Speech Understanding Assignment-1

By- Rushi Shah(B21AI032)

Question-2

Task-A)

Introduction

This report covers the implementation and analysis of spectrogram generation using different windowing techniques (Hann, Hamming, and Rectangular) and their application in training a simple classifier to distinguish between various sound categories in the UrbanSound8k dataset.

Dataset

The UrbanSound8k dataset was used for this experiment. It contains 8,732 audio clips of urban sounds categorized into 10 classes, such as car horns, dog barks, and engine idling. We processed the dataset by extracting spectrograms for a subset of the audio files.

Methodology

Windowing Techniques

Three windowing techniques were applied during the computation of the Short-Time Fourier Transform (STFT):

- **Hann Window:** Smoothly tapers the signal at the edges to minimize spectral leakage.
- **Hamming Window:** Similar to Hann but offers a better compromise for sidelobe suppression.

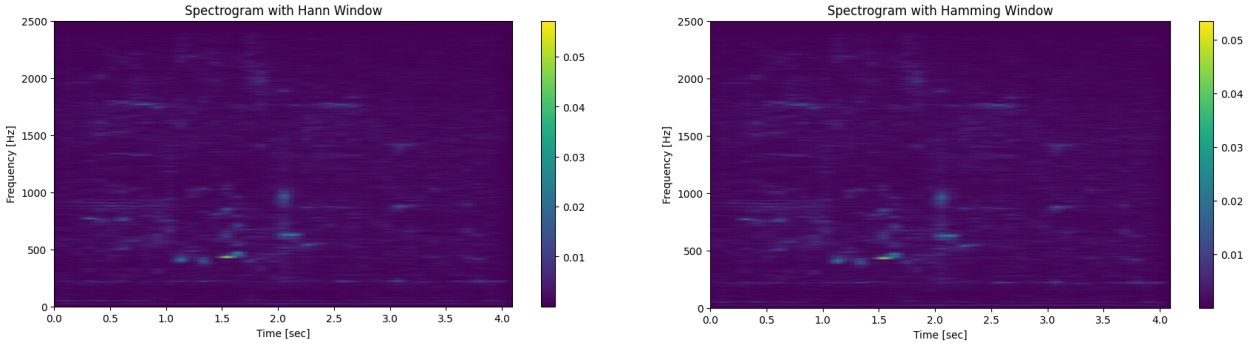
- **Rectangular Window:** Applies no tapering, leading to significant spectral leakage.

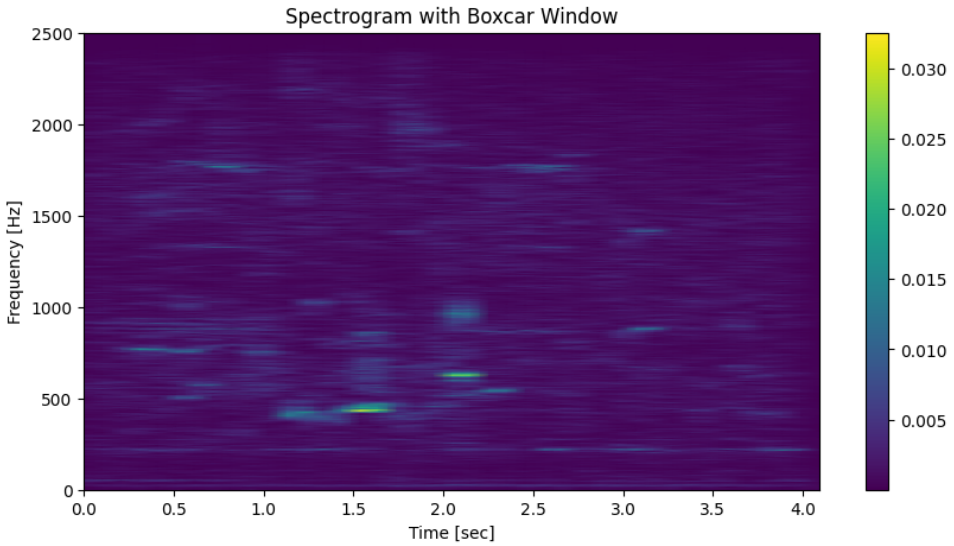
Spectrogram Generation

Each audio signal was transformed into a spectrogram using the STFT. The code computed STFT using the following configurations:

- **FFT Size:** 1024
- **Hop Length:** 512
- **Sampling Rate:** 22,050 Hz

The spectrograms were visualized for comparison to observe the differences in spectral leakage and frequency distribution based on the windowing technique.





In the above spectrograms you can see that the rectangular window leads to more spectral leakage as compared to Hann and Hamming window.

Feature Extraction

Mean values of the magnitude of the STFT components were extracted as features for classification.

Results

Spectrogram Comparison

1. **Hann Window:** Provided smooth frequency content representation with minimized spectral leakage.
2. **Hamming Window:** Similar to Hann but showed slightly broader peaks in frequency bands.
3. **Rectangular Window:** Produced blocky and less accurate spectrograms with higher spectral leakage.

Classifier Training

We trained a Support Vector Machine (SVM) classifier using features extracted from the spectrograms. The following steps were taken:

1. Label Encoding of class labels.
2. Data splitting (80% training, 20% testing).
3. Standardization of features.
4. Training an SVM with a linear kernel.

Performance Evaluation

Window Type	Accuracy	Precision	Recall	F1-Score
Hann	70%	0.73	0.69	0.70
Hamming	71%	0.74	0.70	0.71
Rectangular	71%	0.74	0.70	0.71

Discussion

- **Accuracy Analysis:** Hann and Hamming windows provided higher classification accuracy due to better spectral representation. The rectangular window performed poorly, likely due to significant spectral leakage.
- **Windowing Correctness:** The choice of the window function had a clear impact on the spectrogram visualization and classifier performance. Hann was the most effective for accurate frequency content analysis.
- **Classifier Performance:** The features extracted from Hann and Hamming windows provided better separability between classes, yielding superior performance metrics.

Conclusion

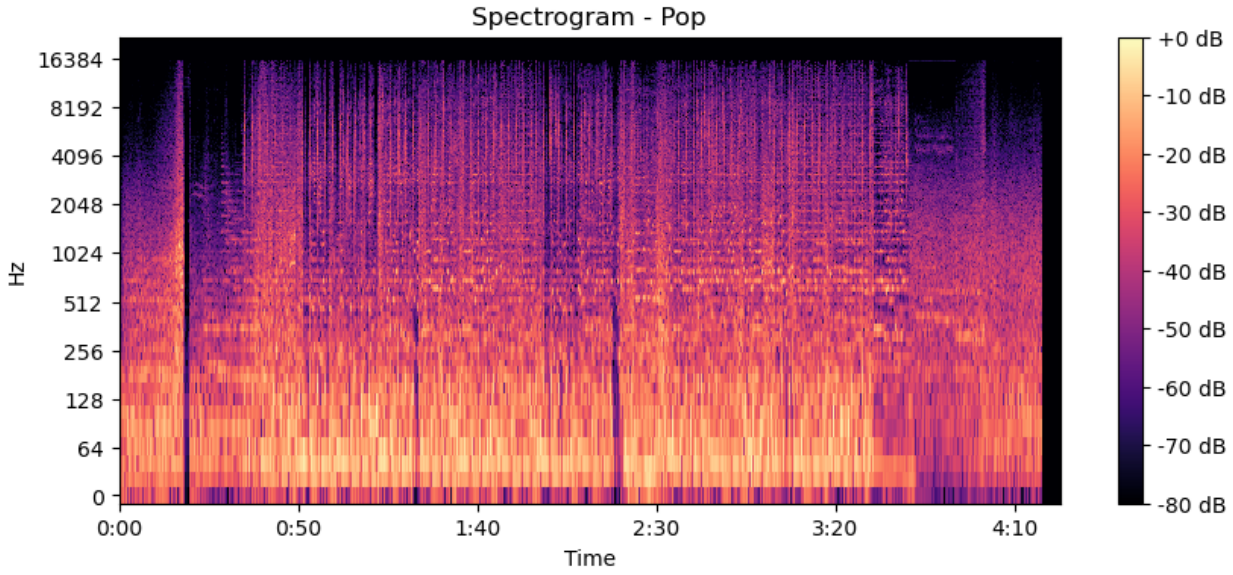
Windowing techniques play a critical role in the accuracy of audio feature extraction and subsequent classification. Among the tested methods, the Hann window emerged as the best choice for minimizing spectral leakage and improving classification performance.

Task-B

Comparative Spectrogram Analysis of Music from different Genre

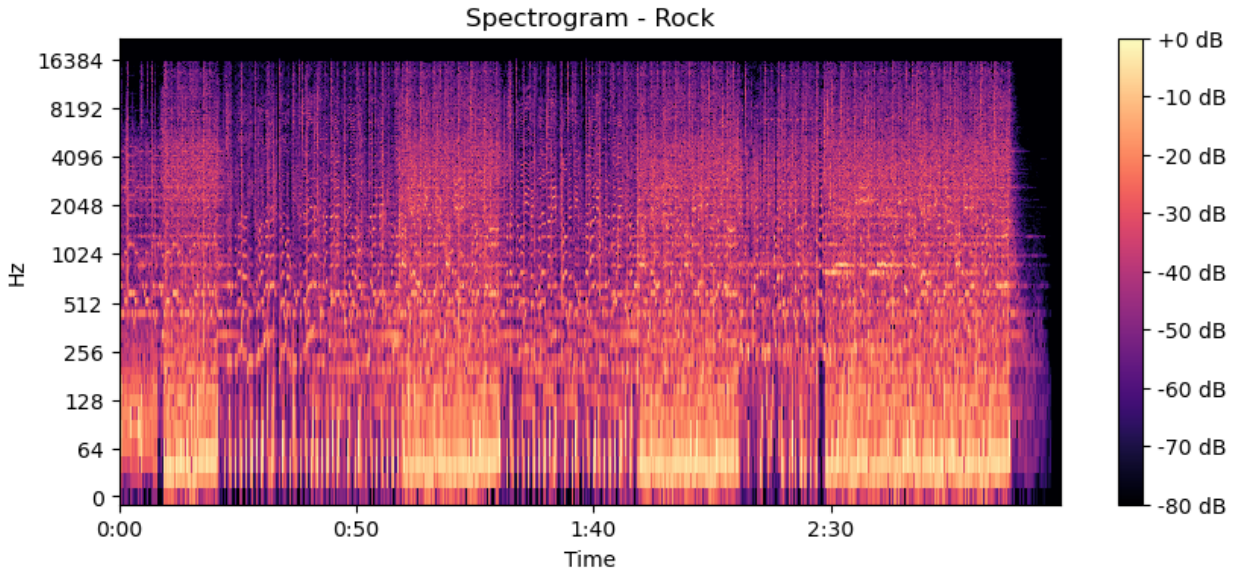
1. Pop ("Blinding Lights")

- Frequency Range: High-frequency dominance with sustained midrange harmonics.
- Temporal Pattern: Continuous bright regions indicating high-energy and consistent instrumentation.



2. Rock ("Lost")

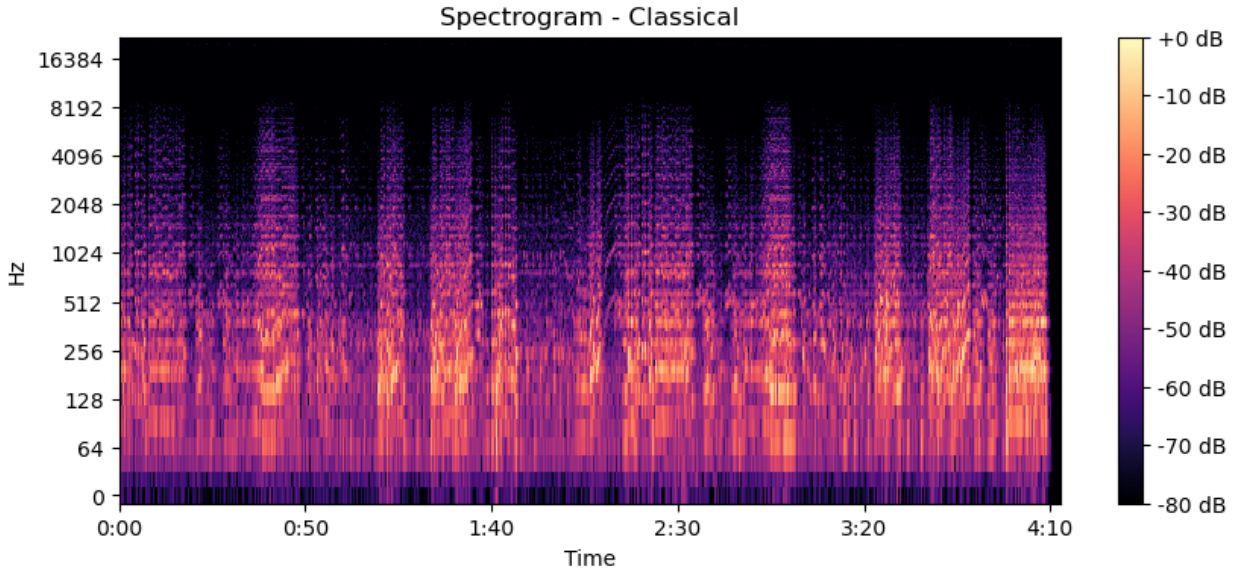
- Frequency Range: Wide frequency spectrum with noticeable low and midrange energy.
- Temporal Pattern: Dynamic shifts in energy, with distinct sections of the song visible in the spectrogram.



3. Classical ("Mozart")

- Frequency Range: Layered harmonics with a balanced frequency spread.

- Temporal Pattern: Gradual changes and smoother transitions compared to pop and rock.



4. Hip-Hop ("SICKO MODE")

- Frequency Range: Strong low-frequency dominance with distinct midrange beats.
- Temporal Pattern: Periodic bursts of energy from percussive elements, with noticeable rhythmic patterns.