*I.J. Image, Graphics and Signal Processing,* **2024, 5, 14-24**
Published Online on October 8, 2024 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2024.05.02

# Speech Enhancement Using Joint Time and DCT Processing for Real Time Applications

**Ravi Kumar Kandagatla***
Department of ECE, Lakireddy Bali Reddy College of Engineering, Mylavaram, Andhra Pradesh, India
E-mail: 2k6ravi@gmail.com
ORCID iD: https://orcid.org/0000-0002-9860-3912
*Corresponding Author

**V. Jayachandra Naidu**
Department of ECE, Sri venkateswara College of Engineering & Technology, Chittoor, India
Email: jayachandra.velur@gmail.com
ORCID iD: https://orcid.org/0009-0006-4965-7793

**P. S. Sreenivasa Reddy**
Department of ECE, Nalla Narasimha Reddy Education Society's Group of Institutions, Telangana, India
Email: sreenivaspanati@gmail.com
ORCID iD: https://orcid.org/0009-0006-6607-2041

**Sivaprasad Nandyala**
Technology Innovation Institute (TII), Abu Dhabi, U.A.E.
Email: sivaprasad.nandyala@tii.ae
ORCID iD: https://orcid.org/0000-0002-0806-8064

Received: 20 January, 2024; Revised: 14 February, 2024; Accepted: 10 March, 2024; Published: 08 October, 2024

**Abstract:** Deep learning based speech enhancement approaches provides better perceptual quality and better intelligibility. But most of the speech enhancement methods available in literature estimates enhanced speech using processed amplitude, energy, MFCC spectrum, etc along with noisy phase. Because of difficult in estimating clean speech phase from noisy speech the noisy phase is still using in reconstruction of enhanced speech. Some methods are developed for estimating clean speech phase and it is observed that it is complex for estimation. To avoid difficulty and for better performance rather than using Discrete Fourier Transform (DFT) the Discrete Cosine Transform (DCT) and Discrete Sine Transform (DST) based convolution neural networks are proposed for better intelligibility and improved performance. However, the algorithms work either features of time domain or features of frequency domain. To have advantage of both time domain and frequency domain here the fusion of DCT and time domain approach is proposed. In this work DCT Dense Convolutional Recurrent Network (DCTDCRN), DST Convolutional Gated Recurrent Neural Network (DSTCGRU), DST Convolution Long Short term Memory (DSTCLSTM) and DST Convolutional Gated Recurrent Neural Network (DSTDCRN) are proposed for speech enhancement. These methods are providing superior performance and less processing difficulty when compared to the state of art methods. The proposed DCT based methods are used further in developing joint time and magnitude based speech enhancement method. Simulation results show superior performance than baseline methods for joint time and frequency based processing. Also results are analyzed using objective performance measures like Signal to Noise Ratio (SNR), Perceptual Evaluation of Speech Quality (PESQ) and Short-Time Objective Intelligibility (STOI).

**Index Terms:** Speech Enhancement, Perceptual Evaluation of Speech Quality, Noise Reduction, Discrete transform, Recurrent Neural Network, Gated Unit, Signal to Noise Ratio.

## 1. Introduction

The goal of speech enhancement is to estimate clean speech from the corrupted speech and hence produces quality speech for listener comfort. However, processing of noisy speech introduces some distortion and to be taken care while

reducing the noise. A tradeoff between noise reduction and quality is always to be considered while using traditional speech enhancement methods like spectral subtraction, wiener filter and so on. All the traditional techniques uses Short Time Fourier Transform (STFT) for processing noisy speech [1,2]. It is noted that [3,4] the STFT based approaches, mask based approaches [5,6] and complex mask [7,8] uses noisy speech spectral amplitudes only and noisy phase remain untouched while processing [9]. At the reconstruction these approaches utilizes noisy speech phase for reconstruction of enhanced speech [3]. Due to complexity in processing noisy speech phase it is neglected in almost many algorithms. In this work DCT [9] and DST based speech enhancement approaches are proposed which considered phase while processing and for better performance.

Recent speech enhancement approaches uses deep learning models for noise reduction and quality improvement. Also mask based techniques like ideal binary mask [5], complex ratio masks [8] are developed for enhanced signal generation [9]. In this work deep neural network model using ideal mask and skip connections is proposed. This work utilizes Ideal Cosine Mask (ICM) and Ideal Sine Mask (ISM) as given in (7) and (9).

The enhanced speech resynthesis is based on phase is developing in time domain using the masks which are phase sensitive. Training strategies using convolution network is discussed using sliding window [10]. But the deep complex networks are introduced for better performance [11]. Complex valued spectrogram based neural network approaches provides benefits of cosine transform and Unet. But the complexities of phase based designs are still poses problems and this paper addresses the inbuilt phase processing transform using DCT. Also addresses the DCT and DST based approaches for speech enhancement.

## 2. State of Art and Insights of Proposed Method

The diagram of the proposed method is shown in Fig. 1. Here dense blocks are used in convolution network. Dense blocks provides better feature map by providing relevant information through different layers of network. Hence the correlation of frames can learn easily. It contains DCTNet, WNet, and a loss based fusion. WNet accepts the framed time domain input and DCTNet uses inbuilt phase and magnitude processing. To improve the quality of enhanced speech DCTNet and Wnet outputs are further processed using loss based fusion.

Input noisy speech is splitted into frames and applied to WNet and DCTNet. Fusion block accepts the output of mentioned networks and noisy speech for further processing [10].

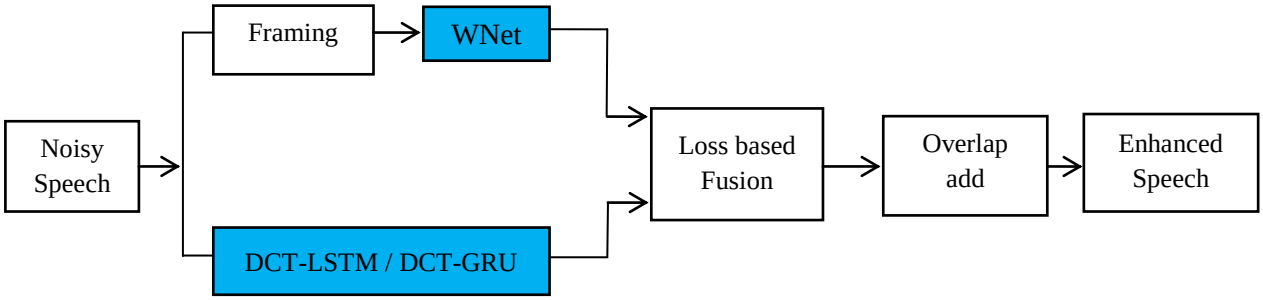



Fig. 1. Simplified Diagram of proposed Method

Gated neural units are used in gradient problem for RNN at the beginning and Long Short-Term Memory (LSTM) is one such variant. It uses the convolution operation [10] which applies convolution filters on input and output feature maps. Wnet uses GTCN as basic building block

DCT based model uses an intermediate processor in which the deep network is utilized. The backend and front end of processor utilizes an decoder and encoder respectively. The flow of encoder-processor-decoder architecture is shown in Fig. 2. Importance of encoder and decoder in proposed method is that the high-level features can easily extracted from input speech and can easily reconstrut the processed signal from lowlevel features at decoder stage. In this work the output of encoder is concatenate with input of decoder layer using skip connection for improved performance in the network.

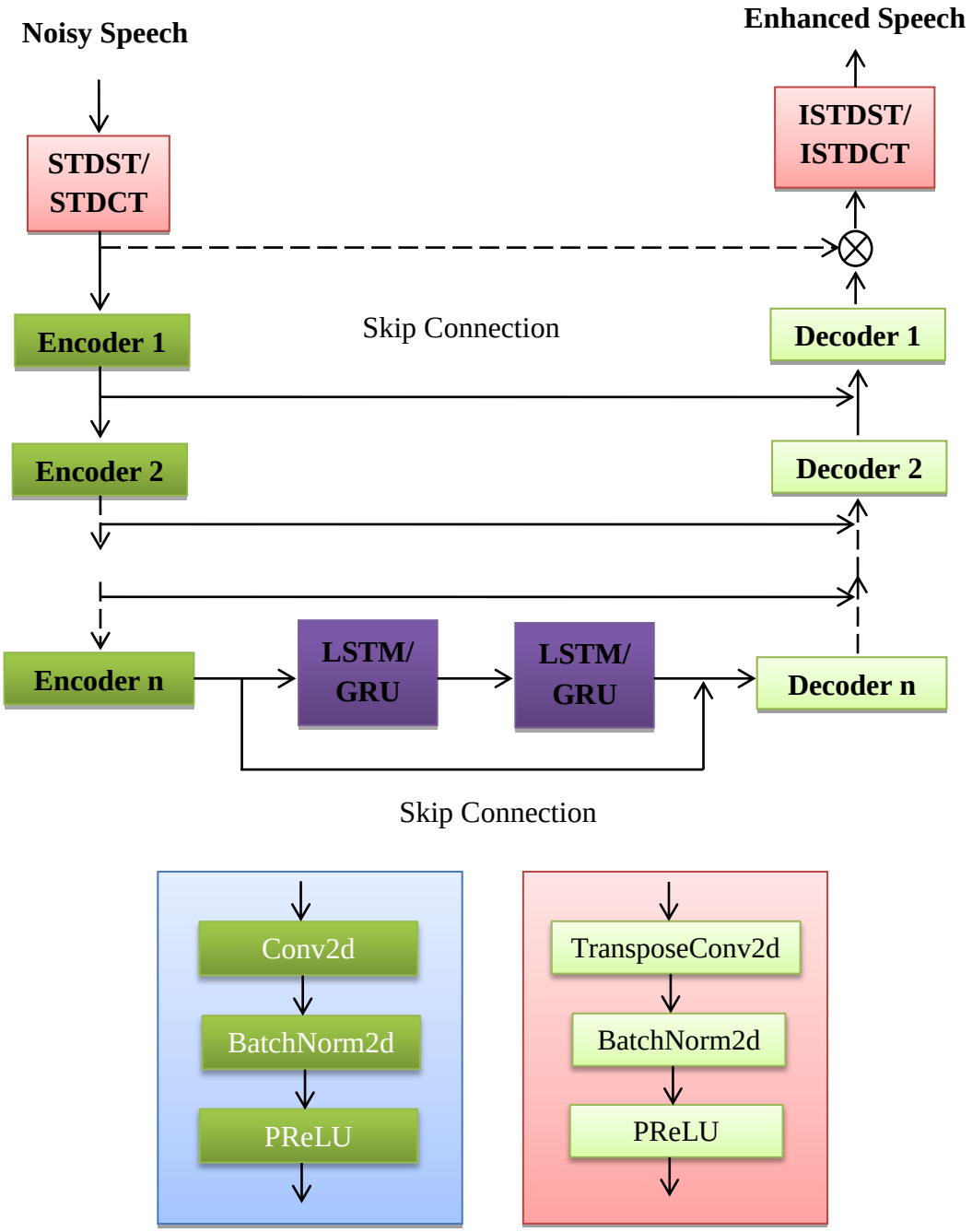



Fig. 2. Proposed encoder-decoder architecture using STDCT and STDST

### *2.1. Discrete domain Input features*

Speech enhancement using neural network uses either time domain or transform domain features [11,12]. For efficient enhancement more calculations and complex network is required to process the signal in low dimension domain like transform domain. Contrast to existed and commonly used DFT transform domain, this work proposes DST and DCT based transform domain for reduced complexity. Complexity of DFT results from its complex nature, wherein real and imaginary components are present [13,14]. Also the complexity in estimating clean speech motivated to process the signal using DST and DCT domains is discussed. The performance of proposed work is compared with different neural networks GCRN [15], SADUNET [16], and MASENET [17].

Performance of deep neural network models are studied using different transforms like DFT, DCT and DST in the literature. The motivation is that each transform has its own ability of energy compaction and to extract the required features from input signal. Speech signal can be decomposed using basis functions of different transforms for frequency related information. DCT transformation is real value based and hence results in reduced calculations as well. DCT and inverse DCT is obtained using (1) and (2) respectively.

$$X_c(k)=\sqrt{\frac{2}{N}}\beta(k)\sum_{n=0}^{N-1}x(n)\cos\left(\frac{\pi k(2n+1)}{2N}\right)\quad k=0,1,...,N-1 \tag{1}$$

$$x(n)=\sqrt{\frac{2}{N}}\sum_{k=0}^{N-1}\beta(k)X_c(k)\cos\left(\frac{\pi k(2n+1)}{2N}\right)\quad n=0,1,..,N-1 \tag{2}$$

Where,

$$\beta(k)=\begin{cases}\frac{1}{\sqrt{2}} & k=0\\ 1 & 1\le k\le N-1\end{cases} \tag{3}$$

Also the DST and IDST transforms are easily represented using (4) and (5) respectively

$$X_k=\sqrt{\frac{2}{N+1}}\sum_{n=1}^{n}x(n)\sin\left(\frac{n\pi k}{N+1}\right)\quad k=1,2,..,N \tag{4}$$

$$x(n)=\sqrt{\frac{2}{N+1}}\sum_{n=1}^{n}X_k\sin\left(\frac{n\pi k}{N+1}\right)\quad n=1,2,..,N \tag{5}$$

*2.2. Masks estimation and target training*

Proposed methods are developed based on estimation of masks during training stage. In this work two masks are used for DST and DCT transform domain deep covolution network [18]. Two masks are discussed in this work, one is Ideal Cosine Mask (ICM) and other is Ideal Sine Mask (ISM) .Cosine mask and sine mask are calculated using (7) and (9) respectively

$$S_{DCT}(t,f)=\frac{S_{DCT}(t,f)}{Y_{DCT}(t,f)}.S_{DCT}(t,f) \tag{6}$$

$$ICM(t,f)=\frac{S_{DCT}(t,f)}{Y_{DCT}(t,f)} \tag{7}$$

Where $S_{DCT}(t,f)$ are the clean speech DCT coefficients and $Y_{DCT}(t,f)$ is the noisy speech discrete cosine transform coefficients. ICM is the ideal cosine mask and *t,f* indicates the time frame and frequency bin respectively.

$$S_{DST}(t,f)=\frac{S_{DST}(t,f)}{Y_{DST}(t,f)}.S_{DST}(t,f) \tag{8}$$

$$ISM(t,f)=\frac{S_{DST}(t,f)}{Y_{DST}(t,f)} \tag{9}$$

Here $S_{DST}(t,f)$ is clean speech short time DST image, $Y_{DST}(t,f)$ is the noisy speech short time DST image and $ISM(t,f)$ is the ideal sine mask, where *t* and*f*indicates the time frame and frequency bin respectively.

In this work proposed methods performance under three activation functions is compared. Sigmoid, PReLU and tanh are used as activation functions in last layer and given the corresponding method numbers as method 1, method 2 and method 3 as follows

**Method-1:**

$$\hat{S}=Sigmoid((method(Y))*Y \tag{10}$$

**Method-2:**

$$\hat{S} = P\,\mathrm{Re}\,LU((method(Y))*Y \tag{11}$$

**Method-3:**

$$\hat{S} = \mathrm{Tanh}((method(Y))*Y \tag{12}$$

### *2.3. Loss Function*

Several loss functions are discussed in the literature for effective working of DNN [19,20]. Most traditional one is the Mean Square Error (MSE) [21,22]. But the characteristics of human hearing are not considered in this loss function [23]. The characteristics based on human hearing are taken into account when Scale Invariant Signal to Noise Ratio (SI-SNR) loss function is considered. Also perceptual based loss function is proposed in literature. Among the available ones the SI_SNR loss function holds well for encoder-decoder based architecture based DNN approaches. In this work SI_SNR (13) is used as loss function.

$$SI\_SNR = 10*\log_{10}\left(\frac{\|s_t\|_2^2}{\|e_n\|_2^2}\right) \tag{13}$$

Where $s_t = \frac{\langle \hat{s}, s \rangle . s}{\|s\|_2^2}$ and $e_n = \hat{s} - s$ . Here noise term in the denominator is obtained by subtracting the clean speech and enhanced speech.

Three loss functions are used in this work. Scale Invariant Signal to Noise Ratio (SI-SNR), Phase Constrained Magnitude (PCM), Time Frequency (TF) loss for optimizing the reconstructed speech [12] .

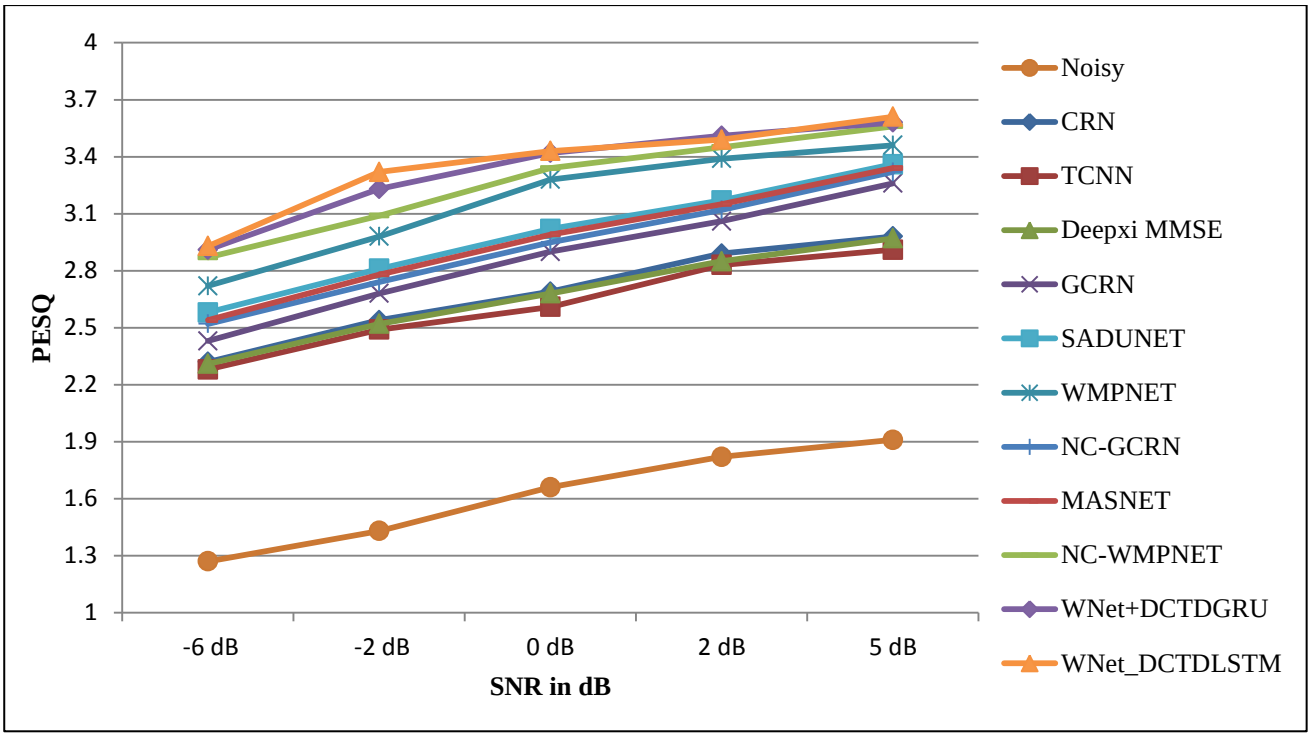
## 3. Results and Experiments

Simulations are performed by considering different noise environments and at different scenarios.

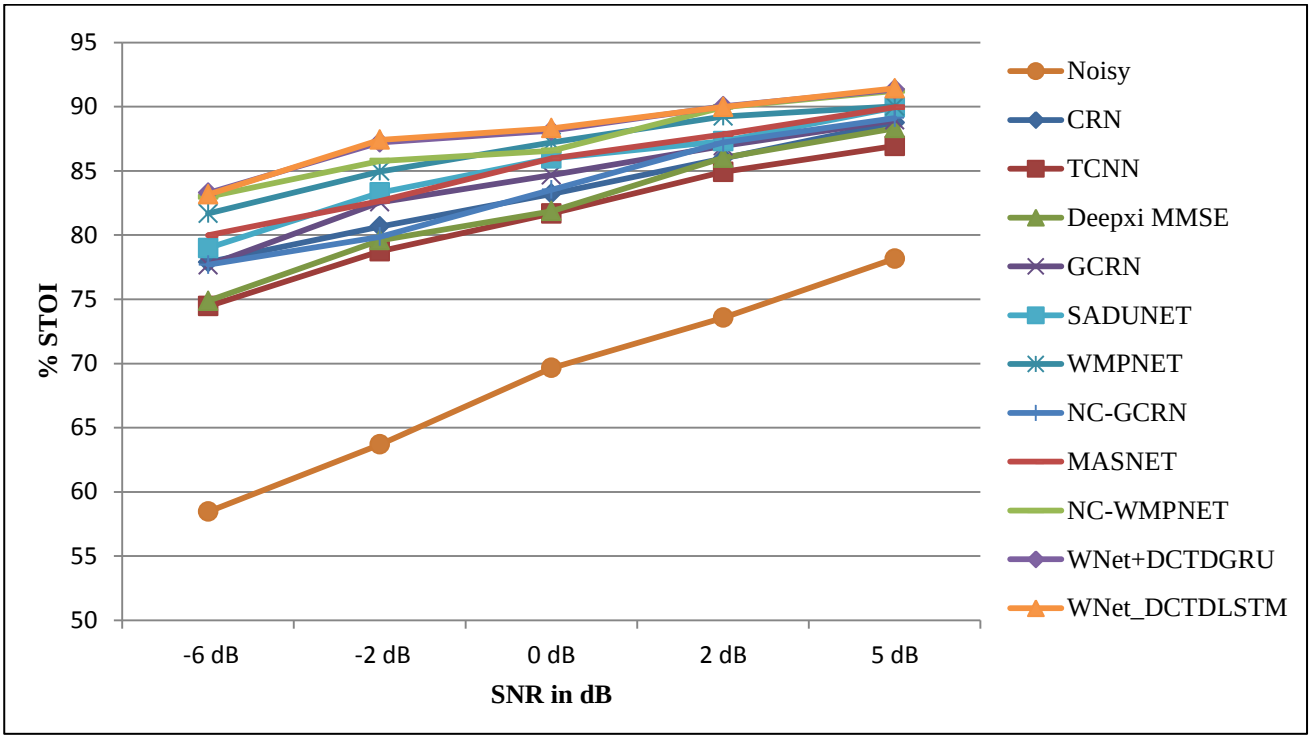
### *3.1. Baseline and result analysis*

Clean speech dataset from NOIZEUS database [24] is taken for training and testing. Noise database is considered for noise samples. Framing is done using hanning window of size 512 point and signal is sampled at 16KHz frequency. During processing short hop time of 8 ms is used. Clean speech signals are corrupted with noise environment at input SNRs of -6 dB, -3 dB, 0 dB, 3 dB and 6dB [23]. Performance of proposed method is analyzed using objective performance measures Perceptual Evaluation of Speech Quality (PESQ), Short Time Objective Intelligibility (STOI) and Signal to Noise Ratio (SNR) as shown in figures. The following baseline methods are compared with proposed method. The methods compared are CRN [12], TCNN [13], DeepXi [14], GCRN [15], SADUNET [16], MASENET [17], Proposed 1 ( WNet+DCTDGRU), Proposed 2 (WNet_DCTDLSTM).

In all these works Adam optimizer is considered. Number of epochs are set to 350 and the first 100 epochs are performed with 0.001 learning rate. This work utilizes different encoder layer and decoder layer channels of 1, 8, 16, 32, 64, 128, 256 and 256 LSTM nodes are choosen as line. The kernel parameter size is set to 5,1 and the stride for various layers as 2,1. Also the GRU nodes of 256 is set in this work. The simulation results of proposed method with baseline methods are given in Fig.3, Fig.4, Fig.5.

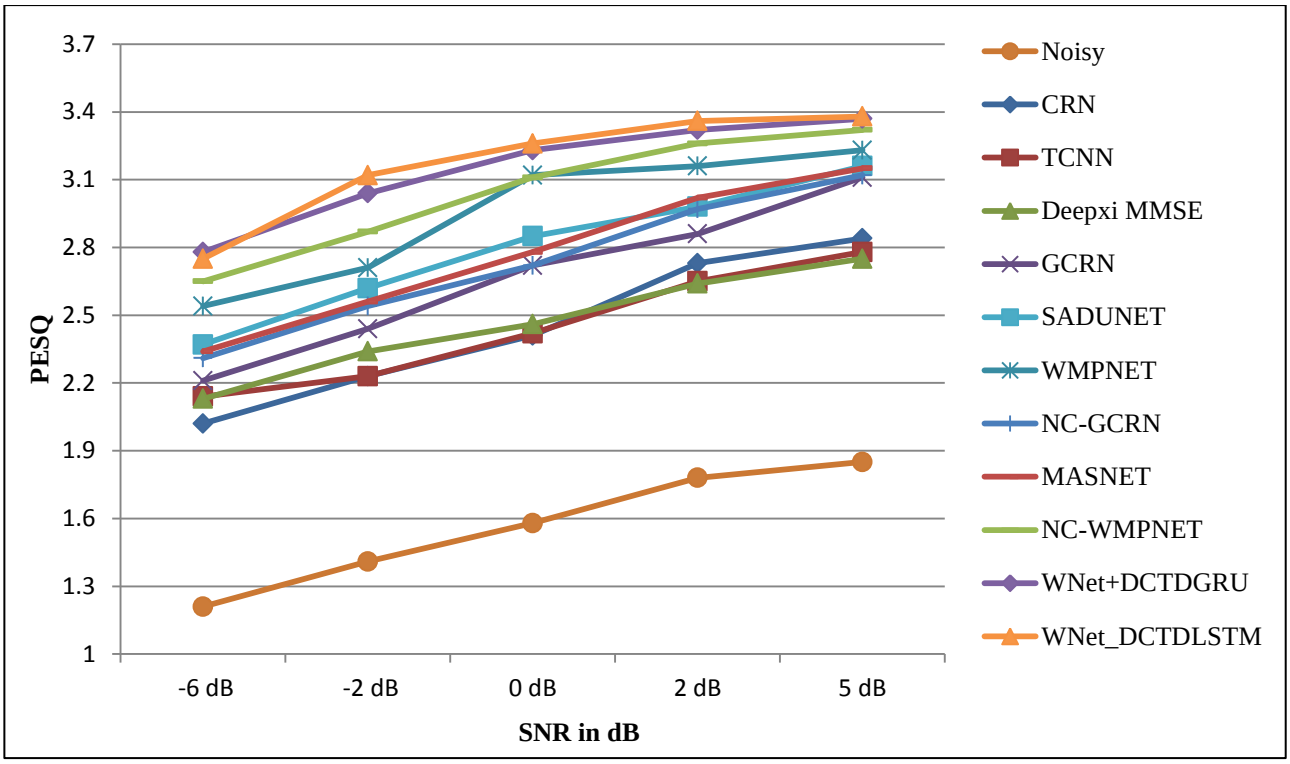



(a)

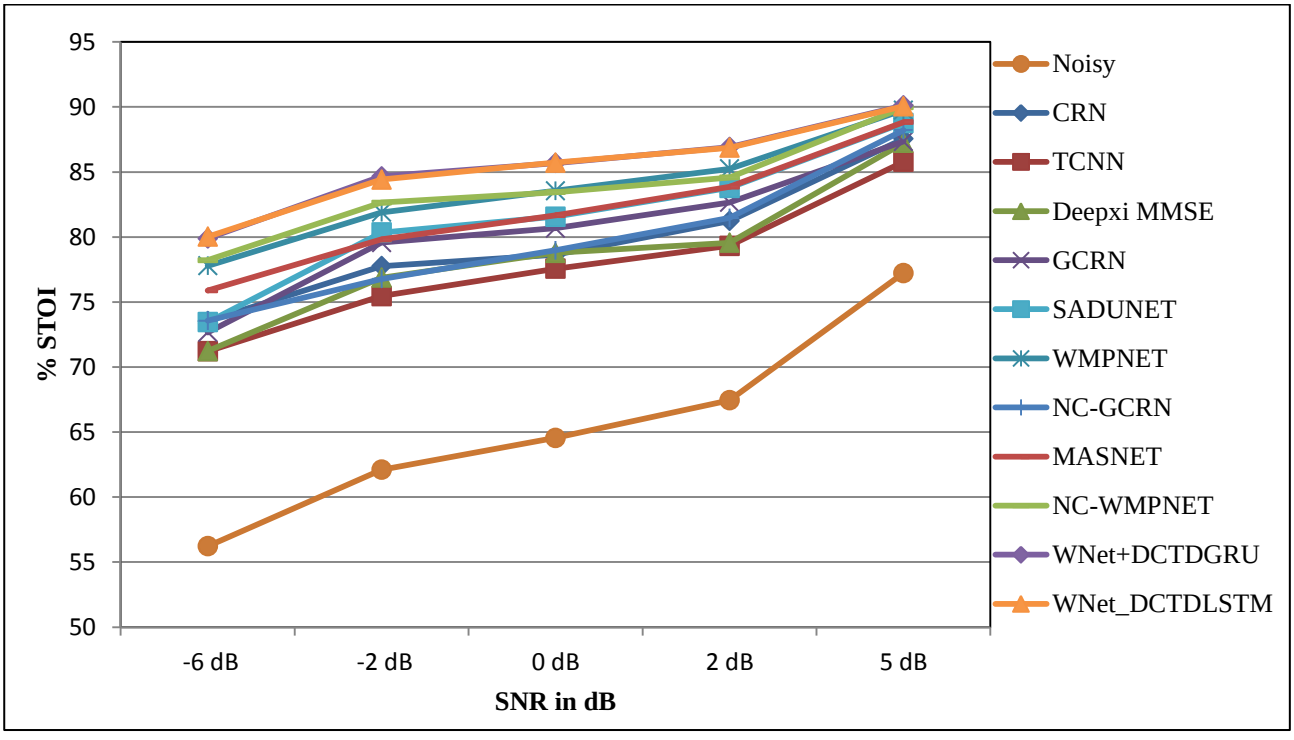



(b)

Fig. 3. Comparison of proposed method with baseline methods using objective performance measures under seen noise environments (a) PESQ (b) %STOI

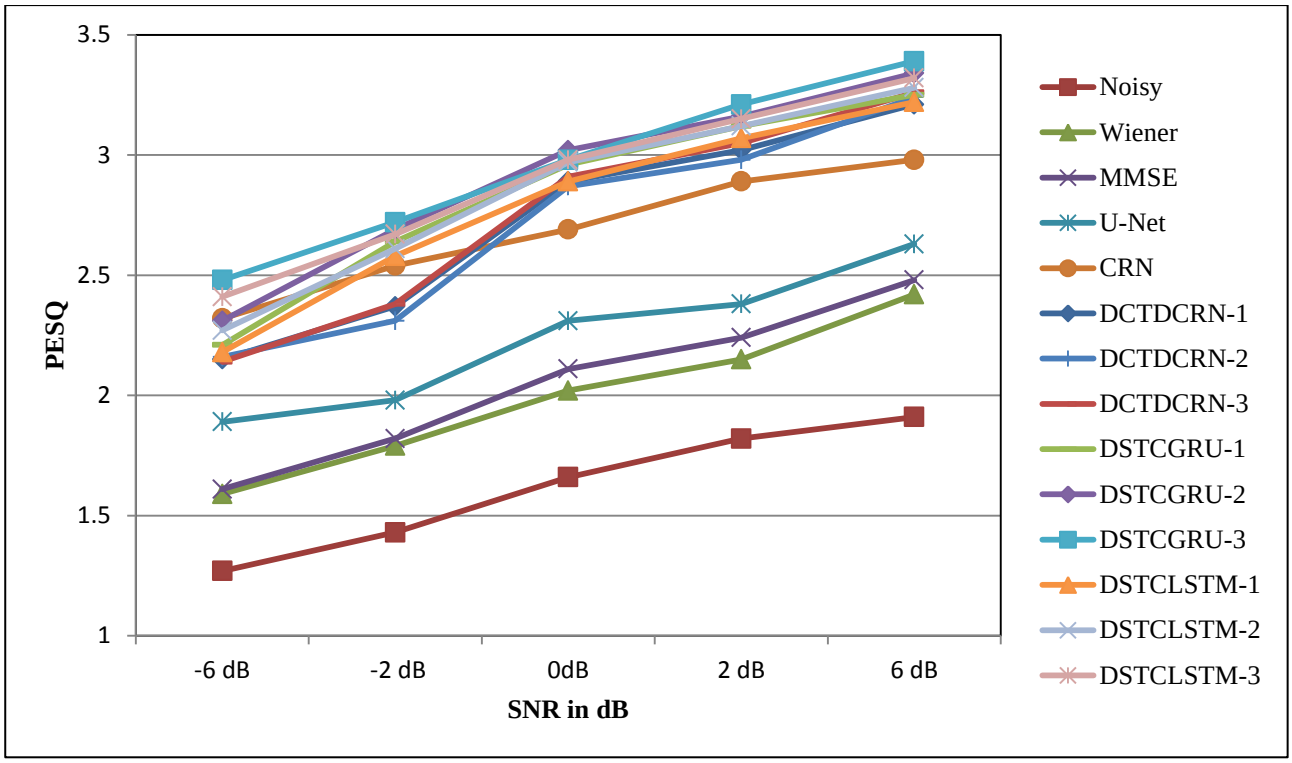



(a)

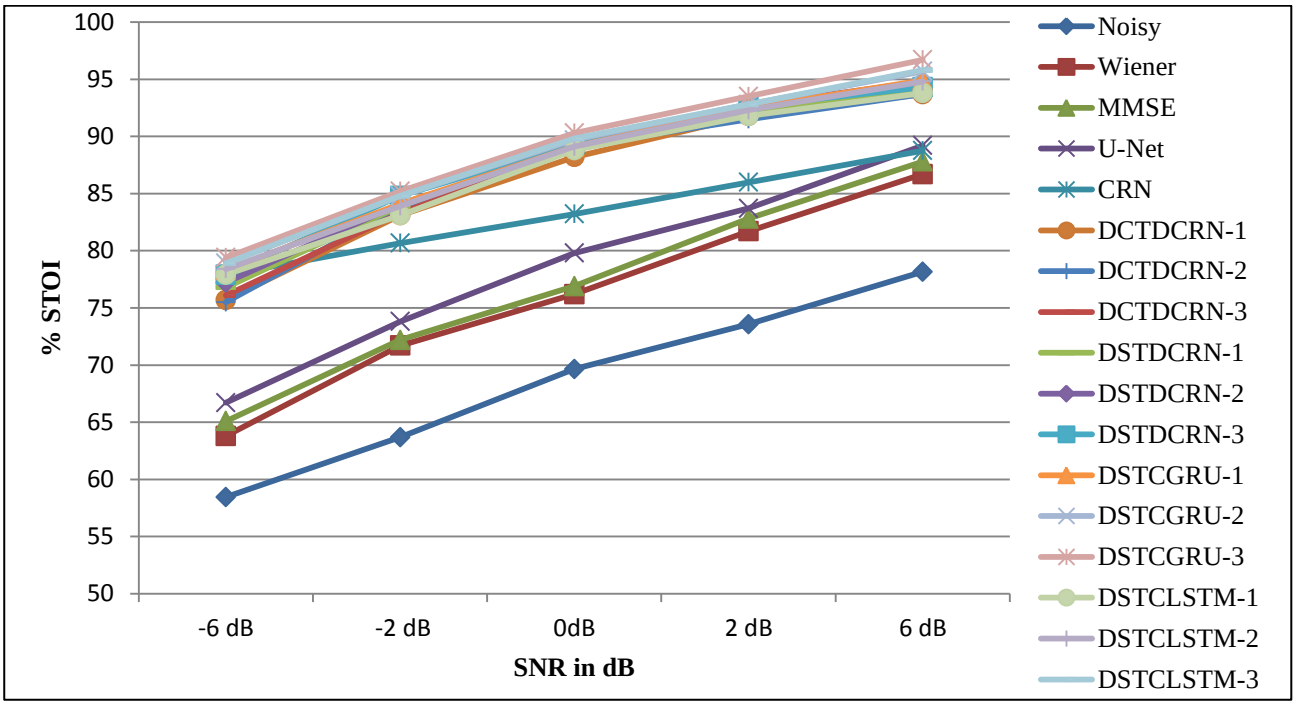



(b)

Fig. 4. Comparison of proposed method with baseline methods using objective performance measures under unseen noise environments (a) PESQ (b) %STOI




(a)




(b)

Fig. 5. Comparision of proposed method with baseline methods using objective performance measures under seen noise environments (a) PESQ (b) %STOI

It is observed from Fig.5 that the DSTCGRU-3 provides a PESQ of 2.48 at low SNR of -6 dB and 3.39 at high SNR of 6 dB. At 0 dB SNR it is observed that DSTCGRU-2 provides a high PESQ of 3.02. It is observed that the DST and DCT based convolutional recurrent neural network provides superior performance than to STFT based convolution neural network CRN and U-net model.

Fig.5. provides the comparison of STOI (In %) measure for existed and proposed methods. It is observed that DSTCGRU-3 provides better results. At low input SNR of -6 dB it is providing a value of 79.4 and a value of 96.7 at high input SNR of 6 dB. At 0 dB SNR DSTCGRU-3 provides a value of 90.3. At 0 dB SNR the methods DCTDCRN-2, DSTDCRN-3 and DSTCGRU-3 provides an STOI of 89.1, 89.3 and 90.3 respectively.

Table 1 provides the comparison of SNR measure for existed and proposed methods. It is observed that DSTCGRU-3 provides better results. At low input SNR of -6 dB it is providing a value of 3.95 and a value of 15.02 at high input SNR of 6 dB. At 0 dB SNR DSTCGRU-3 provides a value of 7.49. At 0 dB SNR the methods DCTDCRN-2, DSTDCRN-3 and DSTCGRU-3 provides an STOI of 5.96, 5.95 and 7.49 respectively

Tab1e 1. SNR Performance Comparison of proposed methods with existing methods

| SNR / Test | **-6 dB** | **-3 dB** | **0dB** | **3 dB** | **6 dB** | **Avg** |
|---|---|---|---|---|---|---|
| Noisy | -5.91 | -2.97 | 0.06 | 3.06 | 5.98 | 0.04 |
| Wiener | -4.86 | -1.68 | 1.14 | 4.47 | 7.05 | 1.22 |
| MMSE | -3.69 | -1.16 | 1.95 | 4.98 | 7.98 | 2.01 |
| U-Net | -2.75 | 1.54 | 3.89 | 7.34 | 9.12 | 3.82 |
| CRN | 1.93 | 4.71 | 6.62 | 9.57 | 11.97 | 6.96 |
| DCTDCRN-1 | 2.23 | 4.32 | 5.76 | 8.32 | 12.67 | 6.66 |
| DCTDCRN-2 | 2.66 | 4.87 | 5.89 | 8.86 | 12.98 | 7.05 |
| DCTDCRN-3 | 2.72 | 4.93 | 5.93 | 8.92 | 13.12 | 7.12 |
| DSTDCRN-1 | 2.41 | 4.43 | 5.82 | 8.46 | 12.78 | 6.78 |
| DSTDCRN-2 | 2.72 | 4.89 | 5.96 | 8.97 | 12.96 | 7.10 |
| DSTDCRN-3 | 2.87 | 5.02 | 5.95 | 8.68 | 12.95 | 7.09 |
| DSTCGRU-1 | 3.73 | 6.98 | 7.24 | 10.78 | 14.68 | 8.68 |
| DSTCGRU-2 | 3.89 | 7.12 | 7.35 | 10.89 | 14.84 | 8.81 |
| DSTCGRU-3 | **3.95** | **7.31** | **7.49** | **10.93** | **15.02** | **8.94** |
| DSTCLSTM-1 | 3.12 | 6.08 | 5.11 | 7.98 | 13.97 | 7.25 |
| DSTCLSTM-2 | 3.28 | 6.32 | 5.42 | 8.11 | 14.15 | 7.45 |
| DSTCLSTM-3 | 3.32 | 6.45 | 5.76 | 8.57 | 14.65 | 7.75 |

## 4. Conclusion

This work proposes a speech enhancement based on deep discrete sine and discrete cosine transform convolutional network. Also the developed techniques are combined with Wnet and joint time domain and DCT based in built phase processing technique is proposed. Experiments are performed on noisy speech samples and the simulation results shows that the proposed methods provide superior results with lesser parameters and calculations. DCT and DST based GRU processing model outperforms the existed approaches at low input and high input SNR corruption. Results show that proposed method shows improved performance in noise reduction (Signal to Noise Ratio) and providing intelligibility (STOI and PESQ) than to existing methods. Joint optimization based on time domain and DCT domain algorithms provides superior performance. This work opens development of different deep transform domain approaches for real time speech enhancement. Hybrid transform domains features can be utilized in deep neural network based algorithms and can be improved further with latest DNN models and transform domain features.

## Acknowledgment

We would like to express our gratitude to the reviewers for their precise and concise recommendations that improved the presentation of the results obtained.

## Conflict of Interest

The authors declare that they have no competing interests.

## References

[1] Y. Xu, J. Du, L.-R. Dai, and C.-H. Lee, “An experimental study on speech enhancement based on deep neural networks,” IEEE Signal processing letters, vol. 21, no. 1, pp. 65–68, 2013
[2] S. Srinivasan, N. Roman, and D. Wang, “Binary and ratio time-frequency masks for robust speech recognition,” Speech Communication, vol. 48, no. 11, pp. 1486–1501, 2006
[3] K. Paliwal, K. Wojcicki, and B. Shannon, “The importance of phase in ´ speech enhancement,” speech communication, vol. 53, no. 4, pp. 465– 494, 2011
[4] Scalart P et al (1996) Speech enhancement based on a priori signal to noise estimation. IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings. 2:629–663
[5] Jansson A, Humphrey E, Montecchio N, Bittner R, Kumar A, Weyde T (2017) Singing voice separation with deep U-NET convolutional networks. Proceedings of the 18th ISMIR Conference, Suzhou, China, 23-27
[6] Tan K, Wang D (2018) A convolutional recurrent neural network for real-time speech enhancement.” in Interspeech. 3229–3233
[7] N. Ahmed, T. Natarajan, and K. R. Rao, “Discrete cosine transform,” IEEE transactions on Computers, vol. 100, no. 1, pp. 90–93, 1974.

[8] Q. Liu, W. Wang, P. J. Jackson, and Y. Tang, "A perceptually-weighted deep neural network for monaural speech enhancement in various background noise conditions," in 2017 25th European Signal Processing Conference (EUSIPCO). IEEE, 2017, pp. 1270–1274
[9] Jassim, W.A., Harte, N.: Comparison of discrete transforms for deep-neural-networks-based speech enhancement. *IET Signal Process*. 16( 4), 438– 448 (2022)
[10] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241
[11] J. Le Roux, S. Wisdom, H. Erdogan, and J. R. Hershey, "Sdr–half-baked or well done?" in ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019, pp. 626–630.
[12] K. Tan, D. Wang, A convolutional recurrent neural network for real-time speech enhancement, in Interspeech (2018), pp 3229–3233
[13] A. Pandey, D. Wang, Tcnn: temporal convolutional neural network for real-time speech enhancement in the time domain, in ICASSP 2019–2019 IEEE International Conference on Acoustics. Speech and Signal Processing (ICASSP) (IEEE, 2019), pp. 6875–6879
[14] Q. Zhang, A. Nicolson, M. Wang et al., Deepmmse: A deep learning approach to mmse-based noise power spectral density estimation. IEEE/ACM Trans. Audio Speech Lang. Process. 28, 1404–1415 (2020)
[15] K. Tan, D. Wang, Learning complex spectral mapping with gated convolutional recurrent networks for monaural speech enhancement. IEEE/ACM Trans. Audio Speech Lang. Process. 28, 380–390 (2019)
[16] X. Xiang, X. Zhang, H. Chen, A nested u-net with self-attention and dense connectivity for monaural speech enhancement. IEEE Signal Process.Lett. 29, 105–109 (2021)
[17] Xiang X, Zhang X, Chen H. A convolutional network with multi-scale and attention mechanisms for end-to-end single-channel speech enhancement.IEEE Signal Process.Lett. 2021;28:1455–9
[18] Xiaoxiao Xiang, XiaojuanZhang,Joint waveform and magnitude processing for monaural speech enhancement,AppliedAcoustics,Volume 200,2022,109077
[19] Ravi Kumar Kandagatla, P.V. Subbaiah,Speech enhancement using MMSE estimation of amplitude and complex speech spectral coefficients under phase-uncertainty,Speech Communication,96,10-27,(2018)
[20] Yecchuri, S., Vanambathina, S. Sub-convolutional U-Net with transformer attention network for end-to-end single-channel speech enhancement. J Audio Speech Music Proc. 2024, 8 (2024)
[21] Kandagatla, R., Subbaiah, P.V. Speech enhancement using MMSE estimation under phase uncertainty. Int J Speech Technol 20, 373–385 (2017)
[22] Yechuri, S., Vanambathina, S. A Nested U-Net with Efficient Channel Attention and D3Net for Speech Enhancement. Circuits Syst Signal Process 42, 4051–4071 (2023)
[23] Kandagatla, R.K., Potluri, V.S. Performance analysis of neural network, NMF and statistical approaches for speech enhancement. Int J Speech Technol 23, 917–937 (2020)
[24] Hu, Y. and Loizou, P. (2007). "Subjective evaluation and comparison of speech enhancement algorithms," Speech Communication, 49, 588-601

## Authors' Profiles

**Ravi Kumar Kandagatla** received his Ph.D from Jawaharlal Nehru Technological University, Kakinada, India in 2021. Currently he is working as Associate Professor in the Department of ECE at Lakireddy Bali Reddy College of Engineering, Mylavaram, India. His research interests include Signal processing, speech enhancement and statistical signal processing.

**V. Jayachandra Naidu** received his M.Tech in Communication Engineering from Vellore Institute of Technology, Vellore, Tamil Nadu, India in 2004. Currently, he is working as an Associate Professor in the Department of ECE at Sri Venkateswara College of Engineering and Technology (Autonomous), Chittoor, Andhra Pradesh, India. His research interests include Computer Vision, Machine Learning, Signal Processing, Embedded Systems and IoT.

**P. S. Sreenivas Reddy** is working as an Associate Professor, Electronics and Communication Engineering, Nalla Narasimha Reddy educational societies group of institutions, Telengana. He has graduated M.E Applied Electronics in the year 2000 and B.E Electronics and Communication Engineering in the year 1998. He has Twenty one years of Academic Experience at under graduate and Post graduate level. His research interests include Nanomaterials, Micro Electronics, and VLSI Design. He is presently doing Ph.D degree in Electronics and Communication Engineering, Dr. M.G.R. Educational and Research Institute, University.

**Sivaprasad Nandyala** currently working as Senior Machine Learning Engineer at Technology Innovation Institute (TII), Abu Dhabi, U.A.E. Dr. Nandyala career includes roles at Eaton Research Labs, Tata Elxsi, Wipro Technologies, Analog Devices and Ikanos communications, covering various technology domains. He has contributed significantly to the field with over 35 research publications and 3 granted patents. Dr. Nandyala earned his Ph.D in speech processing from NITWarangal, India, and was an ERASMUS MUNDUS scholarship recipient for Postdoctoral Research at Politecnico di Milano, Italy.

**How to cite this paper:** Ravi Kumar Kandagatla, V. Jayachandra Naidu, P. S. Sreenivasa Reddy, Sivaprasad Nandyala, "Speech Enhancement Using Joint Time and DCT Processing for Real Time Applications", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.16, No.5, pp. 14-24, 2024. DOI:10.5815/ijigsp.2024.05.02