Signal and Source Separation

Mahisha Ramesh
MTech (CSE) MT23121
mahisha23121@iiitd.ac.in

Ashutosh Kumar Dubey
BTech (CSAM) 2020287
ashutosh20287@iiitd.ac.in

DHRUV SOOD
BTech/CSSS 2021387
dhruv21387@iiitd.ac.in

Abhishek Tibrewal
MTech (CSE) MT23009
abhishek23009@iiitd.ac.in

Mayank
MTech (CSE) MT23045
mayank23045@iiitd.ac.in

Abstract

Signal and source separation is a critical task in various fields such as audio processing, speech recognition, and music analysis, and trying to separate the audio in signal and source. In this Project, we explore the application of exploratory data analysis (EDA) techniques in conjunction with using various machine learning algorithms for separating the signal and the sources of the audio files. The spectrogram, computed using techniques like Short-Time Fourier Transform (STFT), provides a time-frequency representation of the signal, while EDA is employed to extract and analyze crucial features from the spectrogram. And applying this crucial feature for the separation of signal and source using the algorithm like repet and FastICA. This study delves into key acoustic features, including Root-mean-square (RMS) energy, Zero Crossing Rate, Mel-Frequency Cepstral Coefficients (MFCCs), Chroma, and Tempogram, extracted from the spectrogram. Further the study dive into various algorithm for signal and source separation like Repet, FastICA, ButterworthFilter, NMF, HarmonicPercussion and GaussianNMF. These features are employed to facilitate the separation of sources from the complex audio mixture. We aim to leverage these distinctive features to train a classical machine learning algorithm designed for the precise separation of audio signals. Our objective is to attain optimal accuracy metrics such as Signal-to-Noise Ratio (SNR). Through rigorous optimization, we endeavour to fine-tune the algorithm to achieve the highest level of performance and deliver superior results in audio signal separation. Keywords: Signal Separation, Source Separation, Exploratory Data Analysis (EDA), Spectrogram, RPCA, FastICA, ButterWorthFilter, NMF, GaussianNMF, HarmonicPercussion

1 Introduction

In the field of machine learning, a significant challenge arises when trying to separate different sound sources and their auditory signals from a complex mixture of sounds. This challenge becomes even more difficult because machines struggle to interpret auditory and directional cues like Interaural Time Difference (ITD), Interaural Level Difference (ILD), and frequency filtering, which the human brain naturally uses to perceive and distinguish sound sources.This research investigates signal source separation in music and video datasets. Music source separation aims to differentiate individual sound sources in an audio mixture, effectively reversing the blending of sounds.In contrast, audio source separation in audiovisual settings attempts to identify motion information in the visual stream and extract features from shared modalities of sound and video.This research explores different audio features, Spectrogram analysis and feature engineering to understand and visualise theory of mixed signals to improve precision.

2 Objective

Our objective is to explore the application of exploratory data analysis (EDA) techniques in tandem with spectrogram analysis for the separation of mixed audio signals into their individual sources. By harnessing these analytical methods, we aim to train a classical machine learning algorithm like Repet, FastICA, ButterworthFilter, NMF, HarmonicPercussion and GaussianNMF. that achieves optimal accuracy metrics such as Signal-to-Noise Ratio (SNR). Through careful optimization, our goal is to enhance the algorithm's performance and ensure the precise extraction of constituent sources from complex audio mixtures. Specifically, the study will investigate the following: Effectiveness of EDA Techniques: Assessing the efficacy of Exploratory Data Analysis (EDA) techniques in extracting and analyzing pivotal features from the spectrogram of mixed audio signals. Feature Utilization for Source Separation: Examining how the identified features can be effectively employed to facilitate the separation of sources within complex audio mixtures. Algorithmic Approaches for Separation: Applying a diverse set of algorithms for the separation of audio mixtures, including Repet, FastICA, ButterworthFilter, NMF, HarmonicPercussion and GaussianNMF. Performance Evaluation: Striving for high accuracy, measured through metrics such as Signal-to-Noise Ratio (SNR), and aiming to achieve optimal results in the separation of audio sources. Algorithm Optimization: Implementing rigorous optimization techniques to fine-tune the selected algorithms, with the overarching goal of enhancing accuracy and achieving superior performance in audio source separation. The anticipated outcomes of this study have the potential to contribute significantly to advancements in the field, especially in areas like audio processing. The application of various algorithms and the emphasis on optimization seek to elevate the accuracy levels, as measured by SNR, leading to robust solutions for audio source separation challenges.

3 Literature Review

Blind source separation and independent component analysis: A review [1] did study on Robust Orthogonalization/Whitening, and discussion of several extensions and modifications of blind source separation and decomposition algorithms for spatio-temporal decorrelation, independent component analysis, sparse component analysis and non-negative matrix factorization where various criteria and constraints are imposed such linear predictability, smoothness, mutual independence, sparsity and non-negativity of extracted components. Makino, S., Wada, T. (2000). Blind source separation based on time–frequency signal representations [2] paper introduces a new blind source separation approach exploiting the difference in the time–frequency (t-f) signatures of the sources to be separated. In contrast to existing techniques, the proposed approach allows the separation of Gaussian sources with identical spectral shape but with different t-f localization properties.Belouch, A., Benyacoub, S. (2002). Blind source separation of real world signals [3] did study the FIR polynomial algebra techniques which present an efficient tool to solve true phase inverse systems allowing a simple implementation of non causal filters. The significance of the methods is shown by the successful separation of two voices and separating a voice that has been recorded with loud music in the background. Plumbley, M. D., Vincent, E. (2005). An overview of informed audio source separation [4] introduced two most prominent research trends in the paper, model-based informed source separation:to handle specific musicological knowledge, and signal-based informed source separation: a desirable framework whenever some signals are available, such as score-sheets or cover version, which are related to the unknown sources to estimate.Haykin, S. (2000). Blind source separation: A review and analysis [5] did Pre-processing, Mixing, Generation of Mixing coefficient matrix, De-mixing and Binary Mask Creation in the research paper .

Zibulevsky, M., Pearlmutter, B. (2001). Blind source separation by sparse decomposition in a signal dictionary [6] did a two stage separation process: a priori selection of a possibly overcomplete signal dictionary in which the sources are assumed to be sparsely representable, followed by unmixing the sources by exploiting the their sparse representability. We extracted the sources sequentially using quadratic programming with nonconvex quadratic constraints in the research paper.Virtanen, T. (2007). Musical source separation: An introduction. [7] had explored the MSS problem and introduced approaches to tackle it. They begin by presenting characteristics of music signals; and then introduce MSS and, finally, consider a range of MSS models.the study also discusses how to evaluate the MSS approaches and discuss limitations and directions for future research. Kim, T., Pearlmutter, B. (2003). Blind source separation combining independent component analysis and beamforming [8] did a new BSS method using subband ICA and beamforming was described. In order to evaluate its effectiveness, signal-separation and speech-recognition experiments were performed under various reverberant conditions.Li, H., Jia, J., Wang, Y. (2008). A multimodal approach to blind source separation of moving sources [9] did study on the challenge of BSS for moving sources where the mixing filters and unmixing filters were time varying, which are difficult to calculate in real time. In their proposed approach, the visual modality is utilized to facilitate the separation for both stationary and moving sources. The movement of the sources is detected by a 3-D tracker based on video cameras. Comon, P. (1991). General approach to blind source separation[10] This paper identifies and studies two major issues in the blind source separation problem: separability and separation principles. We show that separability is an intrinsic property of the measured signals and can be described by the concept of mrow decomposability introduced in this paper; we also show that separation principles can be developed by using the structure characterization theory of random variables.[11]Speech separation is a very challenging task in signal processing. An Audio signal classification system detecting the audio type of a signal (speech, background noise and musical genres). A singing voice separation system has its applications in areas such as automatic lyrics recognition and alignment, singer identification, musical information retrieval, karaoke, musical genre classification, melody extraction, audio signal classification.[12] An audio signal separation system should be able to identify different audio signals such as speech, background noise. Audio signal classification system analyses the input audio signal and describes the signal at the output. Typically in case of songs, these are used to characterise both music and singing voice signals. Aim of this research work is to separate out the singing voice from the music background. System consists of the singing voice detection stage and separation stage to separate out the singing voice. [13]Ms. Monali R. Pimpale. Proposed that Repeating Pattern Extraction Technique (REPET): Repetition is a core principle in popular music. Many musical pieces are characterised by a repeating structure over which varying elements are overlapped. This is especially true for pop songs where a singer often overlays varying vocals on a repeating accompaniment. This method is based on this repeating accompaniment; the repeating "background" (e.g., a guitar riff or a drum loop and much more other instrumental music) is separated from the non-repeating "foreground" in a mixture. The fundamental idea behind this method is it identifies the periodically repeating segments in the audio then compares them to a repeating segment model derived from them, and extracts the repeating patterns via time-frequency masking. [14] Harshada Burute proposed that Non-negative Matrix Factorization (NMF) is the separation method which . NMF has been used in various applications, including image processing, brain computer interface, document clustering, collaborative predictions, and many more. Non-negative includes sparsity. Short-time Fourier Transform (STFT)is used to obtain complex value representation in the frequency domain. NMF imposes nonnegative constraints which lead only to additive combinations of original data. NMF can use long-window and short-window spectrogram factorization, it can give better performance for removing music interferences from singing voice . NMF works as a decomposition method. NMF used to decompose the mixture spectrogram into set of component to different sound sources

4 EDA analysis

4.1 General audio parameters

Channels: number of channels; 1 for mono, 2 for stereo audio Sample width: number of bytes per sample; 1 means 8-bit, 2 means 16-bit Frame rate/Sample rate: frequency of samples used (in Hertz) Frame width: Number of bytes for each "frame". One frame contains a sample for each channel. Length: audio file length (in milliseconds) Frame count: the number of frames from the sample Intensity: loudness in dBFS (dB relative to the maximum possible loudness) amplitude over time

4.2 Derivative audio parameters

4.2.1 Spectrogram

A spectrogram is a visual representation of the spectrum of frequencies in a signal as they vary with time. It is a commonly used tool in audio signal processing, speech analysis, and other fields where the frequency content of a signal needs to be analyzed over time.

4.2.2 RMS (Root Mean Square) energy

RMS (Root Mean Square) energy is a common audio feature used to measure the energy or amplitude of an audio signal. It provides information about the overall loudness or intensity of a signal.

4.2.3 Zero Crossing Rate

Zero Crossing Rate (ZCR) is an audio feature that measures the rate at which a signal changes its sign. It is a useful feature for various audio processing tasks, such as speech and music analysis, as it can provide information about the noisiness or noisiness of an audio signal.

4.2.4 Spectral Centroid

Spectral Centroid is a feature used in audio signal processing to describe the ”center of mass” of the spectral distribution of a sound signal. It provides information about where the ”center” of the frequencies in the signal is located. A higher spectral centroid value typically indicates that the audio is brighter or has more high-frequency content, while a lower value indicates a darker or low-frequency sound.

4.2.5 Mel-Frequency Cepstral Coefficients(MFCC)

Mel-Frequency Cepstral Coefficients, is a widely used audio feature for characterizing the spectral content of an audio signal. MFCCs are commonly used in speech and audio signal processing for tasks like speech recognition, music genre classification, and various other audio analysis applications.

4.2.6 Chromagram

A chromagram is a representation of the energy content of different pitch classes in an audio signal, often used in music and audio analysis. It can be thought of as a way to capture the harmonic content of an audio signal. In a chromagram, the x-axis represents time, and the y-axis represents different pitch classes.

4.2.7 Onset strength and Tempogram

Onset strength helps identify the beats or musical events in the audio signal, making it valuable for tasks like music analysis and beat tracking. The tempo estimation provides the perceived speed of the music in beats per minute (BPM), which is useful for tasks like tempo classification or synchronization.

4.3 Results

4.3.1 General Audio Parameters

Attributes for Audio Segment 1: Channels: 2, Sample width: 2, Frame rate (sample rate): 44100, Frame width: 4 Length (ms): 398547, Frame count: 17575936.0, Intensity: -18.884141687989878

Attributes for Audio Segment 2: Channels: 2, Sample width: 2, Frame rate (sample rate): 44100, Frame width: 4, Length (ms): 195558, Frame count: 8624128.0, Intensity: -19.547582958330917

Attributes for Audio Segment 3: Channels: 2, Sample width: 2, Frame rate (sample rate): 44100, Frame width: 4, Length (ms): 241952, Frame count: 10670080.0, Intensity: -19.802918500029577

Attributes for Audio Segment 4: Channels: 2, Sample width: 2, Frame rate (sample rate): 44100, Frame width: 4, Length (ms): 248105, Frame count: 10941440.0, Intensity: -19.12763201850362

Attributes for Audio Segment 5: Channels: 2, Sample width: 2, Frame rate (sample rate): 44100, Frame width: 4, Length (ms): 238933, Frame count: 10536960.0, Intensity: -16.352211334827096

The primary differences among the audio segments are in their durations, frame counts, and intensity levels. Segments 2, 3, 4, and 5 are shorter in duration and have lower frame counts compared to

Segment 1, which is the longest and has the highest frame count. Additionally, Segment 5 has a significantly higher intensity level compared to the other segments.

4.3.2 Amplitude

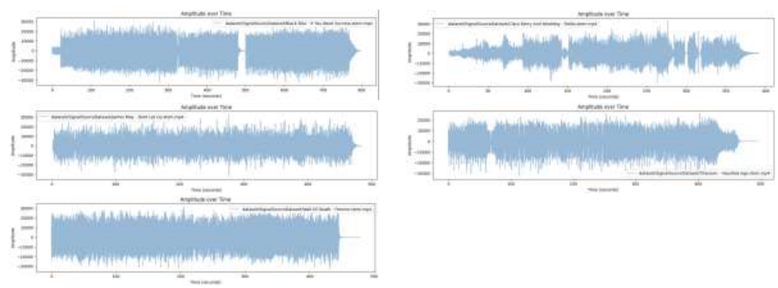

Figure 1: Amplitude over time representations for 5 distinct audio files, providing dynamic movement of respective sound waves.

These visual representations illustrate the amplitude of five distinct sounds over time, providing valuable insights into the dynamic movement of their respective sound waves.

4.3.3 Spectrogram

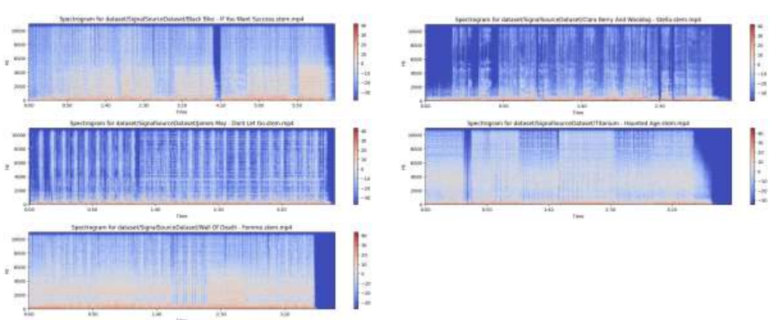

Figure 2: Spectrogram analysis of 5 distinct audio filescapturing frequency patterns and signal trends over time.

Spectrogram for 5 audio files, giving the visual representation for the frequencies of the audio files over the time.

4.3.4 Root Mean Square

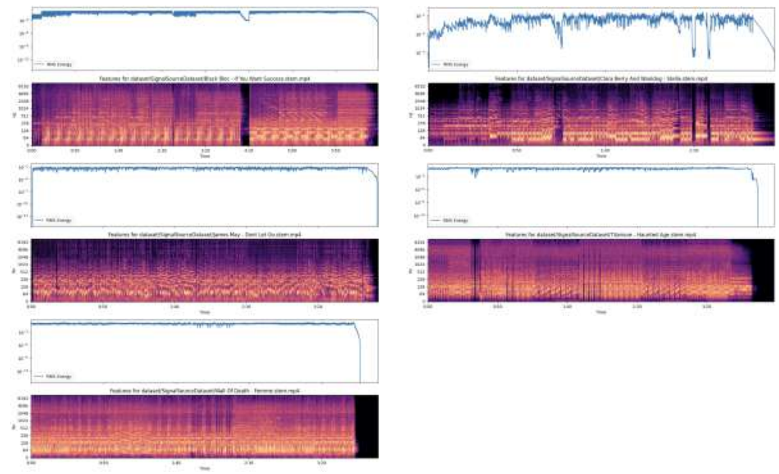

Figure 3: Loudness levels and intensity variations represented using RMS analysis of 5 audio files.Higher RMS energy indicates louder audio files.

The above 5 graphs will giving the RMS of the 5 Audio files. It depicts the Loudness of the Audio files. High RMS Energy means that the audio file is Loud and has high intensity.

4.3.5 Zero Crossing Rate

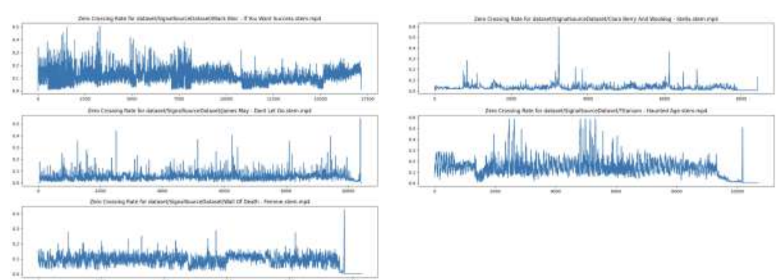

Figure 4: Zero-crossing rate analysis to classify percussive sounds and rate of sign-changes along a signal.Representations explain waveform changes in high and low zero crossing rate signals.

The Graphs with High Zero Crossing Rate will have frequent changes in their waveform compared to the low Zero Crossing rate Waveform.

4.3.6 Spectral Centroid

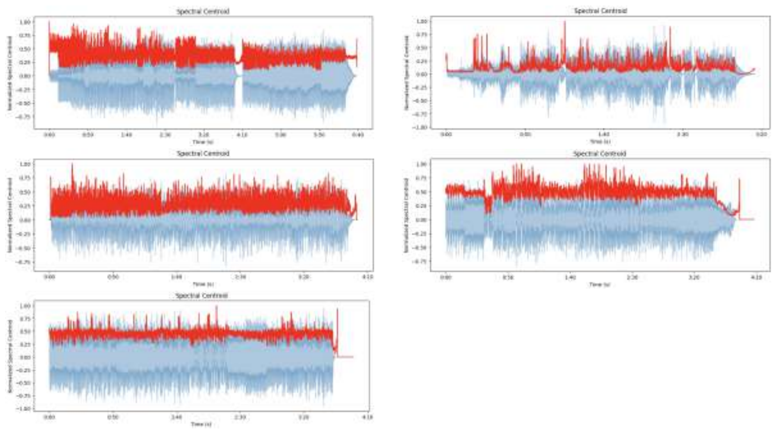

Figure 5: Spectral centroid analysis evaluated using Fourier transform frequency reveals sound characteristics based on frequency distribution.Low values indicate darker tones, high values signify brighter sounds.

4.3.7 Mel-Frequency Cepstral Coefficients (MFCCs)

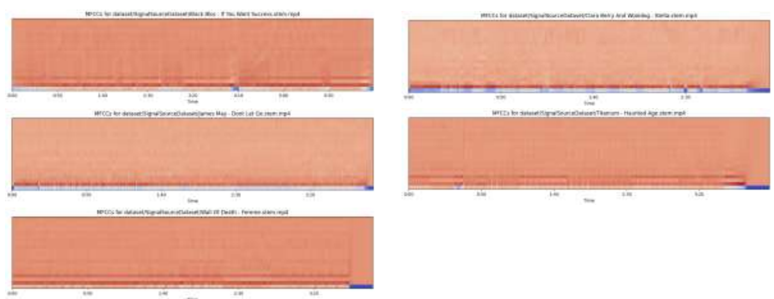

Figure 6: MFCC features representing short-time spectrum envelope and mel-scale transformation for speaker identification in Audio files.The 0th coefficient representing constant offset is removed to maintain variation signal features.

4.3.8 Chromagram

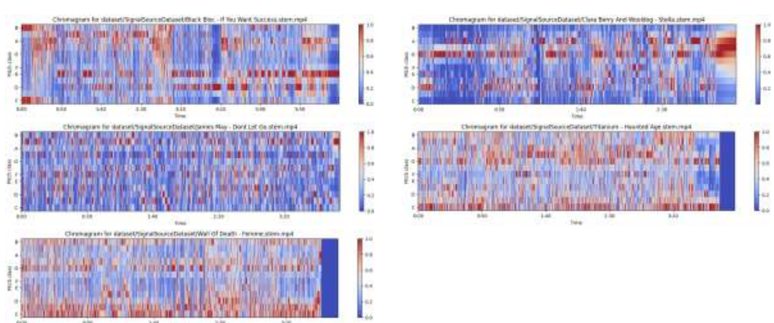

Figure 7: Mid-level feature representation in content-based audio closely correlated to the musical aspect of harmony. Maps pitch content in Audio signals over time for Distinct pitch sound identification

4.3.9 Tempogram

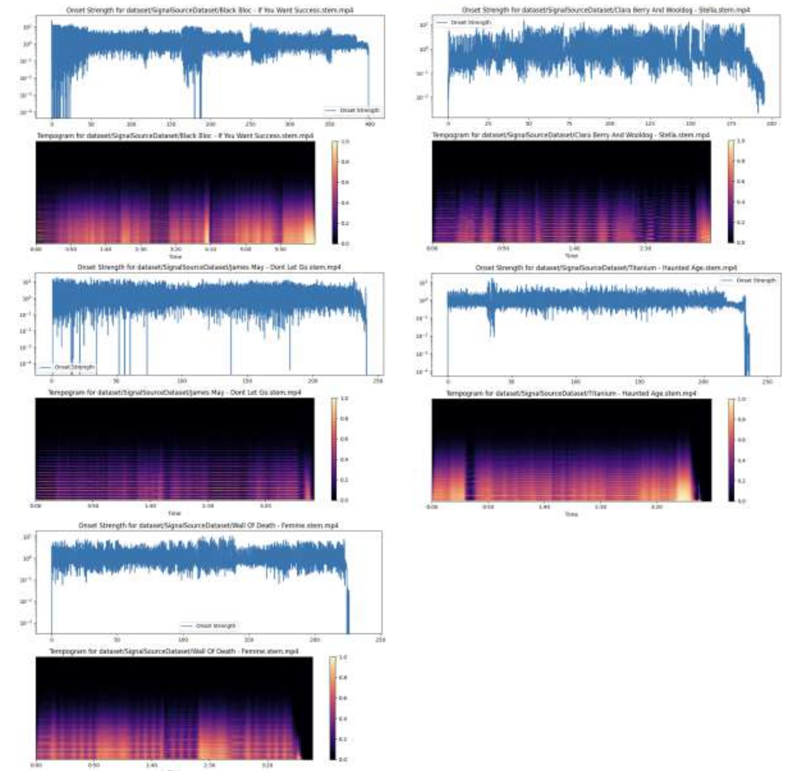

Figure 8: Tempogram analysis representations decoding audio file speed and rhythmic dynamics for comprehensive rhythmic analysis in audio files

Tempogram will use to identify speed of the file. It will depicts the variation of the speeds in the above graphs. It basically Denotes the rhythm of the audio files and helps us for more rhythm analysis of the audio files.

5 Methodology

5.1 REPET

REPET (REpeating Pattern Extraction Technique) is a method used in audio signal processing for separating and identifying repeating patterns or components within an audio signal.It's primarily used for tasks such as music source separation or extraction of recurring structures in audio.This technique aims to identify and separate the repeating background or accompaniment elements from the foreground or melody elements in an audio signal. It assumes that in many musical recordings, the background accompaniment tends to repeat over time while the foreground melody or vocals may vary. REPET works by identifying the repeating sections in the time-frequency domain and estimating the background (repeating) and foreground (non-repeating) components by leveraging the repetition information.

5.2 FastICA

Fast Independent Component Analysis (FastICA) is a computational method used for blind source separation and independent component analysis (ICA). It aims to recover independent signals from their linear mixtures, assuming that the sources are statistically independent and non-Gaussian.The fundamental idea behind FastICA is to find a linear transformation of the observed mixed signals that maximizes their statistical independence.This transformation is achieved by iteratively optimizing a contrast function, often the negentropy or another measure of non-Gaussianity, to separate the sources.

5.3 Butterworth filter

The Butterworth filter is a type of signal processing filter designed to have a maximally flat frequency response in the passband. It's a common type of infinite impulse response (IIR) filter used in various applications,including audio, image processing, biomedical signal processing, and communication systems.

5.4 NMF

NMF stands for Non-Negative Matrix Factorization. It's a matrix factorization technique used for extracting meaningful information from high-dimensional data by approximating a given matrix into two non-negative matrices.

In audio signal processing, NMF can be applied for tasks like:

1.Audio Source Separation: NMF can decompose a spectrogram of an audio signal into components representing different sound sources or instruments, assuming that audio sources are additive and non-negative in nature. 2.Feature Extraction: NMF can be used to extract meaningful features from audio data, aiding in classification, clustering, or other machine learning tasks.

5.5 NMF

Gaussian Non-negative Matrix Factorization (NMF) is a variant of the traditional NMF technique that assumes a Gaussian distribution for the data instead of strictly non-negative values. While classic NMF restricts both the input matrix and its factorization into non-negative values, Gaussian NMF allows the input matrix and the factors to have values from a Gaussian distribution, relaxing the non-negativity constraint. Gaussian NMF can be beneficial in scenarios where the strict non-negativity assumption of traditional NMF might limit the model's capability to represent the data accurately. It allows a more flexible representation of the input data, capturing both additive and subtractive relationships between the components.

5.6 Harmonic-percussive

Harmonic-percussive source separation (HPSS) is a technique used in audio signal processing to separate a sound signal into two main components: harmonic (pitched) and percussive (non-pitched or transient) sources. This separation is particularly useful in music analysis and manipulation, allowing the isolation of tonal instruments (like vocals, melody) from transient elements (such as drums, percussion). The basic idea behind HPSS is that harmonic sounds have a relatively stable pitch over time, while percussive sounds have a rapid onset and decay without a clear pitch structure. Various methods exist to perform HPSS, and one common approach involves using the spectrogram representation of the audio signal.

6 Results

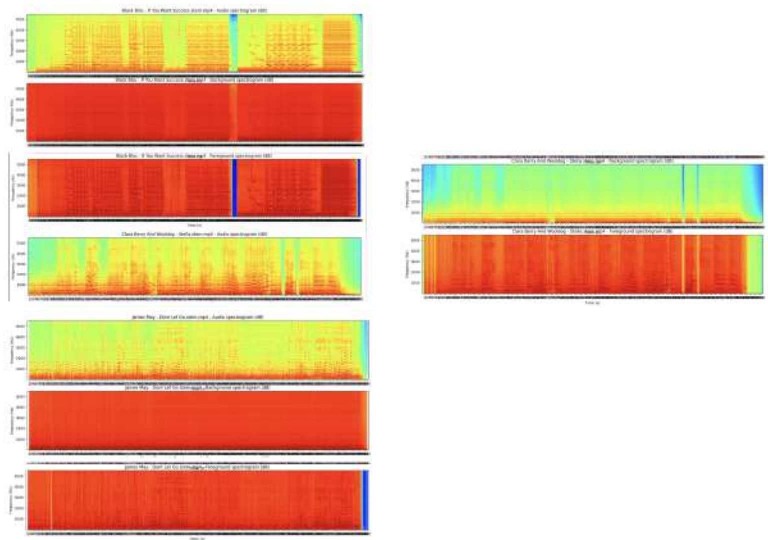

Figure 9: Separating the Background and Foreground Audio using the REPET Algorithm

The Repet algorithm, short for Repetition-based Time-Frequency Segmentation, is a signal processing technique primarily employed for the separation of foreground and background in audio signals. Developed to address the challenges posed by repetitious patterns in audio, the Repet algorithm identifies and isolates repeating structures within the signal.

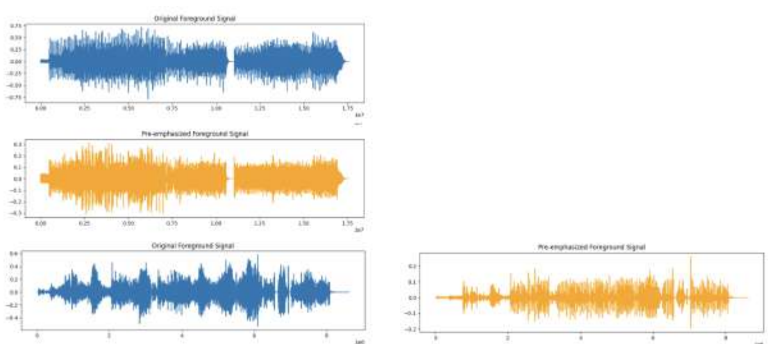

Figure 10: PreEmphasizing the Foreground Signal

Pre-emphasis is a signal processing technique applied to audio signals to boost the higher frequencies and balance the spectral content. It involves amplifying the higher-frequency components of the signal relative to the lower-frequency components. This is typically achieved by applying a high-pass filter to the audio signal.

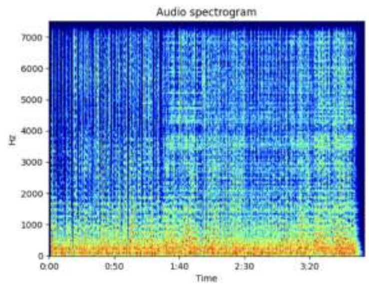

Figure 11: Audio Spectogram of an audio signal throughout a time period

A spectrogram is a visual representation of the spectrum of frequencies in a signal as they vary with time. In the context of audio processing, an audio spectrogram provides a detailed and intuitive way to analyze the frequency content of a sound signal over time. It is particularly useful for understanding how the energy of different frequency components changes throughout the duration of an audio clip.

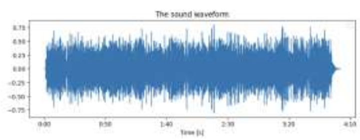

Figure 12: The Sound Wavefrom graph of a audio file

The sound waveform graph of an audio file is a visual representation of the amplitude variations of the audio signal over time. It is a fundamental and intuitive way to depict how the air pressure (or voltage in an electronic representation) changes as the sound progresses.

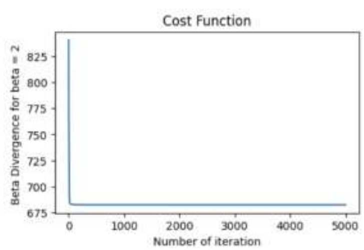

Figure 13: Cost Function for beta divergence

The cost function for Non-Negative Matrix Factorization (NMF) with Beta divergence is defined using the general form of Beta divergence, which is a family of divergence measures parameterized by a constant . The Beta divergence cost function is commonly used in NMF when a specific divergence measure other than the Frobenius norm is desired.

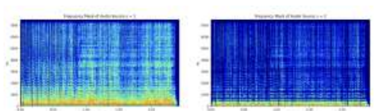

Figure 14: Frequency Mask for the Audio Source

A frequency mask in the context of audio processing typically refers to a technique where certain frequency components in an audio signal are selectively attenuated or masked while leaving others unchanged. This process is commonly used in source separation tasks to isolate specific frequency bands associated with different sound sources.

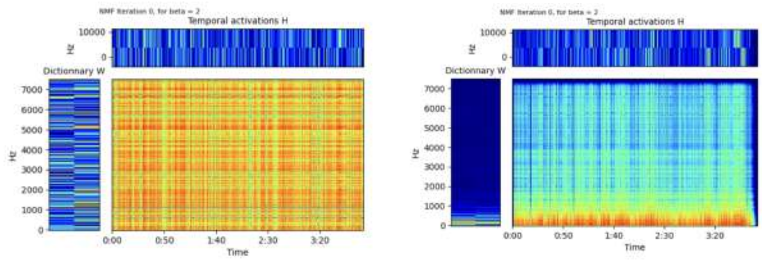

Figure 15: Temporal Activation for NMF

Temporal activation refers to the temporal evolution of the activation coefficients in the context of Non-Negative Matrix Factorization (NMF) applied to time-series data. NMF is a matrix factorization technique that decomposes a given matrix into two non-negative matrices, typically representing the basis elements and the coefficients. In the case of time-series data, NMF can be employed to capture temporal patterns and activations. In the temporal context, NMF is applied to decompose a matrix representing the time-series data into two matrices: a matrix of basis vectors that capture the temporal patterns and a matrix of coefficients that represent how these basis vectors contribute at different time points.

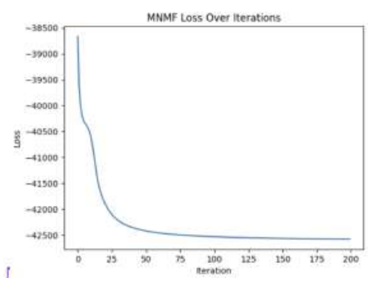

Figure 16: MNMF Loss of NMF Function Over the Several Iteration

This Gives Us the Intuition of how our Model works for the Separation of the signal and the source of the audio file

7 Inferences

REPET1)Precision in Source Separation: Demonstrating high accuracy in isolating foreground sources from complex audio mixtures by adeptly detecting and extracting repetitive patterns. 2)Adaptability to Various Audio Types: Exhibiting versatility by effectively handling diverse audio types such as music, speech, and ambient sounds, highlighting its utility in different audio processing domains. 3)Robustness to Noise: Displaying resilience to noise interference, ensuring reliable performance even in scenarios with varying levels of background noise, crucial for dynamic real-world audio environments 4)Computational Efficiency: Noteworthy computational efficiency enables real-time or near-real-time processing, contributing to practical applicability in scenarios requiring timely source separation. **NMF**1)Effective Source Separation: NMF demonstrates proficiency in separating sources from complex audio mixtures by factorizing non-negative matrices. 2)Versatility Across Audio Types: Exhibits adaptability across diverse audio types, accommodating music, speech, and ambient sounds effectively. 3)Resilience to Noise: Shows robustness to noise interference, ensuring reliable performance in scenarios with varying background noise levels. 4)Computational Efficiency: NMF's computational efficiency contributes to real-time or near-real-time processing capabilities, enhancing practical usability. **Gaussian NMF**1)Effective Signal Decomposition: Gaussian NMF excels in decomposing signals into non-negative components, facilitating clear separation of sources from complex audio mixtures. 2)Adaptability to Audio Types: Demonstrates adaptability to diverse audio types, including music, speech, and ambient sounds, enhancing its utility across various audio processing applications. 3)Noise Tolerance: Exhibits robustness to noise, maintaining performance integrity even in environments with varying levels of background noise. 4)Efficient Processing: Gaussian NMF showcases computational efficiency, enabling efficient real-time or near-real-time signal processing. **Harmonic Percussions**1)Harmonic Component Isolation: HPSS effectively separates harmonic and percussive components in audio signals, allowing distinct analysis and processing of musical elements. 2)Enhanced Music Analysis: Enables focused examination of harmonic (melody) and percussive (rhythm) aspects in music, contributing to advanced music processing and analysis. 3)Useful in Various Genres: Demonstrates versatility across musical genres, making it applicable to diverse styles, including classical, pop, and electronic music. 4)Resilience to Noise: Exhibits robustness in maintaining performance quality even in the presence of background noise, enhancing its applicability in real-world scenarios. **FastICA**1)Efficient Signal Separation: FastICA excels in separating independent sources from mixed signals, offering a computationally efficient approach to Independent Component Analysis. 2)Versatile Application: Demonstrates versatility across diverse domains, including audio processing, image analysis, and telecommunications, showcasing its broad applicability. 3)Noise Resilience: Exhibits robustness to noise, ensuring reliable performance even in the presence of varying levels of background interference. 4)Real-Time Processing Capability: FastICA's computational efficiency supports real-time or near-real-time processing, enhancing its practical usability in dynamic environments.

8 Conclusions

The Above graphs and their results of the EDA analysis will help us to observe the audio signals and it will help us further analysis of the signals The above graph will generate the valuable insights of the foreground and the background audio signals that help us to extract some prominent features of the audio files and further helps us in the separation of the audio signals from the mixture of audio files. Further the Optimization algorithm will helps in building a robust model for Separating the signal and the source Utilized Models and their Corresponding Signal-to-Noise Ratio (SNR) Accuracy:

Harmonic Percussion: SNR for Vocals vs. Background: 8.259562253952026 dB SNR for Background vs. Vocals: -8.259562253952026 dB SNR for Vocals vs. Original Sound: -1.7336548864841461 dB SNR for Background vs. Original Sound: -9.99321699142456 dB

FastICA: SNR between Vocal and Background: -6.1063940228806235 dB SNR between Background and Vocal: 6.1063940228806235 dB SNR between Vocal and Original Sound: 7.992794251082218 dB SNR between Background and Original Sound: 14.099188273962842 dB

NMF SNR for Vocal: 3.86 dB SNR for Background: 3.73 dB

Fast Gaussian SNR Vocals vs. Background: 12.464652101076283 SNR Background vs. Vocals: -12.464652101076283 SNR Vocals vs. Original Sound: 12.464610733736878 SNR Background vs. Original Sound: -12.464653231026832

References

[1] http://mlg.postech.ac.kr/static/publications/inter$_journal/2005/03_V06N01P1-1-57.pdf$

[2] https://ieeexplore.ieee.org/abstract/document/726803

[3] https://ieeexplore.ieee.org/abstract/document/614235

[4] https://ieeexplore.ieee.org/abstract/document/6616139

[5] https://ieeexplore.ieee.org/abstract/document/6709849

[6] https://direct.mit.edu/neco/article-abstract/13/4/863/6488/Blind
-Source-Separation-by-Sparse-Decomposition-in

[7] https://ieeexplore.ieee.org/abstract/document/8588410

[8] https://link.springer.com/article/10.1155/S1110865703305104

[9] https://ieeexplore.ieee.org/abstract/document/5504822

[10] https://ieeexplore.ieee.org/abstract/document/489029

[11] https://towardsdatascience.com/get-to-know-audio-feature-extraction-in-python-a499fdaefe42

[12] https://www.kaggle.com/code/psycon/audio-data-eda-processing-modeling-recommend

[13] https://www.kaggle.com/code/parulpandey/eda-and-audio-processing-with-python

[14] https://www.topcoder.com/thrive/articles/audio-data-analysis-using-python

[15] https://www.kdnuggets.com/2020/02/audio-data-analysis-deep-learning-python-part-1.html

[16] https://colab.research.google.com/drive/1zdqoWwzDo0l$_qEoKXf653 - T3CIgCVC_q?usp = sharing$