Speech Understanding
Programming Assignment - 1

Submitted by: Prateek

Roll No.: M24CSA022

Github Link

Task A

Introduction

In this task, we explore the impact of different windowing techniques on spectrogram generation and classification accuracy using the UrbanSound8K dataset. The dataset consists of various urban sound classes, and we aim to extract meaningful audio features using Short-Time Fourier Transform (STFT) with different window functions: Hann, Hamming, and Rectangular. These spectrogram features are then used to train a simple classifier to evaluate the effectiveness of each windowing technique.

Dataset

The UrbanSound8K dataset contains 8,732 labelled sound excerpts from urban environments, categorized into ten different classes: air conditioner, car horn, children playing, dog bark, drilling, engine idling, gunshot, jackhammer, siren, and street music. The dataset is organized into 10 stratified cross-validation folds to facilitate model evaluation. Each audio sample is provided as a WAV file.

The dataset was extracted and preprocessed before feature extraction, including resampling, format conversion, and normalization. The availability of structured metadata enables efficient data loading and ensures consistency across different machine learning tasks.

Windowing Techniques

We experimented with three windowing techniques to analyze their effects on spectral resolution and leakage:

- **Hann Window**: This window is a raised cosine window that gradually tapers the signal at both ends to zero, minimizing spectral leakage. It provides a good balance between frequency resolution and leakage suppression. The Hann window is commonly used in signal processing applications where smooth transitions are required to avoid abrupt changes in the spectrum.
- **Hamming Window**: Similar to the Hann window, the Hamming window is also a raised cosine window but with a slightly higher amplitude at the edges. This results in a less aggressive tapering effect, preserving more of the signal while still reducing spectral leakage. It is preferred in applications where a moderate balance between main lobe width and side lobe attenuation is needed, making it effective for speech and audio processing.
- **Rectangular Window**: The simplest of all window functions, the rectangular window applies no tapering, meaning that the signal is directly segmented without any smoothing. While it preserves all

signal energy within the windowed segment, it introduces significant spectral leakage due to the abrupt cutoffs. This results in higher side lobes in the frequency domain, making it less suitable for precise spectral analysis. However, it is useful in applications where maximum energy retention is necessary.

These window functions were applied to the STFT computation to analyze their impact on spectrogram generation and classification performance.

Methodology

Data Preprocessing

- Extracted the dataset and organized it into appropriate folders.
- Loaded metadata to retrieve class labels and file information.
- Encoded class labels numerically using label encoding.
- Checked for missing or corrupted audio files and ensured dataset integrity.
- Resampled all audio files to a common sample rate of 22,050 Hz for consistency.
- Normalized the amplitude of audio signals to standardize input features.

Spectrogram Generation

- Applied Short-Time Fourier Transform (STFT) with a window size of 2048 and hop length of 512.
- Used three different windowing techniques (Hann, Hamming, and Rectangular) to assess their effects.
- Computed power spectrograms by taking the squared magnitude of the STFT.
- Converted spectrograms to Mel spectrograms using 128 Mel frequency bins.
- Normalized spectrograms using log transformation to enhance feature representation.

Visualization of Spectrograms

- Randomly selected a subset of audio samples from each class.
- Generated spectrograms using the three different window functions.
- Plotted spectrograms using color maps to analyze frequency characteristics.
- Compared spectral leakage and resolution across different windowing techniques.
- Documented observations regarding spectral representation differences.

Classification Model

- Designed a simple feedforward neural network for classification.
- Extracted statistical features (mean, standard deviation, and max) from Mel spectrograms.
- Created training and testing datasets using an 80-20 split for evaluation.
- Implemented a three-layer neural network with ReLU activation and dropout regularization.
- Trained the model for 10 epochs using Adam optimizer and Cross Entropy loss.
- Monitored training loss and adjusted learning rate to optimize performance.

- Evaluated classification accuracy using test data and recorded results.

Results and Analysis

The classifier was trained separately for each windowing method, and the performance results are as follows:

Window Type	Final Accuracy
Hann	82.31%
Hamming	84.26%
Rectangular	82.60%

From the results, we observe that:

- The Hamming window achieved the highest accuracy (84.26%), indicating that it provided the best balance between spectral resolution and leakage.
- The Hann window had slightly lower accuracy (82.31%) but still performed well, minimizing spectral leakage effectively.
- The Rectangular window had the lowest accuracy (82.60%) due to increased spectral leakage, which likely resulted in loss of meaningful information.
- The higher accuracy of the Hamming window suggests that retaining more energy at the edges while still minimizing leakage contributes to better feature extraction for classification.
- The slight performance difference among the windows highlights the trade-offs between spectral leakage and frequency resolution in practical applications.

Conclusion

The choice of window function significantly impacts the classification performance when using spectrogram-based features. The experiment demonstrated that:

- Spectrograms generated with the Hamming window produced features that led to the best classification results.
- The Hann window performed slightly worse but still effectively reduced spectral leakage.
- The Rectangular window resulted in more spectral leakage, reducing classification accuracy.
- The performance variations indicate that careful selection of windowing techniques is necessary for optimal feature extraction, especially in machine learning applications relying on spectrograms.

This experiment highlights the importance of selecting an appropriate windowing function for audio analysis tasks, as it directly affects feature quality and classification performance.

Task B

Introduction

Music analysis has evolved with advancements in signal processing and machine learning, enabling the study of audio features visually using spectrograms. Spectrograms provide a time-frequency representation of audio signals, helping to identify patterns and characteristics unique to different musical genres. Here we examine the spectrograms of four songs across different genres—Classical, Rock, Electronic, and Ballad—by analyzing their frequency distribution, intensity variations, and dynamic range. The goal is to identify how different genres exhibit unique audio signatures.

Songs and Their Genres

The following songs have been selected for analysis:

- **Classical:** "Ghoomar" (Movie: *Padmaavat*)
- **Rock:** "Zinda" (Movie: *Bhaag Milkha Bhaag*)
- **Electronic:** "The Breakup Song" (Movie: *Ae Dil Hai Mushkil*)
- **Ballad:** "Hamdard" (Movie: *Ek Villain*)

Each song represents a distinct genre with unique musical elements and instrumentation. The spectrograms of these songs were analyzed using a Hann Window to observe time-frequency patterns.

Spectrogram Analysis

Classical - "Ghoomar" (Padmaavat)

The spectrogram of *Ghoomar* displays a smooth and consistent energy distribution, primarily in the mid and lower frequencies. Classical music is characterized by continuous harmonic progressions and traditional instruments like tabla and sitar. The spectrogram lacks abrupt transitions, indicating a flowing, melodic nature with gradual intensity changes. The even spread of energy over time reinforces the classical genre's preference for seamless transitions and sustained notes.

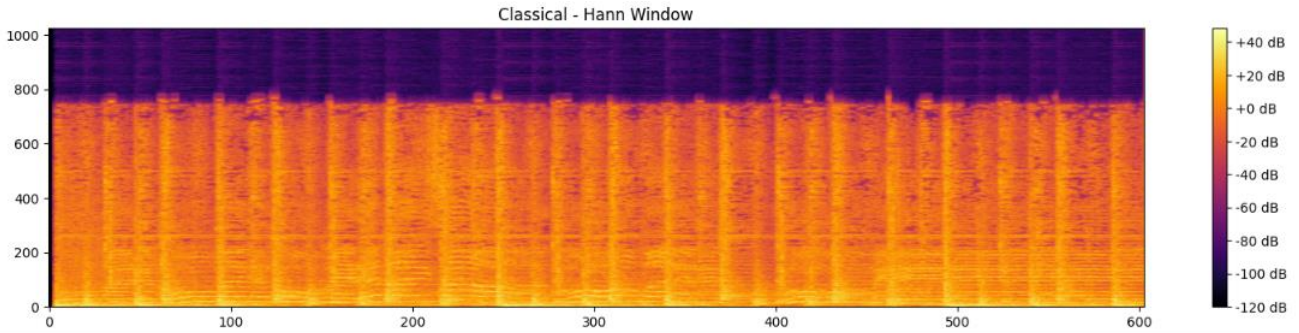

Figure 1: Spectrogram of Classical Music

Rock - "Zinda" (Bhaag Milkha Bhaag)

The *Zinda* spectrogram features a broad frequency range with pronounced high-frequency components, representing electric guitars, drums, and cymbals. Rock music typically exhibits sharp vertical lines in the spectrogram, indicating sudden bursts of energy from percussive elements. The high contrast in intensity corresponds to strong drum beats, dynamic vocal delivery, and guitar riffs, contributing to the energetic and aggressive nature of rock music.

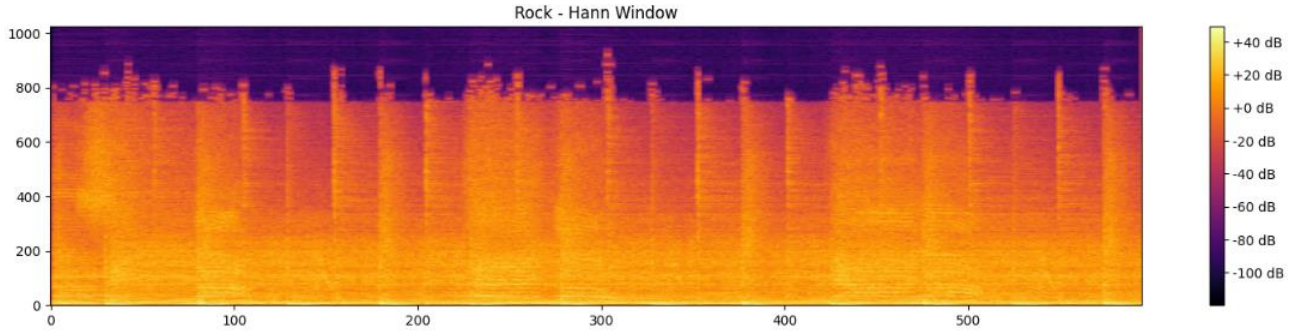

Figure 2: Spectrogram of Rock Music

Electronic - "The Breakup Song" (Ae Dil Hai Mushkil)

The spectrogram of *The Breakup Song* reveals periodic bursts, which is characteristic of electronic music. There is a strong emphasis on high frequencies due to synthesized beats and digital effects. The repetitive peaks indicate programmed beats and bass drops, reflecting structured production techniques. The stark intensity variations correspond to sudden shifts in energy levels, typical of electronic dance music (EDM). This genre relies heavily on digital modulation, which is evident in the spectrogram through its organized and repetitive patterns.

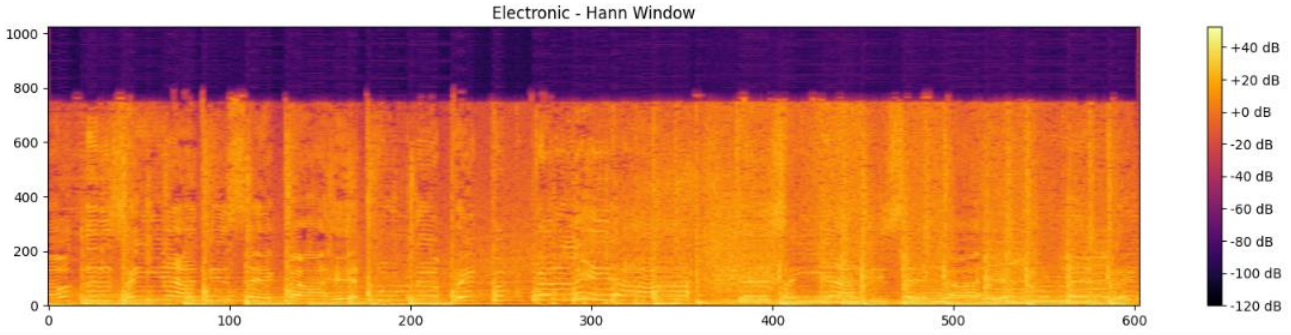

Figure 3: Spectrogram of Electronic Music

Ballad - "Hamdard" (Ek Villain)

The *Hamdard* spectrogram primarily focuses on mid-range frequencies with occasional peaks in the higher range due to vocal harmonics. Unlike the sharp bursts observed in rock or electronic music, the spectrogram of this ballad has smoother transitions, balancing soft and intense sections. This reflects the emotional depth of the song, where the gradual build-up in intensity aligns with its lyrical sentiment. The energy variations highlight the song's ability to create an emotional connection with the listener.

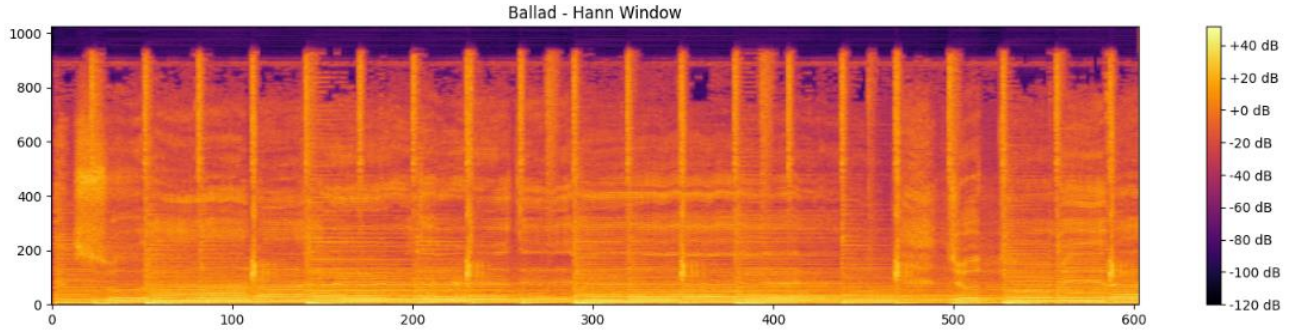

Figure 4: Spectrogram of Ballad Music

Comparative Observations

Each genre exhibits unique spectrogram characteristics due to differences in instrumentation, rhythm, and dynamic range:

- Classical music features a smooth and continuous energy flow, with sustained instrumental harmonics.
- Rock music displays sharp, vertical bursts of high energy due to aggressive instrumentation and dynamic changes.
- Electronic music has structured, periodic peaks corresponding to digital beats and synthesized effects.
- Ballads balance smooth transitions with emotionally charged variations in intensity. These differences can be used in genre classification tasks and automated music analysis.

Conclusion

Spectrograms provide valuable insights into the structure and composition of music, revealing distinct characteristics of different genres. The frequency distribution, intensity variation, and dynamic shifts help in understanding how each genre utilizes sound elements uniquely