

**CSL7770 : Speech Understanding**

**Assignment 1 Task 2(b) Report**

**Task 2(b): Comparative Spectral Analysis of Four  
Songs**

**Name: Anuj Rajan Lalla**

**Roll Number: B22AI061**

**Date of Submission:02-02-2025**

# 1 Introduction

This report is submitted as part of **Assignment 1** in the course on *Speech Understanding*. It addresses **Task 2(b)**: selecting four songs from four different genres and comparing their spectrograms. The objective is to analyze the time-frequency characteristics of each snippet, drawing from fundamental principles commonly employed in speech signal analysis.

All audio material has been sourced from Pixabay.com, and each track was analyzed in the .wav format. The source license and references are included in Section 7.

## 2 Task Description

### Task 2(b) Statement:

*“Select 4 songs from 4 different genres and compare their spectrograms. Analyze the spectrograms and provide a detailed comparative analysis based on your observations and speech understanding.”*

In this report, four songs were selected:

- **Retro City** by Slicebeats (Rock)
- **SCI-FI Background** by SlavaMSound (Electronic/Ambient)
- **Spring Mood** by HD-Studio (Classical)
- **Vinyl** by MiraDar (Jazz)

Each audio track was evaluated by generating spectrograms for both the full audio and a 30-second snippet (from 30 s to 60 s). The results and interpretations are outlined below.

## 3 Methodology

### 3.1 Spectral Analysis

To study the time-frequency distribution of each excerpt, short-time Fourier transforms (STFT) were computed with:

- A 1024-point FFT size
- A 512-sample hop length
- A Hann window function

Amplitude spectra were converted to decibels (dB) to facilitate clearer visualization.

## 3.2 Spectral Features

For the 30 s–60 s segment of each track, the following metrics were extracted:

- **Spectral Centroid (mean):** Approximate “center of mass” of the frequency spectrum.
- **Spectral Bandwidth (mean):** The spread of frequencies around the centroid.
- **Spectral Rolloff (mean):** The frequency below which 85% of total spectral energy resides.
- **RMS Amplitude (mean):** The average signal intensity (loudness).

## 4 Results

### 4.1 Spectrogram Visuals

Below are images showing each track’s full spectrogram and a 30 s–60 s snippet.

**Retro City (Rock)**

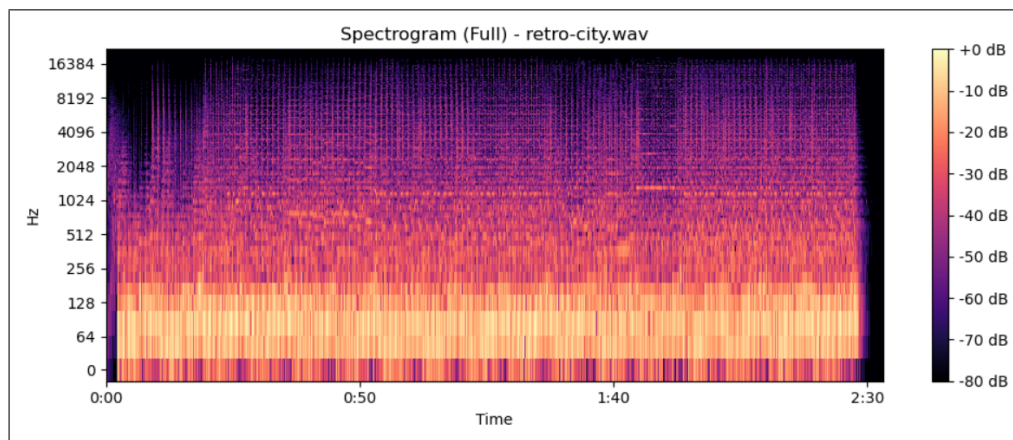


Figure 1: Retro City – Full Spectrogram

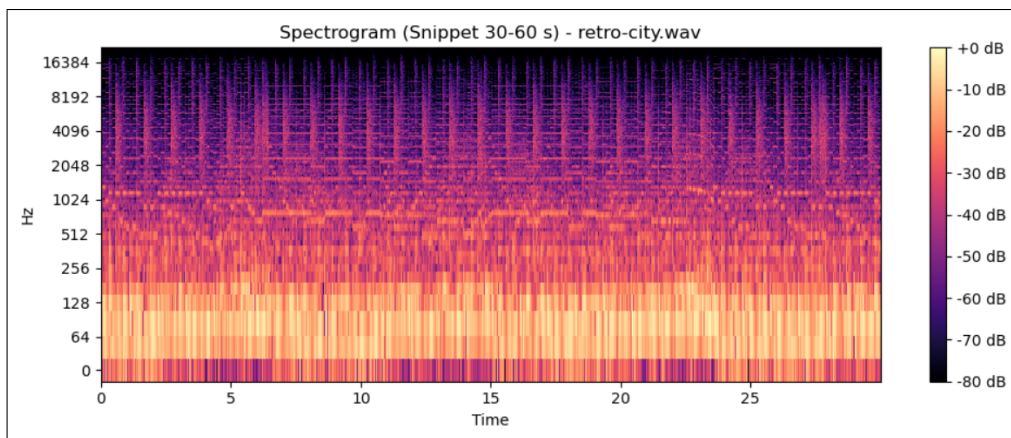


Figure 2: Retro City – Snippet Spectrogram (30–60 s)

### SCI-FI Background (Electronic/Ambient)

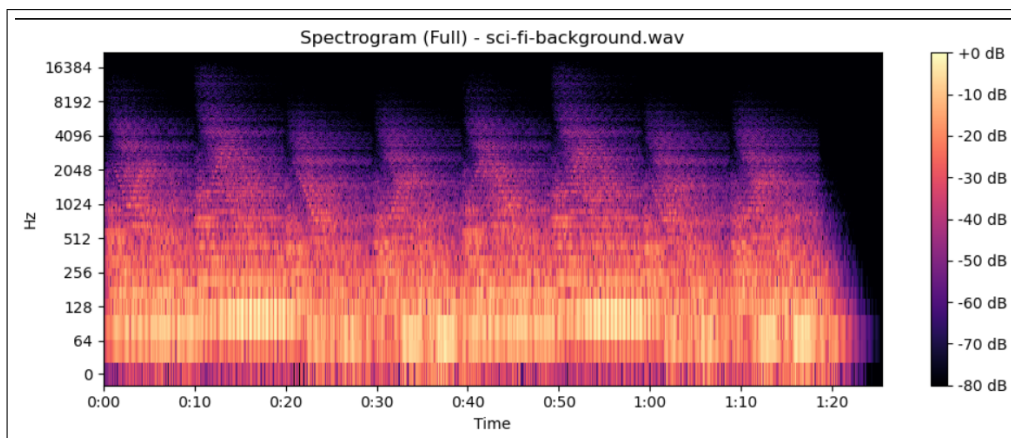


Figure 3: SCI-FI Background – Full Spectrogram

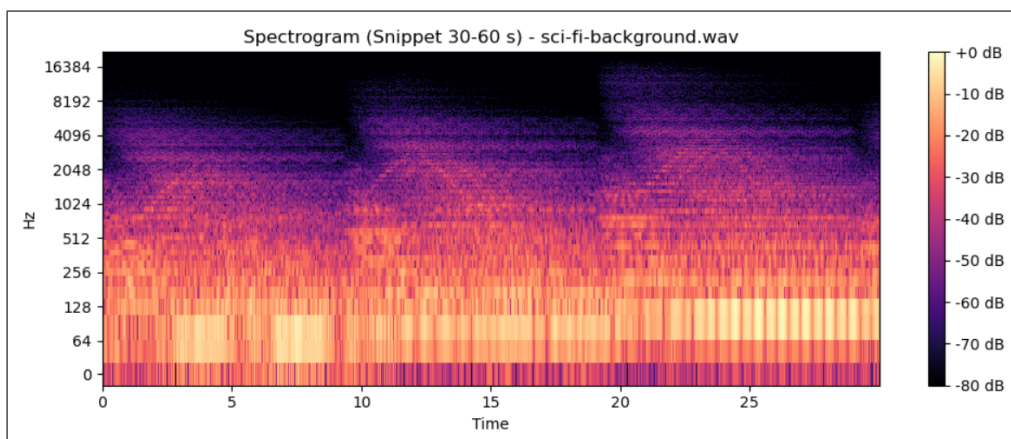


Figure 4: SCI-FI Background – Snippet Spectrogram (30–60 s)

## Spring Mood (Classical)

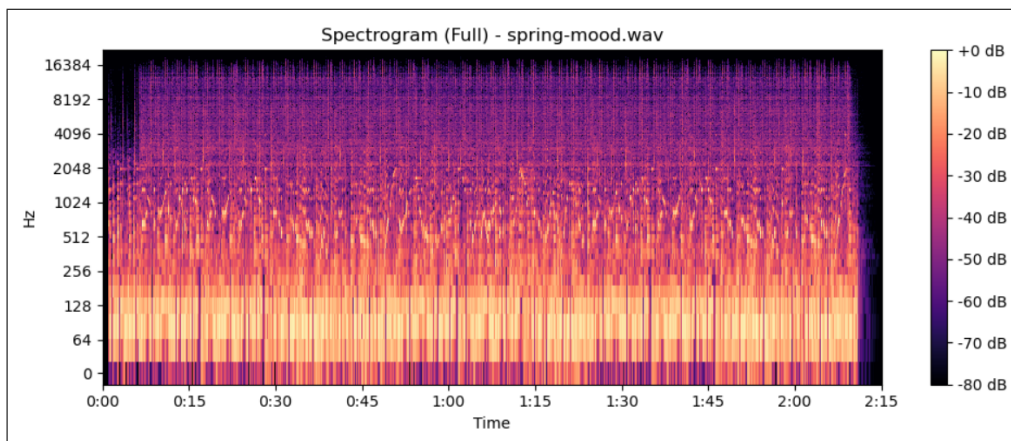


Figure 5: Spring Mood – Full Spectrogram

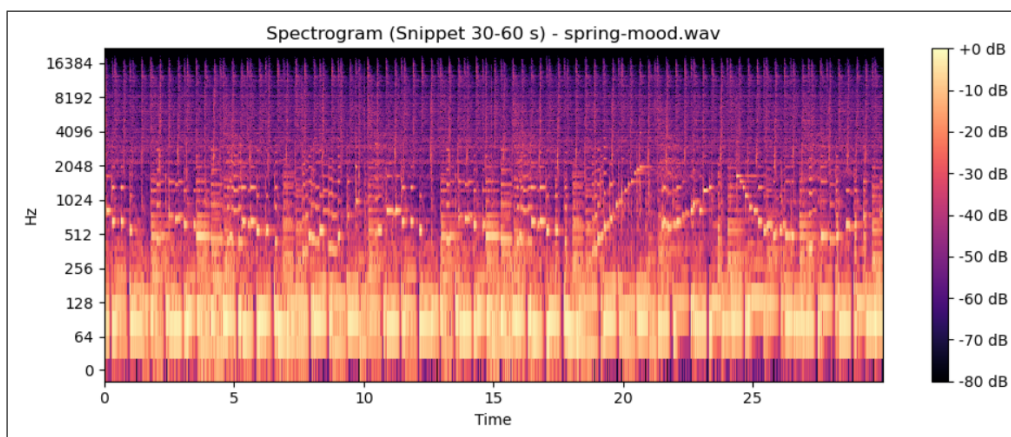


Figure 6: Spring Mood – Snippet Spectrogram (30–60 s)

## Vinyl (Jazz)

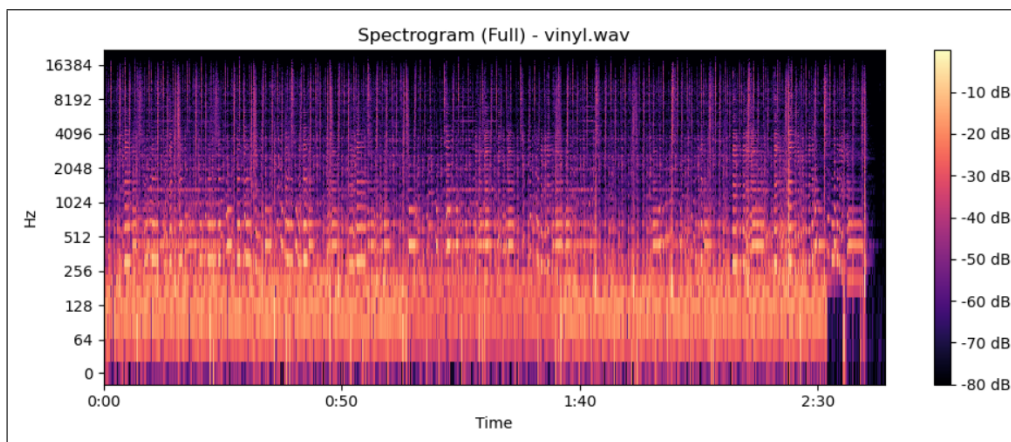


Figure 7: Vinyl – Full Spectrogram

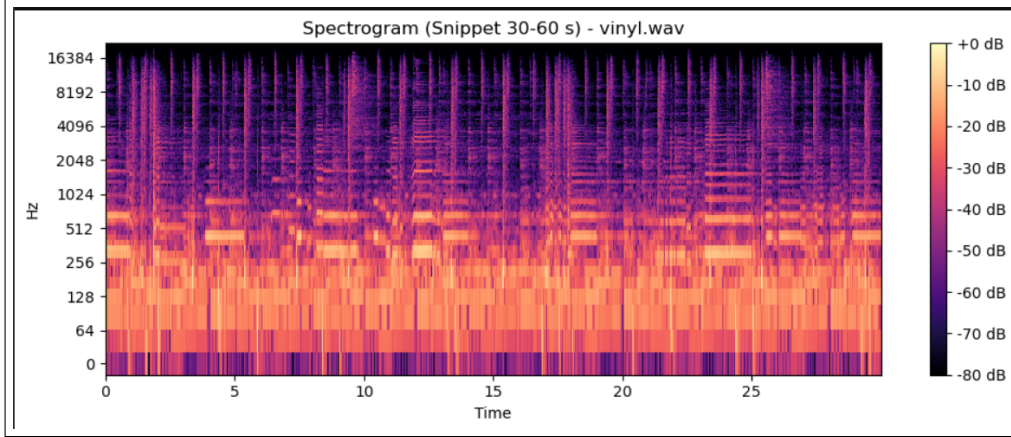


Figure 8: Vinyl – Snippet Spectrogram (30–60 s)

## 4.2 Quantitative Metrics

Table 1 summarizes the mean spectral metrics for the 30 s–60 s region of each track:

Table 1: Mean Spectral Metrics for Each Snippet (30–60 s)

Track	Centroid (Hz)	Bandwidth (Hz)	Rolloff (Hz)	RMS
Retro City	1929	2849	4257	0.169
SCI-FI Background	774	1517	1325	0.020
Spring Mood	2782	3747	6894	0.157
Vinyl	2349	3136	5075	0.089

## 5 Discussion

From Table 1, clear distinctions emerge among the four analyzed snippets:

- **Retro City (Rock):**
  - **Spectral Centroid (1929 Hz):** Indicates a strong presence in the mid-frequency range.
  - **RMS (0.169):** The highest among the four, reflecting robust overall amplitude typical of an energetic production style.
  - **Bandwidth (2849 Hz) & Rolloff (4257 Hz):** Suggest a moderate spread of frequencies extending into the upper range, though not as high as some other snippets.
- **SCI-FI Background (Electronic/Ambient):**
  - **Spectral Centroid (774 Hz):** Significantly lower, indicating emphasis on lower-frequency components and a more ambient, subdued texture.

- **RMS (0.020):** Very low loudness level, supporting a minimalistic or background-oriented production.
- **Bandwidth (1517 Hz) & Rolloff (1325 Hz):** Both are relatively small, reflecting limited high-frequency activity.
- **Spring Mood (Classical):**
  - **Spectral Centroid (2782 Hz):** The highest centroid, pointing to a strong presence in the higher frequency range.
  - **RMS (0.157):** Relatively high, suggesting a more sustained intensity level in this excerpt.
  - **Bandwidth (3747 Hz) & Rolloff (6894 Hz):** The broadest among the samples, highlighting extended high-frequency content.
- **Vinyl (Jazz):**
  - **Spectral Centroid (2349 Hz):** Falls between the extremes of SCI-FI Background and Spring Mood, indicating moderate brightness.
  - **RMS (0.089):** Middle-range loudness, neither as low as SCI-FI nor as high as Retro City.
  - **Bandwidth (3136 Hz) & Rolloff (5075 Hz):** Suggest a balanced distribution of energy across both low and high frequencies.

### Implications for Speech Understanding:

In speech analysis, metrics like spectral centroid, bandwidth, and rolloff help differentiate lower-frequency phonemes from higher-frequency ones. By analogy, a higher centroid/bandwidth here signals stronger high-frequency energy, while a lower centroid denotes emphasis on lower frequencies. RMS amplitude corresponds to overall loudness, analogous to speech volume. Thus, fundamental speech-analysis techniques can provide objective comparisons across different audio genres without requiring specialized musical knowledge.

## 6 Conclusion

The comparative spectrogram and metric analysis shows that each audio snippet exhibits distinctive frequency and amplitude characteristics:

- **Retro City (Rock):** Higher midrange energy and a strong overall amplitude.
- **SCI-FI Background (Electronic/Ambient):** Lower-frequency emphasis with a subdued loudness profile.
- **Spring Mood (Classical):** Substantial high-frequency content and relatively high loudness.
- **Vinyl (Jazz):** Moderate brightness and amplitude, occupying an intermediate position.

By drawing on fundamental methods from speech signal processing (such as STFT and spectral feature calculation), these snippets can be evaluated objectively without requiring genre-specific or instrument-specific knowledge. This highlights the versatility of basic speech analysis tools for a wide range of audio applications.

## 7 References

### Audio Tracks (Accessed via Pixabay):

- *Retro City* by Slicebeats
- *SCI-FI Background* by SlavaMSound
- *Spring Mood* by HD-Studio
- *Vinyl* by MiraDar

**Libraries used:** librosa, numpy, pandas, matplotlib, ffmpeg (for MP3 to WAV conversion)