Signal and Source Separation Report (Group Number-29)

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. In this Project, we explore the application of exploratory data analysis (EDA) techniques in conjunction with spectrogram analysis to separate mixed signals into their constituent sources. 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. 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. These features are employed to facilitate the separation of sources from the complex audio mixture. By harnessing these techniques, we enhance our ability to unveil the underlying structure and content of mixed signals, ultimately contributing to advancements in various fields reliant on signal and source separation. Keywords: Signal Separation, Source Separation, Exploratory Data Analysis (EDA), Spectrogram, Short-Time Fourier Transform (STFT), Root-mean-square (RMS), Zero Crossing Rate, Mel-Frequency Cepstral Coefficients (MFCCs), Chroma, Tempogram.

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

The objective of this project is to explore the application of exploratory data analysis (EDA) techniques in conjunction with spectrogram analysis to separate mixed signals into their constituent sources.

Specifically, the study will investigate the following: The effectiveness of EDA techniques in extracting and analyzing crucial features from the spectrogram of a mixed signal. The use of these features to facilitate the separation of sources from the complex audio mixture. The application of the proposed approach to real-world signal and source separation problems. The successful completion of this study could lead to significant advancements in various fields, such as audio processing, speech recognition, and music analysis. For example, the proposed approach could be used to improve the performance of speech recognition systems in noisy environments or to develop new tools for music transcription and analysis.

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. 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. 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. 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. 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 .

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. 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. 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. 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. 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.

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 Chromogram

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

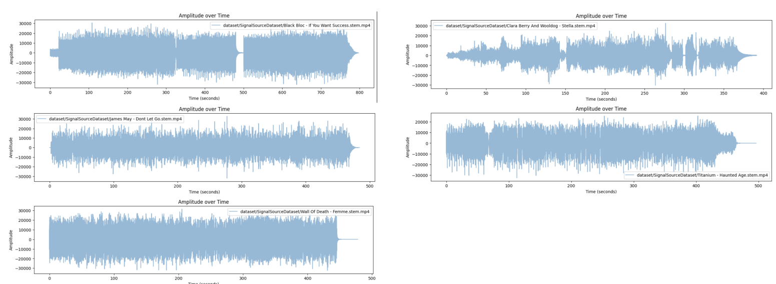

Figure 1: Amplitude over time representations for 5 distinct audio files, providing dynamic movement of respective sound waves.

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

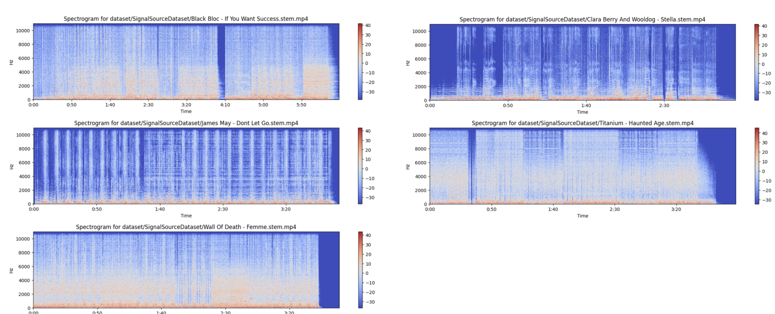

Figure 2: Spectrogram analysis of 5 distinct audio filescapturing frequency patterns and signal trends over time.

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

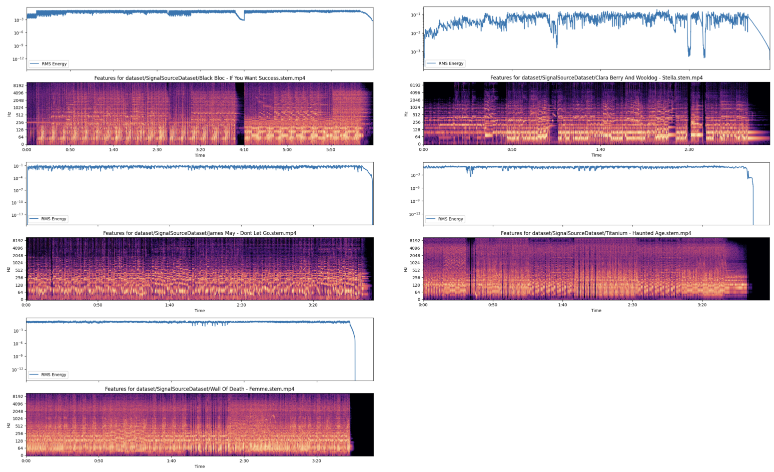

Figure 3: Loudness levels and intensity variations represented using RMS analysis of 5 audio files.Higher RMS energy indicates louder audio files.

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

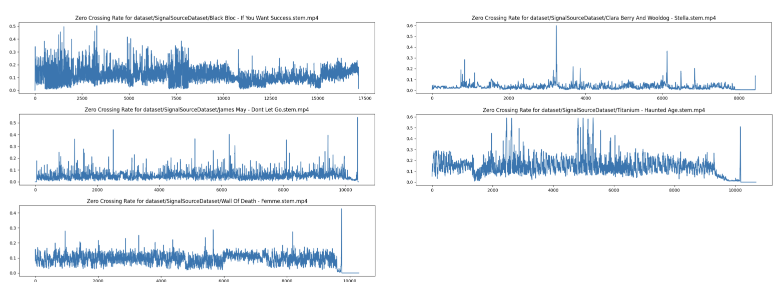

Figure 4: Zero-crossing rate analysis to classify percussive sounds and rate of sign-changes along a signal.Representations explain waveform changes in high and low zero crossing rate signals.

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

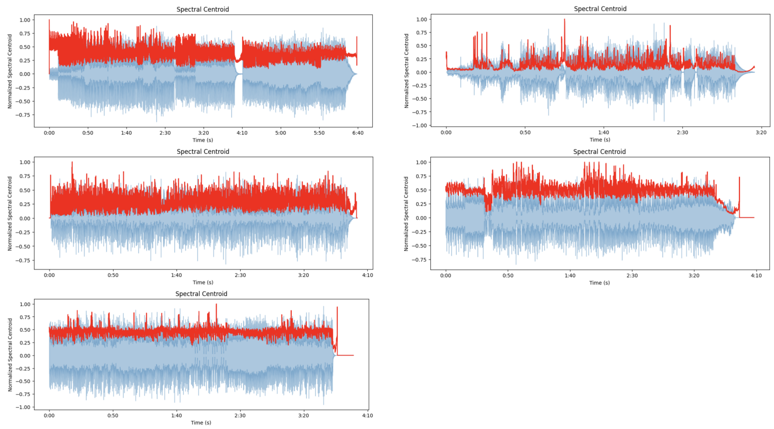

Figure 5: Spectral centroid analysis evaluated using Fourier transform frequency reveals sound characteristics based on frequency distribution.Low values indicate darker tones, high values signify brighter sounds.

4.3.7 Mel-Frequency Cepstral Coefficients (MFCCs)

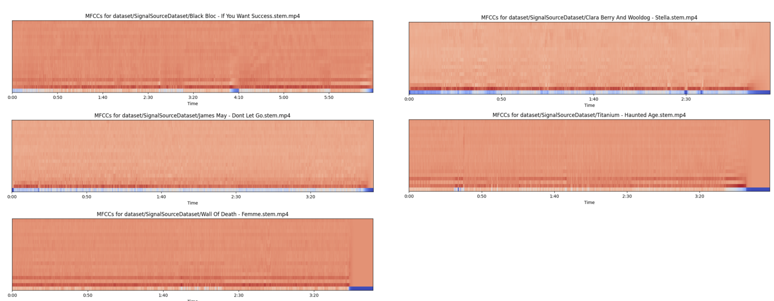

Figure 6: MFCC features representing short-time spectrum envelope and mel-scale transformation for speaker identification in Audio files.The 0th coefficient representing constant offset is removed to maintain variation signal features.

4.3.8 Chromogram

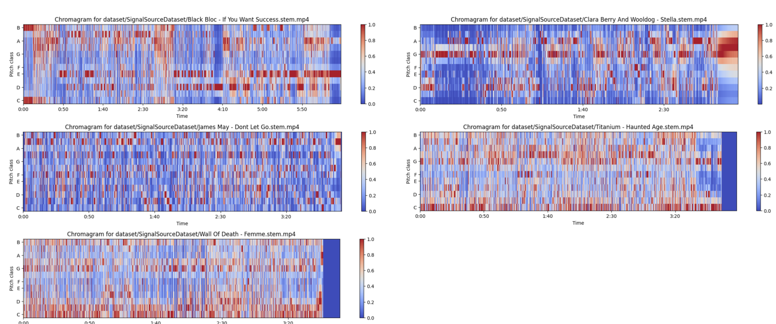

Figure 7: Mid-level feature representation in content-based audio closely correlated to the musical aspect of harmony. Maps pitch content in Audio signals over time for Distinct pitch sound identification

4.3.9 Tempogram

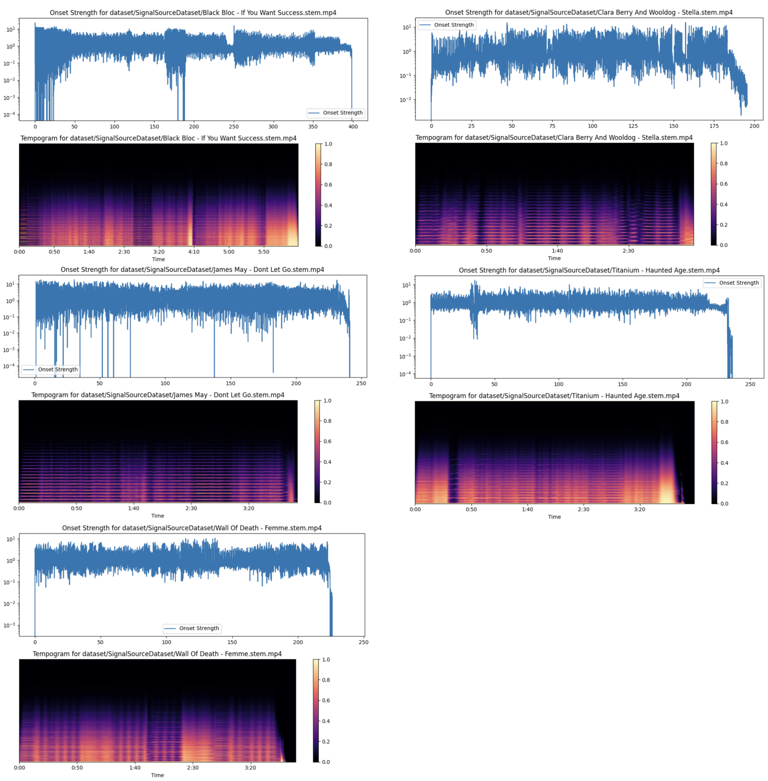

Figure 8: Tempogram analysis representations decoding audio file speed and rhythmic dynamics for comprehensive rhythmic analysis in audio files

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 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 audio signals and help us to extract some prominent features of the audio files and further helps us in the feature engineering and model Training.

References

[1] http://mlg.postech.ac.kr/static/publications/inter$_{j}ournal/2005/03_{V}06N01P1-1-57.pdf$

[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