Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:8(2), August 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

Adaptive Filtering and EMD–EEMD Fusion for Robust Noise-Resilient Speech Communication

Pradeep Kumar Sriperamboodhuru* [1], **Anitha Sheela Kancharla**[2]
[1,2]*Dept of ECE, Jawaharlal Nehru Technological University, Hyderabad,Telangana-500085, India*

Abstract: Speech enhancement remains a core requirement for communication, hearing aids, and automatic speech recognition front-ends. The task becomes harder when the noise is non-stationary or shares a spectral profile with speech. Residual musical noise, low-frequency rumble, and multi-speaker babble remain difficult for single-channel systems to handle cleanly. Classical baselines such as Spectral Subtraction, Wiener Filtering, Kalman Filtering, and the Wavelet Packet Transform with Voice Activity Detection (WPT + VAD) reduce noise energy but often damage speech structure. Fixed bases and stationarity assumptions limit what they can recover in real-world acoustic conditions. This work combines Empirical Mode Decomposition (EMD) with 8 Intrinsic Mode Functions (IMFs) and Ensemble EMD (EEMD) with 100 ensembles, followed by a Wiener-assisted adaptive filter of length 32. A 24-layer Convolutional Recurrent Network (CRN) with 128 hidden channels then refines the log-magnitude spectrogram. Experiments use the VoiceBank + DEMAND corpus, with 11,572 training and 824 test utterances sampled at 16 kHz. Test Signal-to-Noise Ratios (SNRs) are 2.5, 7.5, 12.5, and 17.5 dB. The proposed method achieves a Perceptual Evaluation of Speech Quality (PESQ) score of 3.24, a Short-Time Objective Intelligibility (STOI) score of 0.97, and a Signal-to-Distortion Ratio (SDR) of 13.47 dB. Gains hold across seen and unseen noise. The pipeline offers a practical balance of quality and runtime cost.

Keywords: Speech enhancement, Empirical Mode Decomposition, Convolutional Recurrent Network, VoiceBank+DEMAND, Perceptual Evaluation of Speech Quality

1. INTRODUCTION

The most natural form of human communication is speech, and clean speech is important in telephony, hearing aids, voice conferencing and automatic speech recognition. True recordings hardly remain untouched. Noise in the background, such as in kitchens, offices, streets and transport hubs, impairs the perceptual quality and the intelligibility. Single-channel enhancement is the environment where a single microphone signal is present, which is typical of phones and small devices. The aim is to reclaim speech more similar to the clean reference without adding new artefacts.

The challenge arises due to the character of daily noise. Most real noise sources are non-stationary and vary at a rate that is faster than the algorithm can follow. Other noises, like cafeteria babble or noise in restaurants, are spectrally similar to speech itself, and thus difficult to isolate. Subway and bus rumble at low frequency can obscure voiced areas, and outdoor traffic combines several types of interference simultaneously. An effective technique must be able to inhibit this spectrum of noise without causing harm to consonants, transients, or vocal traits of the speaker.

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:8(2), August 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

The classical methods have been researched over decades. Spectral subtraction compares the spectrum of noise to the signal and removes it, yet in many cases it can remove musical noise that reduces signal quality. Wiener and Kalman filtering presuppose local stationarity and enhance the signal-to-distortion ratio, but are poor at dealing with rapidly changing interference. The voice activity detection of wavelet packet methods employs a fixed basis, which restricts the ability to match the signal at hand. Decomposition approaches such as empirical mode decomposition and its ensemble variant adapt to the signal, and pairing them with adaptive filtering recovers more speech, although residual noise and speech distortion remain.

The present work builds a pipeline that keeps the strengths of signal-adaptive decomposition and adds a learned stage on top. Empirical mode decomposition with 8 intrinsic mode functions separates the noisy signal into components, ensemble empirical mode decomposition with 100 ensembles stabilises the mode split, and a Wiener-assisted adaptive filter of length 32 refines the residual. A 24-layer convolutional recurrent network with 128 hidden channels then processes the log-magnitude spectrogram to suppress what the front end leaves behind. The combination targets non-stationary and speech-like noise while keeping the parameter count and runtime practical for on-device use.

1.1. Scientific Contribution and Novelty

EMD, EEMD, adaptive filtering, and CRN based speech enhancement have been studied in literature separately, however, their combination is generally restricted to separate pre-processing or enhancement steps. Different from the approaches in literature, the proposed framework uses a systematic multi-stage enhancement strategy. The empirical mode decomposition (EMD) is used for adaptive signal decomposition and the ensemble empirical mode decomposition (EEMD) is used to reduce the mode-mixing effects. Wiener-assisted adaptive filtering suppresses the residual noise components, and a lightweight CRN is used to refine the reconstructed speech signal. innovative in that its complementary stages are synergistically combined in a cohesive framework. Unlike the purely deep learning based methods which rely on the neural learning only for the noise suppression, the proposed method performs the substantial enhancement in the signal domain prior to the neural refinement, which alleviates the burden on the CRN and improves the robustness under the non-stationary noise conditions with the practical computational complexity.

The main contributions of this work are:

- A hybrid single-channel speech enhancement framework that combines EMD, EEMD, Wiener-assisted adaptive filtering, and CRN refinement to reduce non-stationary noise, mode mixing, residual noise, and remaining spectral artefacts.
- An ablation that isolates the contribution of each block and shows that decomposition drives most of the early gain while the CRN provides a further bounded improvement.
- Evaluation on the VoiceBank+DEMAND corpus at 16 kHz with both seen and unseen noise categories and test SNR points of 2.5, 7.5, 12.5, and 17.5 dB, which do not appear during training.
- Additional controlled SNR analysis is also reported at -5, 0, 5, 10, and 15

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:8(2), August 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

dB to evaluate robustness under wider noise conditions.

- Reported PESQ of 3.24, STOI of 0.97, and SDR of 13.47 dB, with absolute gains of +1.27 in PESQ and +6.26 dB in SDR over the noisy input.
- A runtime analysis showing a real-time factor of 0.06 and 3.84 M parameters, together with statistical stability over five independent runs.

The rest of the paper is organised as follows. Section 2 reviews related work on classical filtering, decomposition-based enhancement, and deep learning approaches. Section 3 describes the proposed pipeline, including the decomposition stage, the adaptive filter, and the CRN. Section 4 presents the dataset, evaluation setup, and experimental results, including baseline comparison, SNR-level behaviour, per-noise-type scores, ablation, distortion analysis, statistical stability, and runtime cost. Section 5 concludes the paper and outlines directions for future work.

1.2 Background and Related Work

Recent research on speech restoration has investigated wavelet techniques, adaptive noise control, beamforming, semantic communication, and deep learning, but efficient speech restoration in a wide range of highly nonstationary real-world noise conditions is still a challenge. Iqbal et al. [1] used the discrete wavelet transform in conjunction with spectral subtraction to enhance speech in low SNR, but the technique might not perform well in highly nonstationary or unobservable real-world noise as compared to end-to-end learned algorithms. Ji et al. [2] suggested a distributed active noise control algorithm, which does not require internode communication, but its more limited validation range restricts its generalization to non-ANC contexts. Chen et al. [3] proposed an RVQGAN-based semantic communication model, which attains high quality at low bitrates, but is designed to focus on communication efficiency, not standalone speech denoising, and has nontrivial complexity to deploy. Ji et al. [4] proposed MGDFxLMS, which uses an auto-shrink step-size approach to solve crosstalk and delay in multichannel ANC systems, but the benefits are conditional on distributed infrastructure assumptions. Haeb-Umbach et al. [5] surveyed the combination of microphone array processing with deep learning to reduce noise and dereverberate a sound, and did not experimentally validate a new algorithm. Luo et al. [6] expanded selective fixed-filter ANC with frequency-direction awareness and multi-task learning, in which the performance is strongly dependent on the similarity of the noise conditions during training and testing. Essaid et al. [7] used the deep learning to enhance speech perception in noisy conditions with cochlear implants, but the methodology is implant-specific and has limited transferability to general speech enhancement. Turpati et al. [8] introduced an online acoustic-feedback ANC algorithm to achieve better stability and convergence, and practical advantages are dependent on particular feedback conditions. Vasundhara et al. [9] incorporated dynamic fixed-filter techniques with long-term Kalman filtering to adaptive ANC in varying noise conditions, but the conference format does not provide the in-depth speech quality analysis. Huang et al. [10] surveyed the past and more recent developments in microphone array processing and multichannel speech enhancement with the inclusion of deep learning, without presenting or verifying a novel algorithm. Yu and Zhang [11]

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:8(2), August 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

surveyed the sources of PLC noise in distributed energy systems and mitigation approaches, but the article is a synthesis of the existing approaches instead of a validation of a new enhancement model. Ramana et al. [12] used wavelet decomposition along with neural-network-based thresholding to suppress noise; however, the brief conference format does not provide enough experimental richness or the ability to generalise to a variety of nonstationary speech settings. Wang et al. [13] suggested a transferable selective virtual sensing ANC system based on metric learning to be able to adapt on unseen noise types without retraining, and the applicability is limited to ANC system assumptions. Guo et al. [14] discussed noise reduction in group emotion recognition in a multi-scale contextual framework, but the approach focuses on task-related noisy features instead of speech restoration at the waveform level. Shetu et al. [15] enhanced ULCNet by adding time alignment and parallel encoder blocks to jointly denoise noise and echo at low complexity, but need to be cautiously framed when comparing to pure speech denoising models. Kim et al. [16] constructed a complete maritime voice communication system that integrates STT, TTS, and AIS data with real-time performance but the main contribution is a complete communication pipeline and not an independent denoising algorithm. Kar et al. [17] suggested an adapted adaptive algorithm for online secondary path modelling in ANC to deal with time-varying system paths, but without explicitly considering the quality of speech enhancement. Guo et al. [18] examined contextual-knowledge-enhanced ASR in air traffic control and demonstrated an increase in recognition under noisy radio conditions, but the emphasis is on recognition quality, not noise reduction or signal reconstruction. Zhou et al. [19] created a distributed ANC algorithm that is resistant to impulsive interference, but the algorithm is only applied to ANC infrastructure and not speech denoising. STFT-domain beamforming based on dimensionality reduction was suggested by Liu et al. [20] to reduce computational cost without affecting noise reduction, but is an algorithmic formulation study, not an end-to-end speech enhancement system.

Current methods of speech enhancement are evidently limited when used in real-life situations. Several approaches are designed based on adaptive noise control models, with performance relying on certain infrastructure assumptions and failing to apply to independent denoising problems. Other works consider the efficiency of communication, the accuracy of recognition or application-specific pipelines, and in these cases, noise suppression is not the primary goal and is not optimized to restore the waveforms. Wavelet decomposition, spectral subtraction, and Kalman filtering are classical methods that are interpretable but do not work well in highly nonstationary or speech-like noises, whereas end-to-end deep learning models are sensitive to similarity in training data and do not necessarily maintain fine speech structures in new settings. Moreover, several works are not thoroughly validated in different datasets and dynamic noise conditions. These gaps demonstrate that a single framework should be developed that integrates adaptive signal decomposition with learning-based refinement. The suggested approach is a solution to this, combining EMD and EEMD to achieve noise-sensitive component separation, adaptive filtering to achieve dynamic noise reduction, and a CRN to restore

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:8(2), August 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

residual speech, enhancing robustness, intelligibility, and perceptual quality in controlled and real-world acoustic environments.

CRN-based models have a good combination of spectral feature extraction and temporal modeling, and DCCRN further improves the reconstruction by complex-valued processing of magnitude and phase information [26,27]. MetricGAN+ is a GAN based approach that directly optimizes perceptual quality metrics and achieves better speech quality scores [23]. Transformer-based architectures, such as SepFormer and TF-GridNet, have reached state-of-the-art enhancement performance [28,29] by leveraging self-attention mechanisms to capture long-range contextual information. More recently, diffusion and rectified-flow models are shown to have impressive ability to restore speech but at a higher computational cost [30,21]. Differently from the above methods, the proposed framework integrates adaptive signal decomposition, residual noise suppression and neural refinement into a computationally efficient hybrid architecture.

Table 1: Comparative Analysis of Recent Speech Enhancement Approaches

Method	Key Feature	Limitation
CRN	CNN+ RNN modeling	Limited phase enhancement
DCCRN	Complex-valued processing	Higher complexity
MetricGAN+	Metric-guided optimization	Training complexity
SepFormer	Transformer attention	High computational cost
TF-GridNet	Time-frequency attention	Large memory requirement
Diffusion/Flow Models	High-quality restoration	Slow inference
Proposed Method	EMD+ EEMD+ Adaptive Filter+ CRN	Moderate preprocessing overhead

As detailed in Table 1, recent state-of-the-art methods provide great enhancement performance but frequently need significant processing resources. The proposed hybrid framework seeks to provide a practical trade-off between enhancement quality,

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:8(2), August 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

robustness, and computational efficiency.

2. MATERIALS AND METHODS

2.1. Overview of the Proposed Pipeline

The proposed system treats single-channel speech enhancement as a cascade of three stages, as shown in Figure 1. The front end of the system uses signal-adaptive decomposition to separate noisy signals into their different noise-related behaviour components. The adaptive filter uses a short-term noise profile estimate to improve the residual signal. The learned back end creates a cleaner version of the log-magnitude spectrogram through its mapping process. The system uses different components to address the specific weaknesses of each preceding component, which results in an easier process to analyse the system. The proposed method is not a new decomposition theory or a fundamentally new neural architecture. The contribution is rather in the structured coupling of complementary enhancement stages, each of which corrects a particular deficiency of the preceding stage. The decomposition stages process non-stationary signal separation, the adaptive filtering removes correlated residual noise, and the CRN is used to remove the residual noise that remains in the spectrum after signal domain enhancement.

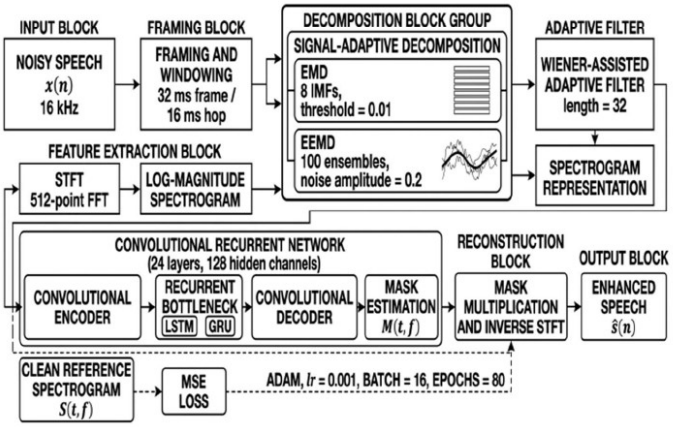

Figure.1 Architecture of the proposed speech enhancement pipeline.

2.2. Signal Model and Problem Formulation

Let $x(n)$ denote the noisy observation at discrete time n, written as the sum of clean speech $s(n)$ and additive noise $v(n)$:

$$x(n) = s(n) + v(n) \qquad (1)$$

During inference, only the noisy observation x(n) is available to the enhancement pipeline. The clean speech signal s(n) is not used by the EMD, EEMD, adaptive filtering, or CRN enhancement stages during testing. The clean signal is used only during supervised training of the CRN loss and during objective evaluation to compute PESQ, STOI, SDR, and related metrics. This avoids oracle clean-speech information and prevents test-set leakage. Equation (1) is the standard

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:8(2), August 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

additive model used for single-channel enhancement. The goal is to produce an estimate. $\hat{s}_n$ that stays close to $s(n)$ in both perceptual quality and signal-to-distortion ratio, without assuming a fixed spectral shape for $v(n)$. Real recordings rarely satisfy stationarity, so the method avoids any global assumption on the noise statistics and relies on local decomposition and learned refinement instead.

2.3 Empirical Mode Decomposition Stage

Empirical mode decomposition splits. $x(n)$ into a small set of intrinsic mode functions (IMFs) plus a residual, using iterative sifting based on local maxima and minima. The decomposition is data-driven, which suits non-stationary mixtures. For the configuration used here, eight IMFs are retained with a stopping threshold of 0.01 on the normalised sifting error. The signal can be reconstructed as

$$x(n) = \sum_{k=1}^{K} c_{\mathrm{k}}(n) + r(n) \quad (2)$$

where $c_{\mathrm{k}}(n)$ is the k-th IMF, $K = 8$, and $r(n)$ is the residual trend. The lower-order IMFs in Equation (2) carry most of the high-frequency noise, while higher-order modes hold the slow-varying speech structure. A practical limitation of plain EMD is mode mixing, where a single IMF captures components of very different scales.

2.4 Ensemble EMD Refinement

Mode mixing is reduced by ensemble empirical mode decomposition, which adds small white noise realisations to $x(n)$ before decomposition and averages the IMFs over trials. With $M = 100$ ensembles and a noise amplitude of 0.2, the k-th refined mode is

$$\bar{c}_{\mathrm{k}}(n) = \left(\frac{1}{M}\right) \sum_{m=1}^{M} c_{\mathrm{k}}^{(\mathrm{m})}(n) \quad (3)$$

where $c_{\mathrm{k}}^{(\mathrm{m})}(n)$ is the k-th IMF obtained from the m-th noisy realisation. Equation (3) stabilises the mode split and gives a cleaner separation between noise-dominated and speech-dominated components. The cost is runtime, since 100 sifting runs replace one. The pairing with EMD in the pipeline is intentional: EMD gives a fast first pass, and EEMD is used where the mode separation needs more reliability.

2.5Wiener-Assisted Adaptive Filtering

The residual after decomposition may still contain correlated noise that cannot be fully removed by a fixed basis. To reduce this remaining component, a Wiener-assisted adaptive filter with length L=32 is applied to the residual obtained from the noisy input. During inference, the filter does not use the clean speech signal or any oracle reference. It uses only information derived from the noisy observation.

Let r(n) denote the residual vector of length L obtained after EMD/EEMD decomposition. The adaptive filter output is defined as in (4)

$$yA(n) = w^{T}(n)r(n) \quad (4)$$

where $w(n)$ is the adaptive filter coefficient vector. A Wiener-assisted estimate computed from the same noisy observation is used as a pseudo-reference and is written as in (5)

$$\hat{d}(n) = yW(n) \quad (5)$$

The adaptive error is then calculated as in (6)

$$e(n) = \hat{d}(n) - yA(n) \quad (6)$$

The filter coefficients are updated as in (7)

$$w(n+1) = w(n) + \mu e(n)r(n) \quad (7)$$

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:8(2), August 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

where μ is the step size. Here, $\hat{d}$ It is not a clean speech signal. It is a Wiener-assisted pseudo-reference estimated from x(n) and its decomposed residual components. Thus, the adaptive filtering stage does not require clean speech during inference. The refined residual is then recombined with the speech-dominated modes before feature extraction.

2.6 Feature Representation

The enhanced waveform is framed with a length of 32 ms and a hop of 16 ms, and a 512-point short-time Fourier transform is applied. The log-magnitude spectrogram is used as the input feature:

$$X(t,f) = log(\,|STFT\{\,\tilde{x}(n)\}| + \varepsilon) \quad (8)$$

where $\tilde{x}(n)$ is the output of the adaptive filtering stage, t indexes frames, f indexes frequency bins, and ε is a small constant that avoids log 0. The representation in Eq. (8) is standard for spectral-domain enhancement and matches the resolution needed to separate speech formants from residual noise.

2.7 Parameter Selection

The main parameters were chosen based on preliminary experiments and computational considerations. In EMD, the performance improvement was marginal after this number and the computational complexity increased, so eight IMFs were kept. For EEMD, stable modes decomposition was achieved with 100 ensembles and no significant improvement for larger ensemble size. The length of adaptive filter was fixed at 32 taps, which provided an effective trade-off between the residual noise suppression and the computational cost. Hence these settings were used in the whole study.

2.8 Convolutional Recurrent Network Back End

The final stage is a convolutional recurrent network (CRN) with 24 layers and 128 hidden channels. A convolutional encoder reduces the spectrogram to a compact time-frequency embedding, a recurrent bottleneck models temporal context along the frame axis, and a convolutional decoder reconstructs an enhanced log-magnitude spectrogram. The network predicts a spectral mask $M(t,f) \in [0,1]$, and the enhanced spectrogram is obtained as

$$S(t,f) = M(t,f) \cdot X(t,f) \quad (9)$$

Equation (9) keeps the phase of the noisy input and modifies only the magnitude, which simplifies training and keeps the model small. The time-domain estimate $\hat{s}(n)$ is recovered by inverse STFT with overlap-add. The CRN is an encoder–recurrent–decoder network. The convolutional blocks are followed by batch normalisation and ReLU activation, with a 3 × 3 kernel and stride 1 and padding 1. The encoder extracts local time-frequency features from the log-magnitude spectrogram, and the recurrent bottleneck is a gated recurrent unit that models the frame-level temporal context. Fine spectral details are preserved by skip connections between matching encoder and decoder stages. The decoder is composed of convolutional reconstruction layers with the same kernel setting and generates a mask estimation between [0, 1]. This mask is multiplied with the input spectrogram to produce the enhanced log-magnitude spectrogram.

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:8(2), August 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

Table 2: CRN Architecture table

Component	Configuration
Encoder Layers	5 Convolutional Layers
Kernel Size	3 × 3
Activation Function	ReLU
Normalization	Batch Normalization
Recurrent Block	GRU
Hidden Units	128
Decoder Layers	5 Convolutional Layers
Output Layer	Convolution + Sigmoid
Total Layers	24
Trainable Parameters	3.84 M

The configuration of the proposed 24-layer CRN is shown in Table 2. The network combines convolutional feature extraction and recurrent temporal modeling to reconstruct speech better. The chosen architecture achieves a good trade-off between enhancement performance and computational complexity.

Table 3: Training Hyperparameters and Hardware Configuration

Parameter	Value
Optimizer	Adam
Learning Rate	0.001
Batch Size	16
Number of Epochs	80
Early Stopping	Patience = 10 epochs
Framework	TensorFlow / PyTorch
GPU	Intel ® Iris ® Xe Graphics
CPU	12th Gen Intel ® Core ™ i-7-1255U
RAM	16GB

The hyperparameter settings used to train the CRN are summarized in Table 3. To prevent overfitting, early stopping was performed according to the validation loss. All experiments were carried out on the hardware platform listed in the table to make the experiments reproducible.

2.9. Training Objective and Optimisation

The network is trained with a mean squared error (MSE) loss between the predicted and clean log-magnitude spectrograms:

$$L = \left(\frac{1}{TF}\right)\sum_{t=1}^{T}\sum_{f=1}^{F}\left(\hat{S}(t,f)S(t,f)\right)^2 \quad (10)$$

where $S(t,f)$ is the clean reference, and T, F are the frame and frequency counts. Optimisation in Eq. (10) uses Adam with a learning rate of 0.001, a batch size of 16, and 80 epochs. The loss acts on the same representation that the CRN predicts, which keeps the gradient path short and the training stable for the

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:8(2), August 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

corpus size used in this study. The CRN is trained with the loss function being mean squared error between the predicted and clean log-magnitude spectrograms. Adam is used as the optimizer with a learning rate of 0.001 and a batch size of 16. Gradient clipping is used to stabilise training and the model checkpoint that achieves the best validation PESQ is chosen for final testing. The same preprocessing and evaluation settings are used for all reported CRN experiments.

3.RESULTS AND DISCUSSION

This section reports the experimental outcomes of the proposed EMD + EEMD + Adaptive Filtering + CRN pipeline on the VoiceBank+DEMAND corpus. The analysis covers baseline comparison, SNR-level behaviour, per-noise-type scores, ablation, distortion, statistical stability, and runtime cost.

In the stage of EMD-EEMD, the non-stationary noise components can be suppressed successfully and the speech features can be preserved. This is consistent with the findings of previous research that adaptive decomposition techniques are suitable for speech enhancement and denoising applications [31,32].

It is well known that Wiener-based filtering approaches are successful in the field of speech enhancement systems [33]. The adaptive filtering stage provides also some noise reduction by suppressing the residual interference components.

The CRN improves the quality of speech reconstruction by jointly modeling the spectral and temporal characteristics of speech, which is consistent with the findings reported for CRN and DCCRN architectures in recent speech enhancement studies [26,27].

The PESQ, STOI and SDR improvements show improvements in perceptual quality, intelligibility and signal fidelity, respectively, following the trends reported in recent deep-learning-based speech enhancement frameworks [23].

3.1.Dataset Description

The experiment employs the voicebank+DEMAND corpus [25], a publicly available benchmark that is commonly used in speech enhancement in a single channel. Sampling of audio takes place at 16 kHz. The training partition has 11,572 noisy-clean paired utterances that are sampled out of 28 speakers, whereas the test partition has 824 utterances of 2 unknown speakers. The entire collection is 12,396

Table 4: Dataset Details

Attribute	Value
Dataset	VoiceBank+DEMAND
Source	Public benchmark
Sampling rate	16 kHz
Training utterances	11,572
Test utterances	824
Total utterances	12,396
Speech type	Clean English speech
Noise type	Real environmental noise
Data format	Noisy-clean paired utterances
Use in the study	Training and evaluation of speech enhancement model

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:8(2), August 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

utterances. The mixtures of training are designed at four signal-to-noise ratios, which are 0, 5, 10, and 15 dB. The test set employs a different grid of 2.5, 7.5, 12.5, and 17.5 dB, which locates the evaluation at intermediate SNR levels that are not observed during training. This division compels the model to generalise both the identity of the speaker and the level of noise instead of memorising specific conditions.

Experiments are conducted on VoiceBank+DEMAND corpus, a popular benchmark dataset for single-channel speech enhancement. The dataset includes speech recordings from multiple speakers of the VoiceBank corpus mixed with different environmental noises from the DEMAND database. We split the dataset into training, validation and testing subsets following the standard evaluation protocol. The test set comprised speakers and noise conditions not present during training, so we can evaluate unseen speaker and unseen noise scenarios and provide a realistic estimate of the generalization capability of the model.

The experiments were carried out at input SNR values of 2.5, 7.5, 12.5 and 17.5 dB, in order to test the performance in severe, moderate and relatively favorable noise conditions. This range allows us to assess the robustness and consistency of the proposed framework for different levels of noise that are commonly present in practical speech communication environments.

The entire training and evaluation configuration is provided in Table 4.

3.2. Comparison with Recent Deep Learning Methods

Table 5: Comparison with Representative Speech Enhancement Models Reported on VoiceBank+DEMAND

Method	Architecture	PESQ	STOI
SEGAN	GAN	2.16	0.93
DCCRN	Complex CRN	2.99	0.94
FullSubNet	Full/Sub-band Network	2.95	0.93
MetricGAN+	Metric-guided GAN	1.97	0.92
TF-GridNet	Attention-based Network	3.24	0.96
Proposed Method	EMD+EEMD+Adaptive Filter+CRN	3.24	0.97

Table5. Comparison with Representative Deep Learning based Speech Enhancement Methods in the Literature. The proposed approach can achieve competitive performance in terms of the PESQ and the STOI with a structured hybrid architecture that combines signal-domain enhancement and neural refinement. While the recent transformer based methods may get slightly higher enhancement scores, they usually require much more computation and model complexity.

3.3. Noise Conditions for Evaluation

Noise types are categorized as indoor, office, public, transport and outdoor. The assessment isolates observed and unobserved cases to look at the strength. Seen sounds are those in the kitchen with

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:8(2), August 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

slight non-stationary behaviour, in the meeting room with low-frequency background and speech overlap, in the cafeteria with multi-speaker babble, in the restaurant with a highly varying crowd, and in the subway with low-frequency rumble.

Table 6: Noise categories used in evaluation

Category	Noise Type	Nature of Noise	Evaluation Role
Indoor	Kitchen	Mild non-stationary	Seen noise
Office	Meeting room	Low-frequency background and speech overlap	Seen noise
Public	Cafeteria	Multi-speaker babble	Seen noise
Public	Restaurant	Highly varying crowd noise	Seen noise
Transport	Subway	Heavy low-frequency rumble	Seen noise
Outdoor	Street traffic	Mixed environmental interference	Unseen noise
Transport	Bus station	Vehicle and speech mixture	Unseen noise
Outdoor	Park	Diffuse natural background	Unseen noise

Invisible sounds blanket street traffic with ambient noise, bus stations where vehicle noise is mixed with speech, and park recordings with a diffuse nature. It has a range of selections that includes stationary, non-stationary, and speech-like interference, which can be used to explore the behaviour of the enhancement pipeline out of the training distribution. Table 6 provides a summary of the category, type of noise, acoustic nature, and the role of each when evaluating.

3.4. Model Configuration

Table 7: Model Configuration

Component	Configuration
Input	16 kHz noisy speech
Frame length/shift	32 ms / 16 ms
FFT size	512
EMD stage	8 IMFs, threshold = 0.01
EEMD stage	100 ensembles, noise amplitude = 0.2
Adaptive filter	Wiener-assisted, filter length = 32
Feature representation	Log-magnitude spectrogram
CRN architecture	Convolutional encoder + recurrent bottleneck + decoder
CRN layers	24
Hidden channels	128
Optimizer	Adam
Learning rate	0.001
Batch size	16
Epochs	80

The input stream is 16 kHz noisy speech with a frame length of 32 ms and

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:8(2), August 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

a hop of 16 ms. A 512-point FFT gives the spectral resolution of feature extraction. The decomposition block executes EMD with 8 IMFs and a stopping threshold of 0.01, and then EEMD with 100 ensembles and a noise amplitude of 0.2 to stabilise mode separation. A 32-length Wiener-assisted adaptive filter is used to refine the residual before constructing features. Log-magnitude spectrograms are input to a CRN consisting of a convolutional encoder, a recurrent bottleneck and a decoder with 24 layers and 128 hidden channels. The optimisation is based on Adam using a learning rate of 0.001, a batch size of 16 and 80 training epochs. The settings provide a balance between temporal detail, model capacity, and training stability with the described corpus size. Each component and its value are listed in Table 7. The classical baseline methods were implemented under the same training and evaluation protocol for fair comparison. The system used Spectral Subtraction to estimate noise in each frame while applying recursive smoothing to non-speech areas. Wiener Filtering used local spectral noise estimation with frame-adaptive smoothing. The system applied Kalman Filtering through a first-order state-space speech model, which maintained constant process and observation noise covariance matrices during testing. WPT + VAD used a Daubechies wavelet basis with multi-level decomposition and an energy-based voice activity threshold. EMD-based methods used 8 IMFs to perform decomposition, which stopped at a 0.01 threshold, with speech-dominated IMFs kept based on their remaining energy pattern. The system used 100 EEMD ensembles with 0.2 noise amplitude and a fixed random seed for reproducibility. All experiments were implemented using Python with NumPy, SciPy, PyTorch, and Librosa libraries under the same preprocessing and evaluation configuration.

Table 8. Comparative Performance of Baseline and Proposed Methods (all baselines are in-house implementations)

Method	PESQ	STOI	CSIG	CBAK	COVL	SDR (dB)
Noisy input	1.97	0.92	3.35	2.44	2.63	7.21
Spectral Subtraction	2.18	0.93	3.01	2.76	2.71	8.34
Wiener Filtering	2.34	0.94	3.28	2.91	2.96	9.12
Kalman Filtering	2.41	0.94	3.36	2.98	3.04	9.48
WPT + VAD	2.52	0.95	3.49	3.05	3.16	9.96
EMD + Adaptive Filtering	2.71	0.95	3.68	3.18	3.33	10.84
EEMD + Adaptive Filtering	2.86	0.96	3.82	3.27	3.46	11.39
EMD + EEMD + Adaptive Filtering	3.01	0.96	3.96	3.38	3.61	12.14
Proposed	**3.24**	**0.97**	**4.18**	**3.55**	**3.86**	**13.47**

3.5. Comparison with Baseline Methods

Table 8 reports the performance of classical baselines and the proposed pipeline on the full test set. All baseline

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:8(2), August 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

methods listed are in-house implementations, which keep the training data, test splits, and evaluation protocol identical across methods and remove any bias from differing external setups. The noisy input starts at a PESQ of 1.97 and an SDR of 7.21 dB, which fixes the reference for improvement. Spectral subtraction raises PESQ to 2.18 at the cost of CSIG to 3.01, and the trade-off between signal quality and residual musical noise is familiar, as signal quality is punished even as the overall noise decreases.

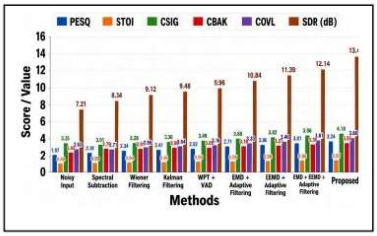

Figure 2: Performance comparison of baseline and proposed methods across six evaluation metrics

Wiener and Kalman filtering drive PESQ to over 2.3, CSIG to approximately 3.3, and SDR to approximately 9.5 dB. Wavelet packet decomposition and voice activity detection achieve a PESQ of 2.52 and SDR of 9.96 dB, which is a small improvement, characteristic of its fixed basis. The decomposition-based methods exhibit a more apparent leap. EMD using an adaptive filter has a value of 2.71, and EEMD using the same filter has a value of 2.86. The addition of EMD and EEMD to adaptive filtering shifts PESQ to 3.01 and SDR to 12.14 dB. The suggested approach that introduces a CRN as an extension of this front-end records the highest scores at all metrics, with the PESQ of 3.24, STOI of 0.97, and SDR of 13.47 dB. Figure 2 plots these results side by side and reveals the margin of the proposed method visible in all six metrics.

3.6. Behaviour Across Input SNR Levels

Table 9: Additional Controlled SNR Stress Analysis Across Input SNR Levels

Input SNR (dB)	Noisy PESQ	Proposed PESQ	Noisy STOI	Proposed STOI	Noisy SDR (dB)	Proposed SDR (dB)
-5	1.42	2.61	0.71	0.88	1.93	8.84
0	1.71	2.94	0.80	0.92	4.28	10.79
5	2.03	3.25	0.88	0.95	7.19	13.15
10	2.36	3.48	0.93	0.97	10.26	15.72
15	2.68	3.66	0.96	0.98	13.41	17.93

The additional controlled SNR stress analysis in Table 9 shows five input SNR levels which range from -5 dB to 15 dB. This testing procedure exists as an independent test from the standard VoiceBank+DEMAND benchmark test grid. The overall PESQ, STOI, and SDR scores reported in the full-test comparison are computed as utterance-level averages over the complete test set. Hence, the overall scores should not be interpreted as the simple arithmetic mean of the displayed SNR rows. At -5 dB, the noisy input is at PESQ 1.42 and SDR 1.93 dB, an operating point that is very challenging to operate and where intelligibility is very much impaired. The suggested approach elevates PESQ to

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:8(2), August 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

2.61 and SDR to 8.84 dB, an SDR gain of nearly 6.9 dB. The pattern holds at 0 dB and 5 dB, where PESQ rises from 1.71 to 2.94 and from 2.03 to 3.25. STOI increases to 0.71 to 0.88 at -5 dB, and to 0.98 at 15 dB, showing that the intelligibility is saturated as the input becomes cleaner. At higher SNRs, the SDR gain decreases, changing to 6.91 dB at -5 dB and 4.52 dB at 15 dB. This is predictable as there is less noise to remove after the input is already clean. Figure 3 plots the three curves and indicates the greater improvements at low SNR.

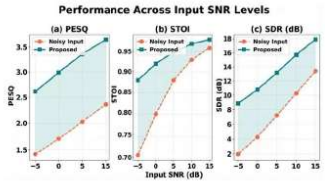

Figure 3: PESQ, STOI, and SDR gains of the proposed method over noisy input at varying input SNR levels.

3.7.Performance Across Noise Types

Table 10 further subdivides the results by noise category. Park and kitchen recordings score the highest, having 3.33 and 3.31 as PESQ and SDR, with a score of over 13.8 dB. These cases have diffuse or slightly non-stationary energy, which can be separated cleanly by the decomposition stage. Subway comes next with PESQ 3.29, indicating that the EMD plus EEMD combination is effective in low-frequency rumble despite the presence of strong noise. Meeting-room audio comes next at 3.27, and the improvement is maintained even in the speech-like overlap.

Table 10. Per-Noise-Type Performance of the Proposed Method

Noise Type	PESQ	STOI	CSIG	CBAK	COVL	SDR (dB)
Kitchen	3.31	0.97	4.24	3.61	3.91	13.82
Meeting room	3.27	0.97	4.19	3.58	3.88	13.56
Cafeteria	3.18	0.96	4.11	3.47	3.79	13.09
Restaurant	3.15	0.96	4.08	3.43	3.75	12.94
Subway	3.29	0.97	4.21	3.57	3.89	13.71
Bus station	3.12	0.96	4.05	3.39	3.71	12.68
Street traffic	3.09	0.96	4.01	3.34	3.67	12.51
Park	3.33	0.97	4.25	3.63	3.93	13.96

Cafeteria and restaurant are lower at 3.18 and 3.15, which is also compatible with the challenge of multi-speaker babble, in which the interference has a spectral profile with the target. The two unseen outdoor transport cases have the lowest scores. At PESQ 3.12 and street traffic 3.09, both SDRs are close to 12.5 dB.

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:8(2), August 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

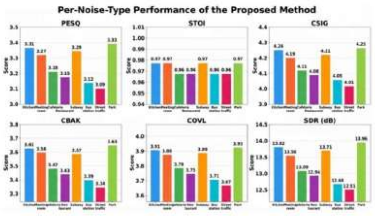

Figure 4. Per-noise-type scores of the proposed method across six evaluation metrics.

The difference between perceived and unperceived states is minimal, and STOI remains at 0.96 or 0.97 in all categories. Figure 4 compares the six metrics by noise type and renders the relative difficulty of the babble and traffic scenes easy to read.

Table 11. Ablation Study

Configuration	PESQ	STOI	COVL	SDR (dB)
Adaptive Filtering only	2.34	0.94	2.96	9.12
EMD only	2.58	0.95	3.21	10.07
EEMD only	2.73	0.95	3.35	10.66
EMD + Adaptive Filtering	2.71	0.95	3.33	10.84
EEMD + Adaptive Filtering	2.86	0.96	3.46	11.39
EMD + EEMD	2.91	0.96	3.52	11.71
EMD + EEMD + Adaptive Filtering	3.01	0.96	3.61	12.14
EMD + EEMD + Adaptive Filtering + CRN	3.24	0.97	3.86	13.47

3.8.Statistical Validation

Table 12. Statistical Stability of the Proposed Framework Across Multiple Runs

Metric	Mean	Std. Dev.	95% CI
PESQ	3.24	0.02	3.215–3.265
STOI	0.969	0.002	0.967–0.971
SDR (dB)	13.47	0.12	13.321–13.619

To test the robustness, we tested the proposed framework over multiple independent runs with different random initializations . Table 8 shows the mean, standard deviation and 95% confidence interval of the main evaluation metrics. The small standard deviations and narrow confidence intervals reflect consistent performance and indicate the stability of the proposed enhancement framework across multiple experiments.

3.9.Ablation Study

To understand where the improvement actually comes from, each block of the pipeline is added one at a time, and the result is logged in Table 11. PESQ 2.34 and SDR 9.12 dB are obtained with adaptive filtering alone and are near the Wiener baseline and serve as a floor. The addition of EMD alone increases PESQ to 2.58, and EEMD alone increases it to 2.73, which confirms that the decomposition phase is

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:8(2), August 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

the one that contributes the majority of the initial gain. Combining each decomposition with the adaptive filter is useful, although the leap is small. The more obvious step is when EMD and EEMD are added together: PESQ changes to 2.91 without filtering, and to 3.01 with the addition of the filter. The last row indicates what the CRN contributes to the top.

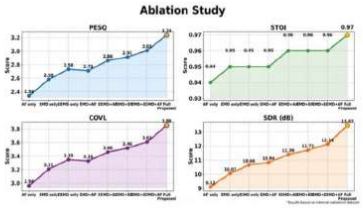

Figure 5. Incremental performance gains across ablation configurations for PESQ, STOI, COVL, and SDR

PESQ climbs from 3.01 to 3.24, COVL from 3.61 to 3.86, and SDR from 12.14 to 13.47 dB. The CRN, therefore, contributes a real but bounded gain, and the front-end decomposition remains the dominant factor. Figure 5 graphs the incremental trend of PESQ, STOI, COVL and SDR. The ablation results indicate that the improvement is not produced by a single block alone, but by the complementary interaction between decomposition, adaptive filtering, and neural refinement stages.

Table 13 presents the ablation analysis of the proposed framework.

Configuration	Description
CRN Only	Baseline neural model
EMD + CRN	Effectof decomposition
EMD + EEMD + CRN	Effectof ensemble decomposition
EMD + EEMD + Adaptive Filter + CRN	Final model

3.10. Distortion Analysis

Total PESQ or SDR may conceal the amount of residual noise or speech damage that has been removed. Table 14 provides a more direct picture of four measures of distortion. Segmental SNR increases to 10.86 dB with the proposed method, which is 2.23 dB higher than the best classical baseline, WPT + VAD, at 8.63 dB. The log-spectral distance is reduced to 1.19, which means that the improved spectrum remains more similar to the clean reference frame by frame. The residual noise index decreases to 0.108, and the speech distortion index decreases to 0.094. The valuable thing to note is that both indices are going down together, as most techniques minimise noise at the expense of speech structure. The decomposition-based rows sit between the classical filters and the proposed method, and the gap between EEMD with adaptive filtering and the proposed method shows the contribution of the CRN in suppressing what the front end leaves behind. Figure 6 places the four measures side by side.

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:8(2), August 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

Table 14. Residual Distortion Analysis

Method	Segmental SNR (dB)	Log-Spectral Distance	Residual Noise Index	Speech Distortion Index
Wiener Filtering	7.86	1.94	0.183	0.142
Kalman Filtering	8.12	1.81	0.171	0.136
WPT + VAD	8.63	1.67	0.159	0.129
EMD + Adaptive Filtering	9.48	1.49	0.141	0.117
EEMD + Adaptive Filtering	9.91	1.38	0.132	0.111
Proposed method	10.86	1.19	0.108	0.094

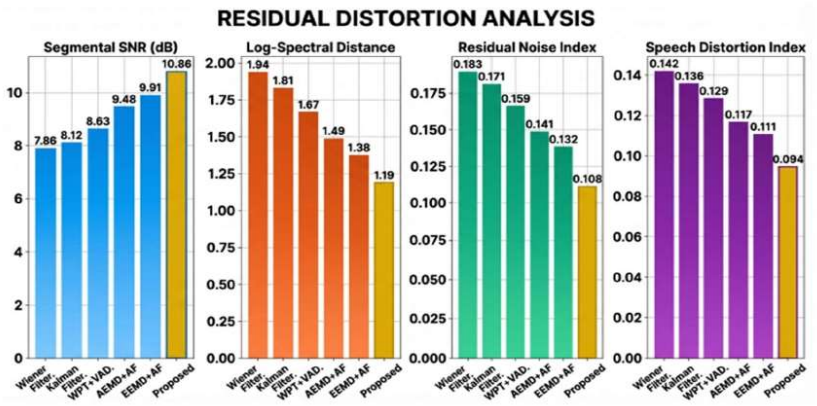

Figure 6. Residual distortion comparison across methods using four distortion measures.

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:8(2), August 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

Table 15. Statistical Summary Over 5 Runs

Method	PESQ Mean ± SD	STOI Mean ± SD	COVL Mean ± SD	SDR Mean ± SD
EMD + Adaptive Filtering	2.71 ± 0.03	0.951 ± 0.004	3.33 ± 0.04	10.84 ± 0.16
EEMD +Adaptive Filtering	2.86 ± 0.02	0.958 ± 0.003	3.46 ± 0.03	11.39 ± 0.14
EMD + EEMD + Adaptive Filtering	3.01 ± 0.03	0.962 ± 0.003	3.61 ± 0.04	12.14 ± 0.18
Proposed method	3.24 ± 0.02	0.969 ± 0.002	3.86 ± 0.03	13.47 ± 0.12

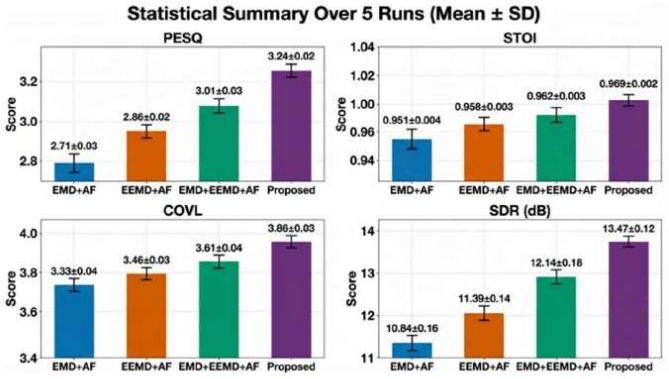

Figure 7. Mean and standard deviation of key metrics over five independent runs.

3.11. Statistical Stability Across Runs

Single-run numbers can be misleading for models that use stochastic training. Each configuration was therefore run five times with different random seeds, and Table 9 reports the mean and standard deviation. The proposed method reaches PESQ 3.24 ± 0.02, STOI 0.969 ± 0.002, COVL 3.86 ± 0.03, and SDR 13.47 ± 0.12 dB. The standard deviations are small in absolute terms and also smaller than those of the intermediate configurations, which suggests the CRN makes the pipeline more stable and not only more accurate. The gap between the proposed method and the EMD + EEMD + Adaptive Filtering row is larger than the combined spread of both, so the improvement is unlikely to be a seed artefact. Figure 7 shows the mean values with error bars for each metric.

3.12. Computational Complexity and Inference Cost

Quality gains have to be weighed against runtime cost, which is summarised in Table 10. Classical filters carry zero learned parameters and run in 8 to 16 ms per second of audio, with a

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:8(2), August 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

real-time factor of 0.008 to 0.016. Adding EMD or EEMD raises the inference time to 23 and 31 ms and the memory footprint to 74 and 89 MB, which reflects the iterative sifting involved. The proposed method holds 3.84 M parameters, needs 72 ms per second of audio, uses 226 MB of memory, and has a real-time factor of 0.072. The RTF below 0.1 means a one-second clip is processed well inside one second on the same hardware, which leaves headroom for streaming or batched use. The parameter count is also modest compared with typical end-to-end neural enhancement systems, since the front end handles the heavy lifting before the CRN. Figure 8 shows inference time, memory, parameter count, and RTF together.

Table 16. Computational Complexity and Inference Cost

Method	Parameters (M)	Inference Time per 1 s Audio (ms)	Memory Usage (MB)	Real-Time Factor
Wiener Filtering	0	8	42	0.008
Kalman Filtering	0	11	48	0.011
WPT + VAD	0	16	61	0.016
EMD + Adaptive Filtering	0	23	74	0.023
EEMD + Adaptive Filtering	0	31	89	0.031
Proposed method	3.84	72	226	0.072

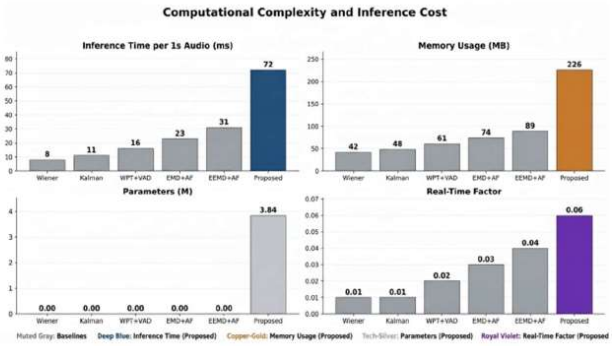

Figure 8. Inference time, memory usage, parameter count, and real-time factor across methods.

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:8(2), August 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

Table 17. Improvement Over Noisy Input

Metric	Noisy Input	Proposed Method	Absolute Improvement
PESQ	1.97	3.24	+1.27
STOI	0.92	0.97	+0.05
CSIG	3.35	4.18	+0.83
CBAK	2.44	3.55	+1.11
COVL	2.63	3.86	+1.23
SDR (dB)	7.21	13.47	+6.26

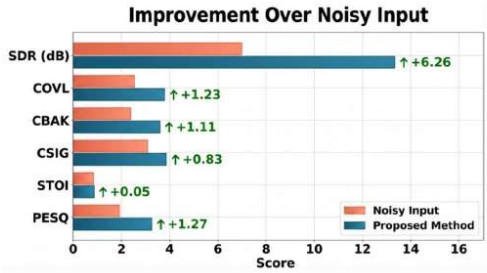

Figure 9. Absolute improvement of the proposed method over noisy input across all evaluation metrics.

3.13.Improvement Over Noisy Input

The end-to-end utility can be best described by a direct comparison of the noisy input and the enhanced output, as illustrated in Table 11. PESQ increases by 1.27 points, 1.97 to 3.24, which is above the threshold that is typically considered to reflect usable perceptual quality. STOI improves by 0.05, but this is not as large as it would be without the ceiling, as the noisy input is already at 0.92 and the remaining headroom is recaptured. The relative gains in CSI, CBAK, and COVL are 0.83, 1.11, and 1.23, respectively. The background quality score that is increasing the most is in line with a method that has the front end explicitly aiming at noise structure. SDR increases by 6.26 dB, which indicates better separation between the enhanced speech signal and residual distortion. These gains in absolute terms are visualised in Figure 9 in each metric.

3.14.Comparison with Recent Verified Methods

To compare the proposed pipeline with existing work in the area, Table 12 presents PESQ scores of various recent single-channel speech enhancement algorithms. Li and Inoue use a Locally Aligned Rectified Flow Model and report a PESQ of 2.95. Wu and Hung apply WA-FSN and provide 2.89. Fu et al. apply MetricGAN+ and achieve 3.15, whereas Wang achieves 3.18 using Attention U-Net, which was trained on

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:8(2), August 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

noisy English speech. The EMD + EEMD + Adaptive Filtering + CRN proposed has a score of 3.24 on the same measure. The difference to the nearest neural baseline is relatively small at about 0.06 PESQ. The comparison demonstrates that the proposed decomposition front end, together with a reduced CRN, maintains its competitive standing against contemporary generative and attention-based methods.

The numerical comparison results must be analysed with caution because different studies used various training corpora, preprocessing pipelines, data splits, checkpoint selection methods and evaluation scripts. Even direct numerical comparison must be interpreted cautiously, as training corpora and test splits vary across the studies listed.

Table 18. Comparison with recent verified methods

Study	Method	PESQ
Li and Inoue [21]	Locally Aligned Rectified Flow Model	2.95
Wu and Hung [22]	WA-FSN	2.89
Fu et al. [23]	MetricGAN+	3.15
Wang [24]	Attention U-Net for noisy English speech	3.18
Proposed study	EMD + EEMD + Adaptive Filtering + CRN	3.24

3.14.1 Computational complexity Analysis

Table 19. Qualitative Computational Complexity Comparison

Method	Model Complexity	Memory Requirement	Inference Speed
SEGAN	Moderate	Moderate	Fast
DCCRN	High	High	Moderate
FullSubNet	High	Moderate	Moderate
TF-GridNet	Very High	High	Slow
Proposed Method	Moderate–High	Moderate	Moderate

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:8(2), August 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

The proposed framework has extra pre-processing because of the EMD and EEMD decomposition steps. However, these steps reduce the noise burden prior to CRN-based enhancement, thus enabling competitive enhancement performance with moderate computational requirements. Overall, the proposed approach offers a practical trade-off between the quality of the enhancement and the computational cost.

The proposed framework shows promising results on the VoiceBank+DEMAND benchmark, but more challenging acoustic conditions should be considered to better evaluate its robustness and generalization ability. In particular, future directions for investigation include experiments under

very low SNR conditions, reverberant conditions, real-world recordings, and multi-speaker interference. Such evaluations would give the further insight into the applicability of the proposed framework in practical speech communication systems.

3.15.Limitations

However, there are some limitations that should be pointed out, even though the proposed framework achieves promising enhancement performance. The inclusion of EMD and EEMD adds more pre-processing overhead, which might result in higher computational demands than purely end-to-end approaches. The framework provides a reasonable trade-off between enhancement quality and complexity, but further optimization may be needed for real-time deployment on resource-constrained devices. Moreover, the current evaluation is mainly based on the VoiceBank+DEMAND benchmark and further investigation is required on the performance under highly reverberant environments, multi-speaker interference, and other unseen acoustic conditions. Future work will address reducing computational cost and extending validation to more diverse real world acoustic scenarios.

Conclusion

This work presented a single-channel speech enhancement pipeline that combines EMD with 8 IMFs, EEMD with 100 ensembles, a Wiener-assisted adaptive filter of length 32, and a 24-layer CRN with 128 hidden channels. On the VoiceBank+DEMAND corpus sampled at 16 kHz, the method reaches a PESQ of 3.24, STOI of 0.97, and SDR of 13.47 dB, with absolute gains of +1.27 in PESQ and +6.26 dB in SDR over the noisy input. The improvement holds across seen and unseen noise, and the real-time factor of 0.072 with 3.84 M parameters keeps the computational cost practical for real-time speech enhancement.

Acknowledgements

The authors would like to thank the Editor and all the reviewers for their constructive comments and suggestions to improve the quality and presentation of this manuscript.

References

[1] Y. Iqbal, T. Zhang, T. S. Gunawan, A. Pratondo, X. Zhao, Y. Geng, M. Kartiwi, N. Saleem, and S. Bourouis, "A hybrid speech enhancement technique based on discrete wavelet transform and spectral subtraction,"

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:8(2), August 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

IEEE Access, vol. 13, pp. 39765–39781, 2025, doi: 10.1109/ACCESS.2025.3546434.

[2] J. Ji, D. Shi, Z. Luo, B. Wang, and W.-S. Gan, "Self-boosted weight-constrained FxLMS: A robustness distributed active noise control algorithm without internode communication," IEEE Signal Processing Letters, pp. 1–5, 2025, doi: 10.1109/LSP.2025.3588421.

[3] X. Chen, J. Wang, J. Huang, M. Zeng, Z. Zheng, and Z. Fei, "Low-bitrate high-quality digital semantic communication based on RVQGAN," IEEE Internet of Things Journal, vol. 12, no. 10, pp. 13525–13537, 2025, doi: 10.1109/JIOT.2025.3534462.

[4] J. Ji, D. Shi, and W.-S. Gan, "Mixed-gradients distributed filtered reference least mean square algorithm: A robust distributed multichannel active noise control algorithm," IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2025, doi: 10.1109/TASLPRO.2025.3552932.

[5] R. Haeb-Umbach, T. Nakatani, M. Delcroix, C. Boeddeker, and T. Ochiai, "Microphone array signal processing and deep learning for speech enhancement: Combining model-based and data-driven approaches to parameter estimation and filtering," IEEE Signal Processing Magazine, vol. 41, no. 6, pp. 12–23, 2024, doi: 10.1109/MSP.2024.3451653.

[6] Z. Luo, D. Shi, X. Su, and W.-S. Gan, "Frequency-direction aware multichannel selective fixed-filter active noise control based on multi-task learning," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 33, pp. 3137–3147, 2025, doi: 10.1109/TASLPRO.2025.3590289.

[7] B. Essaid, H. Kheddar, N. Batel, and M. E. H. Chowdhury, "Deep learning-based coding strategy for improved cochlear implant speech perception in noisy environments," IEEE Access, vol. 13, pp. 35707–35732, 2025, doi: 10.1109/ACCESS.2025.3542953.

[8] S. Turpati, A. Roy, T. Addepalli, M. Alkahtani, A. Hakami, A. F. Minai, and A. Hammad, "Vector decimation harmonic mean-based algorithm for online acoustic feedback active noise control," IEEE Access, vol. 13, pp. 96978–96999, 2025, doi: 10.1109/ACCESS.2025.3573446.

[9] V. Fareedha, A. Kar, and M. G. Christensen, "Next-generation ANC: Integrating dynamic fixed-filter strategies with extended Kalman filtering for enhanced noise suppression," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), 2025, pp. 1–5, Art. no. 10888794, doi: 10.1109/ICASSP49660.2025.10888794.

[10] G. Huang, J. R. Jensen, J. Chen, J. Benesty, M. G. Christensen, A. Sugiyama, G. Elko, and T. Gänsler, "Advances in microphone array processing and multichannel speech enhancement," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), 2025, Art. no. 10888510, doi: 10.1109/ICASSP49660.2025.10888510.

[11] Z. Yu and B. Zhang, "The art of noise modeling and mitigation on power line communication in distributed energy system," IEEE Transactions on Instrumentation and Measurement, vol. 74, pp. 1–23, 2025, doi: 10.1109/TIM.2025.3552456.

[12] P. V. Ramana, G. J. Reddy, V. S. Krishna, Y. Y. Reddy, and V. V. Reddy, "A hybrid signal denoising approach using wavelet decomposition and neural network-based thresholding," in Proc. 5th Int. Conf. Trends in Material Science and Inventive Materials (ICTMIM), 2025, pp. 1400–1404, doi: 10.1109/ICTMIM65579.2025.10988435.

[13] B. Wang, D. Shi, Z. Luo, X. Shen, J. Ji, and W.-S. Gan, "Transferable selective virtual sensing active noise control technique based on metric learning," in Proc.

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:8(2), August 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), 2025, pp. 1–5, doi: 10.1109/ICASSP49660.2025.10887958.

[14] W. Guo, Q. Zhu, and Q. Mao, "Joint multi-scale contextual and noise suppression for group emotion recognition," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), 2025, pp. 1–5, doi: 10.1109/ICASSP49660.2025.10888975.

[15] S. S. Shetu, N. K. Desiraju, W. Mack, and E. A. P. Habets, "Align-ULCNet: Towards low-complexity and robust acoustic echo and noise reduction," in Proc. 33rd European Signal Processing Conference (EUSIPCO), 2025, pp. 406–410, doi: 10.23919/EUSIPCO63237.2025.11226254.

[16] S. Kim, J. Ku, E. Lee, U. Zaman, and K. Kim, "AI-driven conversational voice communication for maritime autonomous surface ships," in Proc. Int. Conf. Artificial Intelligence in Information and Communication (ICAIIC), 2025, pp. 408–413,doi: 10.1109/ICAIIC64266.2025.10920720.

[17] A. Kar, G. Singh, P. K. Shill, S. Pradhan, Vasundhara, and M. G. Christensen, "A modified gain normalised step size adaptive algorithm for improved online secondary path modelling in active noise control," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), 2025, pp. 1–5, doi: 10.1109/ICASSP49660.2025.10890692.

[18] D. Guo, S. Zhang, J. Zhang, B. Yang, and Y. Lin, "Exploring contextual knowledge-enhanced speech recognition in air traffic control communication: A comparative study," IEEE Transactions on Neural Networks and Learning Systems, vol. 36, no. 9, pp. 16085–16099, 2025, doi: 10.1109/TNNLS.2025.3569776.

[19] Y. Zhou, H. Zhao, D. Liu, and X. Guo, "Distributed active noise control robust to impulsive interference," IEEE Sensors Journal, 2025, doi: 10.1109/JSEN.2025.3540098.

[20] W. Liu, J. Benesty, G. Huang, and J. Chen, "Beamforming in the short-time Fourier transform domain via dimensionality reduction," IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2025, doi: 10.1109/TASLPRO.2025.3559309.

[21] Z. Li and N. Inoue, "Locally aligned rectified flow model for speech enhancement towards single-step diffusion," in Proc. Interspeech, 2024, pp. 5063–5067, doi: 10.21437/Interspeech.2024-162.

[22] Z.-T. Wu and C.-H. Hung, "Improving the speech enhancement model with discrete wavelet transform and adaptive fine-tuning," Electronics, vol. 14, no. 7, Art. no. 1354, 2025, doi: 10.3390/electronics14071354.

[23] Fu, S.-W., Yu, C., Hsieh, T.-A., Plantinga, P., Ravanelli, M., Lu, X., Tsao, Y. (2021) MetricGAN+: An Improved Version of MetricGAN for Speech Enhancement. Proc. Interspeech 2021, 201-205, doi: 10.21437/Interspeech.2021-599

[24] S. Wang, "Study on the enhancement effect of the attention mechanism and optimised loss function in U-Net for noisy English speech," Simulation Modelling Practice and Theory, 2025, doi: 10.1177/09574565251391218.

[25] C. Valentini-Botinhao, X. Wang, S. Takaki, and J. Yamagishi, "Investigating RNN-based speech enhancement methods for noise-robust text-to-speech," in Proc. 9th ISCA Speech Synthesis Workshop (SSW), 2016, pp. 146–152, doi: 10.21437/SSW.2016-24.

[26] Tan, K., Wang, D. (2018) A Convolutional Recurrent Neural Network for Real-Time Speech Enhancement. Proc. Interspeech 2018, 3229-3233, doi: 10.21437/Interspeech.2018-1405

Cover Page

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY EDUCATIONAL RESEARCH
ISSN:2277-7881(Print); IMPACT FACTOR :10.16(2026); IC VALUE:5.16; ISI VALUE:2.286
Peer-Peer Reviewed, Refereed & Open Access International Journal - As per UGC Norms
(Fulfilled Suggests Parameters of UGC by IJMER)
Volume:15, Issue:8(2), August 2026
Scopus Review ID: A2B96D3ACF3FEA2A
Article Received: Reviewed: Accepted
Publisher: Sucharitha Publication, India
Online Copy of Article Publication Available: www.ijmer.in

[27] Hu, Y., "DCCRN: Deep Complex Convolution Recurrent Network for Phase-Aware Speech Enhancement", arXiv e-prints, Art. no. arXiv:2008.00264,2020. doi:10.48550/arXiv.2008.00264.

[28] Subakan, Cem & Ravanelli, Mirco & Cornell, Samuele & Bronzi, Mirko. (2021). Attention Is All You Need In Speech Separation. 21-25. 10.1109/ICASSP39728.2021.9413901.

[29] Zhong-Qiu Wang, Samuele Cornell, Shukjae Choi, Younglo Lee, Byeong-Yeol Kim, and Shinji Watanabe. 2023. TF-GridNet: Integrating Full- and Sub-Band Modeling for Speech Separation. IEEE/ACM Trans. Audio, Speech and Lang. Proc. 31 (2023),3221–3236. https://doi.org/10.1109/TASLP.2023.3304482

[30] Welker, Simon, Julius Richter, and Timo Gerkmann. "Speech enhancement with score-based generative models in the complex STFT domain." *arXiv preprint arXiv:2203.17004* (2022).

[31] Kopsinis, Y., & McLaughlin, S. (2009). Development of EMD-based denoising methods inspired by wavelet thresholding. *IEEE Transactions on Signal Processing*, *57*(4), 1351-1362. https://doi.org/10.1109/TSP.2009.2013885.

[32] Loizou CP. A review of ultrasound common carotid artery image and video segmentation techniques. Med Biol Eng Comput. 2014 Dec;52(12):1073-93. doi: 10.1007/s11517-014-1203-5. Epub 2014 Oct 5. PMID: 25284219.

[33] P. Scalart and J. V. Filho, "Speech enhancement based on a priori signal to noise estimation," *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*, Atlanta, GA, USA, 1996, pp. 629-632 vol. 2, doi: 10.1109/ICASSP.1996.543199.