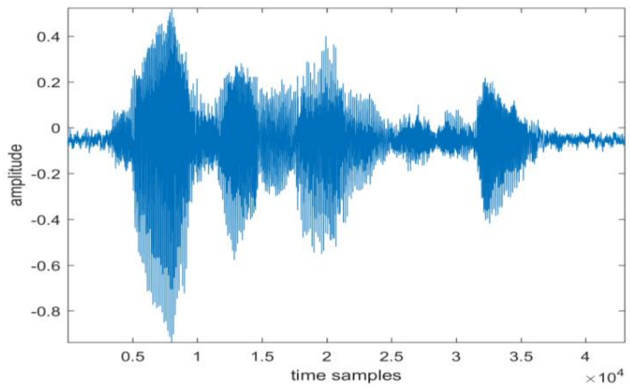


**Figure 6**. Filtered signal of ADALINE enhancement method

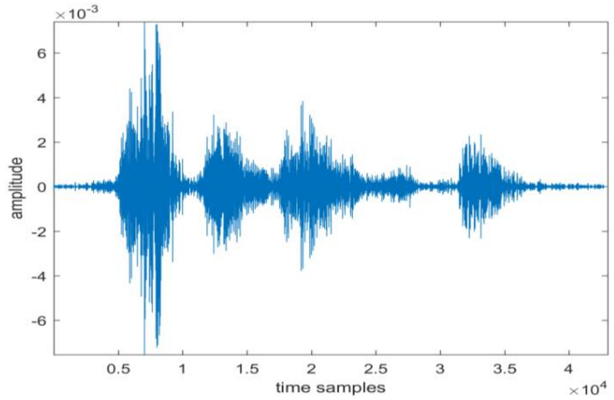


**Figure 7**. Filtered signal of DCT-ADALINE enhnacement method

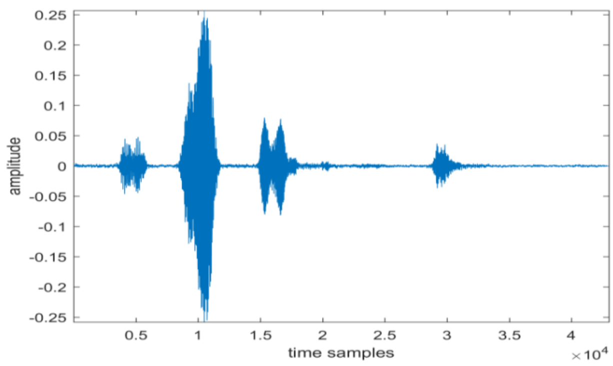


**Figure 8**. Filtered signal of FrDCT-ADALINE enhancement method

To measure the quality assessment of different signal, the Power Spectral Density (PSD) and spectrogram are evaluated. The PSD of clean signal, noisy signal and the ADALINE enhancement method is shown in Figure 9, Figure 10 and Figure 11 respectively. It is noticed that the PSD of ADALINE filtered signal is somehow similar to the target signal. For further improvement another methods have tested. Figure 14 and Figure 15 present the spectrogram of the clean signal and noisy signal. The spectrogram also indicates that the noise is still remains in the ADALINE enhancement method as in Figure 16.

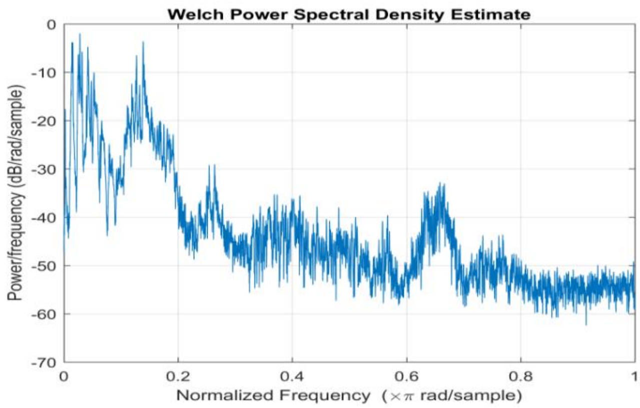


**Figure 9**. PSD of clean signal

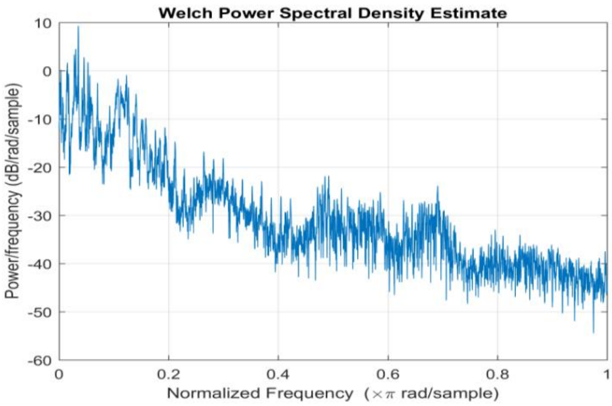


**Figure 10**. PSD of noisy signal

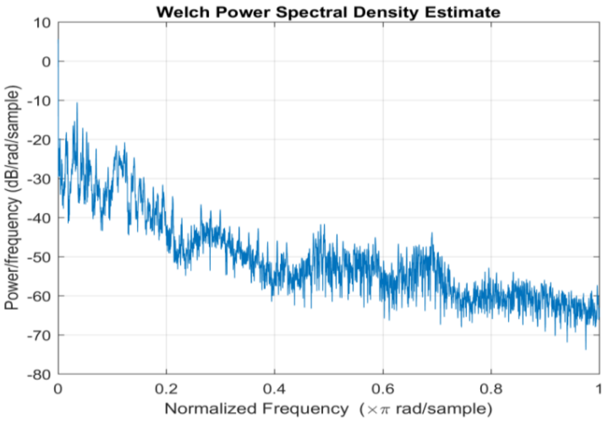


**Figure 11**. PSD of ADALINE enhancement method

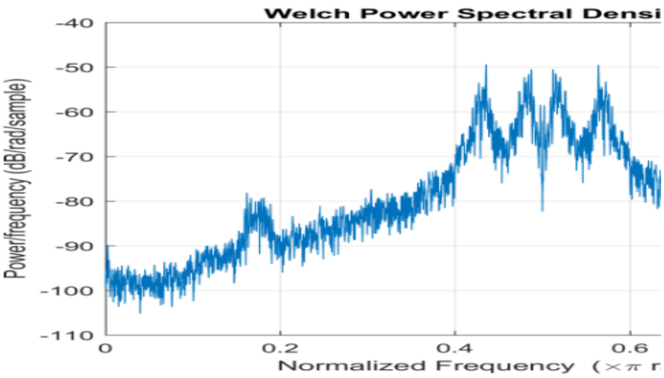


**Figure 12.** PSD of DCT-ADALINE enhancement method

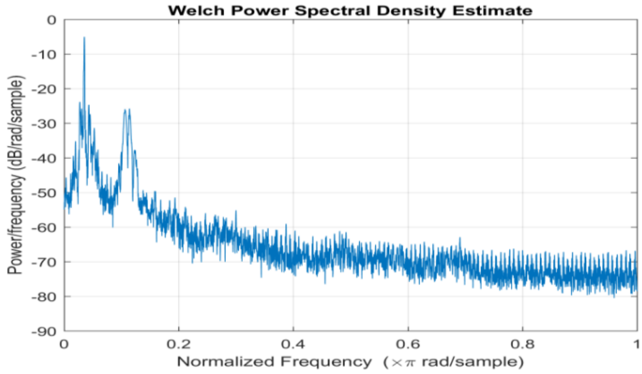


**Figure 13.** PSD of FrDCT-ADALINE enhancement method

The PSD of the FrDCT-ADALINE method (Figure 13) has better result than the DCT method (Figure 12). The spectrogram of the FrDCT (Figure 18) indicates the better result than the other methods. Figure 17 represents the spectrogram of the DCT-ADALINE method consisting of noise.

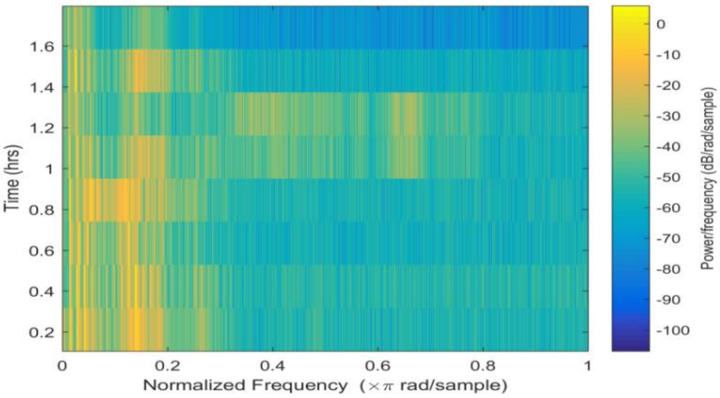


**Figure 14.**Spectrogram of clean speech signal

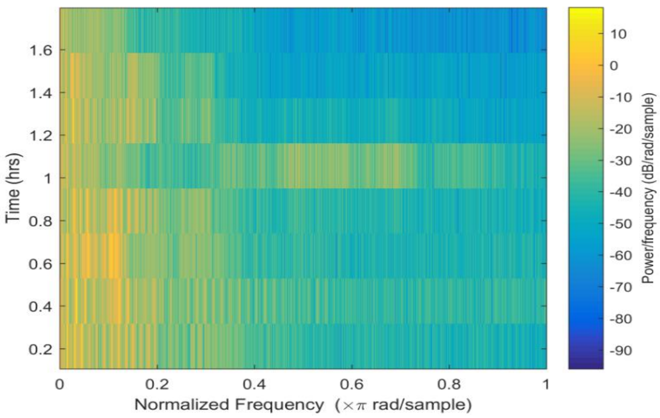


**Figure 15.** Spectrogram of Noisy speech signal

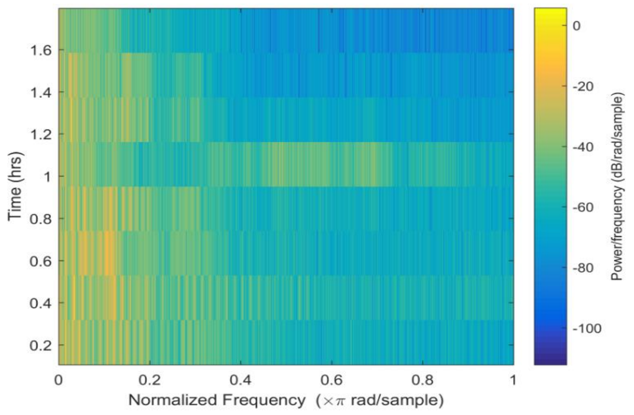


**Figure 16.** Spectrogram of ADALINE enhancement method

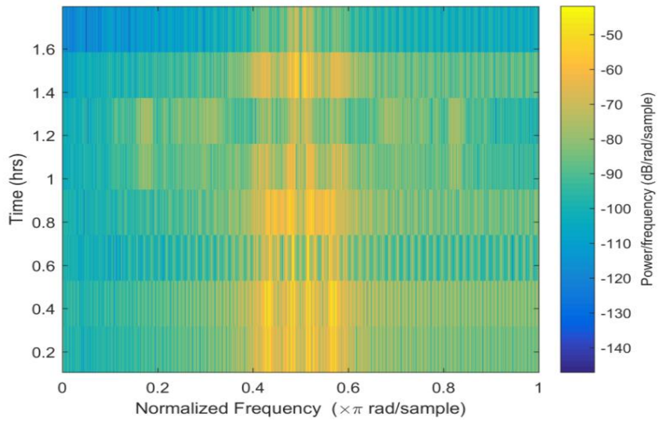


**Figure 17.**Spectrogram of DCT-ADALINE enhancement method

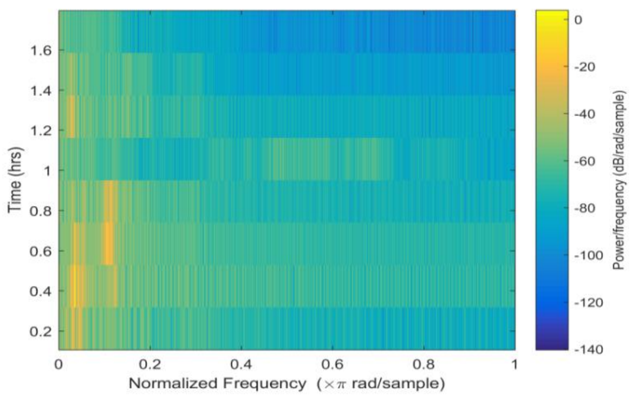


**Figure 18.** Spectrogram of FrDCT-ADALINE enhancement method

Different objective and subjective measures are there to test the speech quality and intelligibilty (Hu, Y., Loizou, P., 2008). In this experiment, the signal-to-noise ratio (SNR) and the Mean-Opinion-Score (MOS) are considered for evaluation of the enhancement methods. Table 1 shows the SNR of different types of noisy signals and Table 2 shows the MOS of noisy signals. The maximum SNR improvement is 6.1042 dB achieved in FrDCT-ADALINE for train noise. But the highest MOS is 4.55 obtained for babble noise.

**Table 1.** SNR Improvement for different types of noise signal

| | SNR before Enhancement (dB) | SNR after Enhancement (dB) | SNR Improvement (dB) |
|---|---|---|---|
| ADALINE | | | |
| Bus Noise | 4.5346 | 6.3652 | 1.8306 |
| Street Noise | 3.4565 | 5.3492 | 1.8927 |
| Train Noise | 6.8760 | 8.5032 | 1.6272 |
| Fan Noise | 3.7659 | 4.8116 | 1.0457 |
| Babble Noise | 2.8762 | 4.6234 | 1.7472 |
| DCT_ADALINE | 4.5346 | 7.6547 | 3.1201 |
| Bus Noise | 3.4565 | 7.8734 | 4.4169 |
| Street Noise | 6.8760 | 9.2560 | 2.3800 |
| Train Noise | 3.7659 | 7.2454 | 3.4795 |
| Fan Noise | 2.8762 | 6.4582 | 3.5820 |
| Babble Noise | | | |
| FrDCT_ADALINE | 4.5346 | 10.0874 | 5.5528 |
| Bus Noise | 3.4565 | 8.9565 | 5.5000 |
| Street Noise | 6.8760 | 12.9802 | 6.1042 |
| Train Noise | 3.7659 | 9.5434 | 5.7775 |
| Fan Noise | 2.8762 | 8.5653 | 5.6891 |
| Babble Noise | | | |

**Table 2.** Mean Opinion Score for different types of noise signal

| | ADALINE | DCT_ADALINE | FrDCT_ADALINE |
|---|---|---|---|
| bus noise | 2.63 | 3.34 | 4.35 |
| street noise | 2.83 | 3.67 | 4.33 |
| train noise | 2.90 | 3.52 | 4.08 |
| fan noise | 2.87 | 3.89 | 4.53 |
| babble noise | 2.44 | 3.55 | 4.55 |

## V. CONCLUSION

The DCT-ADALINE and the FrDCT-ADALINE speech enhancement algorithms are proposed in this work and the results are compared with the ADALINE. Results show that the proposed method is better than the other methods.Due to higher energy compaction and better spectral resolution property, DCT and the FrDCT provide better enhanced signal. The speech signal corrupted by different noise are tested to show the results. The SNR improvement is more perceptible in case of train noise i.e. 6.1042 dB. And the maximum MOS is 4.55 is obtained from the babble noise affected speech signal. Furthermore the listening test is done for all the enhanced signals. FrDCT-ADALINE enhancement method proved to be better for all the tests compared to other algorithms. Better enhanced signals can be achieved in other transforms as well, and thiswillbescoped in future work.

## VI. REFERENCES


[1]. Loizou, P., 2007, Speech Enhancement: Theory and Practice. CRC Press.

[2]. Haykin, S. S., & Haykin, S. S. , 2009, Neural networks and learning machines. New York: Prentice Hall/Pearson.

[3]. Boll,S.F., 1979, Suppression of Acoustic Noise in Speech using Spectral Subtraction.IEEE Transaction ASSP. 113-120.

[4]. Upadhyay, N., Karmakar, A., 2015, Speech Enhancement using Spectral Subtraction-type Algorithms: A Comparison and Simulation Study. Elsevier Procedia Computer Science. 55, 574-584.

[5]. Ram, R Mohanty, M.N.,2016, Performance Analysis of Adaptive Algorithms for Speech Enhancement Applications. Indian Journal of Science and Technology. 9(44).

[6]. Vihari,S., Murthy,A.S., Soni, P., Naik,D.C.,2016, Comparison of Speech Enhancement Algorithms. Procedia Computer Science. 89, 666 – 676.

[7]. Fah,L.B., Hussain, A., Samad,S.A., 2000, Speech Enhancement by Noise Cancellation Using Neural Network. IEEE Conf..

[8]. Kounovsky, T., Malek,J.,2017, Single Channel Speech Enhancement Using Convolutional Neural Network. IEEE International Workshop of Electronics, Control, Measurement, Signals and their Application to Mechatronics. 1-5.

[9]. Daqrouq, K., Abu-Isbeih, I.N., Alfauori, M., 2009, Speech Signal Enhancement Using Neural Network and Wavelet Transform. International Multi-Conference on Systems, Signals and Devices.

[10]. Xia,Y., Wang,J., 2015, Low-Dimensional Recurrent Neural Network-based Kalman Filter for Speech Enhancement.Neural Networks 67, 131–139.

[11]. Goehring,T., Bolner,F., Monaghan,J.J.M., Dijk, B., Zarowski,A., Bleeck,S., 2017, Speech enhancement based on neural networks improves speech intelligibility in noise for

cochlear implant users.Hearing Research 344, 183-194.
[12]. Ozaktas, H.M., Ankan,O., Kutay,M.A., Bozdaki, G., 1996, Digital Computation of the Fractional Fourier Transform. IEEE Transactions on Signal Processing, 44(9).
[13]. Kutay, M.A., Ozaktas, H.M., Arikan, O., Onural, L., 1997, Optimal Filtering in Fractional Fourier Domains. IEEE Transactions on Signal Processing, 45(5).
[14]. Zhenli, W., Xiongwei, Z., 2005, On the application of fractional Fourier transform for enhancing noisy speech. IEEE International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications Proceedings.
[15]. Cariolaro, G., Erseghe, T., Kraniauskas, P., 2002, The Fractional Discrete Cosine Transform. IEEE Transactions on Signal Processing. 50(4).
[16]. Ram, R., Mohanty, M.N., 2017, Design of Fractional Fourier Transform based Filter for Speech Enhancement. IJCTA. 10(07), 235-243.
[17]. Jeeva, M.P.A., Nagarajan, T., Vijayalakshmi, P., 2016, Discrete Cosine Transform-Derived Spectrum based Speech Enhancement Algorithm using Temporal-Domain Multiband Filtering.IET Journal of Signal Processing.
[18]. Ram, R., Mohanty, M.N., 2017, Design of Filter using Fractional-DCTfor Speech Enhancement. Int. Conf on Sustain.able Computing Techniques in Engineering, Science and Management.
[19]. Hu, Y., Loizou, P., 2008, Evaluation of objective quality measures for speech enhancement. IEEE Trans. Audio Speech Lang. Process., 16, 229–238.