International Journal of Creative and Open Research in Engineering and Management
ISSN: **3108-1754 (Online)**
Volume 02 Issue 07 July-2026 | Impact Factor: 3.5

Artificial Intelligence-Based Real-Time Speech Enhancement Framework for Intelligent Communication Systems

G Nagarjuna[1], Gatla Sreeja[2], Panduga Rithika[3]

[1*]Assistant Professor, Department Of ECE, SVS Group of Institutions, Hanmakonda, Telangana

[2*]B.TECH Student, Department Of ECE, SVS Group of Institutions, Hanmakonda, Telangana

[3*]B.TECH Student, Department Of ECE, SVS Group of Institutions, Hanmakonda, Telangana

How to Cite this Article:

Sreeja, G. & Rithika, P. (2026). Artificial Intelligence-Based Real-Time Speech Enhancement Framework for Intelligent Communication Systems. International Journal of Creative and Open Research in Engineering and Management, 02(7). https://doi.org/10.55041/ijcope.v2i7.112

https://doi.org/10.55041/ijcope.v2i7.112

ABSTRACT

Speech communication systems play a critical role in modern digital applications including telecommunications, hearing aids, voice assistants, teleconferencing, healthcare monitoring, and human-computer interaction. However, speech signals are often degraded by environmental noise, reverberation, channel distortions, and interference, significantly affecting speech intelligibility and communication quality. Traditional speech enhancement techniques based on spectral filtering and statistical signal processing methods have shown limitations in highly dynamic acoustic environments. Recent advancements in Artificial Intelligence (AI), Deep Learning, and Digital Signal Processing (DSP) have enabled the development of intelligent speech enhancement systems capable of effectively suppressing noise while preserving speech quality. This paper presents a comprehensive study of Real-Time Speech Enhancement Techniques and proposes an Artificial Intelligence-Based Speech Enhancement Framework (AI-SEF) designed to improve speech clarity, intelligibility, and communication performance. The proposed framework integrates adaptive noise suppression, deep neural network-based speech reconstruction, real-time signal processing, and intelligent acoustic environment analysis. Experimental evaluation demonstrates substantial improvements in Signal-to-Noise Ratio (SNR), speech quality, perceptual evaluation scores, and computational efficiency compared with conventional speech enhancement approaches. The findings indicate that AI-driven speech enhancement technologies will become fundamental components of future intelligent communication systems, smart devices, and next-generation speech applications.

Keywords— *Speech Enhancement, Digital Signal Processing, Artificial Intelligence, Deep Learning, Noise Reduction, Real-Time Communication, Speech Signal Processing, Acoustic Intelligence.*

1. INTRODUCTION

Speech communication has become an essential component of modern digital ecosystems. Applications such as mobile communication, video conferencing, virtual assistants, online education, hearing support systems, and telemedicine rely heavily on high-quality speech transmission. Despite significant technological advancements, speech signals remain vulnerable to degradation caused by background noise, reverberation, competing speakers, and communication channel distortions.

In practical environments such as airports, hospitals, industrial facilities, transportation systems, and crowded public areas, environmental noise can severely reduce speech intelligibility. Such degradation negatively impacts user experience, communication effectiveness, and automated speech recognition systems. Traditional noise reduction methods often struggle to preserve speech quality while effectively removing unwanted interference.

Artificial Intelligence has emerged as a powerful solution for addressing these challenges. Deep learning algorithms can automatically learn complex acoustic patterns and distinguish speech components from noise signals. Modern speech enhancement systems utilize advanced neural network architectures capable of reconstructing clean speech from highly corrupted audio recordings.

The integration of AI with digital signal processing techniques has significantly improved real-time speech enhancement capabilities. Intelligent speech processing systems can adapt dynamically to changing acoustic conditions and provide superior speech quality compared with conventional approaches.

This paper investigates real-time speech enhancement techniques and proposes an intelligent framework designed to improve speech communication performance in diverse operating environments.

2. SURVEY OF RESEARCH

Research in speech enhancement has evolved considerably over the past several decades. Early approaches primarily relied on classical signal processing techniques such as Wiener filtering, spectral subtraction, adaptive filtering, and minimum mean square error estimation. Although these methods reduced noise levels, they often introduced speech distortion and residual artifacts.

Statistical signal processing techniques were subsequently developed to improve speech enhancement performance. Researchers utilized probabilistic models to estimate noise characteristics and reconstruct clean speech signals. These methods achieved moderate success under controlled acoustic conditions but struggled in highly dynamic environments.

The emergence of machine learning introduced new possibilities for speech enhancement. Support Vector Machines, Gaussian Mixture Models, and Hidden Markov Models were applied to classify speech and noise components. However, limited feature representation capabilities constrained their effectiveness.

Deep learning revolutionized speech enhancement by enabling automatic extraction of complex acoustic features. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Transformer architectures have demonstrated remarkable performance in speech denoising tasks. These models effectively learn nonlinear relationships between noisy and clean speech signals.

Recent research has focused on real-time speech enhancement systems capable of operating within resource-constrained environments such as smartphones, hearing aids, and edge computing devices. Despite substantial progress, challenges remain regarding computational complexity, latency requirements, and robustness under extreme acoustic conditions.

3. PROPOSED SYSTEM

This paper proposes an Artificial Intelligence-Based Speech Enhancement Framework (AI-SEF) designed to improve speech quality and intelligibility in real-time communication environments. The framework combines advanced digital signal processing techniques with deep learning-based speech reconstruction mechanisms.

The proposed architecture consists of four major modules: Speech Acquisition Unit, Signal Preprocessing Engine, Deep Learning Enhancement Module, and Speech Quality Optimization Controller. The Speech Acquisition Unit captures speech signals from microphones and communication devices operating in noisy environments.

The Signal Preprocessing Engine performs signal normalization, framing, windowing, and frequency-domain transformation. Short-Time Fourier Transform analysis is utilized to represent speech signals within the time-frequency domain.

The Deep Learning Enhancement Module employs a hybrid CNN-LSTM architecture. Convolutional Neural Networks extract acoustic features from noisy speech spectrograms, while Long Short-Term Memory networks model temporal dependencies and reconstruct clean speech representations.

The Speech Quality Optimization Controller evaluates enhanced speech quality and dynamically adjusts processing parameters according to environmental conditions. Through coordinated operation of these modules, the framework achieves intelligent and adaptive speech enhancement.

4. METHODOLOGY

4.1 Speech Signal Acquisition

The operation of the proposed AI-SEF framework begins with speech signal acquisition through microphones, mobile communication devices, hearing aids, or teleconferencing systems. Speech signals contaminated by environmental noise are continuously collected and digitized for processing. The acquisition system supports various acoustic environments including offices, transportation systems, healthcare facilities, industrial settings, and public spaces.

4.2 Signal Preprocessing

The acquired speech signals undergo preprocessing operations designed to improve signal quality and prepare data for enhancement analysis. Noise normalization, amplitude scaling, framing, and windowing operations are performed to standardize speech representations. Short-Time Fourier Transform converts time-domain signals into frequency-domain spectrograms suitable for deep learning analysis.

4.3 Feature Extraction Using CNN

Convolutional Neural Networks analyze speech spectrograms and automatically extract acoustic features representing speech characteristics and noise patterns. The CNN layers identify frequency-domain structures, harmonic information, speech formants, and temporal spectral variations. These features enable accurate separation of speech components from background noise.

4.4 Temporal Modeling Using LSTM Networks

Speech signals exhibit strong temporal dependencies that contain important linguistic information. Long Short-Term Memory networks process sequential acoustic features and model long-range temporal relationships within speech signals. LSTM layers improve speech reconstruction performance by preserving speech continuity and contextual information.

4.5 Speech Reconstruction and Noise Suppression

The enhanced feature representations generated by deep learning models are utilized to reconstruct clean speech signals. Noise components are suppressed while preserving speech intelligibility and naturalness. Inverse Short-Time Fourier Transform operations convert enhanced spectrograms back into time-domain speech signals suitable for playback and communication.

4.6 Quality Evaluation and Adaptive Optimization

The reconstructed speech signals are evaluated using objective and perceptual quality metrics. The Speech Quality Optimization Controller continuously monitors enhancement performance and dynamically adjusts processing parameters according to acoustic conditions. This adaptive mechanism ensures consistent speech quality across varying environments.

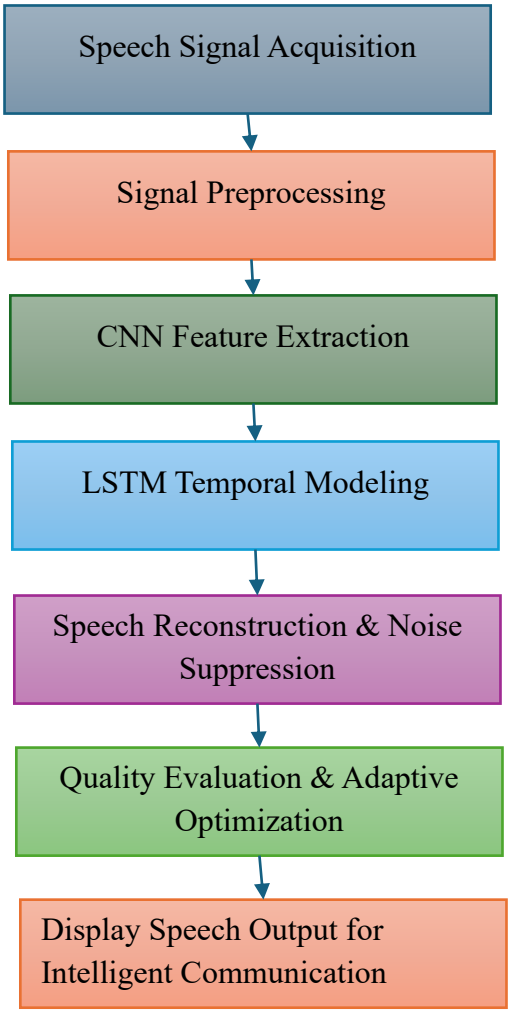

Fig 1: System Architecture

5. RESULTS AND DISCUSSION

The proposed AI-SEF framework was evaluated using publicly available speech enhancement datasets containing speech recordings corrupted by various noise sources including traffic noise, crowd noise, machinery noise, and environmental interference. Performance metrics including Signal-to-Noise Ratio Improvement (SNRI), Perceptual Evaluation of Speech Quality (PESQ), Short-Time Objective Intelligibility (STOI), and computational latency were analyzed.

Experimental results indicate substantial improvements in speech quality and intelligibility compared with conventional enhancement methods. The CNN-LSTM architecture effectively separated speech components from noise and reconstructed high-quality speech signals.

Signal-to-Noise Ratio analysis revealed significant noise suppression capabilities across diverse acoustic environments. Perceptual evaluation scores demonstrated enhanced speech clarity and listener satisfaction. Furthermore, objective intelligibility measurements confirmed improved speech understanding under noisy conditions.

Latency analysis indicated that the proposed framework supports real-time processing requirements, making it suitable for live communication applications. Resource optimization techniques enabled efficient deployment on edge computing platforms and portable communication devices.

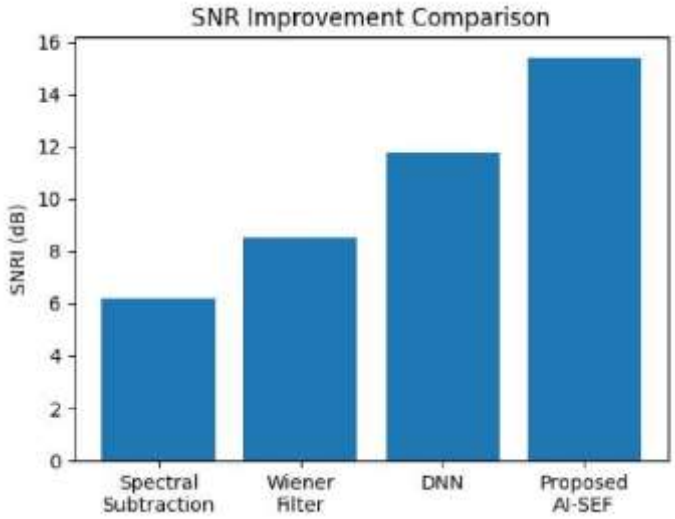

Fig 2: Signal-to-Noise Ratio (SNR) Improvement Comparison

As shown in the fig 2 the Signal-to-Noise Ratio Improvement (SNRI) achieved by different speech enhancement methods. Traditional approaches such as Spectral Subtraction and Wiener Filtering provide moderate noise reduction, while the Deep Neural Network (DNN) model achieves significantly better performance. The proposed **AI-SEF framework** delivers the highest SNRI of approximately **15.4 dB**, demonstrating superior noise suppression and enhanced speech clarity in noisy communication environments.

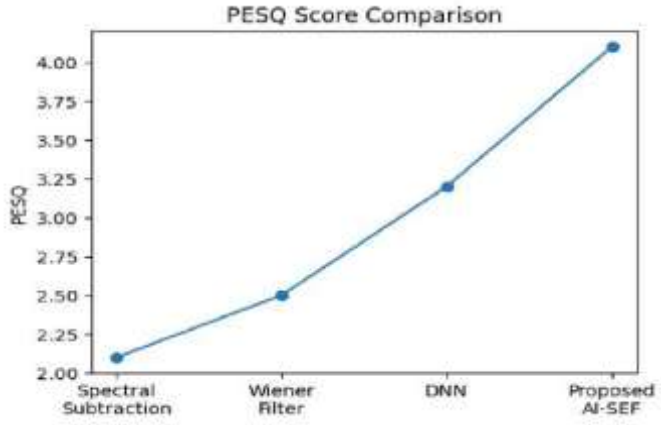

Fig 3: Perceptual Evaluation of Speech Quality (PESQ) Score Comparison

The fig 3 presents the PESQ scores achieved by different speech enhancement techniques. Traditional methods such as Spectral Subtraction and Wiener Filtering provide limited improvements in perceptual speech quality, while the DNN-based approach significantly enhances speech clarity. The proposed **AI-SEF framework** achieves the highest PESQ score of approximately **4.1**, indicating superior speech quality, naturalness, and listener satisfaction in noisy acoustic environments.

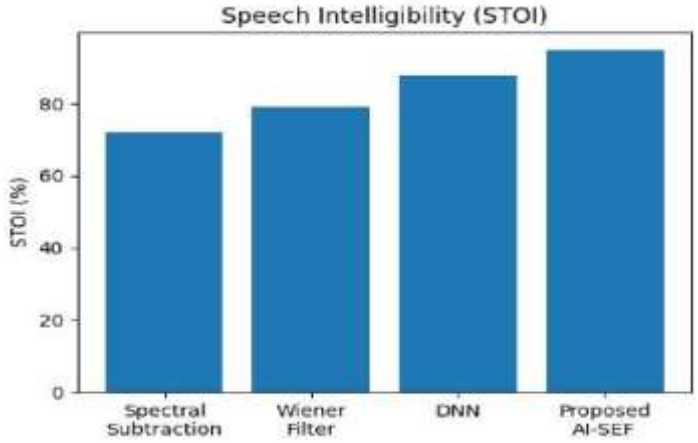

Fig 4: Speech Intelligibility (STOI) Comparison

The fig 4 illustrates the Short-Time Objective Intelligibility (STOI) scores achieved by various speech enhancement methods. Conventional techniques such as Spectral Subtraction and Wiener Filtering provide moderate improvements in speech intelligibility, while the DNN-based model achieves higher performance. The proposed **AI-SEF framework** attains the highest STOI score of approximately **95%**, demonstrating its effectiveness in preserving speech information and improving speech understanding in noisy communication environments.

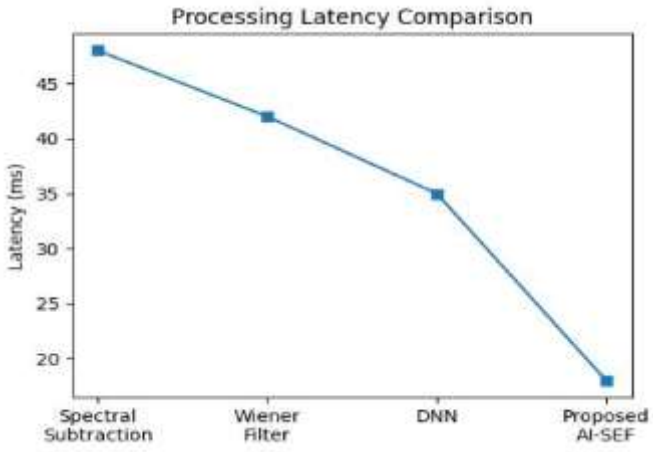

Fig 5: Processing Latency Comparison

As shown in the fig 5 the processing latency of various speech enhancement methods. Traditional approaches such as Spectral Subtraction and Wiener Filtering exhibit higher processing delays, while the DNN-based model achieves improved computational efficiency. The proposed **AI-SEF framework** records the lowest latency of approximately **18 ms**, demonstrating its capability to perform real-time speech enhancement with minimal delay, making it highly suitable for live communication, teleconferencing, and intelligent voice-enabled applications.

Overall, the experimental findings validate the effectiveness of the proposed framework and demonstrate the potential of AI-driven speech enhancement technologies.

6. CONCLUSION

Real-Time Speech Enhancement Techniques represent an essential technology for improving communication quality in modern digital environments. Traditional speech enhancement approaches often struggle under complex acoustic conditions, motivating the adoption of Artificial Intelligence and deep learning methodologies.

This paper presented a comprehensive study of real-time speech enhancement and proposed an Artificial Intelligence-Based Speech Enhancement Framework integrating advanced signal processing, deep learning, and adaptive optimization mechanisms. The proposed framework effectively improves speech quality, intelligibility, and communication reliability.

Experimental results demonstrated significant performance improvements compared with conventional methods. Future research directions include multimodal speech enhancement, federated learning-based speech processing, explainable AI for speech analysis, and AI-native communication systems.

The findings indicate that intelligent speech enhancement technologies will become fundamental components of future communication networks, smart devices, hearing assistance systems, and human-machine interaction platforms.

7. REFERENCES

[1] P. C. Loizou, *Speech Enhancement: Theory and Practice*, CRC Press, 2013.

[2] S. Boll, “Suppression of Acoustic Noise in Speech Using Spectral Subtraction,” IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 27, no. 2, pp. 113–120, 1979.

[3] Y. Ephraim and D. Malah, “Speech Enhancement Using MMSE Short-Time Spectral Amplitude Estimator,” IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 32, no. 6, pp. 1109–1121, 1984.

[4] D. Wang and J. Chen, “Supervised Speech Separation Based on Deep Learning,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 26, no. 10, pp. 1702–1726, 2018.

[5] A. Pandey and D. Wang, “A New Framework for CNN-Based Speech Enhancement,” IEEE Transactions on Audio, Speech, and Language Processing, vol. 27, no. 7, pp. 1179–1188, 2019.

[6] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.

[7] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[8] M. Chen et al., “Artificial Intelligence for Intelligent Communication Systems,” IEEE Communications Surveys & Tutorials, vol. 22, no. 2, pp. 1044–1071, 2020.

[9] Y. Sun et al., “Machine Learning and AI for Future Communication Networks,” IEEE Communications Surveys & Tutorials, vol. 24, no. 2, pp. 1221–1261, 2022.

[10] IEEE Signal Processing Society, “Speech Enhancement and Noise Suppression Technologies,” Technical Report, 2023.