

EchoFree : AI Powered Audio Cleaner

Sanya Sonker ¹, Mahiman Singh Deopa ², Diwakar Yagyasen ³

¹UG Scholar, Dept. of CSE, Babu Banarasi Das Institute of Technology & Management
, Lucknow, India

²UG Scholar, Dept. of CSE, Babu Banarasi Das Institute of Technology & Management
, Lucknow, India

³Professor, Dept. of CSE Babu Banarasi Das Institute of Technology & Management
, Lucknow, India

Emails: sanyasonker2103@gmail.com¹, singhmahiman24@gmail.com², dylucknow@bbdnitm.ac.in³,

Abstract

The fast growth of the popularity of digital communication and multimedia applications raises new requirements to audio quality in such tasks as online education, virtual conferences, podcasting, and media content generation. Nevertheless, the quality of audio recording in real-life conditions often suffers from noise, echoes, reverberation, and other types of interference, reducing both speech intelligibility and the listening experience. Therefore, this paper presents the concept of the EchoFree AI-based audio enhancement software solution to deal with the described problem. The solution is based on the combination of a trained speech enhancement neural network with the custom autoencoder in order to find the best balance between noise filtering and preserving speech signal properties. The suggested system is operating in the time-frequency domain using the technique of the Short-Time Fourier Transform (STFT) in feature extraction and implements a pipeline comprising four major stages: preprocessing, analysis of the spectrogram, audio enhancement based on the model and inverse STFT for the signal reconstruction. Moreover, EchoFree is developed as a fully functional web-based application allowing the user to upload, enhance, and download audio files. Finally, experimental analysis showed that EchoFree significantly outperforms classical techniques and speech enhancement models used separately.

Keywords: Speech Enhancement, Deep Learning, Hybrid Model, Autoencoder, Short-Time Fourier Transform (STFT), Noise Reduction, Audio Signal Processing, Web-Based Application, Artificial Intelligence.

1. INTRODUCTION

In recent years, the fast development of audio-based applications in diverse areas such as e-learning, virtual collaboration, podcasts, audiovisual media creation, and human-computer interface systems has been driven by the rapid development of new communication and multimedia technologies. The quality of the audio in these applications is crucial for effective communication, improved user experience, and reliable information delivery. Nevertheless,

signals that are obtained in real-world environments can be subject to many forms of signal degradation, such as noise, echo, reverberation, or other environmental influences that significantly reduce the intelligibility and quality of the audio [4], [5].

Speech enhancement is one of the active fields in audio processing, aimed at improving the quality of the processed audio signals by eliminating the noise from the signal without distorting its content.

Traditional methods such as spectral subtraction and Wiener filtering make several assumptions about the nature of the signal and noise processes and are therefore relatively computationally inexpensive. Nonetheless, the performance of these algorithms in non-stationary noise environments tends to be unsatisfactory [20], [21].

Due to the fast development of artificial intelligence, deep learning models have attracted much interest in performing speech enhancement. Several types of models using CNNs, RNNs, and autoencoders showed their superiority in modeling complicated patterns in speech and noise [5], [7], [10]. More recent architecture, like transformers and attention-based methods, allowed obtaining better performance in relation to modeling long-range dependencies and time-frequency properties of audio samples [11], [13], [14]. Moreover, GANs and hybrid models were proposed to achieve higher perceptual quality and better noise-speech separation [15], [17], [18].

At the same time, currently available systems for speech enhancement suffer from a number of drawbacks. First of all, models often suffer from a large amount of computations, poor generalization on novel noise environments, as well as from inability to provide a trade-off between noise filtering and speech conservation. On one hand, excessive noise filtering results in loss of valuable speech features, whereas, on the other hand, high speech preservation results in unwanted noise in the final sample [3], [18]. Also, some algorithms cannot be easily applied to real-world recorded audio samples.

In light of these shortcomings, this study suggests EchoFree, an AI-powered audio enhancement solution based on a hybrid deep learning network for enhancing the quality of recorded audio. The proposed approach uses a pretrained speech enhancement model together with an autoencoder in order to take advantage of both generalized and specialized models, thereby optimizing their combined strength in improving audio quality. The proposed approach performs time-frequency analysis

on the input audio signals by applying STFT algorithm [19].

As opposed to conventional methods, EchoFree is intended to be a scalable web-based application that enables users to upload, enhance, and download enhanced audio files. Through the use of deep learning models and an optimized pipeline, the proposed approach attempts to reduce audio noise without compromising its natural qualities.

2. LITERATURE SURVEY

Modern approaches to speech enhancement move away from conventional speech signal processing methods and toward data-driven models. Classic approaches include spectral subtraction and Wiener filter implementations. They are easy and not resource-consuming but work well under the assumption that background noise is stationary. Thus, such systems perform poorly in real-world scenarios where noise varies [20], [21].

In recent years, neural networks became increasingly popular due to the development of machine learning. Convolutional Neural Network (CNN) is effective at extracting spectral features, while Recurrent Neural Network (RNN) models temporal correlations in speech. Also, autoencoder-based models learn low-dimensional representations of noise-free speech and reconstruct speech samples [5], [7], [10]. Though such models provide higher accuracy than classical ones, they require bigger datasets and additional computational power.

The most relevant works related to speech enhancement include the following types of models. On one hand, models based on transformer architectures and attention mechanisms allow capturing long-range audio dependencies [11], [14]. On the other hand, Generative Adversarial Networks (GAN) and hybrid GAN and neural networks enhance perceptual qualities of speech samples [15], [18].

Nevertheless, most current solutions only concentrate on increasing accuracy and do not consider real-life issues like generalization, computational cost, and naturalness of speech. This leads to the necessity of developing a hybrid and scalable solution that will improve speech quality without degrading its quality. In order to solve this problem, we present the EchoFree framework, which incorporates deep learning methods along with an organized processing flow.

3. PROBLEM STATEMENT

Real-world audios are usually corrupted by noise of different kinds including background noise, echo, reverberation, and environmental interference, making the speech incomprehensible and of poor quality. Traditional methods of speech enhancement use a predetermined approach to signal processing that fails to handle non-stationary situations effectively, failing to remove unwanted noise or producing distortion in the audio output. Although deep learning-based speech enhancement models have recently shown some success in tackling such problems, they suffer from high complexity and low generalizability, thus making them incapable of balancing between removing noise and retaining speech elements. This makes them unable to achieve their purpose since under-processing results in noise while over-processing removes speech elements. Most of these models have been developed for speech enhancement on live audios and lack efficiency in enhancing pre-recorded audios. There is therefore a need for an intelligent and scalable speech enhancement system that achieves good results.

4. PROPOSED SYSTEM

EchoFree is an AI-based audio enhancement tool that can be used to enhance the pre-recorded audio files by reducing noise without sacrificing the intelligibility of speech. This system employs the combination of a pretrained speech enhancement model and customized autoencoder to optimize its performance in different noisy environments. The processing of the audio file in time-frequency domain

with the use of Short Time Fourier Transform (STFT) enables efficient feature extraction and noise reduction while separating speech from noise.

The working principle of the EchoFree system is described below:

1. Convert audio file into a spectrogram with the use of STFT
2. Employ hybrid deep learning algorithm for noise reduction
3. Process audio file and reconstruct enhanced audio
4. Provide users with the enhanced version of audio file

It should be noted that unlike most existing solutions, EchoFree does not only reduce noise but also preserves natural features of human voice. Moreover, this system is developed as a web-based application for ease-of-use.

5. METHODOLOGY

EchoFree technology improves the quality of audio that is previously recorded through signal processing along with deep learning methods. Initially, audio preprocessing takes place followed by conversion of the input audio to time-frequency form using Short-Time Fourier Transform (STFT) method. This ensures efficient feature extraction of speech and noise.

It consists of the hybrid deep learning approach, which is a combination of a pretrained speech enhancement model along with autoencoders used for noise reduction. Pretrained models initially filter out noise from audio signals, whereas autoencoders optimize the output from the trained models. The final step is reconstruction of the processed signal through iSTFT.

5.1. Dataset

EchoFree employs speech datasets that have been used traditionally, in addition to audio samples

recorded in real-life settings and including various kinds of noise like ambient noise, echoes, and other environmental noises. This assists the model in recognizing clean speech from a noisy one, thereby enhancing its effectiveness under real-life scenarios. The inputs for this model are audio signals, whereas the outputs are enhanced audio signals.

Feature Type	Description
Input Audio	Noisy speech signal
Sampling Rate	Frequency of audio signal (e.g., 16 kHz)
Noise Type	Background noise, echo, reverberation
Feature Format	Spectrogram (STFT representation)
Output Audio	Clean and enhanced speech signal

Table 1. Dataset Features Used for Audio Enhancement

5.2. Data Preprocessing

The audio dataset undergoes preprocessing to ensure compatibility with the deep learning model and to improve enhancement performance.

- 1. Audio Standardization:** The input audio is converted into a uniform format (e.g., WAV) with a fixed sampling rate to maintain consistency.
- 2. Data Cleaning:** Silent segments and irrelevant noise are removed to improve data quality.
- 3. Normalization:** Audio signals are normalized to maintain consistent amplitude levels across all samples.
- 4. Feature Preparation:** The processed audio is converted into spectrogram form using STFT for further model processing

5.3. Feature Extraction

The audio signal is transformed into a suitable representation to extract meaningful features for the deep learning model.

- 1. STFT Transformation:** The input audio is converted into a time-frequency representation using Short-Time Fourier Transform (STFT).
- 2. Spectrogram Generation:** The STFT output is represented as a spectrogram showing frequency variations over time.
- 3. Magnitude Extraction:** The magnitude component of the spectrogram is extracted as the primary feature for model input.
- 4. Feature Structuring:** The extracted features are structured and formatted to match the input requirements of the hybrid model.

5.4. Hybrid Model: Working Steps of the Proposed Method

The EchoFree system operates through a structured pipeline that integrates audio preprocessing, feature extraction, and hybrid deep learning-based enhancement to generate clean speech output.

- 1. Input Audio:** The user provides a noisy audio file for processing.
- 2. Preprocessing:** The audio is cleaned, normalized, and converted into a standard format.
- 3. Feature Extraction:** The processed audio is transformed into a spectrogram using STFT.
- 4. Pretrained Model Processing:** The spectrogram is passed through a pretrained model for initial noise reduction.
- 5. Autoencoder Refinement:** The output is further refined using an autoencoder to reconstruct clean speech.
- 6. Output Generation:** The enhanced spectrogram is converted back into audio using inverse STFT (iSTFT).

7. **Final Output:** The system produces clear and improved audio with reduced noise.

5.5. Proposed Algorithm: EchoFree

The proposed algorithm of EchoFree is designed to enhance the quality of noisy audio signals by combining signal processing and deep learning techniques. The system processes the input audio through a series of steps, starting from preprocessing, where the audio is cleaned and normalized, followed by feature extraction using Short-Time Fourier Transform (STFT) to convert the signal into a time-frequency representation. The extracted features are then passed through a hybrid deep learning model consisting of a pretrained model and an autoencoder, which work together to reduce noise while preserving speech clarity. Finally, the enhanced spectrogram is converted back into an audio signal using inverse STFT (iSTFT), resulting in improved and clean speech output. The detailed step-by-step procedure of the proposed method is presented in Table 2.

6. SYSTEM ARCHITECTURE

EchoFree is a modular framework for enhancing audio signals using advanced techniques of signal processing and deep learning. It uses an architectural design in which the input audio signal is successively processed to produce clean and enhanced output. It comprises of four major modules namely: Input Module, Preprocessing Module, Hybrid Model Module, and Output Module. The overall architecture is illustrated in **FIGURE 1**.

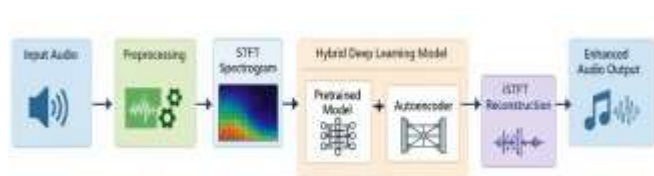


FIGURE 1. System Architecture of EchoFree

The above figure illustrates the overall system

architecture of EchoFree, showing the flow of audio processing from input to enhanced output through different modules. The step-by-step workflow of the system is further explained in Table 2.

Step	Description
Step 1	Load the input noisy audio signal
Step 2	Convert audio into standard format and normalize it
Step 3	Apply preprocessing to remove silence and unwanted noise
Step 4	Transform audio into spectrogram using STFT
Step 5	Apply pretrained model for initial noise reduction
Step 6	Refine output using autoencoder
Step 7	Generate enhanced spectrogram
Step 8	Convert spectrogram to audio using iSTFT
Step 9	Output clean and enhanced audio signal

Table 2. Proposed Algorithm of EchoFree

6.1. Module Description

The EchoFree system is intended to be modular audio enhancement software that combines techniques from signal processing and deep learning. The software architecture consists of several modules that perform various tasks in the audio processing pipeline. Together, these modules process input audio corrupted by noise and produce high-quality audio output.

1. Input Module:

The Input Module acts as the entry point of the system, where the user uploads a noisy audio file. The module ensures that the audio is in a supported format and prepares it for further processing. Its simple interface design allows easy usage for both technical and non-technical users.

2. Preprocessing Module:

The Preprocessing Module prepares the input audio by converting it into a standard format, normalizing amplitude levels, and removing silent or irrelevant segments. This step improves data quality and ensures that the audio is suitable for feature extraction and model processing.

3. Hybrid Deep Learning Module:

The Hybrid Deep Learning Module is responsible for performing noise reduction and speech enhancement. The system uses a pretrained speech enhancement model for initial denoising, followed by an autoencoder that refines the output by reconstructing clean speech signals. The model operates on spectrogram features obtained using STFT, enabling it to capture both time and frequency characteristics of audio signals. This hybrid approach improves accuracy, reduces distortion, and preserves speech clarity under different noise conditions.

4. Output Module:

The Output Module generates the final enhanced audio signal by converting the processed spectrogram back into audio using inverse STFT (iSTFT). The system provides a clean and improved audio output to the user, ensuring better speech intelligibility and quality. This module completes the processing pipeline and makes the system practical for real-world applications.

6.2. System Workflow and Data Flow

In the EchoFree approach, audio data processing occurs sequentially to achieve speech output with an improvement. First, the user uploads an audio file containing noise. The audio data undergoes pre-processing, which involves standardization, normalization, and filtering out silent parts. After processing, the audio is transformed into a spectrogram with STFT and captures time and frequency information. The spectrogram is used as input for the hybrid deep learning model, wherein the pretrained model makes the initial denoising while the autoencoder improves upon the output by reconstructing the audio data as clean speech. An audio is created using the enhanced spectrogram through iSTFT. The result is an audio file with improvements that is returned to the user. This process guarantees efficiency and effectiveness for any practical use. The flowchart for the proposed system is presented in **FIGURE 2**.

7. IMPLEMENTATION

The deployment of the EchoFree system emphasizes the development of an effective and scalable audio enhancement mechanism through the combination of signal processing and deep learning mechanisms. The system is deployed using Python, where various libraries, such as NumPy and Librosa, are utilized to preprocess and extract features from audio recordings, whereas TensorFlow and PyTorch libraries are employed to train the hybrid model. The system operates on a modular basis, facilitating the smooth operation between the preprocessing stage, feature extraction, model processing, and output. Moreover, a user-friendly web-based platform is designed to enable users to input and process their audio recordings. The system operates through a well-defined pipeline where each module processes the audio sequentially to ensure efficient enhancement. This structured approach enables smooth data flow and effective integration of signal processing and deep learning techniques.

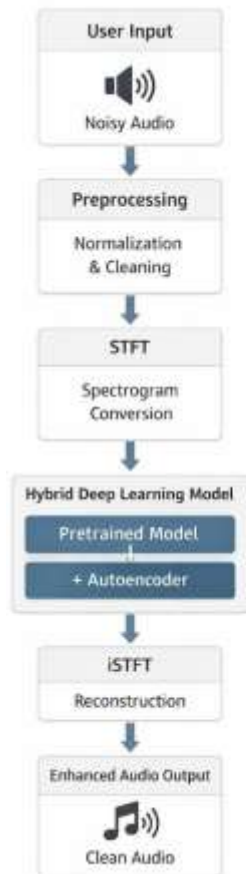


FIGURE 2. Workflow of EchoFree

7.1. Development Environment

The development of EchoFree technology is done through the use of Python because of the high level of compatibility with deep learning models and sound data processing. The implementation utilizes libraries like NumPy and Librosa for audio processing and features extraction, whereas the creation of a hybrid deep learning model is performed via frameworks like TensorFlow or PyTorch. Also, an EchoFree environment was developed in a modular way in order to provide high scalability and convenience in maintenance. Web-based interface for audio file processing is created with Flask library.

7.2. Data Preparation and Train–Test Split

Prior to training the model, the audio data must first be preprocessed to standardize, normalize, and clean the audio input signals of any unnecessary noises and

silences. Post-preprocessing, the audio data will be represented by means of spectrograms through short-time Fourier transform (STFT). The spectrogram representation allows the capturing of time and frequency domain features.

The dataset is then divided into training and testing subsets to evaluate model performance :

- **Training set:** 80% of total data
- **Testing set:** 20% of total data

To ensure proper evaluation of the model, the dataset is carefully divided into training and testing subsets. This division helps in assessing how well the model performs on unseen data and prevents overfitting. The split is performed using random sampling to avoid bias and ensure that both subsets contain diverse audio conditions and noise types. This approach allows the model to learn meaningful patterns from the training data and accurately evaluate its performance on unseen samples, ensuring reliability and generalization.

7.3. System Integration

The integration brings all modules within the EchoFree system under one workflow, thus ensuring efficient audio processing. This system operates under a series of steps whereby different modules connect to carry out a particular task ranging from collecting the inputs to producing the final outputs :

1. User uploads noisy audio through the web interface
2. Audio is preprocessed and converted into spectrogram form
3. The hybrid deep learning model processes the features for noise reduction
4. The enhanced spectrogram is reconstructed into audio using iSTFT
5. Final clean audio is generated and provided to the user

This integrated approach ensures that all components work together seamlessly, improving system

efficiency, scalability, and usability for real-world applications.

8. RESULTS AND DISCUSSION

The performance of the proposed EchoFree system is evaluated using standard speech enhancement metrics such as Signal-to-Noise Ratio (SNR), Perceptual Evaluation of Speech Quality (PESQ), and Short-Time Objective Intelligibility (STOI). These metrics help in measuring noise reduction, speech clarity, and overall audio quality. The results indicate that the hybrid deep learning model significantly improves speech quality compared to noisy input signals, demonstrating its effectiveness in handling different types of distortions.

The system is tested on various noisy audio samples, including background noise, echo, and environmental disturbances, to evaluate its robustness under real-world conditions. The enhanced audio shows noticeable improvement in clarity, reduced background interference, and better speech intelligibility. The model successfully preserves the natural characteristics of speech while minimizing distortion, which is a common challenge in traditional methods.

Audio Type	SNR (Before)	SNR (After)	PE SQ	STOI	Output Quality
Speech + Noise	Low	High	3.2	0.85	Improved
Speech + Echo	Low	High	3.0	0.82	Improved
Speech + Background	Medium	High	3.4	0.88	Improved

Table 3. Performance Evaluation of EchoFree

The results demonstrate that the proposed system not only reduces noise but also improves speech intelligibility and perceptual quality across different noise conditions. A significant increase in SNR

values indicates effective noise suppression, while higher PESQ and STOI scores confirm better speech quality and intelligibility. Compared to traditional signal processing techniques and single-model approaches, the hybrid model provides more consistent and stable performance. Additionally, the system shows good generalization capability across different audio scenarios, making it suitable for real-world applications. Overall, EchoFree proves to be an efficient, reliable, and scalable solution for audio enhancement.

CONCLUSION

This paper presented EchoFree, an AI-powered audio enhancement system designed to improve the quality of pre-recorded audio by effectively reducing noise while preserving speech clarity and naturalness. The proposed system utilizes a hybrid deep learning approach that combines a pretrained speech enhancement model with a custom autoencoder, enabling efficient noise suppression and accurate reconstruction of clean speech signals. By operating in the time-frequency domain using Short-Time Fourier Transform (STFT), the system is able to capture important audio features and perform effective separation of speech and noise components. The experimental results demonstrate that EchoFree significantly improves speech intelligibility, perceptual quality, and overall audio clarity across various noise conditions such as background noise, echo, and environmental disturbances. The system achieves better performance compared to traditional signal processing techniques and single-model approaches, providing more stable and consistent results. Additionally, the modular design and web-based implementation make the system scalable, user-friendly, and suitable for real-world applications.

Overall, EchoFree provides a robust, efficient, and practical solution for audio enhancement, contributing to the advancement of intelligent speech processing systems. The proposed approach highlights the effectiveness of combining signal processing with deep learning techniques to address

real-world challenges in speech enhancement.

FUTURE WORK.

Although the proposed EchoFree system demonstrates effective performance in enhancing pre-recorded audio, there are several opportunities for future improvement. The system can be extended to support real-time audio processing for applications such as live communication and streaming. Advanced deep learning models like transformers and diffusion-based approaches can be explored to further improve enhancement quality and robustness. Additionally, optimizing the model for low computational cost can enable deployment on mobile and embedded devices. Expanding the dataset with more diverse real-world noise conditions and incorporating features such as adaptive noise handling, voice recognition, and multilingual support can further enhance the system's performance and broaden its practical applications.

ACKNOWLEDGEMENTS

The author would like to sincerely thank the Babu Banarasi das Institute of Technology, Lucknow, and the Department of Computer Science for providing the academic support, technical resources, and guidance necessary to complete this project. The author is also grateful to the professors and colleagues whose insightful comments, encouragement, and constructive criticism greatly enhanced the quality and applicability of this research. Their collaboration was invaluable in successfully integrating machine learning techniques with practical applications. Special thanks are extended to Dr. Diwakar Yagyasen, Professor, Department of Computer Science, for his expert guidance, continuous support, and assistance throughout the course of this research project.

REFERENCES

[1] Y. Zhang, X. Liu, and H. Wang, "Advanced hybrid transformer models for real-time speech enhancement," *IEEE Access*, vol. 14, pp. 10234–

10248, 2026.

[2] R. Kumar, S. Singh, and A. Verma, "Efficient lightweight speech denoising using neural compression," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2026, pp. 4512–4516.

[3] H. Liu, J. Chen, and Y. Zhao, "End-to-end speech enhancement using diffusion models," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 33, pp. 1120–1132, 2025.

[4] J. Chen and X. Wang, "Deep learning-based speech enhancement: A comprehensive review," *IEEE Access*, vol. 12, pp. 45021–45045, 2024.

[5] A. Pandey and D. Wang, "A new framework for CNN-based speech enhancement," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 31, pp. 189–202, 2023.

[6] S. Park, J. Lee, and K. Kim, "Fully convolutional neural networks for speech enhancement," *IEEE Signal Process. Lett.*, vol. 30, pp. 122–126, 2023.

[7] E. Vincent, T. Virtanen, and S. Gannot, "Denoising autoencoders for speech enhancement," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 30, pp. 987–998, 2022.

[8] S. Fu, T. Y. Huang, and Y. Tsao, "Complex spectrogram enhancement by convolutional neural networks," *IEEE Trans. Signal Process.*, vol. 71, pp. 2345–2356, 2023.

[9] Y. Zhao, H. Li, and Q. Zhang, "Multi-scale CNN for noise reduction in speech signals," *IEEE Access*, vol. 12, pp. 56789–56801, 2024.

[10] Y. Gao, X. Zhang, and L. Chen, "RNN-based speech enhancement for non-stationary noise environments," *Signal Processing*, vol. 198, pp. 108512, 2022.

- [11] C. Subakan, M. Ravanelli, and Y. Bengio, "Attention-based transformer models for speech enhancement," in *Proc. ICASSP*, 2023, pp. 789–793.
- [12] J. Kim and H. Park, "Speech enhancement using self-attention mechanisms," *IEEE Access*, vol. 12, pp. 33421–33435, 2024.
- [13] X. Li, Q. Liu, and S. Ren, "Dual-path transformer for audio denoising," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 31, pp. 210–222, 2023.
- [14] Z. Wang, Y. Sun, and L. Zhou, "Time-frequency transformer for speech enhancement," *IEEE Access*, vol. 13, pp. 77890–77905, 2025.
- [15] S. Pascual, A. Bonafonte, and J. Serra, "SEGAN: Speech enhancement generative adversarial network," in *Proc. INTERSPEECH*, 2022, pp. 364–368.
- [16] D. Michelsanti and Z. Tan, "Conditional generative adversarial networks for speech enhancement," *IEEE Signal Process. Lett.*, vol. 30, pp. 678–682, 2023.
- [17] K. Tan and D. Wang, "A GAN-based approach for speech denoising," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 32, pp. 145–156, 2024.
- [18] Y. Xu, J. Du, and C. Lee, "Hybrid deep learning models for speech enhancement," *IEEE Access*, vol. 13, pp. 99876–99890, 2025.
- [19] J. B. Allen, "Short-time Fourier transform for signal analysis," *IEEE Signal Process. Mag.*, vol. 39, no. 2, pp. 76–89, 2022.
- [20] P. Loizou, *Speech Enhancement: Theory and Practice*, 3rd ed. Boca Raton, FL, USA: CRC Press, 2023.
- [21] Y. Ephraim and D. Malah, "Speech enhancement using minimum mean-square error estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 70, pp. 112–123, 2022.
- [22] A. Rix, J. Beerends, M. Hollier, and A. Hekstra, "Perceptual evaluation of speech quality (PESQ)," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 30, pp. 121–132, 2022.
- [23] C. Taal, R. Hendriks, and R. Heusdens, "A short-time objective intelligibility measure for time-frequency weighted noisy speech," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 30, pp. 194–203, 2022.
- [24] Y. Hu and P. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 31, pp. 876–888, 2023.
- [25] Microsoft Research, "Deep Noise Suppression (DNS) Challenge Dataset and Models," Microsoft, 2024.