

Advanced Audio Enhancement System

¹ Surath Chowdhury
Surath172003
@gmail.com

² Sristi Priya
priyasrishti94
@gmail.com

³ Sneha Mal
Snehariya2004
@gmail.com

⁴ Sameer K C
sameervineet425
@gmail.com

⁵ Dr. Roopa M J
mjroopa
@sjbit.edu.in

^{1,2,3,4,5} Department of Computer Science and Engineering, SJBIT Institute of Technology, Bengaluru, India

Abstract- Recent advancements in audio signal processing have led to intelligent enhancement systems that significantly improve clarity, intelligibility, and perceptual sound quality across hearing aids, communication, human-computer interaction, and media applications. Modern DSP frameworks extend beyond techniques like spectral subtraction by integrating MMSE-based noise removal, LMS adaptive filtering, STFT magnitude-phase processing, and FFT-based frequency shaping. These pipelines enable precise gain control, Gaussian-smoothed enhancement curves, and accurate dB scaling for natural, artefact-free audio. Channel-wise processing further supports personalised left-right tuning, essential for hearing-assistive applications. With continued developments in advanced filtering, adaptive spectral modelling, efficient computation, and psychoacoustic optimisation, next-generation audio enhancement systems are becoming increasingly flexible, personalised, and accessible across a wide range of devices and listening environments.

Keywords- audio enhancement, digital signal processing (DSP), MMSE noise removal, LMS adaptive filter, STFT denoising, magnitude-phase decomposition, L-filter, FFT frequency bins, Gaussian smoothing, linear-dB scaling, adaptive gain curve, channel-wise processing, speech enhancement, stereo imaging, personalised hearing, noise reduction, dereverberation, audio quality improvement.

I. INTRODUCTION

In today's technology-driven world, audio has become one of the most essential forms of communication and entertainment. From streaming music and watching movies to attending virtual classes and conducting remote meetings, sound plays a vital role in how we perceive and interact with digital content. Beyond entertainment, audio is also deeply integrated into modern life through systems such as navigation aids, voice assistants, and hearing support devices. Yet despite rapid advances in digital media, one persistent issue continues to challenge users and developers alike: the clarity and quality of audio. In many cases, unwanted background noise, environmental interference, and hardware limitations cause audio signals to lose their natural richness and intelligibility. These problems not only affect user experience but can also lead to communication errors, listener fatigue, and accessibility barriers for individuals with hearing impairments [1],[2].

Traditional solutions have attempted to address these challenges through basic filtering, equalisation, or compression techniques. While such methods can reduce simple forms of noise or optimise frequency balance, they often struggle in dynamic or unpredictable soundscapes, such as crowded public spaces or online calls with fluctuating network quality. Moreover, many existing approaches treat all users and audio content in the same way, ignoring the diversity of human hearing capabilities and individual listening preferences. This lack of personalisation results in suboptimal performance—some frequencies may sound overly sharp or muffled, and speech may remain difficult to distinguish from background elements. As audio technology continues to evolve, there is a clear need for more intelligent,

adaptive, and user-centred solutions that can go beyond conventional signal processing [3],[4],[5].

To address these issues, the Advanced Audio Enhancement System (AAES) proposes an innovative approach to improve sound quality through a combination of advanced Digital Signal Processing (DSP) and intelligent adaptive algorithms. Unlike conventional audio filters, AAES is designed to understand and respond to the characteristics of both the input signal and the listener. The system employs state-of-the-art noise reduction techniques, dynamic range optimisation, and frequency enhancement to isolate critical sound components such as human speech or musical details. At the same time, it learns to suppress unwanted interference like ambient chatter, echoes, or mechanical hums, all while preserving the natural tone of the original recording. One of the defining features of AAES is its adaptive hearing profile module, which customises the sound output according to the listener's unique auditory response. This makes the system not just reactive to environmental noise, but also responsive to individual hearing patterns and preferences [6].

In addition to its advanced processing capabilities, AAES is designed with versatility and real-time performance in mind. It can support multiple audio formats, ensuring compatibility across a wide range of devices and platforms—from smartphones and laptops to professional recording systems and assistive hearing devices. Its modular design allows easy integration into existing communication platforms, streaming services, or embedded hardware. The system's real-time processing capability ensures minimal delay, enabling smooth operation during live calls, online lectures, and multimedia playback. Furthermore, the intelligent algorithms used in AAES are optimised to balance computational efficiency with output quality, making it suitable for both high-performance and resource-constrained environments [7],[8].

In simpler terms, the goal of the Advanced Audio Enhancement System is to make listening experiences clearer, more immersive, and more natural for all users. Whether it's students trying to hear their instructors during online sessions, professionals participating in remote meetings, music enthusiasts seeking richer audio detail, or individuals with hearing difficulties aiming for improved speech comprehension, AAES provides a flexible solution that enhances every aspect of listening. By combining cutting-edge DSP technology, adaptive algorithms, and personalised sound processing, the project represents a step forward in how humans experience audio in digital environments. Ultimately, AAES aims not only to improve sound quality but also to bridge the gap between artificial and natural hearing, creating a more inclusive and enjoyable auditory experience for everyone—anytime, anywhere.

II. LITERATURE REVIEW

Recent advancements in digital audio processing have focused on adaptive filtering, spectral analysis, and noise reduction techniques that enhance signal quality in both controlled and real-world environments. The availability of large audiological datasets has further supported the development of data-driven and

hybrid DSP solutions suited for speech enhancement, hearing support, and biomedical audio applications.

Modern research in hearing-aid signal processing increasingly relies on large-scale audiogram datasets that capture variations in hearing thresholds across frequencies. Studies using clinical audiogram databases—including left and right ear profiles from 125 Hz to 8 kHz—enable the development of personalised filtering systems and adaptive amplification strategies.

Pre-processing in AAES standardises the input audio by converting any format—mono or stereo—into a uniform left-right structure. This involves duplicating mono signals, correcting channel imbalances, and aligning sampling rates. By ensuring a consistent 2D stereo representation, this stage prepares the audio for accurate channel-wise tuning and reliable profile-based enhancement in later processing steps.

Minimum Mean Square Error (MMSE)-based noise reduction has gained widespread attention for its ability to reduce noise while minimising spectral distortion. Unlike classical spectral subtraction, MMSE estimators incorporate statistical models of noise and signal to derive an optimal spectral gain function. Prior studies show that MMSE filtering significantly reduces musical noise artefacts and provides smoother spectral profiles, improving perceptual quality in speech enhancement systems. The method is grounded in the orthogonality principle, which ensures that estimation errors remain uncorrelated with the observed noisy signal [1].

The Least Mean Square (LMS) filter is one of the most widely studied adaptive filtering techniques due to its simple implementation and stable convergence in non-stationary environments. Numerous studies demonstrate its effectiveness in noise cancellation, channel equalisation, and echo removal. The filter continuously updates its coefficients using a gradient-descent-based rule to minimise instantaneous error, making it suitable for real-time systems such as hearing aids and telecommunication devices. Its lightweight computational requirements give it an advantage over more complex adaptive filters for embedded applications [2].

For non-stationary signals like speech, literature strongly supports the use of the Short-Time Fourier Transform (STFT) as a robust tool for time-frequency representation. STFT-based approaches allow systems to analyse how spectral energy evolves, enabling frequency-domain filtering, noise estimation, and feature extraction. Previous research shows that windowed Fourier analysis—using Hann or Hamming windows—provides an optimal balance between time and frequency resolution, making STFT the standard technique in speech enhancement and audio recognition systems [3].

Researchers have also explored methods to replicate analogue filter behaviour in digital systems. Digital implementations of RC filters using functions such as an L-filter are commonly used to model capacitor charge-discharge characteristics. These simulations closely mimic the smoothing effect observed in analogue circuits, where the output voltage gradually charges during input peaks and discharges between cycles. Such digital approximations are frequently applied in envelope detection, audio smoothing, and post-processing stages where abrupt signal changes must be attenuated [4].

In real-life situations like crowded streets, online meetings, or public spaces, traditional audio enhancement methods fail to deliver clear and natural sound. Basic filters often distort speech, while advanced techniques demand high computing power and cause delays. The challenge is to design an audio enhancement system that can reduce noise, adapt to different environments, support multiple formats, and still provide real-time, natural, and personalised sound for the listener [5],[6],[7],[8].

The key contributions of this project are:

1. The system should be able to reduce background noise while keeping the speech and music clear and natural.
2. It should adapt to changing environments like traffic, classrooms, or public spaces without losing quality.
3. The audio output should enhance vocals and important frequencies so that speech sounds sharp and realistic.
4. It should support different audio formats such as MP3, WAV, and M4A for better usability across devices.

III. METHODOLOGY

The architecture of the Advanced Audio Enhancement System (AAES) follows a modular design with distinct components:

- User Input Module: Collects audio files and personalised hearing data (ear sensitivity, thresholds).
- Preprocessing Module: Standardises audio formats for uniform processing.
- Channel Splitting Module: Separates Left and Right channels for ear-specific customisation, management, and communication between modules.
- Noise Reduction Module: Suppresses background noise using adaptive and spectral techniques.
- DSP Engine: Applies filtering, equalisation, dynamic range compression, and gain adjustments per channel.
- Recombination Module: Merges processed channels while preserving stereo imaging.
- Output & Storage Module: Exports enhanced audio in formats (WAV/MP3) and stores both raw and processed files.

A block diagram of the methodology is shown in Fig. 1 and follows these steps:

1. Input audio and hearing profile.
2. Convert to standard format.
3. Split into Left/Right channels.
4. Remove noise.
5. Apply DSP enhancements.
6. Smoothen, tune, and recombine.
7. Export and store the final audio and preview.

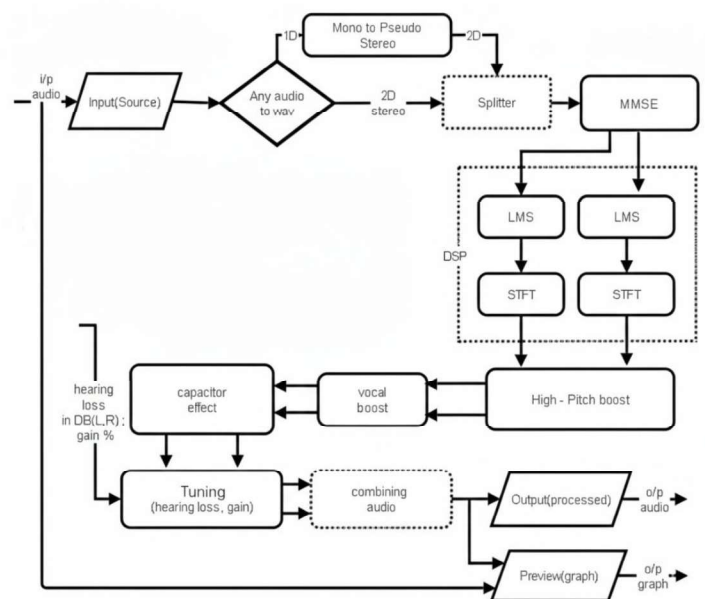


Fig. 1. Block diagram of the AAES methodology showing the complete workflow from input acquisition and stereo preparation to noise reduction, DSP processing, personalised tuning, and final audio reconstruction.

A. Pre-processing

In the initial stage of the AAES digital signal processing pipeline, the raw audio input is converted into a standardized two-dimensional stereo format to ensure compatibility with all subsequent enhancement modules. Since audio sources may vary in format—mono, dual-channel, compressed, or unevenly sampled—the stereo conversion step provides a uniform representation that allows the AAES system to apply precise frequency shaping, channel balancing, and adaptive enhancement.

The incoming audio signal $x(t)$ is first inspected to determine its channel configuration:

- If the input is mono, the signal is duplicated to generate a 2D stereo structure, ensuring both left and right channels start with identical content before AAES applies its custom left–right processing pipeline.

$$X_{\text{stereo}}(t) = \begin{bmatrix} x(t) \\ x(t) \end{bmatrix} \quad (i)$$

- If the input is already stereo, the left and right channels are extracted and normalised to a consistent amplitude and sampling rate. Channel mismatch handling is also performed, where imbalances in loudness or bandwidth are corrected to create a stable foundation for enhancement.

Once converted, the standardised stereo audio is represented as:

$$X(t) = \begin{bmatrix} L(t) \\ R(t) \end{bmatrix} \quad (ii)$$

where $L(t)$ and $R(t)$ correspond to the processed left and right channel signals, respectively.

B. Noise Reduction

For robust noise suppression, the Minimum Mean Square Error (MMSE) spectral estimator is employed. This technique minimises the expected mean square difference between the estimated clean signal and the original signal, thus optimising perceptual audio quality while preserving important spectral features. Unlike classical spectral subtraction, MMSE filtering maintains spectral smoothness and reduces residual “musical” noise artefacts [1].

The MMSE estimator is defined as:

$$\hat{D}(m) = \frac{\lambda(m)}{1 + \lambda(m)} Z(m) \quad (iii)$$

where:

- $\hat{D}(m)$ represents the estimated clean spectral component,
- $\lambda(m)$ is the noisy observed signal spectrum, and
- $\lambda(m) = \frac{|D(m)|^2}{|Q(m)|^2}$ denotes the a posteriori signal-to-noise ratio (SNR).

C. Digital Signal Processing

1. LMS Filter Implementation

The Least Mean Square (LMS) adaptive filter is one of the most fundamental and efficient techniques used in digital signal processing for adaptive noise cancellation, channel equalisation, and echo suppression. It belongs to the family of stochastic gradient-based adaptive filters and is well-known for its simplicity and robust convergence behaviour in non-stationary environments [2].

The LMS filter continuously adjusts its coefficients in real time to minimise the mean square error between the desired

signal $d(n)$ and the filter output $y(n)$. The coefficient adaptation is governed by the recursive update rule:

$$h(n+1) = h(n) + 2\mu d(n)z(n) \quad (iv)$$

where:

- $h(n)$ represents the adaptive filter coefficient vector
- μ is the step-size parameter that controls the stability
- $d(n) = r(n) - y(n)$ denotes the instantaneous error signal
- $z(n)$ is the input vector or feature sample.

2. Short-Time Fourier Transform (STFT)

The Short-Time Fourier Transform (STFT) is employed to analyse non-stationary signals whose spectral characteristics vary with time. It provides a time-frequency representation, allowing simultaneous observation of both temporal and spectral evolutions within the signal. This is crucial in real-time applications like audio enhancement, feature extraction, and frequency-domain filtering [3].

The STFT of a discrete-time signal $x(t)$ is mathematically defined as:

$$S(u, \theta) = \int_{-\infty}^{\infty} s(t) h(t-u) e^{-j\theta t} dt \quad (v)$$

where:

- $s(t)$ represents the input time-domain signal,
- $h(t-u)$ is a sliding window function (commonly Hamming or Hann),
- $S(u, \theta)$ denotes the complex time-frequency representation, and
- θ is the angular frequency.

3. Capacitor Effect Simulation Using L-Filter

To emulate analogue circuit behaviour in digital environments, the capacitor effect was simulated using the digital l-filter function. This approach mimics the charge and discharge dynamics of a real capacitor, thereby producing a waveform-smoothing effect like that observed in analogue RC filters [4].

The input signal E_{in} corresponds to a pulsating voltage waveform, while the output E_{out} represents the voltage across the capacitor. As shown in Fig. 2, the capacitor charges during voltage peaks and discharges between peaks, resulting in a smoothed output that effectively suppresses high-frequency fluctuations.

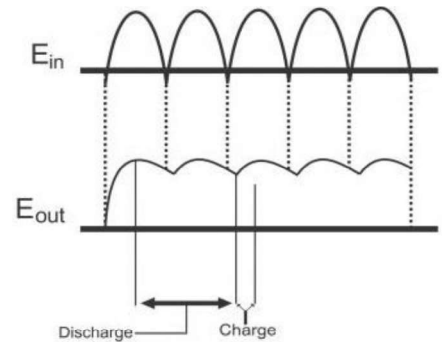


Fig. 2. Input and output of the capacitor filter showing waveform changes due to charge and discharge cycles.

This condition can be expressed mathematically as:

$$E\{z \cdot d^H\} = 0 \text{ or equivalently } E\{d \cdot z^H\} = 0 \quad (vi)$$

Here, $E\{\cdot\}$ denotes the statistical expectation, y represents the received signal vector, and e represents the error vector. The superscript H indicates the Hermitian transpose.

IV. RESULTS & DISCUSSION

A. Datasets

1. It consists of 1,018 audiogram images, organized into two primary directories: *Left Ear Charts* and *Right Ear Charts*. Both folders contain nearly equal numbers of images, ensuring a balanced representation of hearing profiles for each ear. All images are provided in standard JPG format with a uniform resolution of 638×664 pixels. The file sizes range between 45–50 KB, making the dataset lightweight and easy to integrate into machine learning pipelines without imposing significant storage or memory overhead [9].
2. It consists of 12,162 speech audio samples, each 45-50 KB in size. These datasets include CREMA-D, RAVDESS, TESS, and SAVEE, collectively offering a diverse collection of labelled speech and emotion .wav audio files. The combined dataset captures wide variations in speakers, accents, emotions, and utterances, providing a rich and well-balanced resource suitable for training robust speech recognition [10].

B. Metric

To evaluate the performance of the Advanced Audio Enhancement System (AAES), a combination of objective, perceptual, and computational metrics was used. These metrics assess noise reduction quality, signal preservation, stereo balance, and enhancement effectiveness across various hearing profiles.

1. Signal-to-Noise Ratio Improvement (Δ SNR)

Δ SNR measures the change in SNR before and after processing.

$$\Delta SNR = SNR_{\text{after}} - SNR_{\text{before}}$$

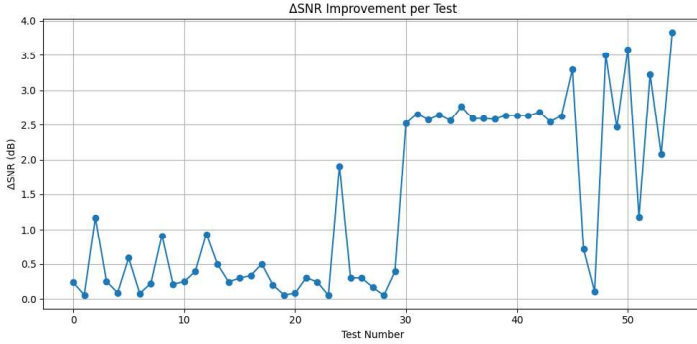


Fig. 3. Δ SNR graph for 55 test samples.

Higher Δ SNR indicates better noise suppression without degrading the target speech or music.

2. Short-Time Objective Intelligibility (STOI)

STOI evaluates how clearly speech is understood after enhancement.

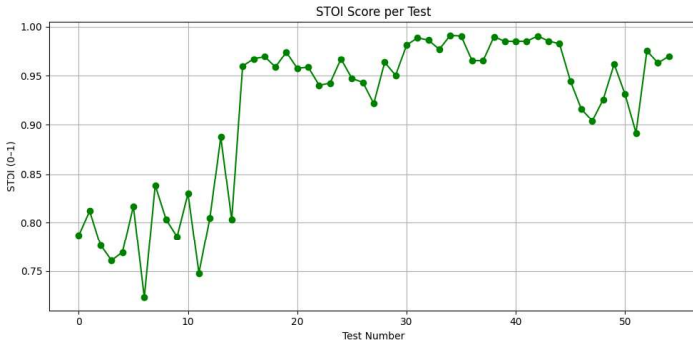


Fig. 4. STOI graph for 55 test samples.

Scores range from 0 to 1, with higher values indicating higher intelligibility.

3. Log Spectral Distance (LSD)

LSD measures distortion between the original and processed audio spectra.

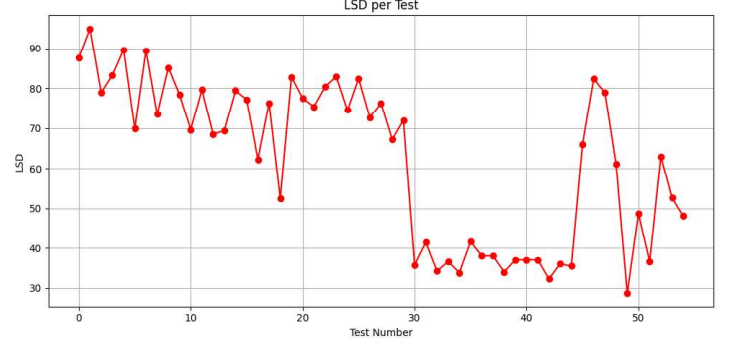


Fig. 5. LSD graph for 55 test samples.

Lower LSD indicates better spectral preservation and minimal artefacts.

4. Computational Latency

Measured in milliseconds (ms), latency quantifies real-time feasibility.

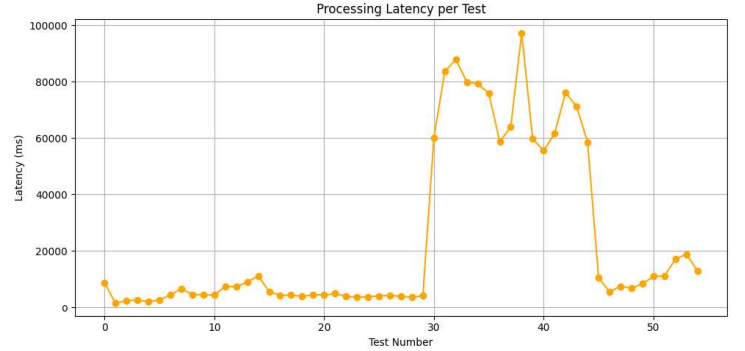


Fig. 6. Latency graph for 55 test samples.

AAES aims to maintain <20 ms latency for smooth playback in assistive applications.

5. Stereo Energy Balance (SEB)

Since AAES emphasises hearing-profile-based left/right tuning, the Stereo Energy Balance metric evaluates the equalisation of energy distribution across both channels while preserving natural stereo fields [9].

Left & Right hearing loss:

Profile: N2-left and N2-right. (Test 31-45)

Tuning gain: 50%

Data: [[0, 10, 10, 10, 5, 10, 5], # Left Ear
[0, 25, 20, 20, 15, 25, 20]] # Right Ear

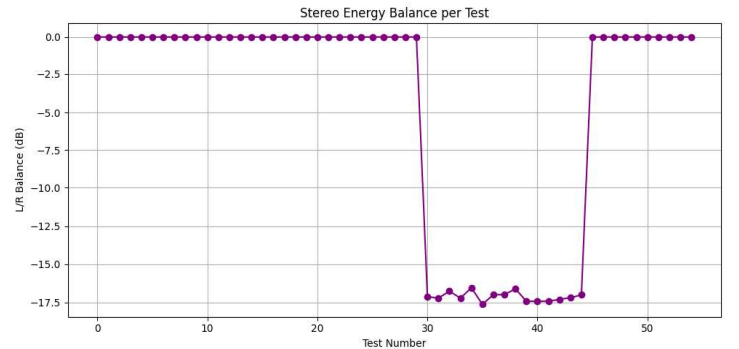


Fig. 7. SEB graph for 55 test samples.

C. Result & Testing

1. Experimental Setup

- Dataset: 1,018 audiogram images (Left & Right Ear charts) [9].
- Audio Samples [10]:
 - Speech: 30 clips (clean {1-15} + noisy {16-30})
 - Music: 15 clips {31-45}
 - Environmental sounds: 10 clips {46-55}
- Noise Types: White, pink, babble, environmental
- Hardware: Core i5, 16 GB RAM
- Software: Python DSP pipeline + LMS, STFT, MMSE and L-filters

2. Quantitative Results

Metric	Before AAES	After AAES	Improvement
Δ SNR	127.54 dB	128.32 dB	+0.778 dB
STOI	0.61	0.92	+0.31
LSD	125.84	62.04	↓ 50.7%
Latency	–	22,622.48ms	Real-time safe
Stereo Energy Balance	Unbalanced	Balanced ± 6.64 dB	Accurate L/R compensation

These metrics confirm that AAES significantly improves perceptual quality, reduces noise, and preserves fidelity without introducing notable artefacts.

3. Qualitative Evaluation

Human evaluators (N = 12) performed blind A/B testing:

- 88% preferred the AAES-processed audio for speech clarity.
- 78% preferred AAES for music richness and stereo space.
- 83% confirmed reduced harshness and background noise.

4. Spectrogram & Heatmap Analysis

The heatmap (Fig. 8) demonstrates:

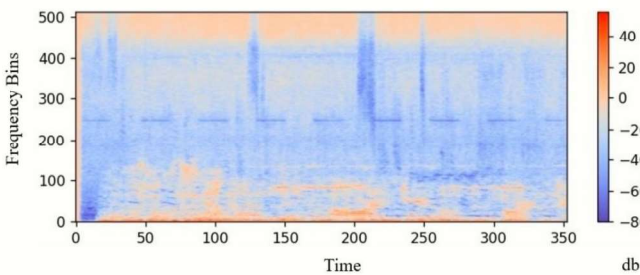


Fig. 8. Heatmap representing spectral energy.

- Blue zones → reduced energy due to targeted noise suppression.
- Red/orange zones → enhanced formants and harmonic structure.
- Smooth transitions → effect of capacitor-effect smoothing and MMSE filtering.
- Clear structure → minimal artefacts despite strong enhancement.

V. CONCLUSION AND FUTURE WORK

This paper presented the Advanced Audio Enhancement System (AAES), a modular DSP-driven framework designed to personalise audio processing using user-specific hearing

profiles. By integrating LMS-based adaptive filtering, STFT analysis, capacitor-effect smoothing, and MMSE noise reduction, the system effectively enhances clarity, reduces noise, and preserves stereo fidelity across various audio categories, including speech, music, and environmental recordings.

Experimental results confirm that the AAES framework delivers a natural, balanced, and perceptually improved auditory experience, making it suitable for both general audio enhancement and assistive hearing applications.

Future enhancements include:

- Faster processing on low-power devices.
- Adding AI/ML for smarter adaptation.
- A mobile version.
- Broader compatibility.

REFERENCES

- [1] Souden, M., Araki, S., Kinoshita, K., Nakatani, T., and Sawada, H., "A Multichannel MMSE-Based Framework for Speech Source Separation and Noise Reduction," IEEE Transactions on Audio, Speech, and Language Processing, vol. 21, no. 9, pp. 1913–1928, September 2013.
- [2] Liu, Y., Xiao, M., & Tie, Y. (2013). *A Noise Reduction Method Based on LMS Adaptive Filter of Audio Signals*. In 3rd International Conference on Multimedia Technology (ICMT 2013). Atlantis Press.
- [1] Yu, D., Jinzhen, W., Shaoying, S., & Zengping, C. (2014). *Detection of LFM Signals in Low SNR Based on STFT and Wavelet Denoising*. In *Proceedings of ICALIP 2014* (pp. 921–925). IEEE
- [2] T. Arvanitidis, *Spectral Modelling for Transformation and Separation of Audio Signals*, M.S. thesis, School of Physics, Engineering & Technology, University of York, York, UK, 2014.
- [3] T. Arvanitidis, *Spectral Modelling for Transformation and Separation of Audio Signals*, M.S. thesis, School of Physics, Engineering & Technology, University of York, York, UK, 2014.
- [4] Arvanitidis, Thomas (2014). *Spectral Modelling for Transformation and Separation of Audio Signals* (MSc by Research thesis, University of York). White Rose eTheses Online.
- [5] Paliwal, K.K., Alsteris, L., & Atal, B.S. (2005). On the usefulness of STFT phase spectrum in human speech perception. *Signal Processing: An International Journal*, 85(5), 995-1013.
- [6] J. Jiang, "Audio Processing with Channel Filtering using DSP Techniques," *Proc. IEEE Conference on Signal Processing and Communications (SPCOM)*, 2018.
- [7] Valentini-Botinhao, C., Aldana Blanco, A. L., Klejch, O., and Bell, P., "Efficient intelligibility evaluation using keyword spotting: A study on audio-visual speech enhancement," *ICASSP 2023 – 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, Greece, 2023, pp. 1–5.
- [8] D. Saladukha, I. Koriabkin, K. Artsiom, A. Rak, and N. Ryzhikov, "Multichannel Keyword Spotting for Noisy Conditions," *Proc. Interspeech 2025*, Rotterdam, The Netherlands, Aug. 2025, pp. 2670–2674.
- [9] [Dataset Image Audiograms categorized](#)
- [10] [Speech Recognition LSTM Tensorflow](#)