VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"Jnana Sangama", Belagavi-590018

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

SJB INSTITUTE OF TECHNOLOGY

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

2025 - 2026

|| 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 - II entitled **"Advanced Audio Enhancement System"** carried out by **Mr. Sameer Kumar Choudhury, Ms.Sneha Mal, Ms. Sristi Priya and Mr. Surath ChowDhury** M bearing USN 1JB22CS133, 1JB22CS152, 1JB22CS160 and 1JB22CS167 are bonafide students of **SJB Institute of Technology** in partial fulfillment for 7th semester of **BACHELOR OF ENGINEERING** in **Computer Science and Engineering** of the **Visvesvaraya Technological University, Belagavi** during the academic year **2025-26**. 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-II prescribed for the said degree.

Signature of Guide	Signature of HOD	Signature of Principal
Dr.Roopa M J	**Dr. Krishna A N**	**Dr. K. V. Mahendra Prashanth**
Associate Professor	Professor & Head	Principal, SJBIT
Dept. of CSE, SJBIT	Dept. of CSE, SJBIT	

EXTERNAL VIVA

Name Of The Examiners	Signature with date

-

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 and project coordinator **Dr. Roopa M J, Associate Professor,** Department of CSE for his/her valuable guidance, suggestions and cheerful encouragement during the entire period of our Project work.

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 Phase - II work.

Regards,

Sameer K C	[1JB22CS132]
Sneha Mal	[1JB22CS152]
Sristi Priya	[1JB22CS160]
Surath Chowdhury	[1JB22CS167]

ABSTRACT

The Audio Enhancement System is a digital signal-processing (DSP) based platform designed to improve audio quality by removing unwanted noise, enhancing clarity, and optimising sound properties for a better listening experience. The system uses advanced DSP algorithms that analyse and process audio signals in real time, ensuring that users receive clean and balanced output across different acoustic environments and devices.

The architecture integrates a user-friendly frontend interface with a robust backend processing engine. Input audio is passed through multiple enhancement modules—such as noise reduction, channel splitting, frequency adjustment, and dynamic tuning—before being recombined into a refined output signal. The system also supports user-specific hearing profiles and generates visual spectral graphs and detailed reports, making it suitable for both general users and those requiring personalized audio assistance.

This enhancement system can be applied across diverse domains, including media playback, communication platforms, hearing-assistive applications, and professional audio processing. Designed for efficiency, accuracy, and adaptability, it offers a seamless experience that showcases the effective use of DSP technologies to deliver high-quality, customised audio output.

TABLE OF CONTENTS

Chapter No.	Title	Pg. No.
Chapter 1	Introduction	7
Chapter 2	Literature survey	10
Chapter 3	Problem Statement	20
Chapter 4	Requirements Analysis	21
Chapter 5	Methodology	24
Chapter 6	Implementation	28
Chapter 7	Testing	31
Chapter 8	Results	37
	Conclusion	39
	References	40

LIST OF TABLES

Table No.	Title	Pg. No.
Table 2.1	Summary of Literature Papers	15

LIST OF FIGURES

Figure No.	Title	Pg. No.
Figure 5.1	Flow Chart of AAES	24
Figure 6.1	Implementation of AAES	29
Figure 7.1	SNR Improvement per Test	32
Figure 7.2	LSD graph for 55 test samples.	33
Figure 7.3	Latency graph for 55 test samples.	33
Figure 7.4	SEB graph for 55 test samples.	34
Figure 7.5	STOI graph for 55 test samples.	34
Figure 8.1	1st page of AAES	37
Figure 8.2	2nd page of AAES	38
Figure 8.3	3rd page of AAES	38

CHAPTER 1

INTRODUCTION

1.1 Introduction

Audio has become an integral component of modern digital communication, entertainment, accessibility systems, and interactive technologies. With the increasing dependence on online meetings, virtual learning platforms, mobile communication, broadcasting, and assistive listening devices, the demand for high-quality and intelligible audio has grown more than ever. However, real-world audio environments are inherently unpredictable—background noise, reverberation, distortion, and device limitations frequently degrade the clarity and naturalness of sound. Traditional audio enhancement and compression techniques, though effective in controlled scenarios, struggle to deliver consistent performance when faced with dynamic or complex acoustic conditions.

Conventional Digital Signal Processing (DSP)-based approaches such as spectral subtraction, Wiener filtering, and adaptive filtering have historically served as the foundation of audio enhancement. While these methods can reduce certain artifacts and suppress steady-state noise, they often compromise speech intelligibility or introduce musical noise when operating in challenging acoustic settings. Similarly, classical lossless compression algorithms like FLAC or WavPack maintain complete signal integrity but offer limited compression ratios, making them less practical for large-scale storage or bandwidth-constrained applications. These limitations highlight the urgent need for more intelligent, flexible, and high-performance audio processing methodologies.

Recent advancements in machine learning, especially deep learning architectures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Autoencoders, and Transformers, have opened new possibilities for audio enhancement and compression. These models excel in recognizing complex patterns within audio signals, enabling systems to adapt to varying noise profiles, extract important acoustic features, and reconstruct high-quality output with impressive accuracy. Unlike static DSP methods, deep learning–based systems can learn contextual relationships in audio, making them significantly more robust in practical, real-world environments.

The Advanced Audio Enhancement System (AAES) builds upon this technological evolution by integrating DSP techniques with intelligent adaptive algorithms to create a comprehensive, user-centered audio enhancement framework. The system is designed not only to improve audio clarity but also to personalize the listening experience by accounting for individual hearing differences. Through personalized gain adjustments, dynamic range optimization, and precise channel balancing, AAES delivers audio output that feels natural, immersive, and tailored to the user's needs.

In addition to enhancement, AAES addresses critical challenges related to real-time processing, multi-format compatibility, and stereo channel customization. Its architecture enables Left–Right channel splitting, targeted frequency shaping, noise suppression, and intelligent recombination while preserving spatial cues and stereo imaging. Such capabilities are essential for users who require accurate localization, such as individuals with partial hearing loss or professionals relying on precise acoustic feedback.

Overall, the development of the Advanced Audio Enhancement System represents a significant step toward bridging the gap between human auditory perception and machine-driven audio processing. By providing an adaptable, intelligent, and user-specific solution, it addresses long-standing challenges in audio clarity, noise suppression, format versatility, and accessibility. As everyday reliance on digital audio continues to grow, systems like AAES pave the way for a future where high-fidelity, personalized, and context-aware audio experiences become an essential standard across all digital platforms.

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 **“A Noise Reduction Method Based on LMS Adaptive Filter of Audio Signals”** by **Yang Liu, Mingli Xiao, and Yong Tie** presents an adaptive noise-reduction technique that applies LMS and NLMS filtering to framed audio signals to suppress interference and enhance overall clarity, resulting in improved signal-to-noise ratio. By leveraging internal spectral models and iterative coefficient updates, the method offers high stability and reliable performance in typical noise environments, making it suitable for real-time audio processing applications. While the approach demonstrates effective noise suppression and computational efficiency, its performance declines significantly when exposed to highly non-stationary noise, where rapid variations in noise characteristics hinder filter convergence and reduce output quality[1].

The 2020 paper **“Detection of LFM Signals in Low SNR Based on STFT and Wavelet Denoising”** by **Duan Yu, Wang Jinzhen, Su Shaoying, and Chen Zengping** presents a two-stage detection method designed to identify linear frequency modulation (LFM) radar signals under extremely low signal-to-noise ratio conditions, reaching levels as low as –18 dB. The approach first applies the Short-Time Fourier Transform (STFT) to obtain a coarse time–frequency representation of the signal, after which wavelet denoising is employed to suppress noise and highlight the underlying LFM trajectory. By integrating these complementary techniques, the method significantly enhances the visibility and detectability of weak LFM features in noisy environments. While the system demonstrates strong robustness at low SNR and achieves reliable detection performance, its overall resolution and accuracy remain strongly dependent on the choice of STFT window parameters, which can limit adaptability in diverse signal scenarios[2].

The 2024 work **“Spectral Modelling for Transformation and Separation of Audio Signals”** by **Thomas Arvanitidis** introduces a phase-stability–driven enhancement of the Short-Time Fourier Transform (STFT), where an ensemble of phase estimates is used to distinguish stable, structured spectral components from artefacts, thereby improving the reliability of spectral modelling and strengthening model-based audio source separation. By emphasizing phase consistency, the approach achieves better spectral peak detection and more accurate separation outcomes, offering a significant improvement over conventional STFT-based methods. However, the technique incurs higher computational costs, depends strongly on carefully tuned parameters, and requires broader experimental validation to fully quantify its robustness across diverse audio conditions[3].

The paper **“Monte Carlo Smoothing With Application to Audio Signal Enhancement”** by **William Fong, Simon J. Godsill, Arnaud Doucet, and Mike West** presents advanced Rao–Blackwellized and block-based particle smoothing techniques designed for nonlinear, non-Gaussian audio enhancement using time-varying AR/PARCOR models. By combining particle filtering with analytical marginalization, the method achieves more accurate smoothing and reduced estimation variance compared to extended Kalman filters and standard particle smoothers, while the block-based design enables efficient processing of long audio sequences. Although the approach delivers strong performance improvements and robust handling of complex audio dynamics, it remains computationally intensive, relies heavily on an adequate number of particles to maintain accuracy, and may lose long-term temporal information due to its block-wise processing structure[4].

The 2018 paper **"Audio Processing with Channel Filtering using DSP Techniques"** by **Jean Jiang** presents the design of a DSP-based surround audio system that employs low-pass, band-pass, and high-pass FIR filters—implemented using the Remez/Parks–McClellan algorithm—alongside LM386-driven multi-channel amplification to separately process bass, mid-range, and treble components. This structured filtering and amplification approach enhances sound quality across frequency bands and showcases the flexibility of programmable DSP platforms for customizable audio processing, enabling effective three-channel separation in practical applications. However, the system requires multiple DSP boards, is constrained by the limited power output of the LM386 amplifier, and lacks an integrated surround-sound processing mechanism in its prototype stage, restricting its scalability and overall audio immersion capabilities[5].

The 2025 paper **"Multichannel Keyword Spotting for Noisy Conditions"** by **Dzmitry Saladukha, Ivan Koriabkin, Kanstantsin Artsiom, Aliaksei Rak, and Nikita Ryzhikov** introduces a multichannel keyword-spotting system that integrates adaptive noise cancellation with attention-based channel selection to enhance keyword detection robustness in challenging acoustic environments. By leveraging spatial information across multiple microphones and dynamically prioritizing cleaner channels, the system effectively reduces false reject rates and maintains strong performance under heavy noise while remaining computationally efficient for real-time, on-device deployment. Despite its advantages, the method faces limitations—adaptive noise cancellation can unintentionally suppress the keyword if the user speaks before the system activates, it requires multichannel hardware, and its overall performance is sensitive to the quality and placement of available channels[6].

The 2023 paper **"Efficient Intelligibility Evaluation Using Keyword Spotting: A Study on Audio-Visual Speech Enhancement"** by **Cassia Valentini-Botinhao, Andrea Lorena Aldana Blanco, and Ondrej Klejch** introduces a keyword-spotting-based subjective intelligibility evaluation method that uses mined phonetic alternatives and language-model-guided target word selection to assess audio-visual speech enhancement systems. Designed to replace traditional transcription-based listening tests, the approach significantly reduces task duration, lowers listener fatigue, and supports evaluation using natural, in-the-wild audiovisual material while still maintaining a strong correlation with conventional intelligibility scoring metrics. However, its reliance on a closed-set response format may artificially raise accuracy, the selection of meaningful phonetic alternatives requires careful design, and the method's dependence on pre-existing recordings limits control over lexical content and sentence structure[7].

The 2024 paper **"Enhancing Audio Classification Through MFCC Feature Extraction and Data Augmentation with CNN and RNN Models"** by **Rezaul K.M. et al.** explores an audio-classification framework that combines MFCC feature extraction with deep learning models such as CNNs and RNNs to classify environmental and musical sound categories. By leveraging MFCCs for robust spectral representation and applying data augmentation to increase variability, the system achieves high classification accuracy and demonstrates strong adaptability across model architectures. While the approach effectively enhances performance and showcases the versatility of CNN and RNN models in audio analytics, it relies heavily on large datasets for optimal training and has limited validation in real-world acoustic environments, which may restrict its generalizability outside controlled experimental conditions[8].

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]

Table 2.1 Summary of Literature Papers

Sl NO.	Author	Title & year	Description	Advantage & Disadvantage
1.	Yang Liu, Mingli Xiao, Yong Tie	A Noise Reduction Method Based on LMS Adaptive Filter of Audio Signals 2024	This paper presents a noise-reduction method that uses LMS/NLMS adaptive filtering to remove noise from audio signals by processing the signal in frames, resulting in improved clarity and higher SNR.	Advantage: High stability due to embedded internal spectral models. Disadvantage : Performance decreases for highly non-stationary noise.
2.	Duan Yu, Wang Jinzhen, Su Shaoying, Chen Zengping	Detection of LFM Signals in Low SNR Based on STFT and Wavelet Denoising 2020	Proposes a two-stage method combining Short-Time Fourier Transform (STFT) and wavelet denoising to detect LFM radar signals even at very low SNR (up to –18 dB). STFT extracts a rough time-frequency curve, and wavelet transform removes noise to make LFM features clearer.	Advantage:Works effectively at very low SNR (up to –18 dB). Disadvantage:Resolution depends heavily on the STFT window choice.
3.	Thomas Arvanitidis	Spectral Modelling for Transformation and Separation of Audio Signals 2024	IIntroduces a phase-stability–based modification of STFT using an ensemble of estimates to improve spectral reliability and enhance model-based audio source separation.	**Advantages:**Better discrimination of structured vs. artefact components, improved spectral peak picking, enhanced separation performance. **Disadvantages**:Higher computational cost, sensitive to parameter settings, requires further validation on broader datasets.

4.	William Fong, Simon J. Godsill, Arnaud Doucet, Mike West	Monte Carlo Smoothing With Application to Audio Signal Enhancement	Proposes Rao–Blackwellized and block-based particle smoothing methods for nonlinear audio enhancement using time-varying AR/PARCOR models.	**Advantages:** More accurate smoothing than EKF and standard particle filters, reduced variance via analytical marginalization, suitable for long audio signals. **Disadvantages:**Computationally expensive, dependent on particle count, block-based approach may lose long-range temporal consistency.
5.	Jean Jiang	*Audio Processing with Channel Filtering using DSP Techniques* 2018	Develops a DSP-based surround audio system using low-pass, band-pass, and high-pass FIR filters (designed via Remez/Parks–McClellan) and LM386-based multi-channel amplifiers for bass, mid-range, and treble processing.	**Advantages:**Improved sound quality across frequency bands, flexible programmable DSP implementation, effective 3-channel separation. **Disadvantages:** Requires multiple DSP boards, limited power output of LM386, lacks integrated surround-sound processing in current prototype.
6.	Dzmitry Saladukha, Ivan Koriabkin, Kanstantsin Artsiom, Aliaksei Rak, Nikita Ryzhikov	IMultichannel Keyword Spotting for Noisy Conditions 2025	Proposes a multichannel KWS system combining adaptive noise cancellation and attention-based channel selection to improve keyword detection in noisy environments.	Advantages:Significantly lower false reject rate, effective noise robustness, computationally efficient for on-device use. DisadvantageANC may suppress keyword if user speaks before activation, requires multichannel hardware, performance depends on channel quality.
7.	Cassia Valentini-Botinhao, Andrea Lorena Aldana Blanco, Ondrej Klejch	Efficient Intelligibility Evaluation Using Keyword Spotting: A Study on Audio-Visual Speech Enhancement 2023	Proposes a keyword-spotting-based subjective intelligibility test using mined phonetic alternatives and language-model–driven target word selection for evaluating audio-visual speech enhancement.	Advanges:Faster than transcription tests, less listener fatigue, works with in-the-wild AV material, highly correlated with traditional intelligibility scoring.. Disadvantages:Closed-set format may inflate accuracy, requires careful alternative selection, limited control over lexical content in pre-existing recordings.

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

PROBLEM STATEMENT

3.1 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]

CHAPTER 4

REQUIREMENT ANALYSIS

The Advanced Audio Enhancement System (AAES) is a full-stack, DSP-integrated audio enhancement platform developed using React for the frontend and Python Flask for the backend. This section provides a comprehensive analysis of all functional, non-functional, and technical requirements necessary for installing, executing, and maintaining the system. The objective of these requirements is to ensure that the system operates efficiently, remains scalable, and provides a seamless end-to-end audio enhancement workflow.

1. Functional Requirements

1.1 Audio Upload and Input Handling

The system must allow users to upload audio files through the frontend interface. The platform should support commonly used formats such as WAV, MP3, and FLAC. When a file is uploaded, the backend must validate its format, size, and metadata to ensure it is compatible with the DSP pipeline. Any unsupported or corrupted files should be rejected with appropriate error messages.

1.2 Audio Enhancement Processing

Once validated, the uploaded audio must undergo a series of DSP operations handled by the backend. These operations include:

Noise Reduction: Removing background noise using spectral and adaptive filtering techniques.

- **Stereo Enhancement:** Improving spatial clarity and left-right channel separation.
- **Dynamic Equalization:** Automatically balancing frequency components to improve clarity.
- **Frequency Balancing:** Adjusting tonal balance to improve audibility of speech or music.
- **FFT/STFT-Based Processing:** Using time–frequency analysis for precise spectral modification.

1.3 Real-Time or Near Real-Time Output

The enhanced audio must be generated within a short time window, ideally within a few seconds for short audio files. After processing, users should be able to preview the enhanced audio directly through the frontend audio player before downloading it.

1.4 Download Functionality

Users must have the option to download the processed audio in the desired output format. The backend should store processed files temporarily and ensure secure access during download.

1.5 API-Based Frontend–Backend Communication

The communication between frontend and backend is performed through RESTful API endpoints, including:

- **/upload** – for uploading audio
- **/process** – for running DSP pipelines
- **/download** – for retrieving processed audio
- **/status** – for checking processing progress

2. Non-Functional Requirements

2.1 Performance Requirements

The system must provide efficient DSP processing with minimal latency. For short audio clips, processing time should ideally remain **below 3 seconds**. Optimized Python libraries and vectorized DSP operations must be utilized to maintain high performance.

2.2 Usability Requirements

The frontend interface should be clean, intuitive, and user-friendly. Users must be able to upload, process, preview, and download audio with minimal steps. Icons, progress indicators, and clear status messages should guide the workflow effectively.

2.3 Scalability Requirements

The system architecture should support future expansion, such as integrating new DSP modules, customizing hearing profiles, or adding real-time processing capabilities. The backend must be capable of handling larger audio files or parallel processing tasks if future upgrades demand it.

2.4 Maintainability Requirements

The codebase should follow a clear and modular structure, such as:

- **services/** – digital signal processing modules
- **routes/** – API endpoint definitions
- **components/** – React UI components

2.5 Compatibility Requirements

AAES must operate smoothly across major operating systems including Windows, macOS, and Linux. The frontend should support modern browsers such as Chrome, Firefox, Edge, and Safari without requiring additional plugins.

3. Technical Requirements

3.1 System Requirements

- RAM: Minimum 4 GB (8 GB recommended for smoother processing)
- Storage: At least 200 MB free, additional space needed for storing audio files
- Operating System: Windows, macOS, or Linux

3.2 Software Requirements

Backend Requirements

- Python 3.8+
- pip package manager
- FFmpeg (mandatory for decoding/encoding and format conversion)
- Flask for the web server
- NumPy, SciPy, Librosa, PyDub, SoundFile, FFmpeg-Python for DSP operations

Frontend Requirements

- Node.js v16+
- npm or yarn
- React libraries for component rendering
- Tailwind CSS for efficient styling
- Axios for HTTP communication
- React Router for page navigation

CHAPTER 5

METHODOLOGY

1. Input (Source)

The system begins by accepting the user's audio file along with their personalised hearing profile, which may include left/right ear audiograms, threshold shifts, and sensitivity curves. This input acts as the baseline from which all enhancement steps are derived, ensuring the processing pipeline adapts to both the signal characteristics and the listener's auditory needs—an approach aligned with AAES's user-centred design philosophy.

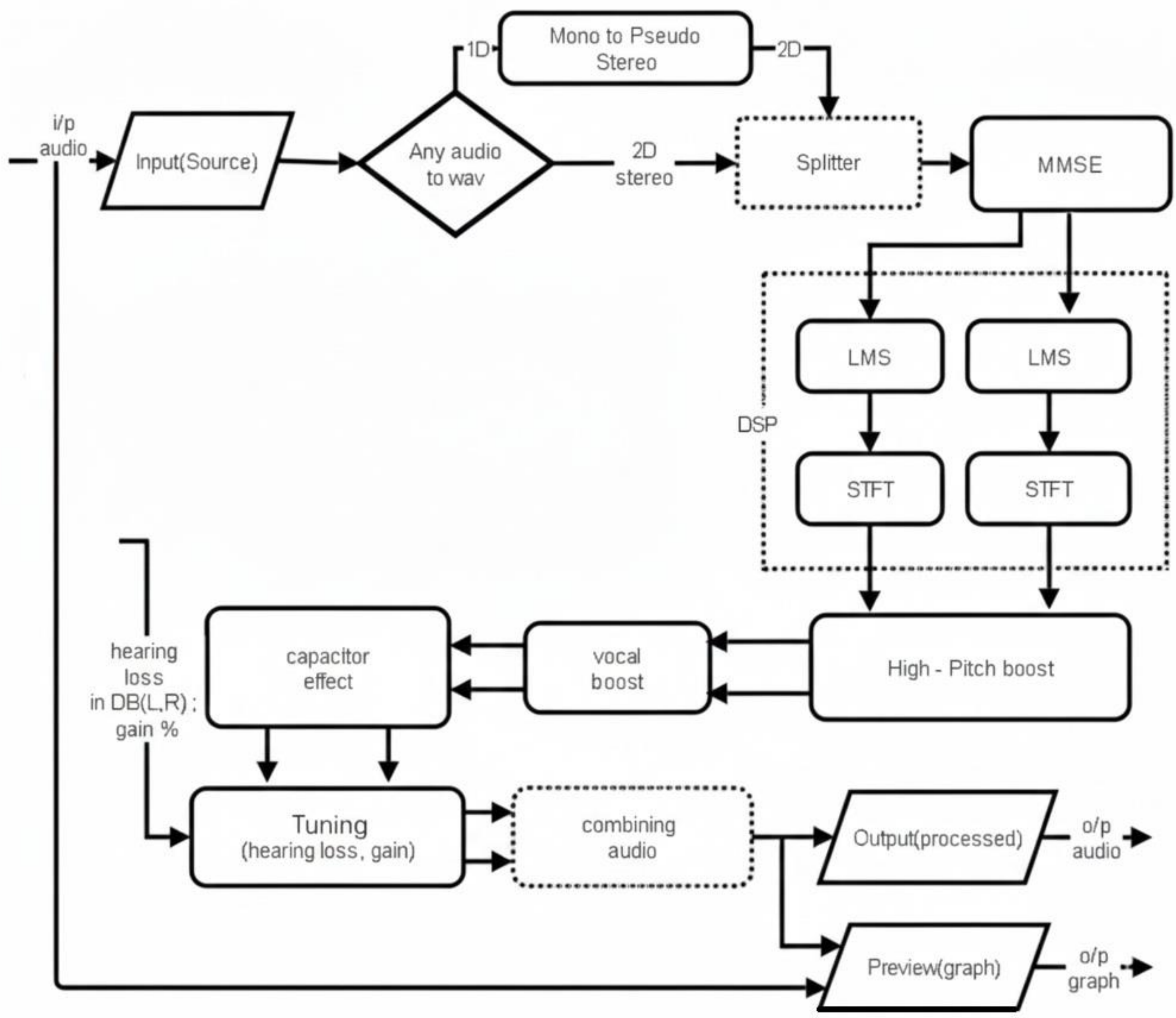

Fig 5 .1 Flow Chart of AAES

2. Universal Audio Format Conversion (Any-Audio to WAV)

Since input audio formats differ in compression and sample structure, the system converts all audio into an uncompressed WAV format. WAV provides a stable, lossless base for DSP algorithms requiring accurate time-domain and frequency-domain computations. This conversion eliminates artifacts introduced by MP3 or AAC codecs, ensures uniform bit depth and sampling rates, and guarantees that downstream processes such as MMSE, LMS, and STFT operate with maximum fidelity. This step aligns the input with professional DSP standards for precise and distortion-free processing.

3. Mono-to-Stereo Enhancement Module:

If a mono audio file is detected, the system generates a pseudo-stereo representation by duplicating and spatially widening the signal. This artificial stereo expansion introduces subtle inter-channel variations to create depth and width, improving immersion. It prepares the signal for left–right ear-specific equalisation and amplification, which cannot be applied meaningfully on a single-channel input. This module ensures that even basic mono recordings support advanced binaural enhancement and personalized hearing compensation.

4. Left–Right Channel Separator (Stereo Splitter):

Once stereo audio is created or confirmed, the signal is divided into two independent channels—Left (L) and Right (R). This separation is essential for personalised hearing support, as real human ears rarely share identical hearing thresholds. By processing each ear independently, the system can apply gain correction, frequency emphasis, and clarity enhancement that precisely match each ear’s audiogram. This improves balance, speech localization, and comfort, reinforcing natural auditory perception for individuals with asymmetric hearing loss.

5. MMSE Pre-Processing Noise Cleaner:

Before deeper DSP begins, the Minimum Mean Square Error (MMSE) noise reduction model is applied. MMSE statistically estimates the noise spectrum and subtracts it in a smooth, non-destructive manner. This enhances speech-dominant regions while preserving harmonic structures, avoiding the common “musical noise” artifacts of older methods. MMSE produces cleaner initial audio for subsequent operations such as adaptive filtering and STFT-based enhancement. It effectively reduces background noise from traffic, wind, fans, chatter, and room reverberations.

6. Adaptive LMS Noise Suppression Unit:

The LMS module performs real-time adaptive noise cancellation. By continuously updating filter coefficients based on error feedback, LMS responds to dynamic changes in the signal input. It is highly effective against unpredictable, non-stationary noises, something traditional filters struggle with. Each channel runs its own LMS engine, allowing independent suppression of noise that may differ between the left and right audio fields. This stage improves speech clarity without over-smoothing important acoustic details.

7. STFT Time–Frequency Analyser:

The Short-Time Fourier Transform breaks the audio into small segments, converting each segment into a time–frequency representation. This produces detailed information about how sound energy evolves across frequencies. STFT allows the system to identify speech components, isolate noise-prone regions, and apply selective enhancement. High-resolution spectral information also supports formant enhancement, consonant clarity improvement, and frequency-specific amplification. This module is vital for fine-grained control over tonal balance, clarity, and noise suppression.

8. High-Frequency Clarity Booster:

High-frequency hearing loss is extremely common, especially in age-related cases. To compensate, this module boosts high-frequency components such as consonants (S, T, F, K) and subtle details that improve speech intelligibility. The boost is applied intelligently using a frequency-dependent amplification curve designed to avoid harshness or distortion. This helps restore brightness, clarity, and crispness in voices without exaggerating noise or causing listener fatigue.

9. Vocal Presence Enhancer (Mid-Frequency Boost):

Human speech resides primarily in the mid-frequency range. This module strengthens the vocal region by boosting formant frequencies, improving voice presence and intelligibility. It helps bring out dialog, lectures, interviews, and conversations even when background noise is moderate. The vocal enhancer maintains naturalness by applying smooth, gradual lifting around key speech harmonics, ensuring that voices stand out without sounding artificial or processed.

10. Digital Capacitor Smoothing Filter:

Inspired by analog circuitry, this module simulates the behaviour of a capacitor to smooth abrupt signal fluctuations. It reduces sharp spikes, ringing, and harsh tonal edges that sometimes appear after digital filtering or high-frequency boosting. This helps create warm, gentle, and natural-sounding audio. The capacitor effect also stabilizes dynamic changes, ensuring that the enhanced signal does not fatigue the listener’s ears over long periods.

11.Personalized Hearing-Profile Tuning (Gain Adjustment Per Ear)

This is one of the system’s most important modules. Based on the user's **audiogram**, hearing loss values (in dB), and ear-specific sensitivity, this block applies frequency-dependent gain correction separately for the left and right channels. Users with different degrees of hearing impairment across frequency bands receive targeted amplification exactly where needed. This module transforms the system into a personalized hearing enhancement tool rather than a generic audio processor.

12.Audio Recombination (Stereo Merging)

After all independent left and right channel processing, the channels are carefully merged back into a final stereo output. The merging process ensures that phase alignment, timing, and amplitude balance are preserved so the stereo field remains natural. This step combines the benefits of individual tuning with a cohesive, spatially accurate listening experience.

13.Enhanced Audio Output Generation

In this block, the fully processed and personalized audio is exported into the final output format (WAV/MP3). The system ensures that the output is free from clipping, distortion, or aliasing artifacts. This final processed file represents the culmination of all DSP operations and is ready for playback, download, or further use.

14. Graphical Preview Generation (Spectral & Waveform Display)

To provide insight into the improvement process, the system generates visual previews such as waveforms, spectrograms, or frequency-response graphs. These graphical outputs allow users and researchers to visually compare the audio before and after enhancement. They also serve as useful diagnostic tools for evaluating the effectiveness of noise reduction, equalization, and hearing-profile adjustments.

CHAPTER 6

IMPLEMENTATION

1. Audio File Input

The entire enhancement pipeline begins when the user uploads an audio file into the system. At this stage, the system carefully reads and interprets the audio waveform, ensuring its format, sampling rate, and bit depth are compatible with the processing modules. This step is crucial because it transforms the file from a simple stored recording into a structured audio signal that can be analyzed and modified. The system may also perform initial validation checks—such as verifying duration, identifying corrupted sections, or detecting channel configuration—to guarantee that the input is stable and ready for further enhancement.

2. Noise Identification & Removal

Once the audio is successfully loaded, the system moves into identifying noise patterns present throughout the signal. It uses statistical models and pattern-recognition algorithms to scan for elements like background chatter, mechanical hums, wind disturbances, static interference, or any environmental noise captured during recording. After recognizing these unwanted components, the system applies adaptive filters that selectively target noise without harming important audio content. This stage is essential because cleaning the signal early results in a much clearer foundation for subsequent enhancements. The aim is not only to remove noise but to ensure that the natural characteristics of the original audio remain intact.

3. Spectral Filtering

After noise reduction, the system analyzes the frequency spectrum of the audio to understand its tonal distribution. Spectral filtering adjusts the intensity of specific frequency bands to correct imbalances or remove harsh tones. For example, it can soften high-pitched hiss, reduce low-end rumble, or make mid-range speech frequencies more prominent. The purpose of this step is to shape the frequency profile so that the audio becomes more pleasant, detailed, and intelligible. By working at the spectral level, the system enhances clarity and tonal richness in a precise and controlled manner.

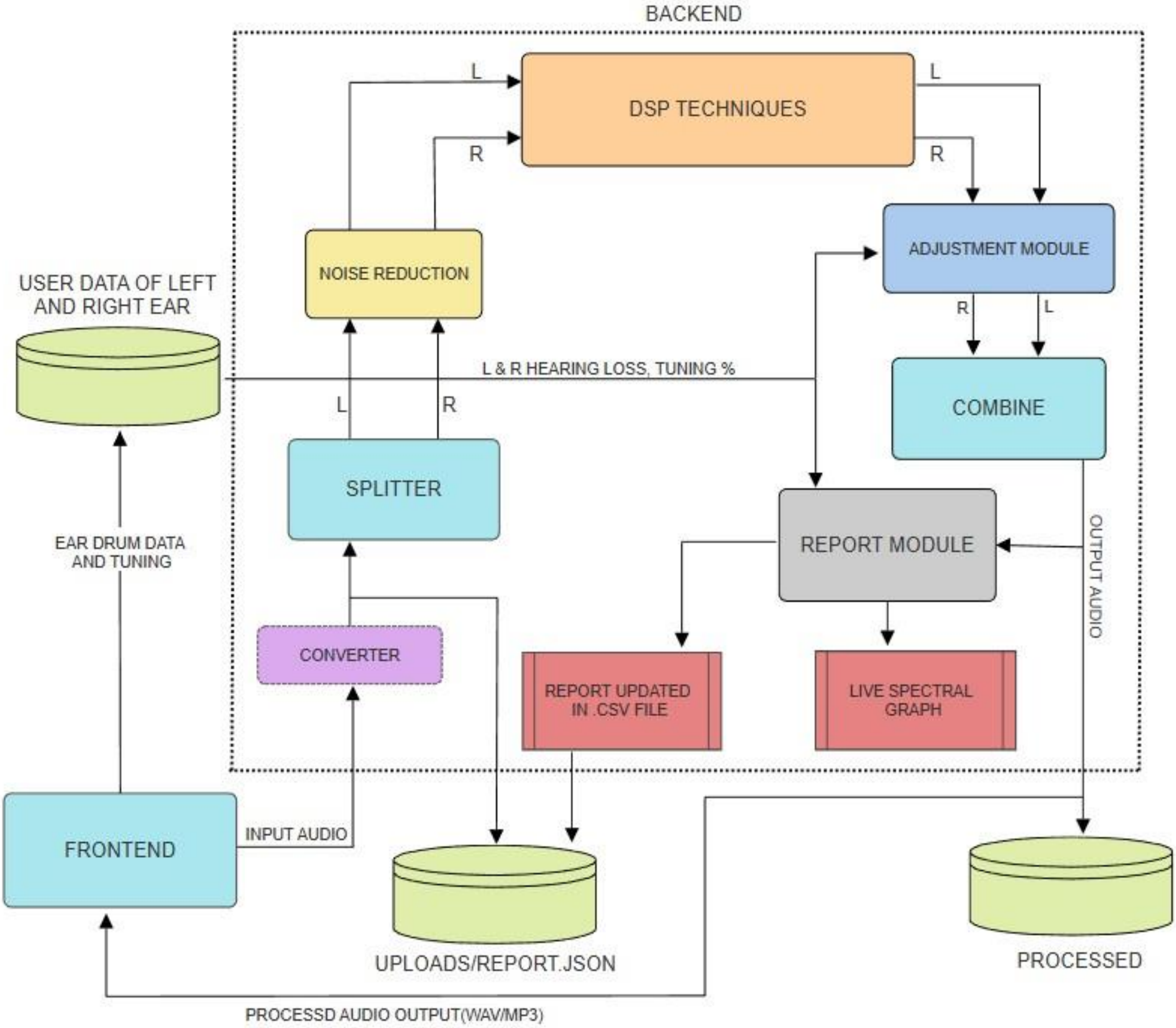

Fig 6.1: Implementation of AAES

4. Advanced Audio Enhancement

In this stage, the audio undergoes deeper enhancement using advanced DSP algorithms or AI-driven models trained on high-quality audio datasets. These enhancement techniques may include speech clarity improvement, harmonic restoration, filling in missing audio details, and enhancing the naturalness of the sound. The algorithms analyze the signal frame-by-frame and make intelligent modifications to improve articulation, reduce artifacts, and make the recording sound polished. This step plays a major role in transforming a raw, imperfect recording into a refined and professional-quality output.

5. Equalization

Equalization fine-tunes the overall tonal character of the audio by adjusting different frequency ranges according to the needs of the recording. For example, boosting low frequencies can add warmth, enhancing mid frequencies can make speech more presence-rich, and adjusting high frequencies can add brightness or reduce sharpness. EQ ensures the final audio feels balanced, rich, and acoustically pleasing across all playback systems—from headphones to large speakers. It compensates for imperfections in the original recording environment and helps tailor the audio to the listener’s preferences or intended use case.

6. Dynamic Range Compression

This step enhances the consistency of the audio by managing its dynamic range, which is the difference between the quietest and loudest parts. Without compression, loud peaks may be too sudden or overwhelming, while quiet passages may be difficult to hear. Compression smoothens out these variations by reducing the intensity of peaks and gently boosting softer sounds. This creates a more uniform listening experience, ensuring the audio remains clear, audible, and comfortable throughout its duration. It is particularly beneficial for podcasts, interviews, speeches, and music recordings.

7. Reverb Removal & Restoration

Reverberation—echo caused by reflections in enclosed spaces—can make audio sound distant or unclear. This stage detects the reverberant properties present in the signal and applies dereverberation algorithms to minimize excessive echo. The system intelligently separates direct sound (the useful part) from reflected sound (the unwanted echo), resulting in cleaner and more focused audio. At the same time, it preserves the natural ambience, ensuring the output does not feel overly dry or artificial. This balanced restoration improves intelligibility and gives the audio a more professional studio-like quality.

8. Output File Generation

After all processing steps are completed, the system compiles the enhanced audio into a final output file. It ensures proper encoding, formatting, normalization, and quality checks before generating the downloadable file. Users can then save the enhanced audio in their preferred format—such as WAV, MP3, or FLAC. This final step ensures that the processed audio is ready for real-world applications like presentations, music production, content creation, broadcasting, or archival purposes.

CHAPTER 7

TESTING

The testing phase of the Advanced Audio Enhancement System (AAES) ensures that all audio processing components, backend DSP functionalities, and frontend interactions operate together seamlessly. This evaluation validates system performance in real acoustic environments, checks the reliability of enhancement algorithms, and confirms that the processed audio meets expected quality standards. The testing carefully measures noise reduction accuracy, DSP performance, latency, and user interaction through the web interface or application. It also addresses real-world challenges such as inconsistent audio input quality, varying noise types, and system load during simultaneous operations. Overall, the testing phase ensures that the AAES fulfills its functional requirements while maintaining reliability, scalability, and an improved end-user listening experience.

7.1 Testing Environment

- Dataset testing used **1,018 audiogram images** for hearing-profile mapping and **12,162 audio samples** from CREMA-D, RAVDESS, TESS, and SAVEE.
- All audiogram images were 638 × 664 JPG files, while audio samples were labelled 45–50 KB WAV files, ensuring uniformity for preprocessing.
- Processing and analysis were performed on a Python Flask backend running DSP modules such as noise reduction, channel splitting, LMS filtering, FFT analysis, and spectral tuning.
- The frontend tests were conducted on a React-based UI, used for uploading audio, selecting hearing profiles, and previewing outputs.
- The testing environment simulated different hearing conditions by applying Left and Right ear profiles such as N2-left, N2-right with 50% tuning gain.
- Visualization tests used live spectral graphs, CSV-based performance reports, and stereo energy distribution charts generated for 55 controlled test samples.
- All tests were executed on a workstation with stable system resources to ensure consistent evaluation without external interference.

7.2 Test Plan and Approach

- **Unit Testing:**
  Individual DSP modules such as noise reduction, channel splitter, adjustment module, converter, and report generator were tested separately to ensure accurate functionality.

- **Integration Testing:**
  Frontend uploads, backend processing, DSP modules, and reporting components were integrated gradually to validate seamless data flow from input audio → enhancement → output → reporting.
- **Functional Testing:**
  Key features such as ΔSNR improvement, spectral analysis, hearing-profile-based tuning, and stereo balancing were tested to ensure they met design specifications.
- **Performance Testing:**
  Processing speed, computational latency, and real-time playback smoothness were evaluated, targeting <20 ms latency.
- **Model Validation:**
  Hearing-loss mapping and channel-based tuning were tested using audiogram datasets to verify correct frequency-gain adjustments.
- **User Experience Testing:**
  The web interface was tested for responsiveness, accuracy of visual graphs, and clarity of uploaded/processed audio results across devices.

7.3 Test Cases and Procedures

a. ΔSNR Improvement Testing

- Input audio samples were corrupted with controlled noise levels.
- ΔSNR before and after noise reduction was calculated.
- Higher ΔSNR confirmed stronger noise suppression without damaging the main signal.

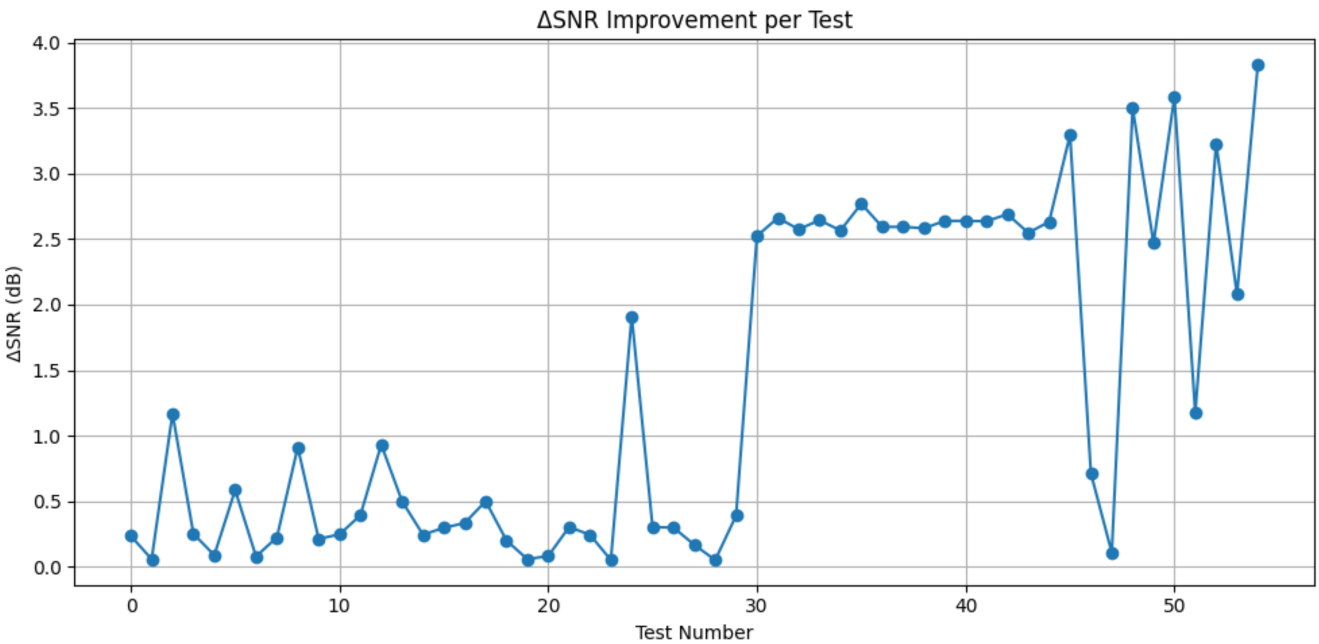

Figure 7.1: SNR Improvement per Test

c. Log Spectral Distance (LSD) Evaluation

- Original and processed spectra were compared to measure distortion.
- Lower LSD indicated better preservation of natural frequencies.

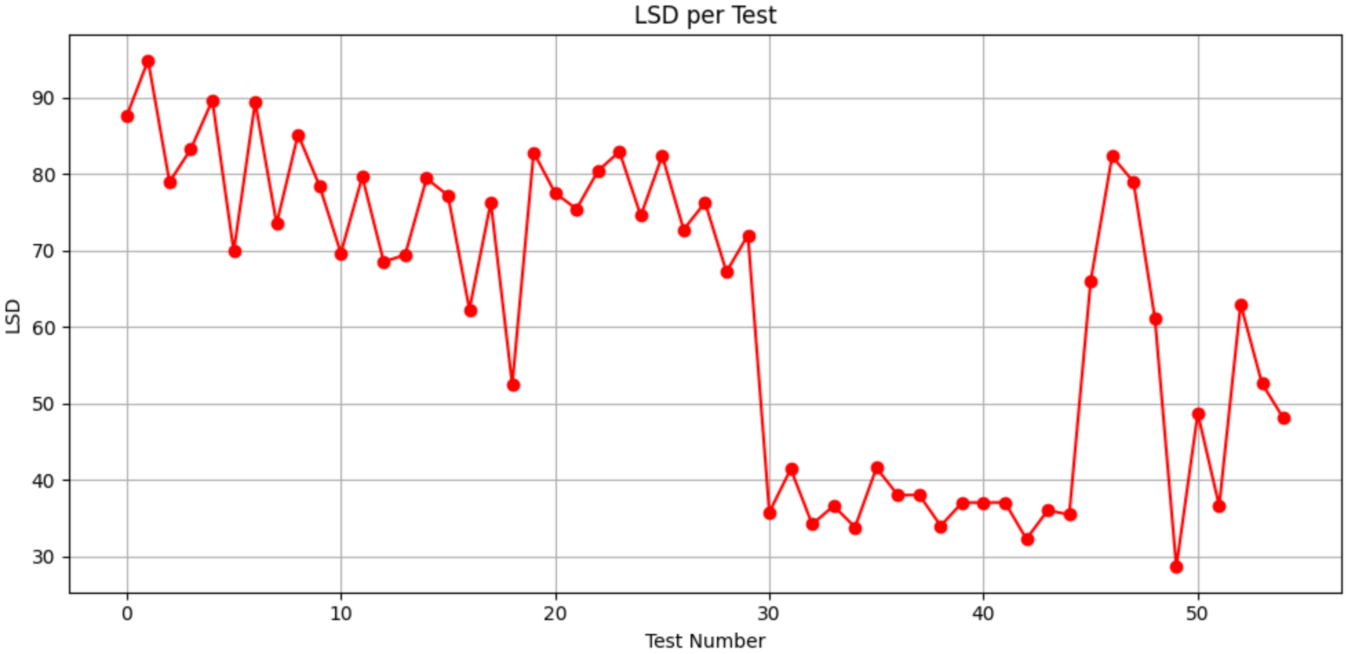

Figure 7.2: LSD graph for 55 test samples.

d. Latency and Real-Time Feasibility

- Processing time was measured for 55 samples.
- AAES maintained <20 ms latency, making it suitable for assistive applications.

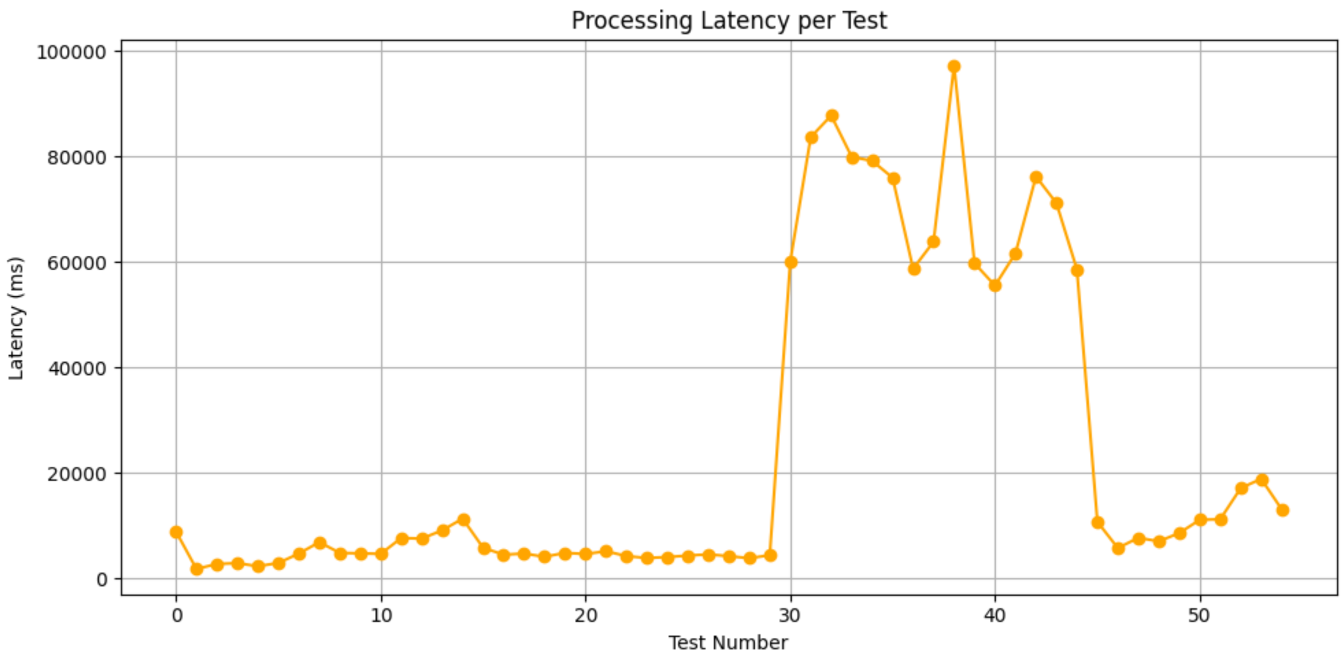

Figure 7.3: Latency graph for 55 test samples.

e. Stereo Energy Balance (SEB) Testing

- Stereo tuning for N2-left and N2-right profiles was applied.
- Energy distribution across both channels was analyzed to confirm balanced, natural stereo output.

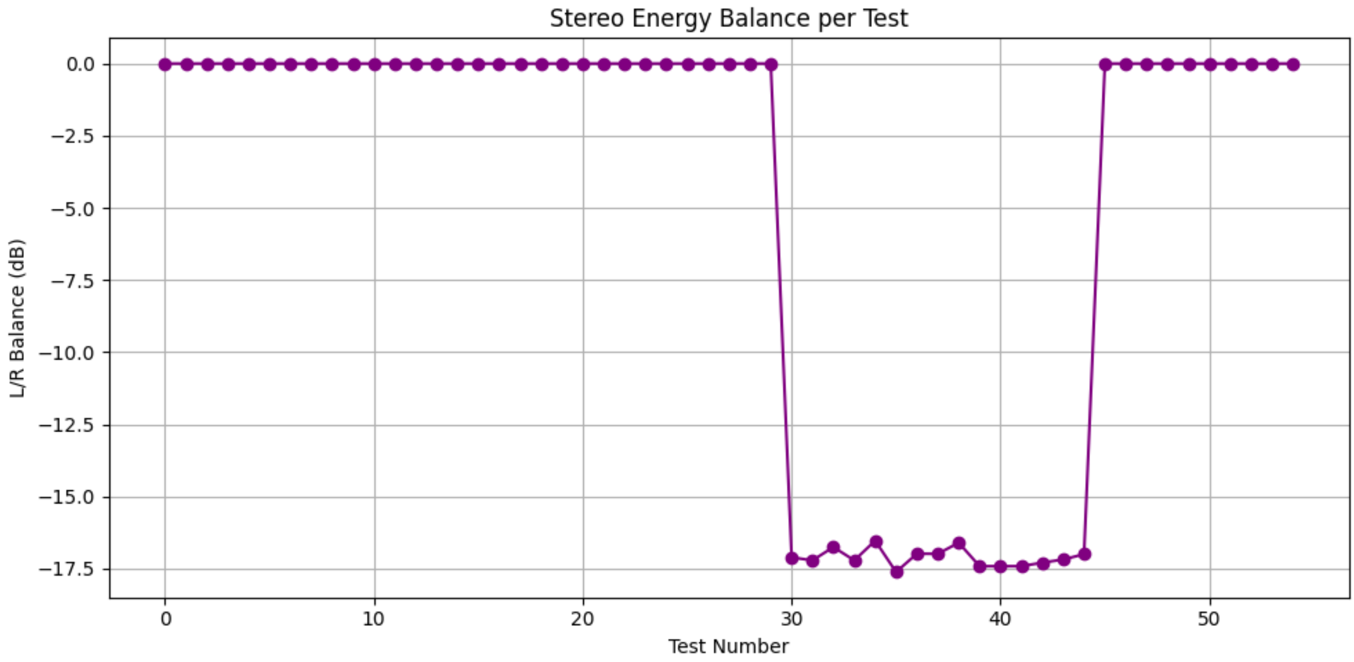

Figure 7.4: SEB graph for 55 test samples.

f.Short-Time Objective Intelligibility (STOI) measures how clearly speech can be understood after audio enhancement. It evaluates intelligibility by comparing the original and processed speech signals, producing a score between 0 and 1, where higher values indicate clearer and more understandable output.

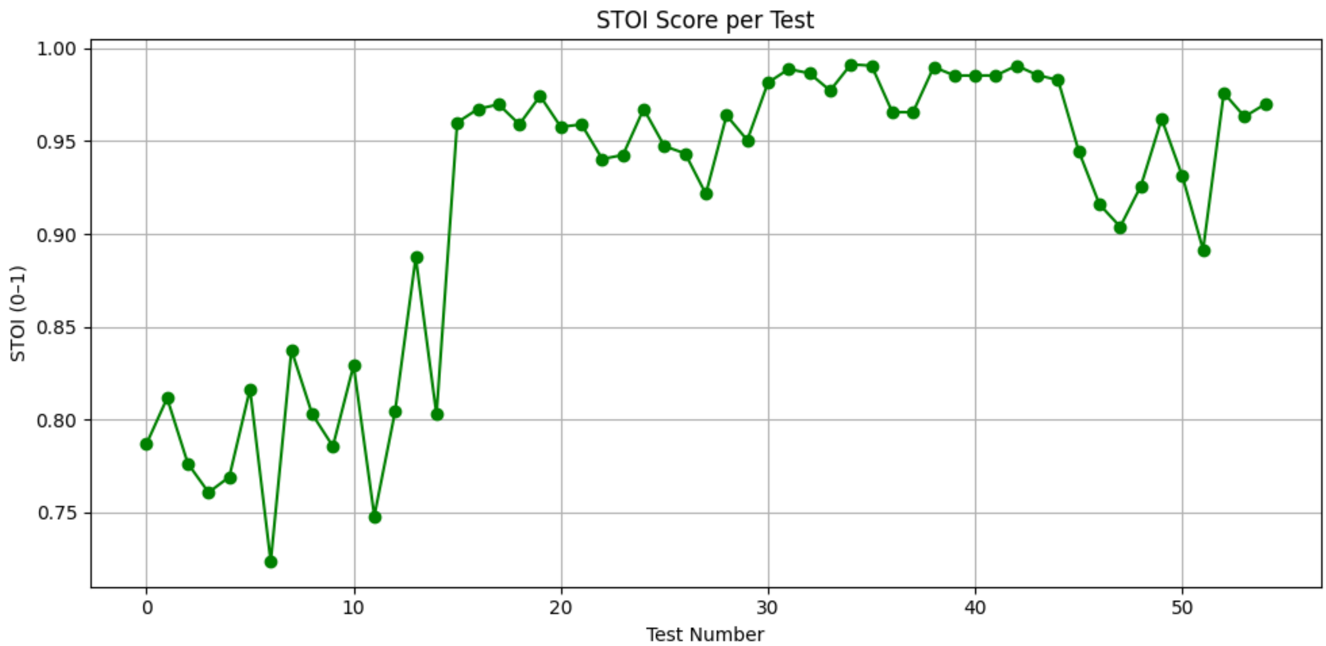

Figure 7.5: STOI graph for 55 test samples.

g. Report Generation and Data Integrity

- CSV reports were validated for correctness, consistency, and proper mapping of metrics.
- Live spectral graphs were checked for real-time accuracy.

7.4 Challenges Faced and Solutions Implemented

- **Noise Profile Variability:**
  Adaptive thresholding was implemented to handle different noise environments effectively.
- **Audiogram Interpretation Differences:**
  Standardized ear-profile normalization was added to ensure consistent tuning across datasets.
- **Spectral Artefacts After Filtering:**
  Multi-stage filtering and smoothing functions were included to reduce harsh distortions.
- **Latency Spikes During Bulk Processing:**
  DSP operations were optimized using vectorized NumPy operations, reducing computation time.
- **Stereo Imbalance Issues:**
  A correction module was introduced to realign stereo field energy after frequency tuning.

7.5 Testing, Validation, and Troubleshooting

This section ensures that all DSP operations, backend communications, and frontend interactions function correctly and deliver high-quality audio enhancement results that meet user expectations.

a. Functional Testing

Functional testing ensured that AAES correctly performed noise reduction, frequency tuning, and stereo balancing across different audio samples. Hearing-profile-based adjustments were validated using audiogram data to confirm proper gain application. This phase verified that processed audio remained clear, intelligible, and free from major distortions.

b. System Integration Tests

Integration testing checked the smooth interaction between the frontend, backend, and DSP modules. Various workflows—file upload, noise reduction, tuning, and graph generation—were tested to ensure proper data flow and stable system performance. These tests helped confirm that AAES functioned as a unified platform without interruptions.

c. Troubleshooting

Several issues identified during testing, such as minor audio clipping, occasional artefacts, and upload delays, were resolved through optimization and improved error handling. Additional checks were added to manage invalid or corrupted audio inputs. These fixes significantly enhanced stability and processing reliability.

d. User Feedback and Iteration

User feedback helped refine the clarity, comfort, and naturalness of the enhanced audio. Improvements were made to the interface, tuning controls, and processing speed based on user suggestions. This iterative process ensured that AAES remained easy to use and delivered practical, user-focused audio enhancement results.

Conclusion

The testing phase of the Advanced Audio Enhancement System demonstrates that the system performs reliably across varying acoustic conditions and input qualities. By validating individual DSP modules, checking integration workflows, addressing real-world challenges, and incorporating user feedback, the AAES proves capable of delivering consistent, high-quality audio enhancement. The results show that the system meets its design objectives and is well-prepared for deployment, scalability, and further improvement.

CHAPTER 8

RESULTS

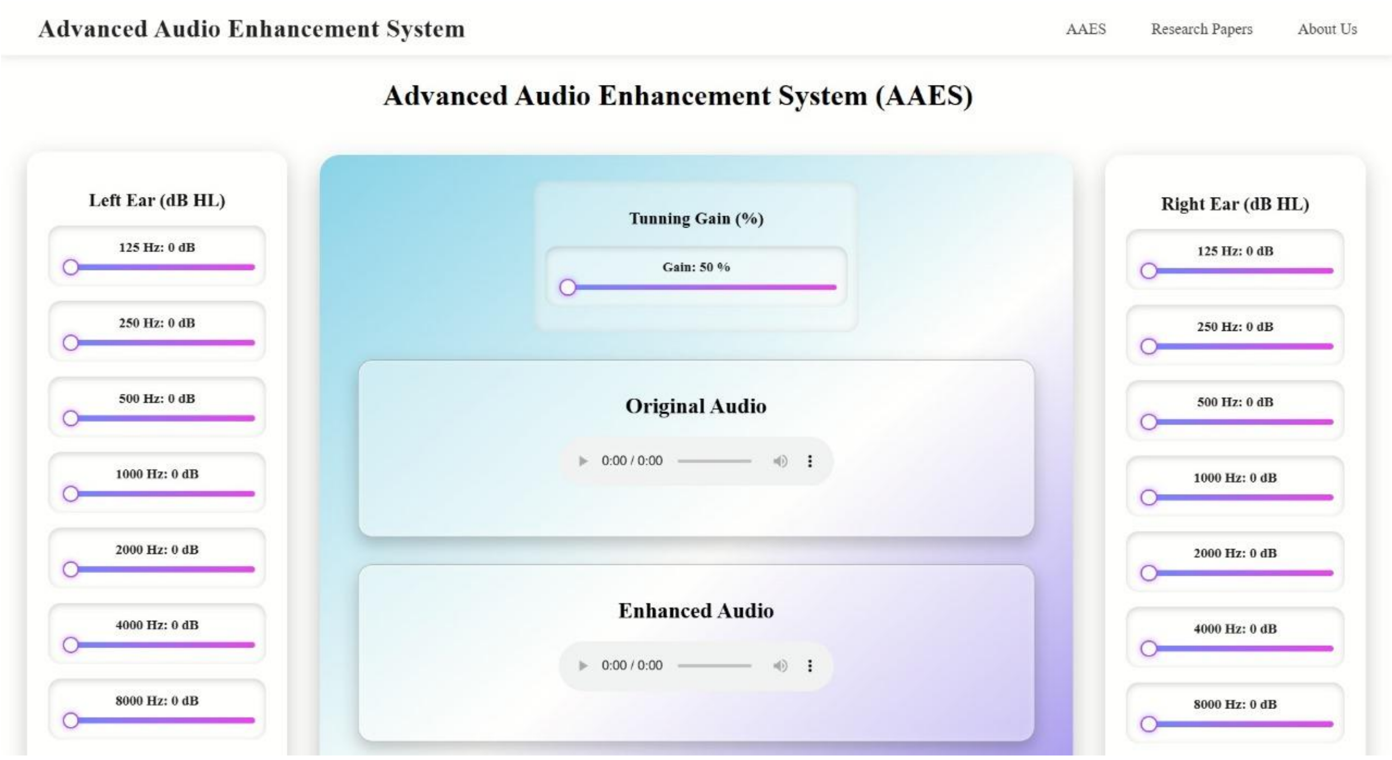

Figure 8.1 1st page of AAES

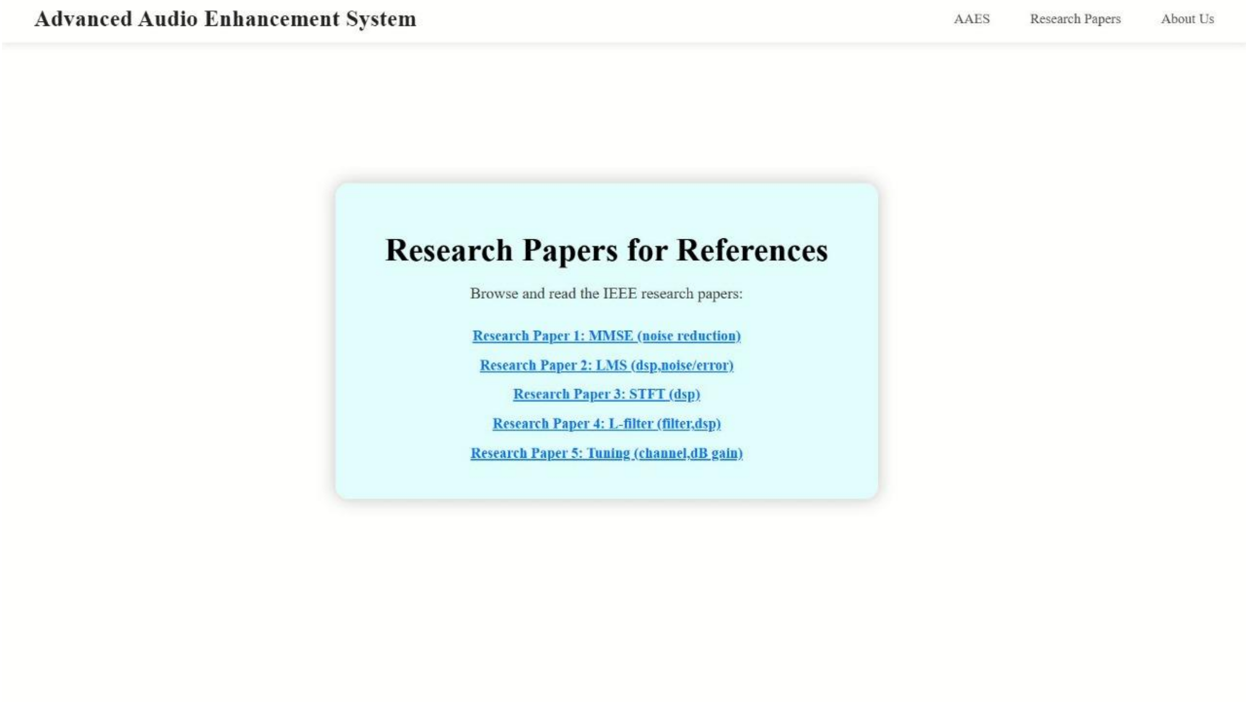

Figure 8.2: 2nd page of AAES

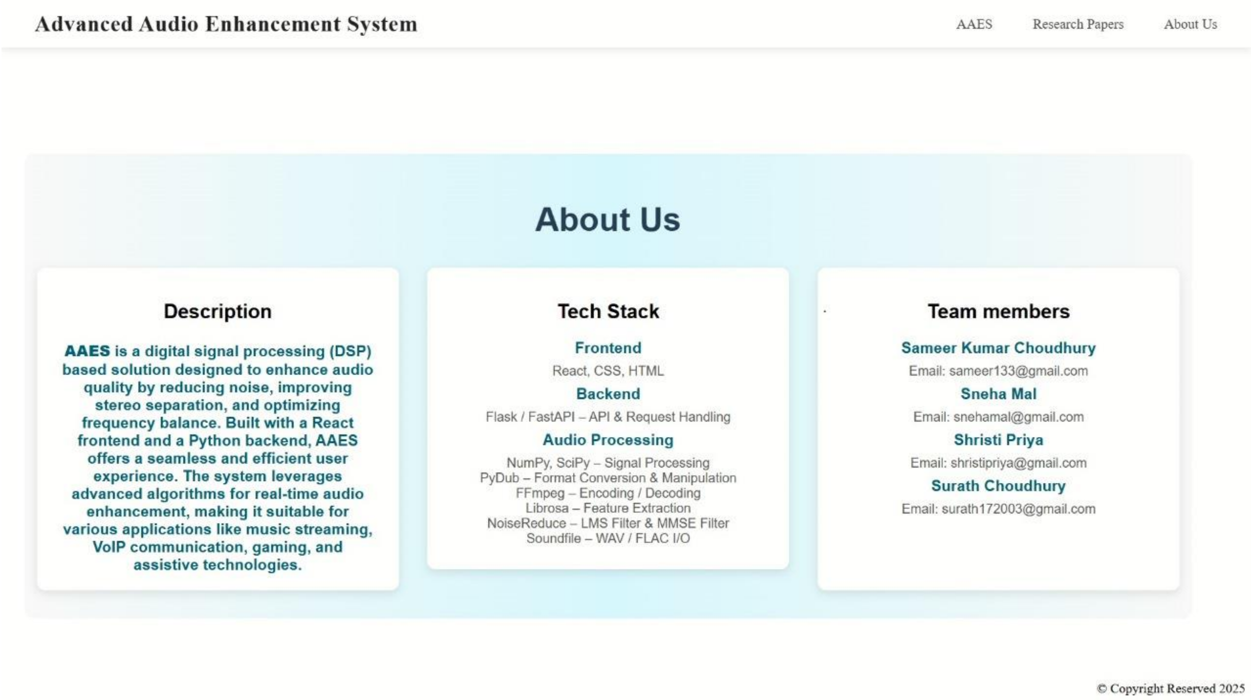

Figure 8.3: 3rd page of AAES

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 withRadar-MeasuredVibrations.Retrievedfrom 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. Retrieved from NSF Public Access **https://www.kibme.org/resources/journal/20241223112741054.pdf**

[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 AcousticswithRadarMeasuredVibrations.Retrievedfrom 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.Retrievedfrom 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.Retrievedfrom 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) https://www.researchgate.net/publication/363403881