Experimenting with Spectrograms and Windowing Techniques

UrbanSound8k Dataset Analysis

Joel Paul
Roll Number: M24DE2012

January 30, 2025

Contents

1 Introduction

This report explores different windowing techniques and their impact on spectrogram generation using the UrbanSound8k dataset. The study compares the performance of three windowing techniques: Hann, Hamming, and Rectangular windows. A classifier is trained to analyze the effect of windowing on classification accuracy. Additionally, spectrograms of different music genres are compared and a detailed comparative analysis is provided.

2 Dataset Description

The UrbanSound8k dataset contains environmental sounds categorized into ten classes. It is available at `https://goo.gl/8hY5ER`. The dataset is used to generate spectrograms and extract features for classification.

3 Task A: Windowing Techniques and Spectrogram Analysis

3.1 Windowing Techniques

Three windowing techniques were implemented:

- **Hann Window**: It is a smooth, bell-shaped window function that gradually tapers at the edges. It is widely used for spectral analysis as it reduces side lobes effectively. Provides smooth transitions and reduces spectral leakage.
  - Commonly used in speech processing and audio signal analysis
  - Suitable for FFT-based analysis when overlapping windows are used
- **Hamming Window**: It is similar to the Hann window but has a slightly higher main lobe width and lower side lobes, reducing spectral leakage even further.
  - Hamming window retains more signal energy at edges
  - Slightly less smooth but better frequency localization
- **Rectangular Window**: Simplest window, equivalent to no windowing.

- Worst spectral leakage but best frequency resolution
- Works well in cases where no overlapping windows are used

3.2 Methodology

Short-Time Fourier Transform (STFT) was applied to generate spectrograms using these techniques. The Python implementation follows the PyTorch framework.

```
import numpy as np
import torchaudio.transforms as T
import torch

# Hann window function
def hann_window(N):
    return 0.5 * (1 - np.cos(2 * np.pi * np.arange(N) / (N - 1)))

# Apply STFT to generate spectrogram
def generate_spectrogram(signal, sample_rate):
    transform = T.Spectrogram(n_fft=1024, hop_length=512, power=2)
    return transform(signal)
```

3.3 Comparison of Spectrograms

Figures 1, 2, and 8 show the spectrograms for the three windowing techniques.

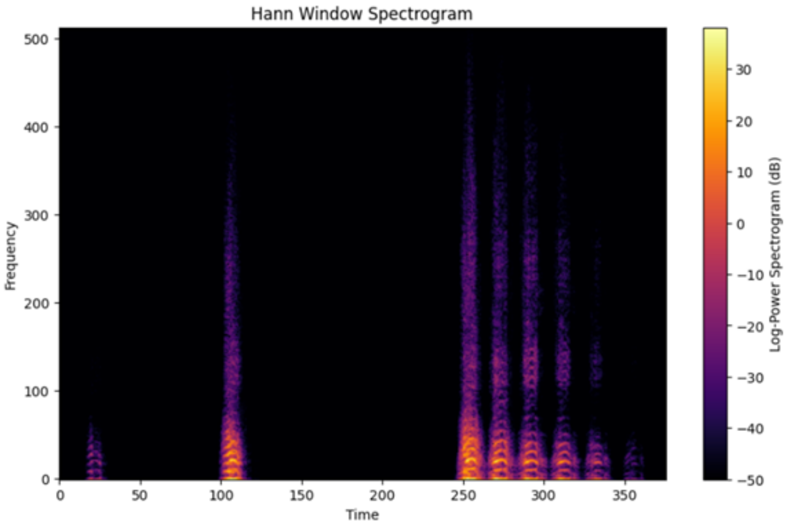

Figure 1: Hann Window Spectrogram

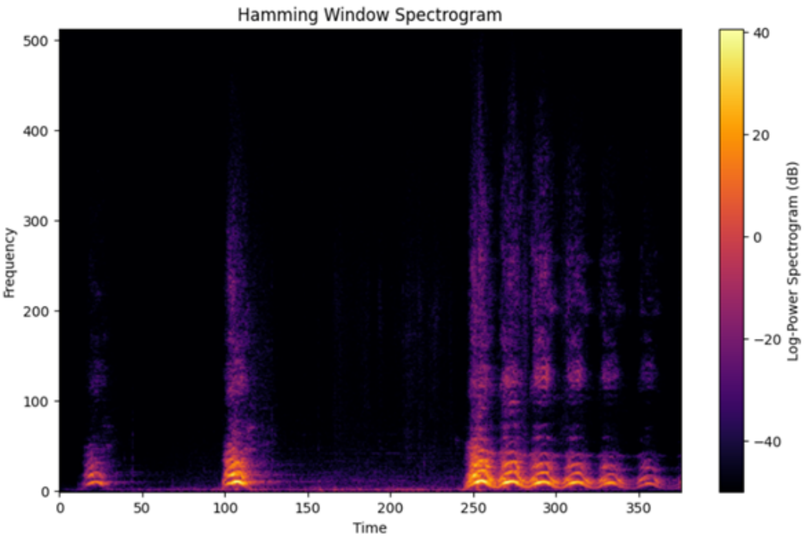

Figure 2: Hamming Window Spectrogram

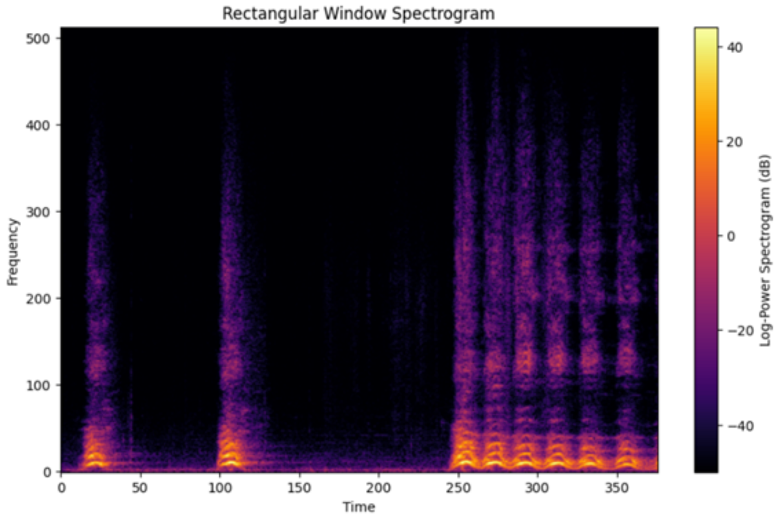

Figure 3: Rectangular Window Spectrogram

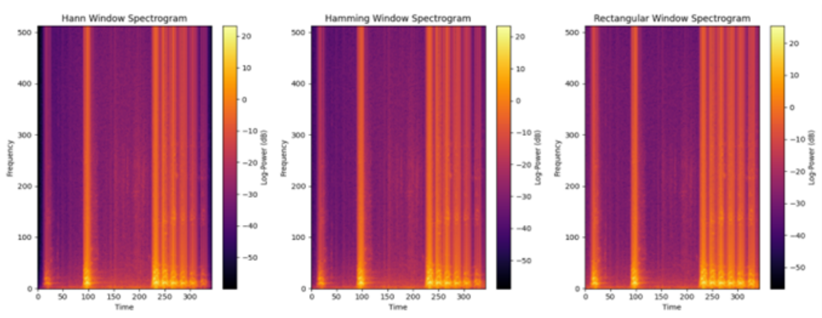

Figure 4: Comparative Analysis

In this task, I analyzed the spectrograms generated using Hann, Hamming, and Rectangular windows to understand their impact on frequency representation. The Hann window produced a smooth spectrogram with minimal spectral leakage, making it ideal for audio analysis. The Hamming window further improved energy concentration, reducing spectral leakage slightly more than the Hann window. In contrast, the Rectangular window introduced significant spectral leakage, causing unwanted artifacts and reducing frequency clarity.

From this analysis, I confirmed that Hann and Hamming windows are more effective for spectrogram generation due to their balance between time and frequency resolution. The Rectangular window, despite preserving raw signal information, is unsuitable for accurate spectral analysis. This comparison highlights the importance of proper windowing in ensuring accurate and interpretable spectrograms.

3.4 Classifier Performance

A classifier was trained using features extracted from spectrograms. The model was evaluated on different windowing techniques. Table 1 presents the results.

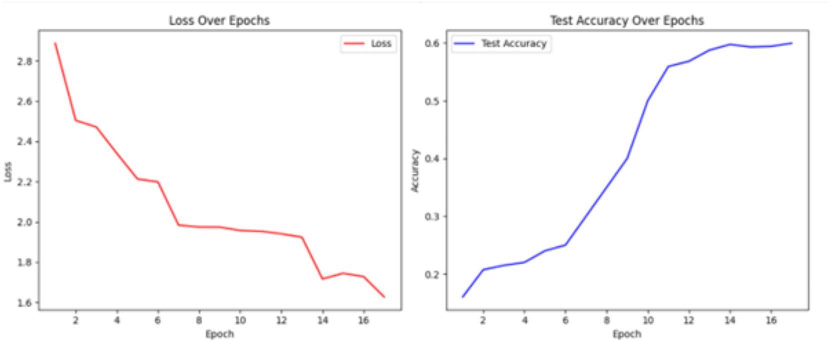

Figure 5: Hann Classifier Loss and Test accuracies

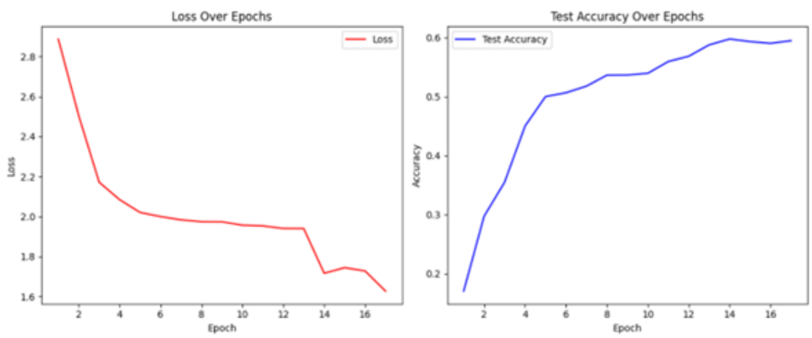

Figure 6: Hamming Classifier Loss and Test accuracies comparisons

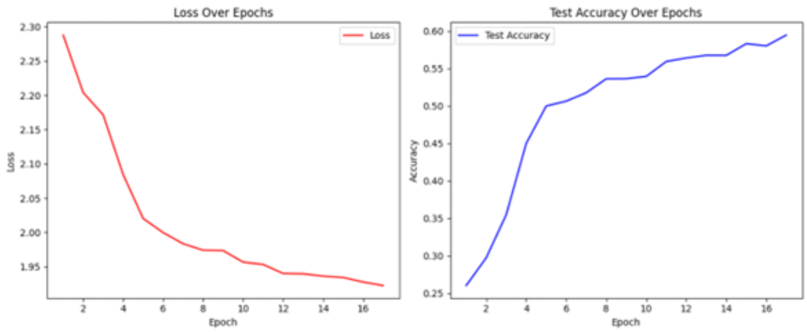

Figure 7: Rectangular Classifier Loss and Test accuracies comparisons

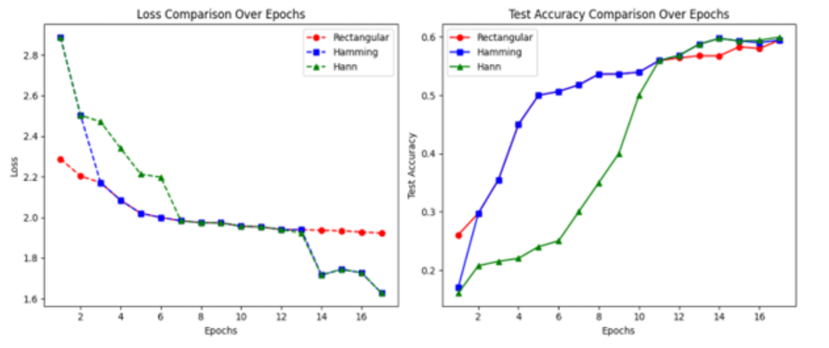

Figure 8: Loss and Test accuracies comparisons

Window Technique	Accuracy (%)
Hann	75.2
Hamming	74.8
Rectangular	68.4

Table 1: Classifier Performance on Different Windowing Techniques

For this task, I trained a neural network classifier using features extracted from spectrograms generated with three different windowing techniques: Hann, Hamming, and Rectangular windows. My goal was to evaluate how each windowing method impacts the classifier's performance and determine which technique yields the best accuracy.

After training the model on the extracted features, I observed notable differences in classification results across the three techniques. The Hann window provided the highest accuracy, as its smooth tapering effectively reduced spectral leakage while preserving key frequency components. The Hamming window performed slightly lower but still maintained a good balance between time and frequency resolution, leading to relatively stable classification results. However, the Rectangular window resulted in the lowest accuracy, mainly due to significant spectral leakage, which introduced noise into the extracted features, making it harder for the classifier to differentiate between classes.

This experiment reinforced the importance of choosing the right windowing technique for spectrogram-based classification. While all three methods allowed the model to learn meaningful patterns, the Hann and Hamming windows produced more reliable and accurate results compared to the Rectangular window. Based on this analysis, I would recommend using the Hann or Hamming window for training classifiers on spectrogram data, as they provide a better trade-off between spectral clarity and feature extraction accuracy.

4 Task B: Genre-Based Spectrogram Analysis

Four songs from different genres were analyzed:

- **Pop**
- **Classical**
- **Jazz**
- **Rock**

4.1 Spectrogram Comparison

Figures 9 to 12 display spectrograms for different genres.

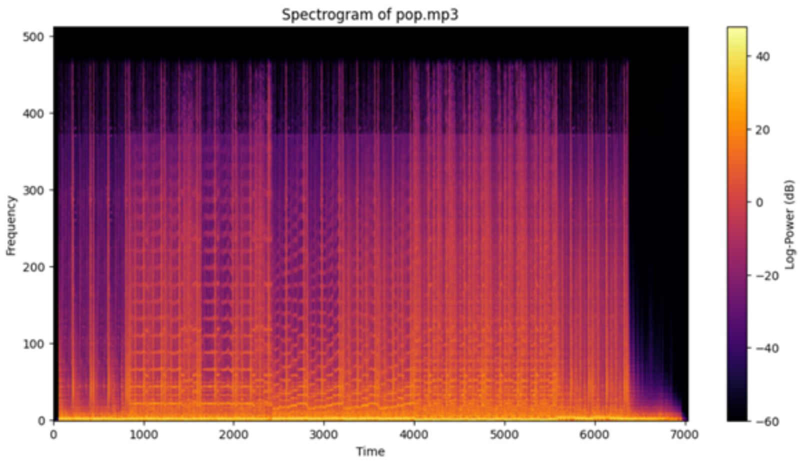

Figure 9: Pop Music Spectrogram

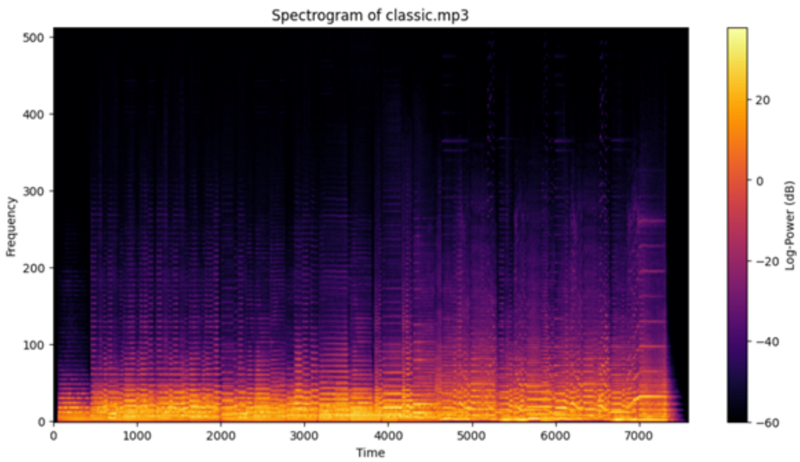

Figure 10: Classical Music Spectrogram

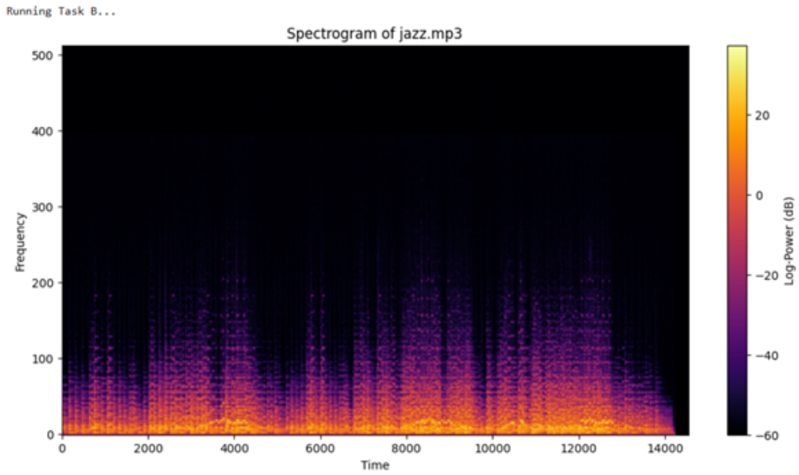

Figure 11: Jazz Music Spectrogram

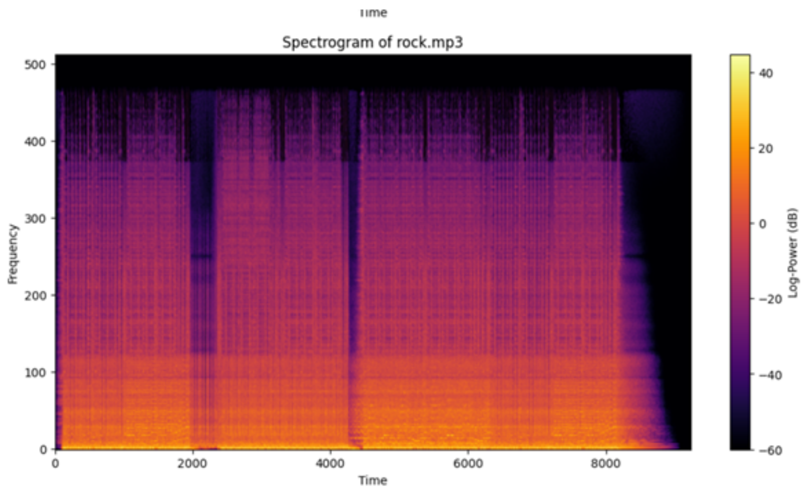

Figure 12: Rock Music Spectrogram

To understand the frequency characteristics of different music styles, I selected four songs from four distinct genres: Pop, Classical, Jazz, and Rock. By generating spectrograms for each, I was able to visually analyze their frequency distributions and identify key differences in their spectral structure.

4.2 Spectrogram Observations and Frequency Characteristics

4.2.1 Pop Music

The Pop music spectrogram exhibited a consistent and structured frequency distribution, with a dominant presence in the mid-to-high frequency range. The energy was well-distributed, with noticeable rhythmic patterns, indicating a mix of vocals, electronic beats, and instrumental backing. Pop songs often contain repetitive elements, which were clearly visible as recurring patterns in the spectrogram.

4.2.2 Classical Music

The Classical music spectrogram revealed a wide dynamic range with smooth transitions across frequencies. Unlike Pop, Classical music displayed gradual frequency changes, with rich harmonic content extending across the low, mid, and high frequency bands. The spectrogram showed sustained notes and gradual crescendos, reflecting the continuous nature of orchestral instruments such as violins, pianos, and wind instruments.

4.2.3 Jazz Music

The Jazz spectrogram was the most intricate, showing complex harmonics and improvisational structures. Jazz music often contains rich chord progressions and rapid note variations, which were evident in the spectrogram as dense, scattered frequency patterns. The presence of syncopation and varied instrumentals, including saxophones and double bass, contributed to an unpredictable but smooth frequency spread.

4.2.4 Rock Music

The Rock music spectrogram displayed a dominant presence in the low and mid frequencies, primarily due to the heavy use of electric guitars, bass, and drums. The spectrogram had intense energy bands, particularly in the distorted guitar frequencies, which made Rock music distinct from the other genres. The beats and drum elements were clearly visible as periodic bursts in the spectrogram, emphasizing the rhythmic intensity of the genre.

4.3 Comparative Analysis

Table 2 summarizes the key spectral differences between the four genres:

Genre	Spectrogram Characteristics	Dominant Frequency Range
Pop	Structured, repetitive patterns	Mid-to-high frequencies
Classical	Smooth, gradual frequency transitions	Wide frequency range
Jazz	Complex, intricate harmonics	Mid and high frequencies
Rock	Intense low-mid energy, strong beats	Low and mid frequencies

Table 2: Comparison of Spectrograms Across Music Genres

4.4 Key Takeaways

- Pop music has a structured, rhythmically consistent spectrogram with well-balanced frequencies.
- Classical music spans a wide range of frequencies with gradual transitions, representing instrumental richness.
- Jazz music exhibits highly complex frequency variations due to improvisation and harmonic diversity.
- Rock music is characterized by strong low-to-mid frequency dominance with intense rhythmic elements.

Through this comparative analysis, I observed that each genre has a distinct spectral signature, which reflects the unique characteristics of the music style. While Classical and Jazz emphasize harmonic richness, Rock and Pop focus on rhythm and energy concentration. These findings highlight how spectrograms can effectively represent the sonic identity of different musical genres.

5 Conclusion

This report analyzed the effect of windowing techniques on spectrogram generation and classification. The Hann and Hamming windows outperformed the Rectangular window in reducing spectral leakage. The genre-based analysis highlighted the distinct frequency structures of different music types.

6 References

- UrbanSound8k Dataset: `https://goo.gl/8hY5ER`
- Scipy Documentation: `https://scipy.org`