International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

# Transform Based Speech Enhancement for Auditory Signals Based on Iterative Weiner Filtering

**V.R.Balaji, M.Karpagam, J.Sathiya Priya**

***Abstract:** One of the important features of Speech processing is speech enhancement. In a noisy environment, speech enhancement plays a vital role. Many research works are being done in speech enhancement methods in recent years but still, it can't be attained. It mainly depends on Speech intelligibility which can improve the speech quality. In this research work, signal representation is considered and the various transforms are applied and compared. The analysis is done with the help of two parameters and the results are compared. Here the enhancement process is focused on using Advanced DCT (ADCT) and Discrete fractional Cosine transform. The ADCT has the advantage of energy compaction and flexible window switching. Iterative Wiener Filtering is used for filtering the coefficients. Pitch Synchronous Analysis (PSA) is combined for finding the exact pitch period.*

***Index terms:** Advanced Discrete Cosine Transform, Discrete fractional Cosine transform, Iterative Wiener Filtering, Pitch Synchronous Analysis*

## I. INTRODUCTION

The intelligibility of the speech signal is improved by speech enhancement techniques. Voice signals are normally altered by environmental factors. So Speech enhancement techniques are needed for all types of multimedia applications. It helps in improving the hearing ability of voice signals. Different methods were incorporated to make the audibility of the speech signal to be clear. So speech enhancement is one of the active research areas.

Various transforms can be used for speech enhancement. First, the filters are selected based on the speech quality and then transform coefficients are determined so that the speech enhancement process can be done. In the final stage, inverse coefficients are obtained to get the required speech signal. The algorithms will not be able to contain the speech energy, so it is hard to filter the noise coefficients. The various analysis is to be done in the transforms for analyzing the speech signal.

DCT based algorithms provide effective ways to improve speech intelligibility. The Fourier transform along with the spectral subtraction algorithm was widely used to improve the quality of speech. The speech enhancement process involves the Discrete Cosine Transform since it depends on a Fourier-related transform. The advantage of using only real numbers in DCT is better than DFT. DFT relies on complex numbers. Compared to DFT, DCT provide a higher level of advantage in speech enhancement.

**Revised Manuscript Received on 30 July 2019.**

* Correspondence Author

**V.R.Balaji***, Associate Professor, Department of Electronics and Communication Engineering, Sri Krishna College of Engineering and Technology, Coimbatore, TN, India.

**M.Karpagam**, Professor, Department of Electronics and Communication Engineering, Sri Krishna College of Engineering and Technology, Coimbatore, TN, India

**J.Sathiya Priya, J.Sathiya Priya** was earlier working as an Assistant Professor in the Department of Electronics and Communication Engineering. Coimbatore, TN, India



## II. LITERATURE SURVEY

In speech signal processing, the pitch of the signal plays a vital role. The speech signal is split into two sounds namely voiced and unvoiced. Based on the pitch period, the speech information will be manipulated.

The speech signal is re-sampled using segments of speech and then truncation is done based on zero padding . DCT provides energy compaction rather than DFT [1]. The energy comparison is minimum in this method when compared with other methods. So DCT provides better results rather than basic algorithms. Pitch synchronization is used to reduce the discontinuities by employing a windowing technique and with the help of pitch period.

## III. SPEECH ENHANCEMENT PROCESS BASED ON ADCT

The proposed system involves various blocks as shown in Figure 1. Initially, the Noisy speech signal is taken as input for processing and transformation was applied on the input signal. Iterative Wiener filtering is used for the filtering process [3]. The speech frames are split into voiced/unvoiced sounds which can be represented as a signal. Time shifting of the speech signal is done based on the voiced/unvoiced decision. A suitable windowing technique is to be applied for the signal. ADCT uses sampling done by overlapping at 50% interval. The coefficients will remain the same before and after taking transformations [7]. The signals are considered as a frame and it is overlapped on which ADCT will perform its function.

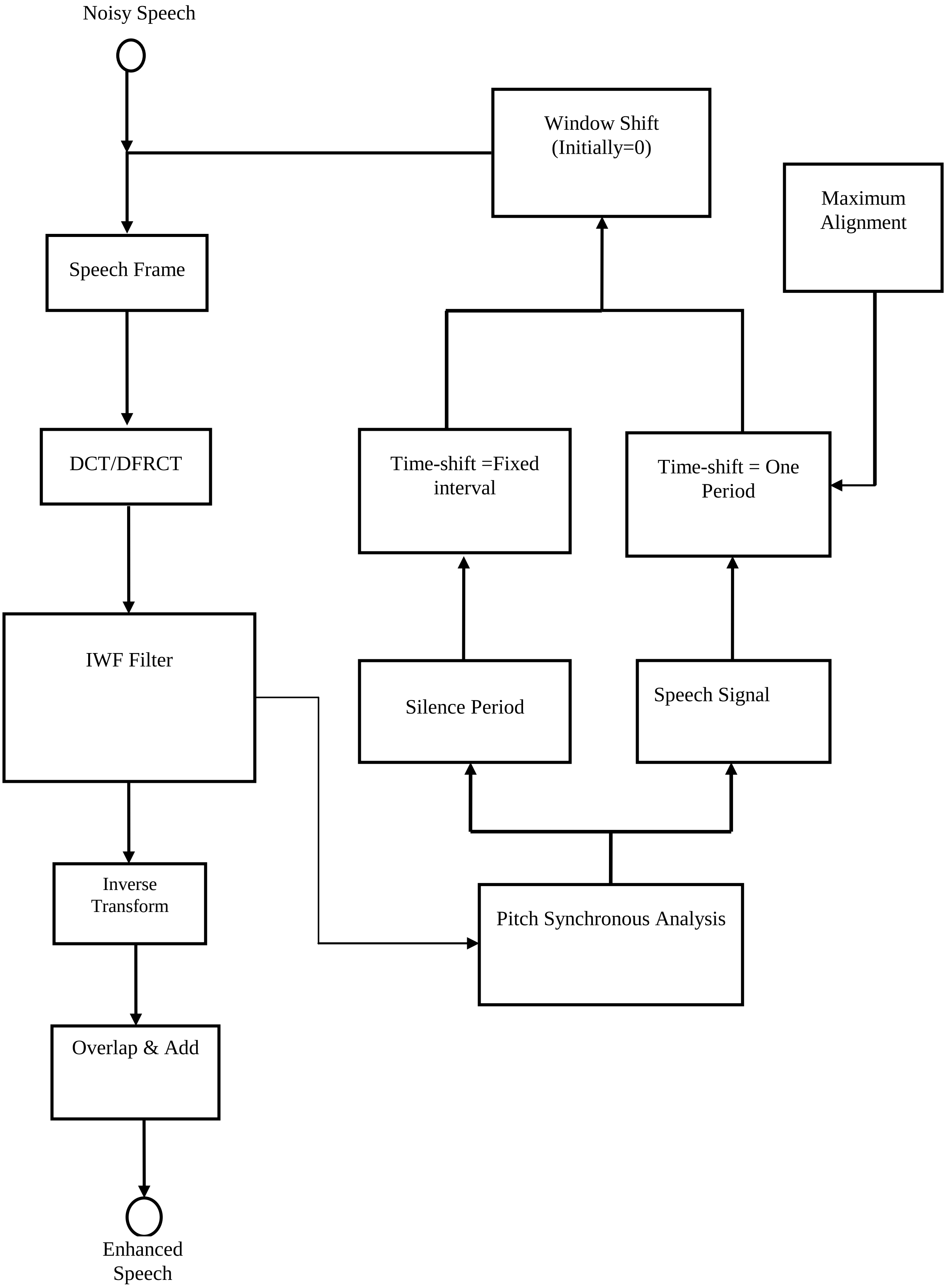


Figure 1: Architecture

The ADCT is given by:

$$a_r = \sum_{k=0}^{2N-1} \tilde{a}_k \cos\left[\pi \frac{(k+(N+1)/2)(r+1/2)}{N}\right] \quad where\ r = 0 \ldots\ldots N-1 \tag{1}$$

$$\tilde{a}_k = \frac{2}{N}\sum_{r=0}^{N-1} a_r \cos\left[\pi \frac{(k+(N+1)/2)(r+1/2)}{N}\right] \tag{2}$$

$$where\ \ k = 0 \ldots..2N-1$$

where $\tilde{a}_k = h_k a_k$ is a correlated signal $a_k$ , is the input signal, $h_k$ is the window function[8]. Limitations are to be considered for perfect reconstruction):

$$h_k = h_{2N-1-k} + {h_k}^2 + {h_{k+N}}^2 = 1 \tag{3}$$

The window will be selected based on attenuation and the ability to reconstruct the signal [9]. Here sine window is used which is given as:

$$h_k = \sin[\pi(k+1/2)/2N] \tag{4}$$

where k = 0 ,1, …, 2N – 1

## IV. SPEECH ENHANCEMENT PROCESS BASED ON DFRCT

The discrete fractional cosine transform is represented by

$$v(k) = \propto(k)\sum_{n=0}^{N-1} x(n)\cos\left(\frac{\pi(2n+1)}{2N}\right), 0 \le k \le N-1 \tag{5}$$

where $x(n)$ represents the input signal and *v(k)* represents the output signal

$$\propto(0) = \sqrt{\frac{1}{N}}$$

$$\propto(k) = \sqrt{\frac{2}{N}}\ for\ 1 \le k \le N-1$$

If the signal is double, the Fourier transformation is given by:

$$y[n] = x[n] + x[2N-n-1] \tag{6}$$

Let $$C_x(k) = e^{-\frac{i2\pi k}{4N}} Y(k) \tag{7}$$

where $Y(k)$ is DFT Transform. The inverse Discrete Fourier Transform is defined as:

$$C_x(k) = \cos\left(\frac{\pi k}{2N}\right)\left[F_{2N}\{x(n)\}[k] + F_{2N}^{-1}\{x(n)\}[k]\right] \tag{8}$$

## IV. ITERATIVE WEINER FILTER

Based on a priori SNR, the minimum Mean Square Error has to be calculated. It is calculated between the two signals. One signal is the desired signal and the other is the estimated signal. The a priori SNR is calculated with the help of a decision-directed approach.
The noisy signal which is the combination of both noise signal and the speech signal is denoted as $x = s + d$. This method is used to estimate the signal. In non-speech regions, the noise power spectral density (PSD) estimate is revised when the noisy signal in the background gets varied. But here the limitation is to classify the speech signal into speech or non-speech signals. Also, the noise estimate is needed to be updated based on the non-speech duration [10].

The noise signal kept on getting changed in the non-speech region. So it may lead to an inaccurate calculation in the filter. To avoid this dynamic technique has to be realized which must be adaptive in nature. By means of iteration, the spectrum of the signal is calculated. Wiener filter can calculate the optimum signal for each frame. This can be done by means of signal subtraction [11]. The spectrum can be calculated by estimating the signal. The noise signal will vary less than the speech signal. The noise estimate is calculated by taking the average of all the frames.

## V. EXPERIMENTAL RESULTS

The results were taken with the help of various speech signals taken from TIMIT database. For experimentation, 8Khz frequency signal is used. The noise may be Pink, Babble and Factory noise. The speech segments had a duration of 313.998s including the silence period. The Voiced speech is approximately 50% of the speech segments.
This comparison compares ADCT based IWFPS technique and DFrCT using the IWFPS technique using two objective measures. They are evaluated using parameters like Segmental SNR and Perceptual Evaluation of Speech Quality. The DFrCT using IWFPS technique provides higher accuracy when compared to ADCT using IWFPS technique.

**Table 1: Δ SegSNR Results Comparison**

| Noise Type | SNR (dB) | Δ SEGSNR | |
|---|---|---|---|
| | | ADCT using IWFPS | DFrCT using IWFPS |
| Pink | 0 | 6.85 | 6.25 |
| | 5 | 6.36 | 6.06 |
| | 10 | 5.85 | 5.55 |
| | 15 | 4.73 | 4.03 |
| Babble | 0 | 8.75 | 8.25 |
| | 5 | 8.35 | 8.05 |
| | 10 | 7.11 | 7.00 |
| | 15 | 6.95 | 6.15 |
| Factory 1 | 0 | 11.85 | 11.05 |
| | 5 | 11.67 | 10.66 |
| | 10 | 10.82 | 10.12 |
| | 15 | 9.84 | 9.24 |

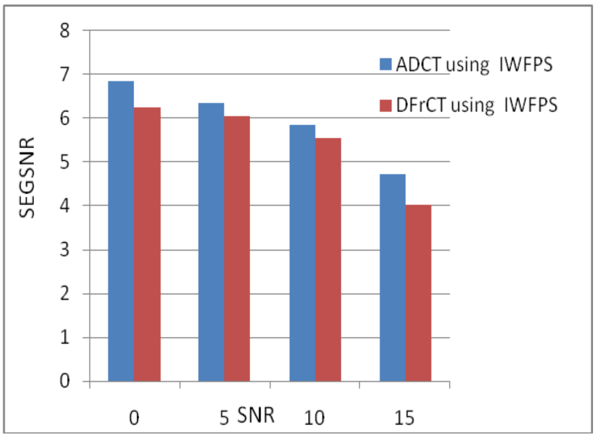

**Figure 2: Comparison of ΔSEGSNR results for Pink Noise**

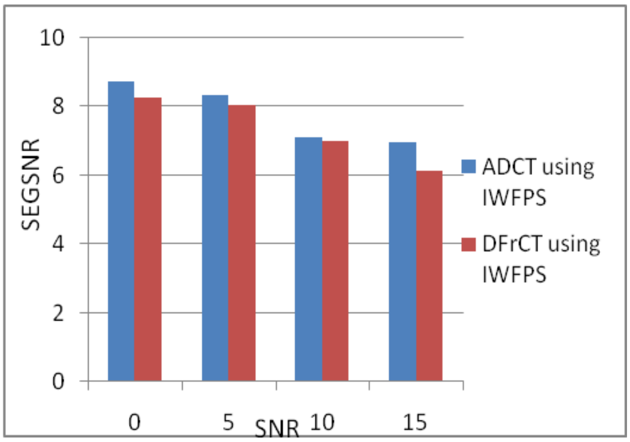

**Figure 2: Comparison of ΔSEGSNR results for Babble Noise**

**Table 2: Δ PESQ Results Comparison**

| Noise Type | SNR (dB) | Δ PESQ ($\times 10^{-1}$) | |
|---|---|---|---|
| | | ADCT using IWFPS | DFrCT using IWFPS |
| Pink | 0 | 7.82 | 6.16 |
| | 5 | 7.03 | 6.48 |
| | 10 | 6.62 | 6.11 |
| | 15 | 6.14 | 5.92 |
| Babble | 0 | 8.05 | 7.54 |
| | 5 | 7.27 | 6.44 |
| | 10 | 6.24 | 5.42 |
| | 15 | 5.03 | 4.55 |
| Factory 1 | 0 | 5.66 | 5.22 |
| | 5 | 4.45 | 4.06 |
| | 10 | 2.99 | 2.09 |
| | 15 | 1.97 | 1.65 |

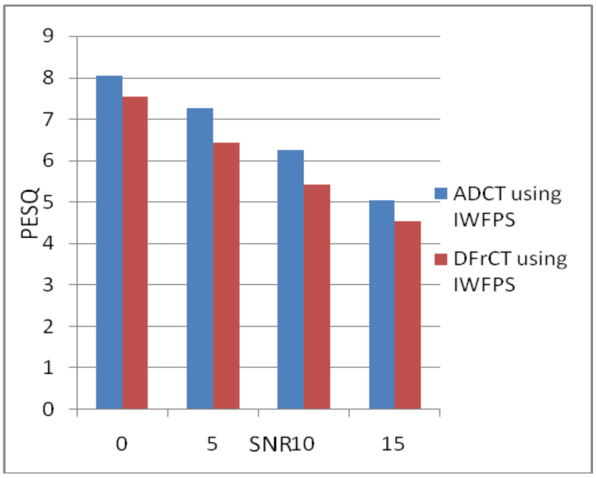

**Figure 3: Comparison of ΔPESQ results for Pink Noise**

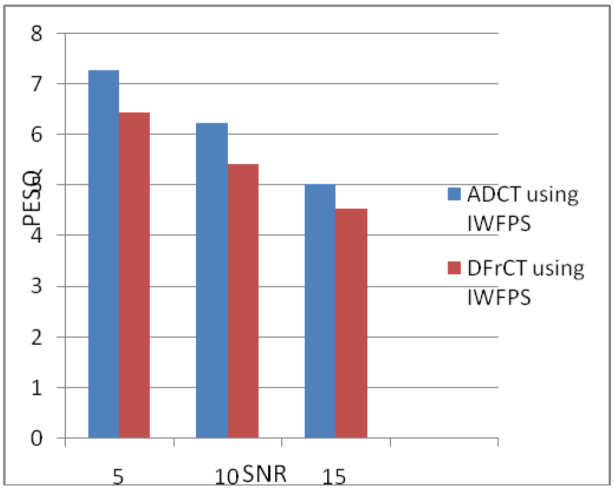

**Figure 3: Comparison of ΔPESQ results for Babble Noise**

## VI. CONCLUSION

In this research work, speech enhancement methods were analyzed and compared for different types of noise. The speech signal is divided into frames and two different transforms are applied to analyze the output. The comparison shows that DFRCT provides enhanced speech when compares with ADCT based IWFPS method. For better performance, Iterative Wiener filtering is used for filtering. Based on the pitch period, time shift is done based on autocorrelation function. The objective measures segmental SNR and PESQ are compared for various noise types using this technique.

## REFERENCES:

1. Y. Zhao, Z.-Q. Wang, and D. L. Wang, “A Two-Stage Algorithm for Noisy and Reverberant Speech Enhancement,” in IEEE International Conference on Acoustics, Speech and Signal Processing, 2017, pp. 5580–5584.

2. X. Zhang, Z.-Q. Wang, and D. L. Wang, “A Speech Enhancement Algorithm by Iterating Single- and Multi-Microphone Processing and its Application to Robust ASR,” in IEEE International Conference on Acoustics, Speech and Signal Processing, 2017, pp. 276–280.
3. Z. Zhang, J. Geiger, J. Pohjalainen, A. E.-D. Mousa, W. Jin, and B. Schuller, “Deep Learning for Environmentally Robust Speech Recognition: An Overview of Recent Developments,” arXiv preprint arXiv:1705.10874, May 2017.
4. Z.-Q. Wang, X. Zhang, and D. Wang, "Robust TDOA Estimation Based on Time-Frequency Masking and Deep Neural Networks,” in Proceedings of Interspeech, 2018.
5. J. Du and et al., “The ustc-iflytek system for chime-4 challenge,,” in CHiME-4 workshop, 2016.
6. A. Veit, M. J.Wilber, and S. Belongie, "Residual networks behave like ensembles of relatively shallow networks,” in Advances in Neural Information Processing Systems, 2016, pp. 550–558.
7. .L.-S. D. A. via Teacher-Student Learn, "Large-scale domain adaptation via teacher-student learn,” in Interspeech 2017. ISCA,2017.
8. alaji V R, Sathiya Priya.J 2018 “Enhancement of Speech Signal Using Modified Binary Mask Based Algorithm for Vehicular Noise”, Journal of Advanced Research in Dynamical & Control Systems, Vol. 10, 12-Special Issue, 2018
9. A. Nugraha, A. Liutkus, and E. Vincent, "Multichannel audio source separation with deep neural networks,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 24, no. 9, pp. 1652–1664, September 2016.
10. S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov, “A Consolidated Perspective on Multi-Microphone Speech Enhancement and Source Separation,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, pp. 692– 730, 2017.
11. Balaji V R 2017 “A Discrete Fractional Cosine Transform Based Speech Enhancement System using DUET algorithm through Adaptive Kalman Filter Combined With Perceptual Weighting Filter”, Perspectivas Em Ciencia Da Informacao, ISSN No:1413-9936
12. Balaji V R 2018, “A Comparison of Compression Sensing Algorithm and DUET algorithm for Advanced DCT based Speech Enhancement System for Vehicular noise", International Journal of Pure and Applied Mathematics, Volume 119 No. 12 2018, 1385- 1394
13. Bhushan C. Kamble Speech Recognition Using Artificial Neural Network –A Review , Int'l Journal of Computing, Communications & Instrumentation Engg. (IJCCIE) Vol. 3, Issue 1 (2016) ISSN 2349-1469 EISSN 2349-1477
14. Lekshmi.K.R#1, Dr.Elizabeth Sherly, Automatic Speech Recognition using different Neural Network Architectures – A Survey,International Journal of Computer Science and Information Technologies, Vol. 7 (6) , 2016, 2422-2427
15. Khaled M. AlhawitiAdvances in Artificial Intelligence UsingSpeech Recognition, World Academy of Science, Engineering and Technology,International Journal of Computer and Information Engineering,Vol:9, No:6, 2015

## AUTHORS PROFILE

**Dr.V.R.Balaji** has completed his Ph.D from Anna University, Chennai in September 2015. He is currently working as an Associate Professor in the Department of Electronics and Communication Engineering. His area of research includes Speech signal processing, VLSI and Image Processing. He has published various research papers in reputed journals. He is a member of various professional bodies.

**Dr.M.Karpagam** has completed her Ph.D from Anna University, Chennai in April 2015. She is currently working as Professor in the department of Electronics and Communication Engineering, Sri Krishna College of Engineering and Technology, Coimbatore, India. Her areas of interest include Wireless Adhoc Sensor Networks and Image Processing.

**J.Sathiya Priya** was earlier working as an Assistant Professor in the Department of Electronics and Communication Engineering. Her area of research includes VLSI and image processing. She has published various research papers in reputed journals.