Modeling Steerable Sweet Spot with Machine Learning

We experience sound in a three-dimensional way, and when done right, it can make the unreal feel real. - Hans Zimmer

Immersion is the essence of experiencing sound without limitations.

The "sweet spot" is the defined position in a home theatre where the audio experience is at its best, based on speaker placement and room acoustics. Traditionally, only a few seats in the centre of the room enjoyed this optimal sound, while other positions offered a less immersive experience.

Given the advances in digital audio processing, modern well-defined sweet spots create immersive surround sound experiences in a wider range, allowing for a larger audience to share the best experience.

Now with AI/ML we have taken a leap ahead in the 'sweet spot' experience creation, intelligently and dynamically shifting sound across different channels. In an optimal setup, one can move sound around, as it would in a movie theater, making you feel as though you are "in" the action.

In our previous article on [Immersive Sweet Sound Spot](), we discussed how the sweet spot is positioned within a controlled space, as illustrated in Fig. 1: Static Sweet Spot. This setup delivers immersive sound for a stationary position, utilizing an audio data sensor to calibrate the decibel levels and adjusting the real-time audio based on the listener's position.

Fig 1: Static Sweet Spot

Fig 2: Dynamic Audio Zone

The objective here, however, is to extend this static sweet spot, enabling a more expansive and unrestricted audio experience by utilizing precisely measured and calibrated audio data, coupled with continuous Machine Learning which will enable and create a steerable surround sound environment**.** This improvement in the surrounding sound transcends traditional limitations, providing truly immersive audio experience for all listeners within the dynamic zone, regardless of where they are seated.

How to achieve a steerable sweet spot:

This approach adjusts in real-time based on the listener's position and the room's acoustics, to adapt without changing the default speaker placements, modulating audio levels according to listener's distance. The existing static sweet spot serves as a reference, allowing for precise adjustments without altering the original setup.

By measuring dynamic zone, speaker levels, distances, and latencies for fine-tuning the audio to an ideal volume and at precise time to enhance the overall audio experience.

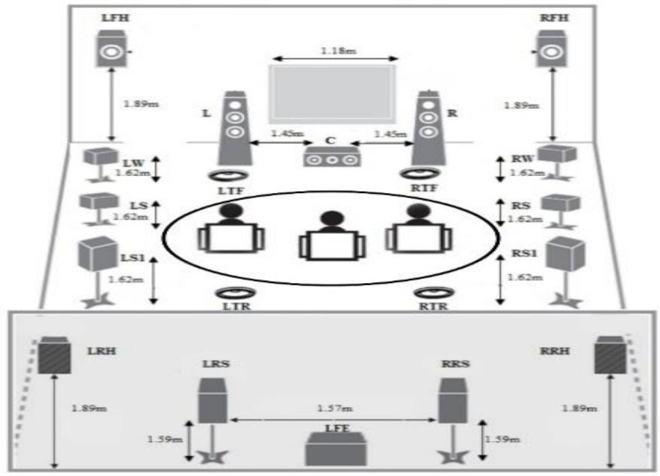


Fig 3: Speaker and listener positioning in the dynamic zone

Steerable positioning audio enhancement using Machine Learning

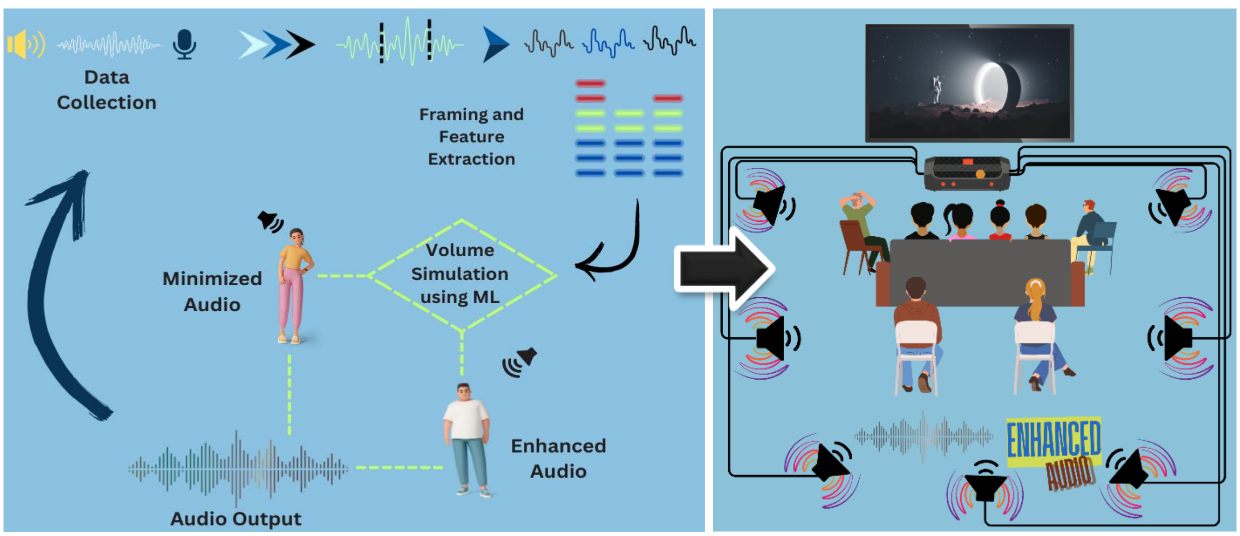


Fig 4: Process flow: Steerable positioning audio enhancement using ML

Audio analysis is a process of transforming, exploring, and interpreting audio signals recorded by digital devices, aiming at understanding sound data. It applies to a range of technologies, including state-of-the-art Machine Learning algorithms.

Process of audio analysis using ML:

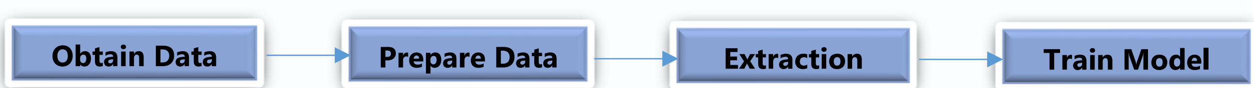


Fig 5: Process of Audio Analysis

Obtaining datasets for ML model training:

Audio data represents analog sounds in a digital form, preserving the main properties of the original audio. In the field of audio analysis, there are diverse types of datasets available, including commercial, expert-curated datasets, and open-source libraries. Here, we have specifically utilized the environmental sounds to simulate the real-world acoustic in theater setup for obtaining and analyzing audio data as they preserve all the properties of the audio.

Data preparation for ML:

As we know, sound is a wave of vibrations that travel through air and by considering its characteristics such as Amplitude, Time period, Frequency of the audio, the required data can be obtained to analyze for enhancing the audio output like fine tuning, noise filtering, normalization, equalization etc.

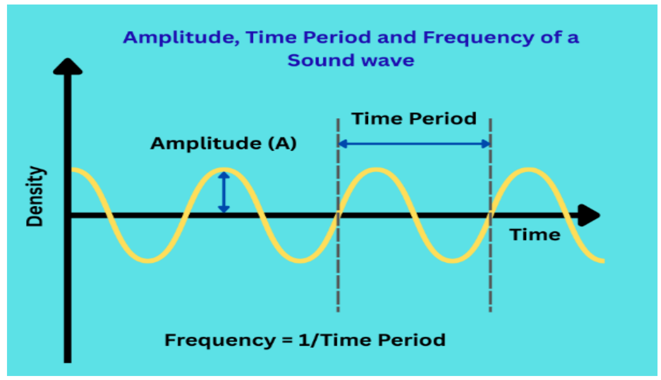


Fig 6: Characteristics of audio

To train the model, we simulate speaker positions and volume adjustments for each audio file, combining them with audio features to create a training dataset. The audio data is then framed into short, overlapping time segments for analysis, as shown in *Fig. 7*: *Framing*.

To mitigate the spectral leakage, windowing is applied to each frame for smooth transitions between frames as shown in *Fig. 8: Windowing*.

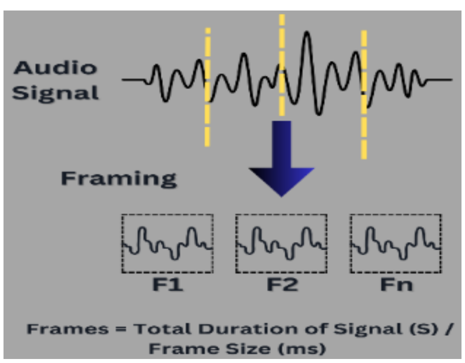


Fig 7: Framing

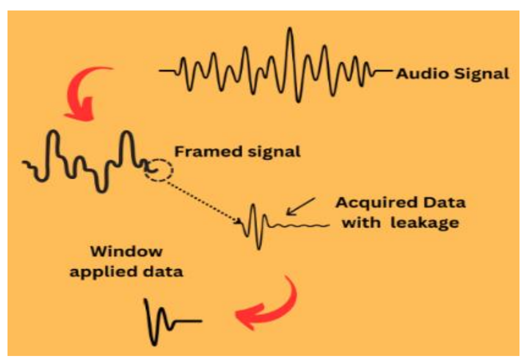


Fig 8: Windowing

Feature extraction:

This process involves extracting features in terms of Frequency Domain, Time Domain, Frequency and Time Domain represented in spectrogram.

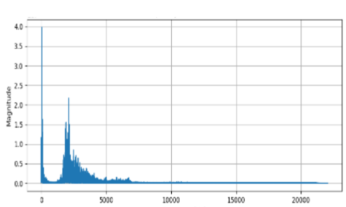


Fig 9: Frequency Domain

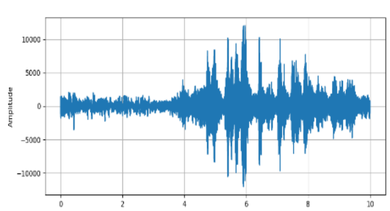


Fig 10: Time Domain

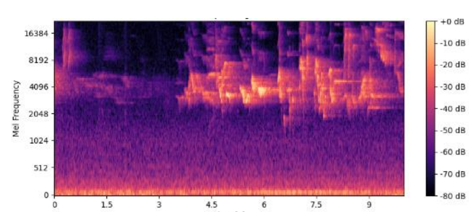


Fig 11: Time and Frequency Domain

How audio feature extraction is done:

The first step of the pipeline is the extraction of relevant features from the input audio files. These features include Mel Spectrogram and Mel-Frequency Cepstral Coefficients (MFCC).

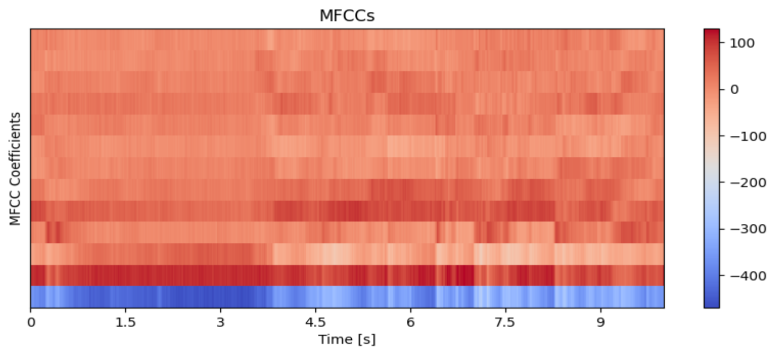


Fig 12 :MFCC

The Librosa library is used to load audio files and extract features, with mean values computed across time to create a compact representation of each sample. Below is the audio datasets used to train the model.

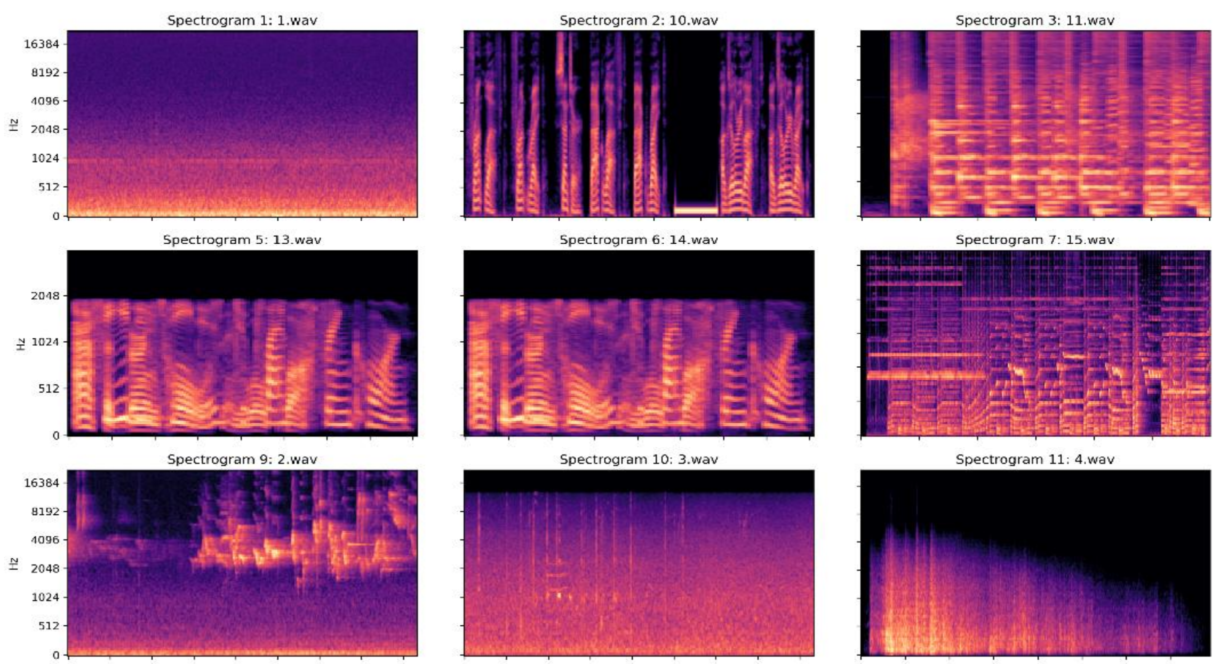


Fig 13: Audio datasets Extraction

Simulating speaker positions and volume adjustments:

In real-world scenarios, the volume of a speaker's output is influenced by distance and angle of the speaker relative to the listener. To simulate this, we generated labelled positions for speakers, each characterized by a distance and an angle. The volume adjustment is then computed based on a linear scaling model, where the distance affects the volume level.

Apart from volume adjustments, to get the quality audio, we apply filtering techniques that eliminate hums, echoes, and environmental sounds. By isolating relevant frequencies and suppressing irrelevant ones, we preserve key features such as pitch, rhythm, and tone and it further reduces static and distortion, improving feature extraction accuracy.

Audio enhancement involves training Machine Learning models to improve sound quality by reducing noise and enhancing clarity. The Random Forest Regressor handles complex relationships to adjust audio parameters, while hyperparameter tuning optimizes model accuracy. A key application is optimal volume control, ensuring clear and balanced audio across various environments as shown in *Fig 14: Audio Enhancement with Machine Learning: Optimizing Clarity and Volume*.

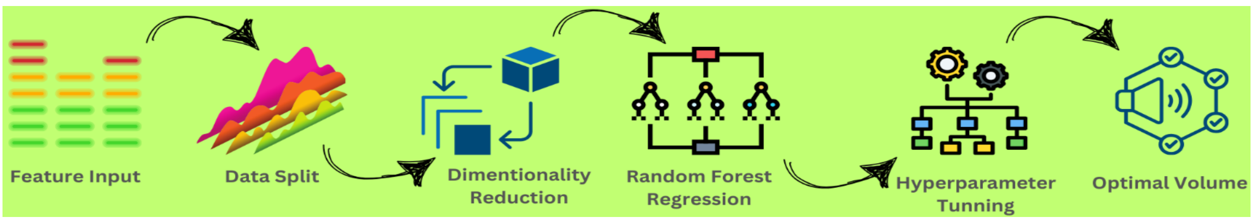


Fig 14: Audio Enhancement with Machine Learning: Optimizing Clarity and Volume

Audio volume prediction, enhancement, and optimization workflow:

A Random Forest Regressor predicts volume adjustments from extracted audio features and simulated positions. The dataset is split into training and testing sets, with Principal Component Analysis (PCA) reducing dimensionality.

Once trained, the model predicts volume adjustments for new audio files based on speaker positions. The predicted scaling factor is applied to adjust the audio amplitude, and the enhanced audio is saved as a new file. The pipeline's main function, shown in *Fig. 15: Workflow of ML*, trains the model, performs predictions for predefined speaker positions, enhances the audio, and saves the results in new files.

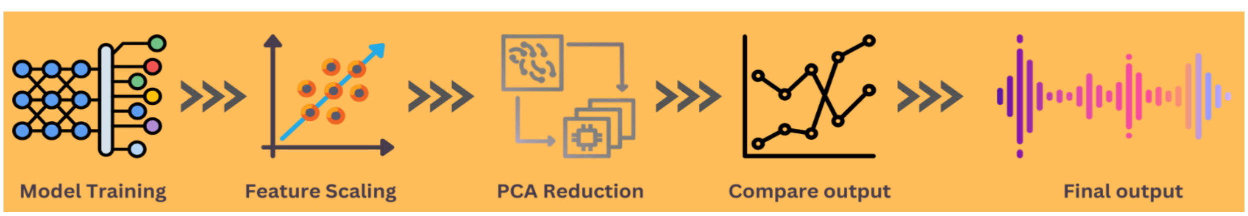


Fig 15: Workflow for ML

Comparing Actual vs Predicted values in trained and tested data using Machine Learning

The data undergoes preprocessing, including feature scaling and PCA for dimensionality reduction. After splitting into training and test sets, the model learns from the training data and is evaluated on the test set. The predicted values are compared to actual values as shown in *Fig 16: Training data and test data* to assess the model's accuracy and generalization.

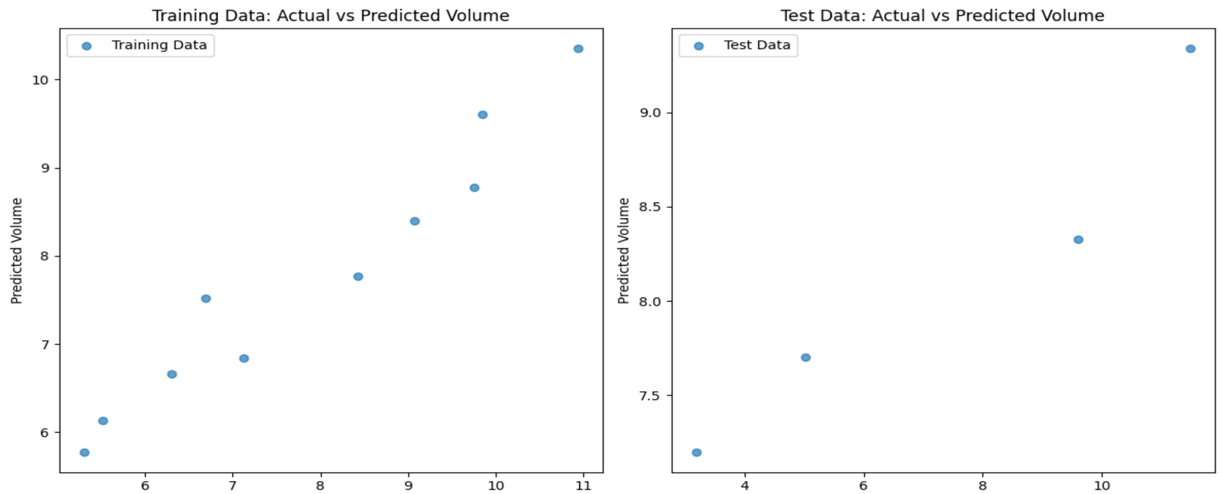


Fig 16: Training data and test data

RESULTS and DISCUSSION

The results of the model training and predictions are evaluated based on the Mean Squared Error (MSE) and R-squared metrics. The Random Forest model demonstrated reliable performance in predicting the required volume adjustments, with minimal error across different speaker positions. The enhancement process successfully adjusted the audio volumes based on the predicted scaling factors, which could be applied in various real-world audio applications such as sound system calibration or virtual reality environments.

The waveforms are also plotted in distinct colors for visual comparison, as shown in *Fig. 17: Waveform comparison*. This visualization helps to compare the audio functions to provide both numerical and visual analysis, making it useful for comparing original and modified audio files.

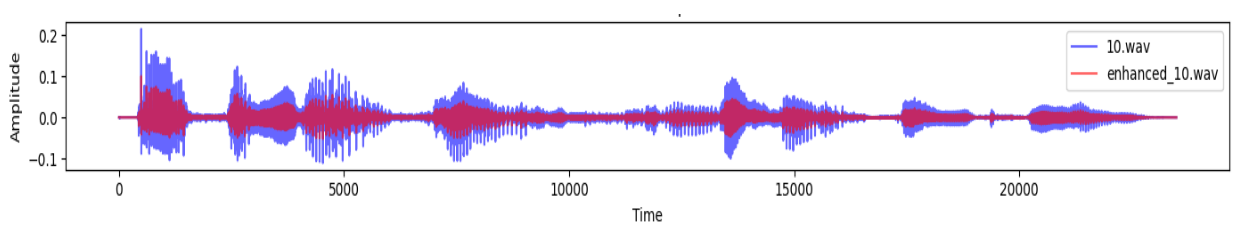
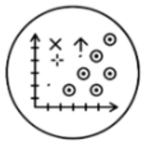


Fig 17: Waveform comparison

Benefits of ML implementation:

Limitations of ML implementation:



Dependence on Randomized Data

Limited Generalization of Complex Environment

Computational Complexity

Comparing ML-Based Method to Conventional Method

Parameters	ML-Based Steerable Method	Conventional Steerable Method
Accuracy and Duration	High due to real-time analysis and related adjustments. Reduces setup and recalibration time through automatic adjustments.	Performance is lower, as the system depends on fixed parameters and lacks real-time adaptation, requiring manual recalibration and intervention for adjustments.
Configuration Options and Enhancements	Offers greater flexibility and dynamically adapts to various room setups, with real-time fine-tuning based on listener positioning.	Limited flexibility for dynamic changes and requires manual fine-tuning with reduced real-time responsiveness.
Complex settings and Adaptation	Effectively handles complex acoustics through dynamic learning, continuously adapting to maintain optimal performance.	Struggles with complex acoustics without manual recalibration, as it lacks automatic learning and adaptation, relying instead on manual intervention
Scalability	Highly scalable and can adapt to larger or more complex setups.	Limited scalability; requires manual adjustment for larger setups.
Listener's Experience	More intuitive, personalized, and seamless as the system learns.	Requires more user input and frequent adjustments for optimal experience.
Performance	Performs reliably in dynamic terrain and offers low latency in optimized systems, however, without proper training, it may risk overfitting and cause occasional computational delays.	Remains stable in fixed environments with consistently low latency, it lacks adaptability to changing conditions and doesn't support real-time adjustments.
Integration with Technologies	Easily integrates with IoT and smart home systems, supports multi-source audio.	Limited integration with other technologies or smart systems.
Customization and Personalization	Highly customizable, adapts to individual listener preferences.	Fixed customization options, manual adjustments for personalization.

Conclusion:

The implementation of a steerable sweet spot using Machine Learning represents a significant advancement in audio technology. This promising approach enhances audio volume adjustments by simulating speaker positions and utilizing Machine Learning to make accurate predictions.The model's ability to predict and apply volume adjustments effectively enhances audio clarity and realism. Further improvements will focus on improving model accuracy and integrating real-time dynamic tracking to further refine sound quality. In future, this dynamic zone can be trained and tested for dynamic positions, contributing to even greater improvements in audio quality. This ongoing refinement will ensure a more immersive and adaptable audio experience, pushing the boundaries of sound technology powered by Machine Learning.

DISCLAIMER

The steerable sweet spot using ML for spatial surround sound system set up at home, audio rooms, theatres, and virtual reality will bring revolution in immersive audio experience.

Jasmin Engineering Support Team strives to achieve accuracy in attaining enhanced audio with Machine Learning algorithms integrated for predicted adjustments in real-time.

With current popularity of AI implementation in every application, stay tuned for upcoming developments and groundbreaking solutions in the realm of advanced audio engineering.

Contact Details

Jasmin Infotech Private Limited (HQ)
Plot 119 Velachery Tambaram Road
Pallikaranai, Chennai 600100
India

MS. JEYASUDHA / MR. NARENDRAN OVI

M: +91 89251 09996
narendran.ovi@jasmin-infotech.com