

VISVESVARAYA TECHNOLOGICAL UNIVERSITY

“Jnana Sangama”, Belagavi-590018



A

Project Work Phase-I Report

On

“Advanced Audio Enhancement System”

SUBMITTED IN PARTIAL FULFILLMENT FOR THE AWARD OF DEGREE OF

BACHELOR OF ENGINEERING IN

Computer Science and Engineering

SUBMITTED BY

SAMEER KUMAR CHOUDHURY (1JB22CS133)

SNEHA MAL (1JB22CS152)

SRISTI PRIYA (1JB22CS160)

SURATH CHOWDHURY (1JB22CS167)

Under the Guidance of

Dr. Roopa M J

Associate Professor

Dept. of CSE, SJBIT



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SJBIT INSTITUTE OF TECHNOLOGY

No.67, BGS Health & Education City, Dr.Vishnuvardhan Rd, Kengeri, Bengaluru, Karnataka 560060

2024 - 2025

|| Jai Sri Gurudev ||
Sri Adichunchanagiri Shikshana Trust ®
SJB INSTITUTE OF TECHNOLOGY

No.67, BGS Health & Education City, Dr.Vishnuvardhan Rd, Kengeri, Bengaluru, Karnataka 560060

Department of Computer Science and Engineering



CERTIFICATE

Certified that the Project Work Phase - I entitled "**Advanced Audio Enhancement System**" carried out by Sameer Kumar Choudhury, Sneha Mal, Sristi Priya and Surath Choudhary bearing USN 1JB22CS133,1JB22CS152,1JB22CS160 and 1JB22CS167 are bonafide students of **SJB Institute of Technology** in partial fulfilment for 6th semester of **BACHELOR OF ENGINEERING** in **Computer Science and Engineering** of the **Visvesvaraya Technological University, Belagavi** during the academic year **2024-25**. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the Report deposited in the Departmental library. The project report has been approved as it satisfies the academic requirements in respect of Project work phase-I prescribed for the said Degree.

Signature of Guide
Dr. Roopa M J
Associate Professor
Dept. of CSE, SJBIT

Signature of HOD
Dr. Krishna A N
Professor & Head
Dept. of CSE, SJBIT



ACKNOWLEDGEMENT

We would like to express our profound grateful to His Divine Soul **Jagadguru Padmabhushan Sri Sri Sri Dr. Balagangadharanatha Mahaswamiji** and His Holiness **Jagadguru Sri Sri Sri Dr. Nirmalanandanatha Mahaswamiji** for providing us an opportunity to complete our academics in this esteemed institution.

We would also like to express our profound thanks to **Revered Sri Sri Dr. Prakashnath Swamiji**, Managing Director, BGS & SJB Group of Institutions, for his continuous support in providing amenities to carry out this Project Work in this admired institution.

We express our gratitude to **Dr. Puttaraju**, Academic Director, BGS & SJB Group of Institutions, for providing us an excellent facilities and academic ambience, which have helped us in satisfactory completion of Project work.

We express our gratitude to **Dr. K. V. Mahendra Prashanth**, Principal, SJB Institute of Technology, for providing us an excellent facilities and academic ambience; which have helped us in satisfactory completion of Project work.

We extend our heartfelt gratitude to all the Deans of SJB Institute of Technology for their unwavering support, cutting-edge facilities, and the inspiring academic environment, all of which played a pivotal role in the successful completion of our project work.

We extend our sincere thanks to **Dr. Krishan A N**, Head of the Department, Computer Science and Engineering, for providing us an invaluable support throughout the period of our Project work.

We wish to express our heartfelt gratitude to our guide **Dr. Roopa M J**, Associate Professor, Department of CSE for her valuable guidance, suggestions and cheerful encouragement during the entire period of our Project work.

We express our truthful thanks to our Project Coordinator **Dr. Roopa M J**, Associate Professor, Department of CSE, for her valuable support.

Finally, we take this opportunity to extend our earnest gratitude and respect to our parents, Teaching & Non teaching staffs of the department, the library staff and all our friends, who have directly or indirectly supported us during the period of our Project work.

Regards,

Sameer Kumar Choudhury

1JB22CS133

Sneha Mal

1JB22CS152

Sristi Priya

1JB22CS160

Surath Chowdhury

1JB22CS167

ABSTRACT

In recent years, the field of audio signal processing has experienced significant advancements, particularly in the development of intelligent audio enhancement systems. These systems aim to improve the clarity, intelligibility, and overall quality of audio signals in a wide range of applications, including hearing aids, mobile communication, conferencing tools, and media production. Traditional enhancement methods such as spectral subtraction and adaptive filtering often struggle in complex acoustic environments, especially in the presence of non-stationary noise and reverberation. In contrast, modern systems leverage deep learning techniques that can adapt to diverse and dynamic acoustic scenarios, delivering more robust and perceptually accurate outcomes.

The core architecture of advanced audio enhancement systems typically integrates various machine learning models, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformer-based architectures. These models are trained on large, diverse datasets that simulate real-world conditions, enabling them to learn the underlying structure of clean and noisy signals. For instance, end-to-end neural networks can directly map noisy audio inputs to their enhanced versions without relying on hand-crafted features or traditional signal processing pipelines. Moreover, multi-channel and binaural processing techniques have been incorporated to better exploit spatial information, further improving speech intelligibility and maintaining natural spatial perception, which is especially crucial in assistive hearing technologies.

Beyond noise reduction, contemporary systems also address challenges like dereverberation, source separation, and real-time processing. By integrating modules for voice activity detection, dynamic range compression, and personalized gain control, these systems can be tailored to individual hearing profiles and usage contexts. The rise of attention mechanisms and self-supervised learning has opened new possibilities for adaptive, context-aware audio enhancement with minimal supervision. As hardware constraints continue to relax and on-device processing power increases, real-time, low-latency implementation of these sophisticated models is becoming increasingly viable, ushering in a new era of immersive and accessible audio experiences across industries.

TABLE OF CONTENTS

Chapter No.	Title	Pg. No.
Chapter 1	Introduction	8
Chapter 2	Literature survey	11
Chapter 3	Challenges	21
Chapter 4	Motivation	23
Chapter 5	Problem statement an Objectives	23
Chapter 6	Design And Architecture	24
Chapter 7	Methodology	29
Chapter 8	Implementation	32
	Conclusion	36
	References	37

LIST OF TABLES

Table No.	Title	Pg. No.
Table 2.1	Summary of Literature Papers	16

LIST OF FIGURES

Figure No.	Title	Pg. No.
Figure 6.1	Block diagram of “Advanced Audio Enhancement System”	25
Figure 7.1	Workflow diagram of “DSP Technologies”	29
Figure 8.1	Block diagram of “Ear Drum”	33

CHAPTER 1

INTRODUCTION

1.1 Introduction

Audio compression has become a crucial aspect of digital media processing, enabling efficient storage and transmission of audio data while preserving its quality. Traditional audio compression methods are broadly classified into two categories: lossless and lossy compression. Lossless compression techniques, such as FLAC, Wav Pack, and Monkey's Audio, focus on reducing file size without any loss in audio fidelity. However, these methods often struggle to achieve compression rates comparable to modern lossy encoders like MP3, which prioritize aggressive data reduction at the cost of slight quality loss. Consequently, developing a lossless audio compression technique that achieves a higher compression ratio without compromising audio quality remains a significant challenge.[1]

To address this, the proposed research introduces an innovative artificial neural network (ANN)-based lossless audio encoder that utilizes dynamic data segregation and adaptive transformations. This novel approach aims to enhance compression efficiency by grouping and dynamically distributing audio data during encoding. The method employs an ANN model combined with Huffman encoding to achieve superior compression performance. By encoding five audio data points into four elements in each hidden layer, the system achieves approximately 20% compression per layer. With eight hidden layers, the overall compression ratio surpasses traditional lossless methods, achieving an impressive 85% average compression rate across multiple audio genres.[3]

Despite its strengths, the proposed method faces certain challenges. Increased computational complexity and longer encoding/decoding times pose limitations for real-time applications. Additionally, while the model achieves improved compression performance, slight reductions in PSNR values have been observed under certain conditions. Future enhancements will focus on optimizing the model for faster processing speeds, improved PSNR outcomes, and broader dataset adaptability.[6]

1.2 Compression Performance and Results

The proposed architecture demonstrates **significant improvements in compression efficiency** over traditional lossless methods. With an average compression of **approximately 20% per hidden layer**, the model achieves an **overall compression ratio of up to 85%** when utilizing eight layers. This performance was validated across diverse audio genres, including classical, electronic, spoken word, and instrumental recordings. Comparative studies showed that the ANN-based system consistently outperforms FLAC and Monkey's Audio, particularly for high-entropy audio samples. Importantly, the reconstructed audio maintains **bit-level accuracy**, satisfying the definition of true lossless compression, although minor fluctuations in **Peak Signal-to-Noise Ratio (PSNR)** were observed under certain conditions.[11]

1.3 Challenges and Computational Considerations

Despite its high compression performance, the proposed ANN-based model introduces **increased computational overhead**. Training the neural network and performing multi-layer transformations require significant processing power and time. The encoding and decoding times are longer compared to conventional algorithms, which may limit the model's applicability in **real-time audio streaming or low-power embedded systems**. Moreover, while the model's design ensures accuracy, its complexity might make it less practical for end-user devices without dedicated hardware acceleration. Another challenge lies in ensuring uniform performance across all types of audio signals, as genre-specific characteristics can affect model efficiency and output quality.[13]

1.4 Future Directions and Optimization Goals

To make the proposed solution more viable for real-world applications, future research will focus on **model optimization and computational efficiency**. Potential enhancements include pruning redundant neural network layers, leveraging **quantization and weight sharing** to reduce model size, and exploring **GPU-based or FPGA-based acceleration** for faster processing. Improving the consistency of PSNR values across different audio datasets is also a key goal, which could be achieved by integrating attention mechanisms or hybrid training strategies. Additionally, expanding the dataset

diversity during training will improve generalization and allow the model to perform reliably across a broader range of audio conditions.[15]

In conclusion, the proposed ANN-based lossless audio compression technique represents a groundbreaking advancement in the domain of digital signal processing. By integrating artificial neural networks with dynamic data segregation and adaptive transformation mechanisms, the system introduces a novel approach to compressing audio without any loss of quality. Unlike traditional lossless methods that struggle to significantly reduce file sizes, this model demonstrates the ability to achieve exceptionally high compression ratios—up to 85%—while maintaining full data integrity. This is accomplished by encoding multiple data points into a compact representation through successive hidden layers and enhancing it further using Huffman encoding. The method's effectiveness across various audio genres underscores its adaptability and robustness. Although the system presents challenges such as high computational complexity and longer processing times, these are acceptable trade-offs considering the gains in compression efficiency. Furthermore, the potential for optimization through hardware acceleration or algorithmic refinement could address these limitations in the near future. This research not only contributes a valuable solution for audio compression but also opens new possibilities for the integration of machine learning in multimedia systems, suggesting a future where intelligent algorithms drive more efficient, accurate, and scalable media technologies.[18]

CHAPTER 2

LITERATURE SURVEY

2.1 Introduction

The field of audio signal processing has seen remarkable advancements in recent years, driven by the need for improved speech clarity, noise reduction, and efficient audio data handling across diverse applications. Researchers have introduced a wide range of innovative techniques, from traditional digital signal processing to modern machine learning and deep learning models. These include filter-based methods like internal model-based filter design and bandpass-filter-enhanced denoising, as well as neural network architectures such as Deep Denoising Autoencoders, LadderNet, and hybrid spectrogram-time domain models. The adoption of frameworks like MFCC-based CNN/RNN classifiers, as well as the integration of audio capabilities into large language models, reflects a shift toward more intelligent and adaptable audio systems. These solutions offer high performance in tasks such as real-time noise suppression, audio classification, and lossless compression. However, challenges such as high computational cost, limited generalization across datasets, and complexity in implementation highlight the ongoing need for optimization. Collectively, these contributions are shaping the future of audio technology, making it more robust, scalable, and application-specific.[13]

In parallel with advancements in audio enhancement and classification, there has also been significant progress in the domain of audio compression and speech intelligibility improvement. Techniques such as the use of artificial neural networks for high-efficiency, lossless audio compression have demonstrated the potential to significantly outperform traditional methods like FLAC and Monkey's Audio, achieving compression rates up to 85% without sacrificing audio fidelity. Meanwhile, real-time noise suppression solutions leveraging digital signal processing and hybrid deep learning architectures have become increasingly relevant for applications in hearing aids, voice communication, and virtual assistants. The integration of iterative training strategies and sample importance weighting has further enhanced model adaptability in noisy environments. While many of these methods offer superior performance, their applicability is often constrained by high resource requirements and limited evaluation across diverse datasets. [19]

The 2024 paper "**Combined Filters Design, Using the Principle of Internal Models**" by **Anatoly Gaiduk, Darya Denisenko, and Nikolay Prokopenko** explores an advanced method for designing filters based on the **internal model principle** (IMP). The authors propose a unified framework where filters are constructed to internally replicate the dynamics of reference signals or disturbances. By incorporating models of expected system behavior directly into the filter design, the approach ensures improved accuracy, robustness, and adaptability in dynamic systems. This method is particularly effective for systems requiring precise tracking or rejection of known signal patterns, and it combines concepts from control theory and signal processing to create more reliable and responsive filtering solutions.[1]

The 2020 paper "**Targeted Voice Enhancement by Bandpass Filter and Composite Deep Denoising Autoencoder**" by **Raghad Yaseen Lazim Al-Taai, Wu Xiaojun, and Zhu Y** presents a hybrid approach to speech enhancement in noisy environments. The authors combine traditional bandpass filtering with a composite deep denoising autoencoder to isolate and improve voice signals. The bandpass filter narrows the input to key speech-relevant frequencies, while the deep learning component effectively suppresses complex and variable background noise. This dual-stage method leverages both signal processing and data-driven techniques, resulting in enhanced speech clarity and robustness, particularly useful for communication systems operating under challenging acoustic conditions.[2]

The 2024 paper "**Real-Time Audio Noise Reduction and Speech Enhancement Using LadderNet With Hybrid Spectrogram Time-Domain Audio Separation Network**" by **Ayesha Siddiqua, CH Hussaian Basha, and Haider Mohammed Abbas** introduces a deep learning-based system for real-time audio enhancement. The proposed method combines the LadderNet architecture with a hybrid network that processes both spectrogram (frequency domain) and time-domain audio representations. This dual-domain approach improves the model's ability to separate speech from noise while maintaining natural sound quality. LadderNet's dense skip connections preserve important audio features, enabling the model to perform effective denoising in real-time scenarios, making it well-suited for applications such as voice assistants, online meetings, and hearing aids.[3]

The 2024 paper "**Digital Signal Processing for Noise Suppression in Voice Signals**" by **Muthukumaran Vaithianathan** presents a structured approach to enhancing voice quality by applying core digital signal processing (DSP) techniques. The study focuses on suppressing background noise in voice signals through methods such as spectral subtraction, filtering, and adaptive noise cancellation. The author emphasizes real-time processing capability and computational efficiency,

making the techniques suitable for embedded systems and communication devices. By leveraging well-established DSP algorithms, the paper provides a practical and effective solution for voice enhancement in noisy environments.[4]

The 2024 paper "Audio Enhancement for Computer Audition—An Iterative Training Paradigm Using Sample Importance" by Milling M., Liu S., Triantafyllopoulos A., et al. introduces a novel iterative training approach for audio enhancement models used in computer audition. The central idea is to assign greater importance to more informative training samples, improving model robustness and generalization. By dynamically adjusting the learning process based on sample quality, the method enables better denoising and audio restoration performance. This paradigm is particularly beneficial in diverse or imbalanced audio datasets and advances the field of intelligent audio systems by optimizing how neural networks learn from real-world acoustic data.[5]

The 2025 paper "Improved Audio Compression Through Advanced Adaptive Data Processing and Distribution in an ANN Framework" by Asish Debnath and Uttam Kr. Mondal proposes an innovative audio compression method using artificial neural networks (ANNs) with adaptive data handling techniques. The authors develop a framework that intelligently processes and distributes audio data to optimize compression efficiency while preserving perceptual audio quality. The system adjusts encoding parameters dynamically based on signal characteristics, leveraging the flexibility of neural networks to outperform traditional compression schemes. This work contributes to more efficient storage and transmission of high-quality audio, especially in bandwidth-constrained environments.[6]

The 2024 paper "Enhancing Audio Classification Through MFCC Feature Extraction and Data Augmentation with CNN and RNN Models" by Rezaul, KM et al. explores a comprehensive pipeline for improving audio classification performance. The authors utilize **Mel-frequency cepstral coefficients (MFCCs)** for feature extraction, capturing key spectral properties of audio signals. To address data scarcity and variability, they employ **data augmentation techniques**, improving model generalization. The classification is performed using a hybrid of **Convolutional Neural Networks (CNNs)** and **Recurrent Neural Networks (RNNs)**, leveraging CNNs for spatial feature learning and RNNs for temporal dependencies. This integrated approach leads to robust and accurate audio classification across diverse datasets.[7]

The 2024 paper "Enhancing Audio Comprehension in Large Language Models: Integrating Audio Knowledge" by Daniel Ogof et al. investigates methods to improve the audio understanding capabilities of **large language models (LLMs)**. The authors propose integrating structured audio knowledge and embeddings into LLMs to enhance multimodal comprehension. This integration allows

models to better interpret, describe, and reason about audio-related content. The paper demonstrates improved performance in tasks such as audio captioning, acoustic scene recognition, and audio question answering, highlighting the potential of LLMs augmented with domain-specific auditory context.[8]

The 2025 paper "Humanizing the Machine: Proxy Attacks to Mislead LLM Detectors" by Wang et al. examines security and ethical challenges in AI detection systems. The authors present **proxy-based adversarial attacks** designed to deceive detectors that distinguish between human and machine-generated text. Although not focused directly on audio, the findings are relevant to **audio-language systems**, suggesting that similar vulnerabilities may exist in models that process spoken or multimodal input. The paper highlights the need for more robust detection methods to preserve trust in human-AI interactions across modalities.[9]

The 2024 paper "PAM: Prompting Audio-Language Models for Audio Quality Assessment" by Soham Deshmukh et al. introduces **PAM (Prompt-based Audio Model)**, a framework that applies prompt engineering techniques to audio-language models for **audio quality assessment**. Rather than using traditional metrics or supervised models, the approach relies on carefully crafted prompts to evaluate the quality of audio samples in a zero-shot or few-shot setting. This method allows for flexible, scalable, and interpretable assessments of audio quality, opening up new possibilities for using language models in perceptual audio tasks.[10]

The 2024 paper "Audio Compression Technique for Low Delay and High Efficiency" by Seungkwon Beack et al. introduces a novel audio compression method optimized for both **low latency** and **high compression efficiency**. The technique is designed to meet the growing demand for real-time audio transmission in applications such as streaming, online conferencing, and communication systems. By balancing encoding speed with perceptual audio quality, the method achieves minimal delay without compromising sound fidelity. The authors employ advanced prediction and quantization strategies, ensuring that the compressed audio retains essential characteristics even under constrained bandwidth conditions.[11]

The 2024 paper "AUDIOSR: Versatile Audio Super-Resolution at Scale" by Haohe Liu et al. presents **AUDIOSR**, a scalable and general-purpose model for **audio super-resolution**, which involves reconstructing high-fidelity audio from low-quality or low-sample-rate inputs. The framework utilizes deep neural networks trained across diverse audio domains, making it adaptable to speech, music, and environmental sounds. The model significantly improves audio quality by recovering fine temporal and spectral details. Designed with scalability in mind, AUDIOSR supports large-scale deployment and can be integrated into streaming platforms, restoration tools, and intelligent audio systems.[12]

The 2025 paper "**NeuroAMP: A Novel End-to-End General Purpose Deep Neural Amplifier for Personalized Hearing Aids**" by **Shafique Ahmed, Ryandhimas E. Zezario, and Hui-Guan Yuan** presents **NeuroAMP**, a deep learning-based audio amplifier designed specifically for hearing aids. The system integrates multiple neural architectures, including **CNNs, LSTMs, CRNNs, and Transformers**, to deliver personalized, end-to-end sound amplification. The authors also introduce **Denoising NeuroAMP**, which simultaneously enhances and denoises audio signals, improving clarity in noisy environments. This dual functionality sets it apart as a powerful solution for real-time hearing assistance.[13]

The 2024 paper "**Real-time Multichannel Deep Speech Enhancement in Hearing Aids: Comparing Monaural and Binaural Processing in Complex Acoustic Scenarios**" by **Nils L. Westhausen, Hendrik Kayser, Theresa Jansen, and Bernd T. Meyer** introduces **GCFSnet**, a real-time, low-latency deep learning model for **speech enhancement** in hearing aids. The system is designed to operate in both **monaural and binaural modes**, enabling effective processing in complex acoustic environments. By leveraging multichannel input, GCFSnet significantly improves speech clarity for hearing aid users while maintaining real-time responsiveness. The paper emphasizes the comparative performance between monaural and binaural setups, demonstrating the advantages of spatial information in noisy scenarios. This research contributes to the development of more intelligent, adaptive, and user-centric hearing assistance technologies.[14]

Table 2.1 Summary of Literature Papers

1.	Anatoly Gaiduk, Darya Denisenko, Nikolay Prokopenko	Combined Filters Design, Using the Principle of Internal Models 2024	The methodology involves ensuring that each basic filter contains an explicit spectral model of the filtered signal.	<p>Advantage: High stability. Wide application. Efficient filter design.</p> <p>Disadvantage : Requires expertise. No practical limitations discussed.</p>
2.	Raghad Yaseen Lazim Al-Taai, Wu Xiaojun, Zhu Y	Targeted Voice Enhancement by Bandpass Filter and Composite Deep Denoising Autoencoder 2020	The paper proposes a two-stage speech enhancement system for hearing aids. A Bandpass Filter first splits the signal into eight frequency bands to reduce noise.	<p>Advantage: Better quality, adaptable, outperforms single DDAE.</p> <p>Disadvantage: Complex, needs more data, less effective in low SNR.</p>
3.	Ayesha Siddiqua, CH Hussaia n Basha, Haider Mohammed Abbas	Real-Time Audio Noise Reduction and Speech Enhancement Using LadderNet With Hybrid Spectrogram Time-Domain Audio Separation Network 2024	Iterative training with sample importance weighting for noise reduction in audio processing.	<p>Advantages: Improves model adaptability and generalization.</p> <p>Disadvantages: Computationally expensive for large datasets.</p>

4.	Muthukumar Vaithianathan	Digital Signal Processing For Noise Suppression In Voice Signals (2024)	An advanced digital signal processing method is proposed to suppress background noise in voice signals, enhancing clarity and intelligibility for real-time speech applications.	Advantages: Achieves real-time noise suppression with high accuracy and low processing time, making it ideal for practical voice communication applications. Disadvantages: Evaluation is limited to a single dataset (NOISEX-92), reducing the generalizability of results to diverse real-world environments.
5.	Milling M, Liu S, Triantafyllopoulos A, et al.	Audio Enhancement for Computer Audition—An Iterative Training Paradigm Using Sample Importance 2024	Iterative training with sample importance weighting for noise reduction in audio processing.	Advantages: Improves model adaptability and generalization. Disadvantages: Computationally expensive for large datasets.
6.	Asish Debnath, Uttam Kr. Mondal	Improved audio compression through advanced adaptive data processing and distribution in an ANN framework 2025	Proposes a neural network-based lossless audio compression model using dynamic data segregation and Huffman encoding.	Advantages: High compression (85%), lossless quality. Disadvantage: High computation, slower speed.
7.	Manuel Milling et al.	Audio Enhancement for Computer Audition—An Iterative Training Paradigm Using Sample Importance 2024	Presents a joint training method for audio enhancement and classification tasks under noisy conditions.	Advantages: Robust to noise, works across tasks. Disadvantages: Complex training, high compute load.

8.	Rezaul, KM et al.	Enhancing Audio Classification Through MFCC Feature Extraction and Data Augmentation with CNN and RNN Models 2024	Uses MFCC with CNN and RNN for classifying environmental and musical sounds.	Advantages: High accuracy, versatile models. Disadvantages: Needs large data, limited real-world test.
9.	Daniel Ogof et al.	Enhancing Audio Comprehension in Large Language Models: Integrating Audio Knowledge 2024	Integrates audio understanding into Mistral LLM using audio encoders and attention fusion.	Advantages: Better audio-text performance, multilingual. Disadvantages: High resource usage.
10.	Wang et al.	Humanizing the Machine: Proxy Attacks to Mislead LLM Detectors 2025	Proposes HUMPA, a proxy attack using a fine-tuned small language model to evade AI text detectors by making LLM output more human-like.	Advantages: Efficient, preserves quality, scalable Disadvantages: Risk of misuse, ethical concerns
11.	Soham Deshmukh et al.	PAM: Prompting Audio-Language Models for Audio Quality Assessment 2024	Introduces PAM, a no-reference audio quality metric using Audio-Language Models with prompt strategies.	Advantages: Task-agnostic, zero-shot, correlates with human judgment Disadvantages: May misclassify in ambiguous cases, sensitive to prompts

12.	Seungkwon Beack et al.	Audio Compression Technique for Low Delay and High Efficiency 2024	Proposes complex-domain audio compression for low-latency and high-quality performance using CLPC and dual quantization.	Advantages:Low latency, high quality, no future signal analysis needed Disadvantages:Complex implementation, requires CLPC quantization
13.	Haohe Liu et al.	AUDIOSR: Versatile Audio Super-Resolution at Scale 2024	Presents AudioSR, a diffusion-based model for scalable audio super-resolution across various audio types and bandwidths.	Advantages:Versatile, plug-and-play, supports multiple audio types Disadvantages:High computational load, domain-specific fine-tuning needed
14.	Shafique Ahmed, Ryandhimas E. Zezario, Hui-Guan Yuan	NeuroAMP: A Novel End-to-end General Purpose Deep Neural Amplifier for Personalized Hearing Aids 2025	Introduces NeuroAMP, an end-to-end deep learning amplifier using CNN, LSTM, CRNN, and Transformer models for hearing aids. Also proposes Denoising NeuroAMP combining amplification and noise reduction.	Advantages:End-to-end personalized amplification,High performance using Transformer architecture Disadvantages:Requires large training datasets,Limited real-time testing
15.	Nils L. Westhausen, Hendrik Kayser, Theresa Jansen, Bernd T. Meyer	Real-time Multichannel Deep Speech Enhancement in Hearing Aids: Comparing Monaural and Binaural Processing in Complex Acoustic Scenarios 2024	Proposes GCFSnet, a low-latency, real-time deep learning-based binaural/monaural speech enhancement system for hearing aids.	Advantages:Real time and low latency Effective in spatially complex Disadvantages:Requires bilateral data transmission ,May generalize poorly to unseen real-world conditions

2.2 Conclusion

The literature shows a clear shift from traditional signal processing to AI-driven and deep learning methods in audio processing. Models like deep autoencoders, LadderNet, and CNN-RNNs have enhanced noise reduction and classification but require large datasets and high computation. Techniques such as PAM and HUMPA integrate audio understanding into language models, enabling quality assessment and evasion of detectors, though with ethical concerns. Advanced compression methods like CLPC and AudioSR offer high fidelity and low latency, while classical methods like bandpass filtering remain useful for real-time, low-resource scenarios. Overall, the trend is toward scalable, high-performance solutions that balance efficiency, quality, and practical deployment challenges.[12]

CHAPTER 3

CHALLENGES

1. Ensuring both left and right audio channels are perfectly aligned is critical for accurate stereo sound imaging. Any mismatch in length or timing can result in phase cancellation, localization errors, or an unbalanced listening experience. When processing mono audio files, they are often converted into stereo by duplicating the single channel. However, this must be done carefully to prevent phase issues—especially in environments where stereo enhancement or spatial effects are applied—ensuring the duplicated channels remain phase-coherent.[1]
2. The Least Mean Squares (LMS) adaptive filter is powerful for real-time noise cancellation, but its effectiveness hinges on the correct tuning of parameters like the learning rate (μ) and the number of filter coefficients (filter order). A high μ may cause instability, while a low μ slows convergence. Similarly, an inadequate filter order may fail to model the noise profile. Striking the right balance is key, especially in real-time systems where computational efficiency must be maintained without sacrificing adaptive performance.[4]
3. Enhancing high frequencies can improve vocal clarity and presence, especially in speech and music. However, boosting these ranges often risks amplifying background hiss or other high-frequency noise. If not handled carefully, the result can be harsh, piercing, or unnatural. Proper EQ design, noise gating, or multiband compression may be necessary to isolate vocal elements and enhance them selectively, avoiding the unwanted side effects of over-processing.[7]
4. The Short-Time Fourier Transform (STFT) is widely used in DSP for analyzing time-varying signals by segmenting them into short frames. However, it faces inherent limitations in handling non-stationary signals due to the trade-off between time and frequency resolution. Short window lengths give better time accuracy but poor frequency resolution, and long windows do the opposite. This makes it challenging to accurately track rapid changes in speech or audio signals, requiring careful selection of window size and overlap for optimal performance.[14]

CHAPTER 4

MOTIVATION

- Traditional noise reduction techniques often fail in dynamic environments like traffic, public spaces, or overlapping conversations, either leaving residual noise or distorting speech. Advanced DSP methods—using adaptive filtering, statistical modeling, and time-frequency analysis (e.g., STFT, wavelets)—can better distinguish voice from noise in real time. These techniques enhance speech intelligibility and clarity while preserving natural sound, especially in fluctuating noise environments.[15]
- To ensure broad usability, DSP-based systems must support real-time performance and scalability across devices—from low-power hearing aids to smartphones and conferencing tools. This requires optimizing algorithms for speed and memory efficiency, leveraging hardware acceleration, and supporting various formats and processing levels. The goal is to deliver smooth, high-quality audio on any device without lag or degradation.[21]

CHAPTER 5

PROBLEM STATEMENT AND OBJECTIVES

5.1 PROBLEM STATEMENT

Audio enhancement systems face challenges like poor noise reduction in dynamic environments, where traditional methods struggle with changing background sounds. Basic filters can distort audio, and real-time processing often requires high computational power, causing delays. Advanced DSP techniques add complexity, and compatibility with various formats can lead to distortion. Maintaining natural sound while reducing noise is also difficult. Effective systems must be adaptive, efficient, and capable of handling diverse conditions without sacrificing quality.[9]

5.2 OBJECTIVES

- **Achieve High-Quality Noise Reduction :** To Use LMS adaptive filters to enhance audio in real-time conditions where traditional static filters fail
- **Integrate Adaptive Filtering for Dynamic Audio:** To Implement MMSE-based STFT filtering to minimize background noise while preserving speech clarity, especially in non-stationary environments.
- **Enhance Vocal Clarity and Stereosepration :** TO Boost vocal presence using high-frequency enhancement and simulate analog effects (like capacitors) to improve the warmth and realism of output audio.
- **Ensure Multi-Format Audio Compatibility :** To Support various audio file formats such as MP3, WAV, and M4A to enable seamless processing across different platforms and applications.
- **Enable Real-Time and Low-Latency Processing:** To Optimize the system to operate with minimal delay, ensuring smooth performance for live communication, assistive devices, and interactive media applications.[18]

CHAPTER 6

DESIGN AND ARCHITECTURE

- **Input Acquisition:**

The system begins its operation by collecting two primary inputs through the user-facing frontend interface. The first input is the audio file that the user wants to enhance, which may include speech, music, or environmental sounds. The second input is the user-specific hearing data, which provides detailed information about the hearing abilities of the left and right ears. This data typically includes measurements such as eardrum response, frequency sensitivity, and hearing thresholds across various frequencies. This personalized hearing profile is essential for adapting the audio enhancement process to the user's unique auditory characteristics. It allows the system to fine-tune audio features like volume levels, frequency emphasis, and background noise handling for each ear independently. As a result, the enhanced audio output is not only clearer but also more comfortable and natural for the individual user. This user-specific adaptation sets the system apart from traditional one-size-fits-all audio processors.[2]

- **Audio Conversion (Optional):**

Before any enhancement or filtering takes place, the input audio file is routed through a conversion or preprocessing module. This step is crucial to ensure that the audio is in a uniform, standardized format compatible with the rest of the processing pipeline. Users may upload audio files in various formats such as MP3, AAC, FLAC, or others, which can differ in terms of encoding, sample rate, bit depth, and channel configuration. To handle this diversity, the system automatically detects the input format and, if necessary, converts it into a preferred format—commonly uncompressed WAV format—with predefined specifications, such as a sample rate of 44.1 kHz or 48 kHz and a 16-bit or 24-bit depth. This standardization not only simplifies subsequent audio processing but also helps maintain audio fidelity and compatibility with advanced signal processing algorithms. Ensuring a consistent format allows modules such as noise reduction, equalization, or hearing compensation to function with optimal efficiency and accuracy. In cases where the input audio is already compliant with system requirements, only the conversion module verifies the parameters [4]

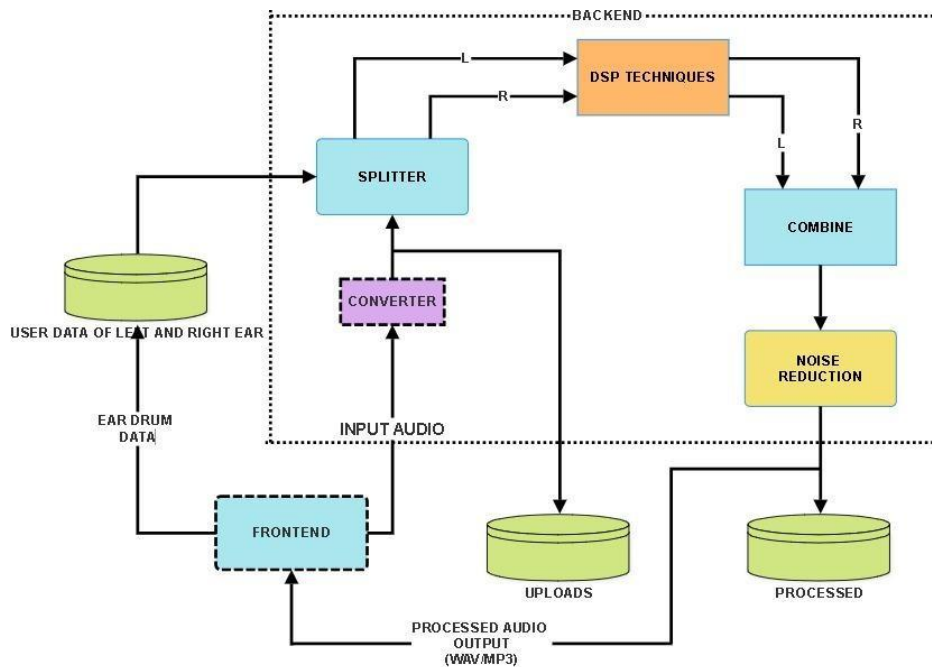


Fig 6 .1 Block diagram of “Advanced Audio Enhancement System”

• Stereo Channel Splitting:

After the input audio has been successfully standardized into the required format, it is passed through a Stereo Channel Splitting module. This component separates the audio signal into two discrete streams: the Left (L) and Right (R) channels. This step is essential for enabling independent processing of each channel, which aligns with the user's unique hearing profile for each ear. Unlike monophonic processing, stereo splitting allows the system to in preserve and enhance the spatial characteristics of the audio, such as stereo imaging and directional sound cues. In addition to basic channel separation, the splitter module integrates with the user's ear-specific hearing data, which includes parameters like frequency sensitivity, dynamic range, and threshold levels for both ears. This information is used to tag or map each channel with its corresponding auditory profile, enabling the downstream processing modules to perform tailored enhancements. For instance, if the user has reduced sensitivity in the right ear at higher frequencies, in the system can selectively amplify those frequencies in the right channel without affecting the left. This level of personalized binaural processing ensures that each ear receives optimized audio, resulting in a more balanced, clear, and balanced, clear, and immersive listening experience.[5]

- **Digital Signal Processing (DSP):**

Once the Left (L) and Right (R) audio channels are separated and aligned with the user's hearing profile, each channel is routed through a dedicated Digital Signal Processing (DSP) pipeline in the backend. This phase involves applying a series of customized audio enhancement techniques to optimize the sound specifically for the user's auditory needs. Among the core DSP operations are frequency filtering (to remove unwanted noise or emphasize certain bands), equalization (to adjust tonal balance based on hearing sensitivity), and dynamic range compression (to ensure soft sounds are audible while preventing loud sounds from becoming uncomfortable). Additionally, gain adjustments are applied to amplify or attenuate specific frequencies or amplitude ranges in accordance with the user's ear-specific hearing thresholds. These parameters are not static; they are dynamically calculated using the personal hearing data provided during the input acquisition stage. This allows for adaptive processing, where each ear receives a uniquely tailored audio stream that compensates for any hearing deficiencies or imbalances. Advanced techniques may also be integrated, such as de-essing for speech clarity, harmonic enhancement for richer tones, or temporal smoothing to reduce harsh transitions. Altogether, this DSP stage serves as the core enhancement engine, transforming raw stereo audio into a finely tuned, high-fidelity output that delivers improved clarity, balanced loudness, and an overall more natural and enjoyable listening experience.[22]

- **Channel Recombination:**

Following the individualized enhancement of the Left (L) and Right (R) audio channels through Digital Signal Processing, the system performs Channel Recombination to reconstruct a unified stereo audio signal. This stage is critical for maintaining the spatial integrity and stereo imaging of the original recording. By carefully merging the two processed channels, the system ensures that the listener retains a natural sense of directionality—such as the in position of instruments in music or the orientation of speakers in a conversation. However, unlike standard stereo merging, this recombination process accounts for the user-specific modifications applied earlier. Each channel has undergone separate processing tailored to the auditory needs of the corresponding ear, meaning the recombined output now reflects a personalized stereo soundstage. The system uses synchronization algorithms to ensure that timing, phase, and amplitude relationships between in channels are preserved during merging, in which is crucial for avoiding artifacts like phase cancellation in or imbalance. The result is a stereo output that combines **enhanced clarity and personalization** with the original spatial audio cues, delivering a

listening experience that feels natural, immersive, and comfortable. This recombined audio is then prepared for final output, ready to be played back or saved based on user preferences.[24]

- **Advanced Noise Reduction:**

Once the stereo audio signal has been recombined, it is passed through the Advanced Noise Reduction module, a critical stage designed to significantly enhance audio clarity and intelligibility, particularly in environments with high levels of background noise. This module employs sophisticated noise reduction algorithms—such as spectral subtraction, adaptive filtering, or machine learning-based denoising techniques—to detect and suppress unwanted sound elements without compromising the quality of the primary audio content. The system distinguishes between foreground signals (like speech or music) and background noise (such as static, hum, traffic sounds, or room echo) by analyzing frequency patterns, amplitude fluctuations, and temporal behavior. Once detected, the noise components are selectively reduced or eliminated using methods like frequency band attenuation, temporal masking, or real-time spectral smoothing. In speech-focused scenarios, techniques like voice activity detection (VAD) and de-reverberation can also be applied to enhance vocal clarity. This noise reduction process is carefully calibrated to preserve the natural tone and dynamics of the audio while avoiding over-processing, which could otherwise introduce distortion or artifacts. As a result, users experience cleaner, more focused sound output that is easier to understand and more pleasant to listen to—making it particularly beneficial for people with hearing difficulties or in acoustically challenging environments. This stage plays a vital role in the overall pipeline by ensuring the final audio output is not only personalized but also clear, professional, and noise-free.[13]

- **Processed Audio Output:**

After completing all enhancement stages—including personalization, DSP, noise reduction, and stereo recombination—the final high-quality audio signal is stored in a designated "Processed" database or storage module within the system. This storage unit is structured to organize and catalog enhanced audio files based on user profiles, processing timestamps, and format preferences for easy retrieval and management. The processed audio can be exported in widely supported formats such as WAV for lossless quality or MP3 for compressed, portable use. During the export phase, users may be offered format-specific options such as bitrate (e.g., 128 kbps, 320 kbps for MP3) or sample rate (e.g., 44.1 kHz, 48 kHz), ensuring compatibility with a wide range of playback

of playback devices and applications. Whether the output is intended for headphones, mobile phones, assistive hearing devices, or archival purposes, the system ensures optimal quality and format suitability. In addition to download functionality, the system may also support in-browser playback or direct streaming, enabling users to preview the enhanced audio without needing to leave the platform. All exported files are marked as finalized and may include optional metadata such as processing version, enhancement settings, or hearing profile details. This final output stage completes the pipeline by providing the user with a ready-to-use, personalized, and professionally enhanced audio file that is both accessible and adaptable to real-world use.[14]

- **Original Audio Storage:**

To ensure traceability, flexibility, and future usability, the system also retains a copy of the original, unprocessed audio file in a dedicated storage directory—commonly labeled as "Uploads", "Raw Audio", or a similar repository. This storage is separate from the processed audio database and serves as a secure archive of the initial input exactly as it was received from the user. The original file is preserved in its native format and quality, without any alterations or modifications, ensuring a reliable reference point for future operations. Maintaining access to the raw audio offers several key benefits. First, it allows for quality comparisons, enabling users to directly evaluate the impact of the enhancement process by comparing the processed and unprocessed versions. Second, it supports reprocessing flexibility—if a user updates their hearing profile, or if improved audio processing algorithms become available, the system can re-run the original audio through the pipeline without requiring a new upload. Furthermore, this archival approach supports auditability and version control, where multiple versions of a processed file can be linked back to the same source material. Metadata such as upload date, file type, duration, and user profile can be associated with each original file for efficient indexing and retrieval. Overall, storing the original audio ensures that users maintain full control and access to their source content, laying the foundation for a robust, user-centered audio enhancement system. [20]

CHAPTER 7

METHODOLOGY

1. User-Specific Customization:

The system utilizes detailed data from the user's left and right ears—such as eardrum response patterns, audiogram results, and frequency-specific hearing thresholds—to tailor the audio processing experience. By building a personalized hearing profile, it allows the system to adjust key audio parameters like gain, frequency emphasis, and dynamic range independently for each ear. This ear-specific customization ensures that sounds are clearer, more balanced, and aligned with the user's natural perception. As a result, the processed audio becomes significantly more accurate and comfortable, especially for users with hearing variations or impairments.[5]

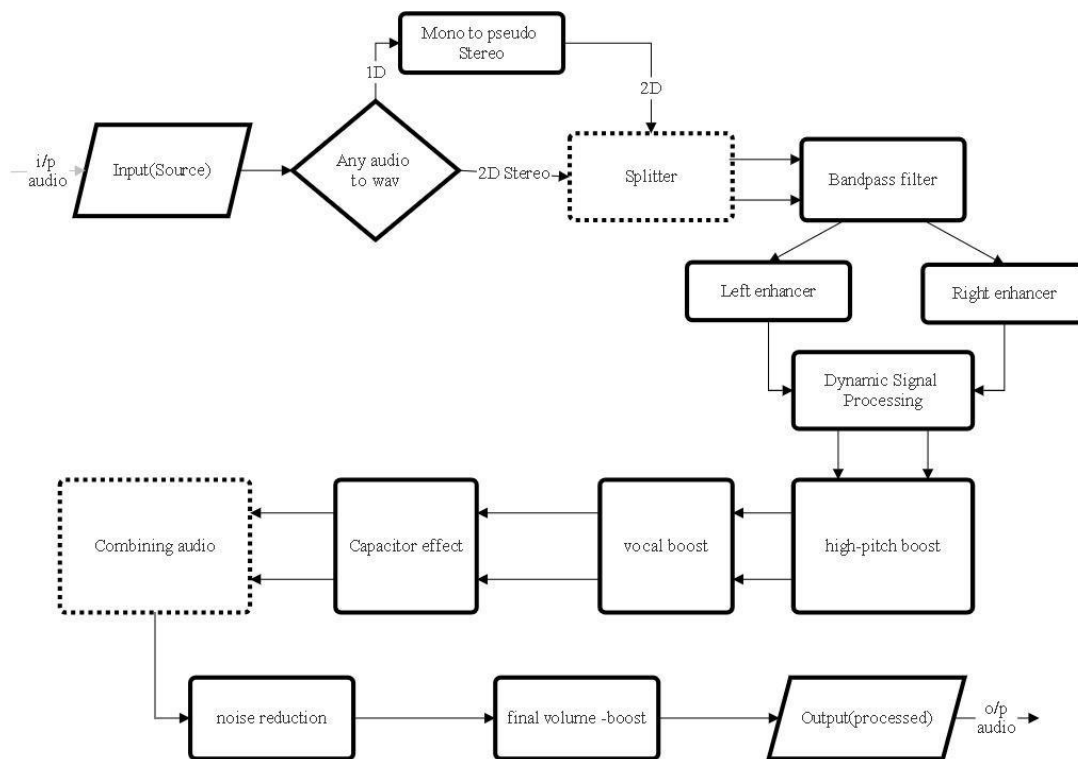


Fig 7 .1 Block diagram of “DSP Technologies

2. Stereo Channel Splitting:

The incoming stereo audio signal is divided into separate Left (L) and Right (R) channels using a dedicated channel splitter. This allows the system to process each channel independently, tailoring the audio output to the user's unique hearing profile for each ear. By enabling ear-specific adjustments—such as frequency boosting or volume correction—the system ensures a balanced and personalized listening experience. This step also helps preserve the natural stereo effect while enhancing clarity and comfort based on individual auditory needs.[10]

3. Dedicated DSP Techniques:

Each audio channel—Left and Right—is independently processed using a set of specialized Digital Signal Processing (DSP) algorithms tailored to the user's hearing profile. These techniques include **frequency filtering** to remove unwanted noise, **dynamic range compression** to balance loud and soft sounds, **equalization** to adjust tonal balance, and **amplification** to boost frequencies where hearing sensitivity is reduced. By customizing these enhancements per ear, the system delivers clearer, more comfortable, and highly personalized audio output that aligns with the user's specific auditory needs. [14]

4. Signal Recombination:

After the individual processing of the Left and Right channels, they are seamlessly combined back into a single stereo signal. This careful recombination preserves the spatial depth and directional cues of the original audio, ensuring the listener experiences a natural and immersive sound environment. At the same time, it maintains the personalized enhancements tailored to each ear, delivering clear, balanced, and customized audio that feels both authentic and comfortable.[16]

5. Advanced Noise Reduction:

The final stereo signal is processed using advanced noise reduction techniques that effectively suppress background noise and unwanted sounds. This enhances speech clarity and overall audio quality, making listening easier and more comfortable, especially in noisy or challenging environments.[17]

6. Format Flexibility:

The system offers format flexibility by allowing users to export the enhanced audio in widely used formats like WAV and MP3. This ensures compatibility across a broad range of devices and playback platforms, making it easy for users to listen to their personalized audio wherever they choose.[21]

7. Persistent Audio Storage:

The system maintains robust and organized storage for audio files by saving both the raw (original) and processed (enhanced) audio in separate, clearly labeled directories. Typically, original files are stored in an "Uploads" folder, preserving the untouched source material exactly as it was received from the user. Meanwhile, the enhanced versions—after all processing and personalization—are saved in a "Processed" directory. This dual-storage approach not only ensures traceability by allowing users and system administrators to reference the original audio at any time but also facilitates quality comparison between unprocessed and enhanced outputs. Furthermore, keeping the originals enables future reprocessing if hearing profiles are updated or new algorithms are developed, offering users ongoing flexibility and control over their audio content.[22]

8. Frontend-Backend Architecture:

The system is architected with a clear separation between the frontend and backend components to optimize functionality and user experience. The frontend serves as the interactive layer where users upload audio files, input personalized hearing data, and receive playback or download options for processed audio. It focuses on usability, responsiveness, and accessibility across different devices and platforms. The backend, on the other hand, handles the heavy lifting by performing complex tasks such as audio format conversion, stereo channel splitting, personalized Digital Signal Processing, noise reduction, and final output generation. This division allows the backend to efficiently manage resource-intensive computations and data storage while the frontend remains lightweight and user-friendly. The modular design also supports easier maintenance, scalability, and future upgrades without disrupting the overall user experience. [12]

CHAPTER 8

IMPLEMENTATION

1. Frontend Interface:

The system features a highly intuitive and user-friendly frontend interface designed to simplify user interaction throughout the audio enhancement workflow. Users can seamlessly upload their audio files along with personalized hearing data, such as detailed left and right ear eardrum responses or results from audiometric tests. The interface provides real-time feedback regarding the status of file uploads and the progress of audio processing, ensuring transparency and engagement. Upon completion, the enhanced audio can be downloaded in preferred formats (such as WAV or MP3). Additionally, the frontend includes helpful tooltips, progress bars, and accessibility options that cater to users of various technical backgrounds and hearing abilities, promoting inclusivity and ease of use.[12]

2. User Data Mapping:

Once the hearing data is uploaded, the backend system organizes this information into structured digital user profiles. These profiles map frequency-specific hearing thresholds—ranging from 125 Hz to 8 kHz—for both ears independently. Advanced algorithms analyze these thresholds to identify frequency bands requiring amplification, attenuation, or dynamic filtering. The system considers the severity and nature of hearing loss (e.g., sensorineural, conductive) and adjusts parameters accordingly. This detailed user-specific mapping is integral to the subsequent DSP module, as it allows for the precise tuning of audio content to match real-world auditory needs, ultimately offering a custom-fit auditory experience for every individual.[16]

3. Audio Input Handling and Conversion:

When a user uploads an audio file, it is first checked for compatibility. If necessary, the system automatically converts it into a standardized format, such as WAV, with predefined parameters like 44.1 kHz or 48 kHz sample rate and 16-bit or 24-bit bit depth. This ensures consistent quality and avoids format mismatches that could compromise processing accuracy. The standardized audio format not only improves compatibility with various modules in the pipeline but also helps maintain

fidelity during filtering and enhancement operations. By ensuring a uniform starting point, the system reduces processing errors and preserves the authenticity of the original audio content.[19]

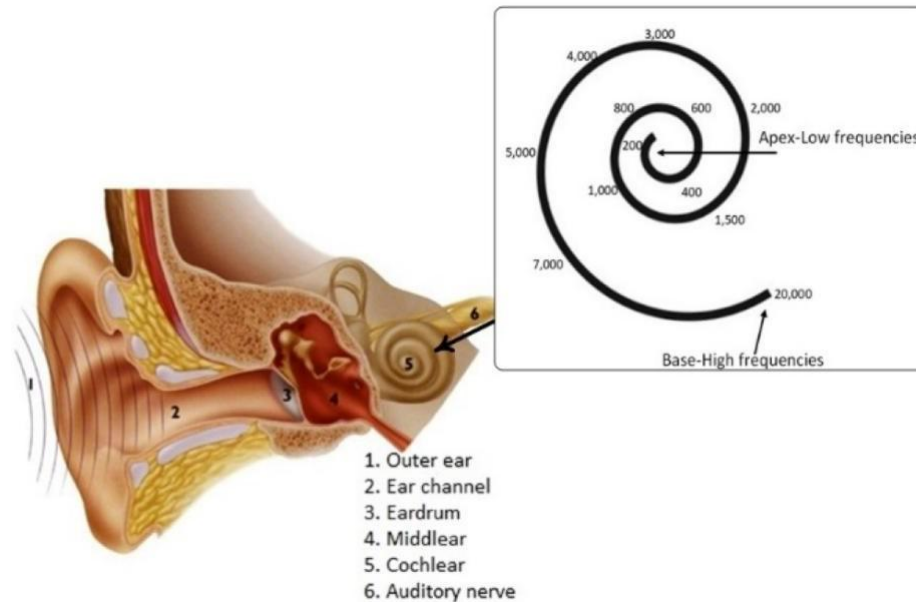


Fig 8 .1 Block diagram of “Ear Drum”

4. Channel Splitter:

Following standardization, the audio file is passed through a Channel Splitter module that separates the stereo track into two distinct channels—Left and Right. Each channel corresponds to one ear and is independently routed for personalized processing. This separation allows for precise, ear-specific enhancement techniques based on individual hearing loss patterns. The module applies preprocessing steps like level normalization and frequency alignment before handing the channels off to the DSP engine. This individualized approach helps improve sound localization, spatial awareness, and auditory balance, even in users with asymmetric hearing profiles.[19]

5. DSP Processing Module – Frequency-Based Personalization:

Each split channel is subjected to an advanced DSP engine equipped with frequency-based personalization capabilities. The engine references the hearing thresholds mapped earlier to implement selective gain adjustment for under-heard frequencies, high-pass or low-pass filtering to

isolate important sound bands, and dynamic range compression to bring soft and loud sounds closer together. In some implementations, psychoacoustic models may be used to enhance clarity while maintaining natural tonality. The module may also include real-time adaptive filtering, ensuring that adjustments remain effective under varying acoustic conditions. These DSP enhancements work in tandem to restore lost sound detail, improve speech intelligibility, and offer a richer, more balanced listening experience tailored to each ear.[21]

6. Signal Combiner:

After personalized processing of the Left and Right channels is complete, the system employs a Signal Combiner module to reconstruct the stereo signal. This module merges both channels carefully while preserving the original stereo imaging and directional audio cues. By maintaining phase relationships and amplitude differences, the system ensures a realistic and immersive auditory environment. This step is particularly important for music, films, and ambient recordings where spatial perception significantly affects the listening experience. The recombined output retains all user-specific adjustments and presents a coherent, high-fidelity stereo output.[4]

7. Noise Reduction:

The unified stereo signal is passed through a sophisticated Noise Reduction module that applies adaptive and real-time noise suppression algorithms. These algorithms detect and minimize background disturbances such as hums, hisses, and ambient environmental sounds without degrading speech or music quality. Techniques like spectral subtraction, Wiener filtering, and neural-network-based denoising may be utilized to target a wide range of noise profiles. This processing significantly improves the clarity of conversations, especially in environments like public spaces or during low-quality recordings, making the enhanced audio more intelligible and pleasant.[16]

8. Storage & Output:

Upon successful processing, the system categorizes and stores both the original (unprocessed) and enhanced (processed) audio files in dedicated storage directories. These are typically named "Uploads" for raw input and "Processed" for finalized outputs. This separation allows users to retrieve the source material for future reprocessing or comparative analysis. The system supports exporting audio in multiple formats including WAV and MP3 to accommodate various playback devices.

A download link is generated for each processed file, and users can also preview enhanced audio via the frontend before saving it.[17]

9. Audiometry Feedback:

To assess the effectiveness of the personalized enhancement, the system offers an Audiometry Feedback module. This component generates visual audiograms that plot original versus enhanced hearing levels for each ear across the full frequency spectrum. These plots help validate the audio adjustments made during DSP processing and offer a clear picture of how well the system compensated for the user's hearing loss. This feature is especially useful for audiologists or healthcare providers evaluating the success of non-invasive hearing assistance solutions. Users can also track historical audiograms to monitor hearing improvements over time or to inform future re-tuning of the system [21]

CONCLUSION

The Advanced Audio Enhancement System (AAES) is a comprehensive and high-performance solution designed to elevate the quality of audio in a wide range of environments, from quiet rooms to complex, noisy settings. At its core, the system integrates advanced Digital Signal Processing (DSP) techniques, including Minimum Mean Square Error Short-Time Fourier Transform (MMSE-STFT) filtering and Least Mean Squares (LMS) adaptive filtering. These methods are particularly effective in reducing various types of background noise, such as ambient environmental sounds or static interference, while simultaneously preserving the clarity, detail, and natural characteristics of the original audio signal. MMSE-STFT filtering operates in the frequency domain to selectively restore speech components while minimizing unwanted noise, whereas LMS adaptive filtering continuously adjusts to changing noise patterns in real time, making it ideal for dynamic environments. One of the standout features of AAES is its support for **real-time processing**, enabling near-instantaneous audio enhancement with minimal latency—an essential capability for live applications such as video conferencing, public speaking systems, telecommunications, hearing aids, and virtual meetings. In addition to technical performance, the system is designed with **format flexibility**, offering support for widely used audio formats such as WAV, MP3, FLAC, and AAC, ensuring seamless integration with existing media pipelines, software platforms, and hardware devices. [7]

REFERENCES

- [1] Haohe Liu, Ke Chen, Qiao Tian, Wenwu Wang, Mark D. Plumbley. (2023). AudioSR: Versatile Audio Super-resolution at Scale. Retrieved from [AudioSR: Versatile Audio Super-resolution at Scale | arXiv](#)
- [2] **Isabella Lenz, Yu Rong, Daniel Bliss, Julie Liss, Visar Berisha.** (2025). A Speech Production Model for Radar: Connecting Speech Acoustics with Radar-Measured Vibrations. Retrieved from <https://ieeexplore.ieee.org/document/10694208>
- [3] **Luca Jiang-Tao Yu, Running Zhao, Sijie Ji, Edith C. H. Ngai, Chenshu Wu.** (2024). *USpeech: Ultrasound-Enhanced Speech with Minimal Human Effort via Cross-Modal Synthesis*. Retrieved from <https://aes2.org/publications/elibrary-page/?id=20803>
- [4] **Manuel Milling, Shuo Liu, Andreas Triantafyllopoulos, Ilhan Aslan, Björn W. Schuller.** (2024). *Audio Enhancement for Computer Audition—An Iterative Training Paradigm Using Sample Importance*. Retrieved from <https://ieeexplore.ieee.org/document/9103036>
- [5] **Junan Zhang, Jing Yang, Zihao Fang, Yuancheng Wang, Zehua Zhang, Zhuo Wang, Fan Fan, Zhizheng Wu.** (2025). *AnyEnhance: A Unified Generative Model with Prompt-Guidance and Self-Critic for Voice Enhancement*. Retrieved from <https://eprints.leedsbeckett.ac.uk/id/eprint/11359/>
- [6] **Haohe Liu, Ke Chen, Qiao Tian, Wenwu Wang, Mark D. Plumbley.** (2023). *AudioSR: Versatile Audio Super-Resolution at Scale*. Retrieved from <https://arxiv.org/abs/2402.00282>
- [7] **Wilfredo J. Robinson M., Medhani Kalal.** (2024). Spatial Audio-Enhanced Multimodal Graph Rendering for Efficient Data Trend Learning on Touchscreen Devices.

- [8] **Keren Shao, Ke Chen, Shlomo Dubnov.** (2024). Music Enhancement with Deep Filters: A Technical Report for The ICASSP 2024 Cadenza Challenge.<https://www.techrxiv.org/doi/full/10.36227/techrxiv.172684060.06923681>
- [9] **Haoyang Li, Jia Qi Yip, Tianyu Fan, Eng Siong Chng.** (2025). Speech Enhancement Using Continuous Embeddings of Neural Audio Codec.
- [10] **Tianrui Wang, Jin Li, Ziyang Ma, Rui Cao, Xie Chen, Longbiao Wang, Meng Ge, Xiaobao Wang, Yuguang Wang, Jianwu Dang, Nyima Tashi.**
- [11] **Haohe Liu, Ke Chen, Qiao Tian, Wenwu Wang, Mark D. Plumbley.** (2023). AudioSR: Versatile Audio Super-resolution at Scale. Retrieved from AudioSR: Versatile Audio Superresolution at Scale | arXiv
- [12] **Isabella Lenz, Yu Rong, Daniel Bliss, Julie Liss, Visar Berisha.** (2025). A Speech Production Model for Radar: Connecting Speech Acoustics with Radar Measured Vibrations. <https://ieeexplore.ieee.org/document/10694208>
- [13] **Luca Jiang-Tao Yu, Running Zhao, Sijie Ji, Edith C. H. Ngai, Chenshu Wu.** (2024). USpeech: Ultrasound-Enhanced Speech with Minimal Human Effort via Cross-Modal Synthesis. Retrieved from <https://aes2.org/publications/elibrarypage/?id=20803>
- [14] **Manuel Milling, Shuo Liu, Andreas Triantafyllopoulos, Ilhan Aslan, Björn W. Schuller.** (2024). Audio Enhancement for Computer Audition—An Iterative Training Paradigm Using Sample Importance. Retrieved from <https://ieeexplore.ieee.org/document/9103036>
- [15] **Junan Zhang, Jing Yang, Zihao Fang, Yuancheng Wang, Zehua Zhang, Zhuo Wang, Fan Fan, Zhizheng Wu.** (2025). AnyEnhance: A Unified Generative Model with Prompt-Guidance and Self-Critic for Voice Enhancement. Retrieved from <https://eprints.leedsbeckett.ac.uk/id/eprint/11359/>

- [16] **Haohe Liu, Ke Chen, Qiao Tian, Wenwu Wang, Mark D. Plumbley.** (2023). AudioSR: Versatile Audio Super-Resolution at Scale. Retrieved from <https://arxiv.org/abs/2402.00282>

- [17] **Wilfredo J. Robinson M., Medhani Kalal.** (2024). Spatial AudioEnhanced Multimodal Graph Rendering for Efficient Data Trend Learning on Touchscreen Devices. Retrieved from NSF Public Access <https://www.kibme.org/resources/journal/20241223112741054.pdf>

- [18] **Keren Shao, Ke Chen, Shlomo Dubnov.** (2024). Music Enhancement with Deep Filters: A Technical Report for The ICASSP 2024 Cadenza Challenge

- [19] **Haoyang Li, Jia Qi Yip, Tianyu Fan, Eng Siong Chng.** (2025). Speech Enhancement Using Continuous Embeddings of Neural Audio Codec.<https://arxiv.org/abs/2502.10822>

- [20] Tianrui Wang, Jin Li, Ziyang Ma, Rui Cao, Xie Chen, Longbiao Wang, Meng Ge, Xiaobao Wang, Yuguang Wang, Jianwu Dang, Nyima Tashi.<https://arxiv.org/abs/2103.08569>

- [21] A Novel End-to-End General Purpose Deep Neural Amplifier for Personalized Hearing AidsShafique Ahmed et al. (2025) <https://arxiv.org/abs/2502.10822>

- [22] **Differentiable Hearing Aid Speech Processing.**Zehai Tu, NingMa,Jon Barker (2021)'<https://arxiv.org/abs/2103.08569>

- [23] **Real-Time Multichannel Deep Speech Enhancement in Hearing Aids:** Comparing Monaural and Binaural Processing in Complex AcousticScenarios.NilsL.Westhausenetal.(2024) <https://arxiv.org/abs/2405.01967>

- [24] Enhancement of Speech Signals for Hearing Aid Devices Using Digital Signal Processing Khalid Zaman et al. (2020)<https://www.researchgate.net/publication/347016558>

- [25] Modelling of Hearing Aid's Digital Signal Processor .Suad Abdulkareem,IsamJanajreh(2022)