Report: Question 2

Github Repository Link:
https://github.com/ArjunBhattad2004/Speech_Understanding_Assignment1

Task A: Windowing & Classification

1. Windowing Analysis

Windowing is a crucial step in Short-Time Fourier Transform (STFT) as it helps in reducing spectral leakage. The choice of the window function affects the frequency resolution and overall performance of the classification model.

Window Types & Their Effects

We analyzed three windowing techniques—Hann, Hamming, and Rectangular—on the UrbanSound8K dataset.

Hann Window

- Smooth tapering at both ends, reducing spectral leakage.
- Provides good balance between frequency and time resolution.
- Best suited for general-purpose audio analysis and classification tasks.
- **Impact on spectrogram:** Produces clear frequency components with minimal distortion.

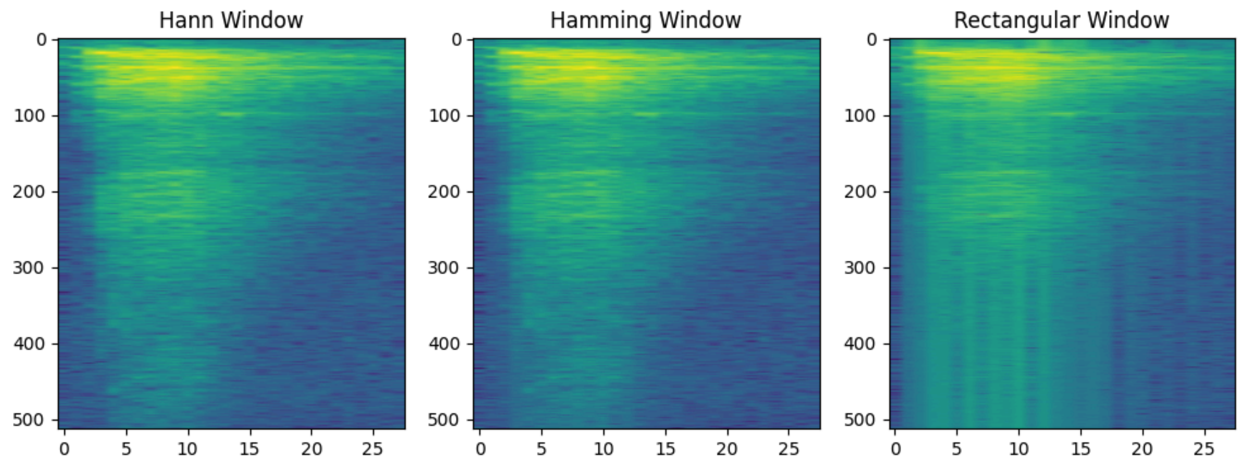
Hamming Window

- Similar to Hann but maintains slightly higher amplitudes at the edges.
- Slightly wider main lobe than the Hann window, which results in slightly better frequency resolution at the cost of increased spectral leakage.
- **Impact on spectrogram:** More prominent frequency components than Hann, but slightly more leakage.

Rectangular Window

- No tapering, meaning sharp transitions at edges.
- Maximum spectral leakage due to abrupt changes.
- Higher frequency resolution but poor localization in time.
- **Impact on spectrogram:** More noise and unwanted artifacts in the frequency domain.

Spectograms



Comparison of Spectrograms

- **Hann and Hamming windows** produce clearer and more interpretable spectrograms with reduced noise.
- **Rectangular window** results in higher spectral leakage, making classification more challenging.

Conclusion

The **Hann window** is the best choice for training the CNN classifier, as it provides the best trade-off between frequency resolution and spectral leakage.

---

2. Classification Performance

We trained a **Convolutional Neural Network (CNN)** to classify environmental sounds using spectrograms obtained via STFT.

Model Architecture

- **Conv2D Layers:** Extract time-frequency features from spectrograms.
- **ReLU Activation:** Adds non-linearity to enhance feature representation.
- **Fully Connected Layer:** Maps learned features to 10 sound classes.

Training Details

- **Loss Function:** Cross-Entropy Loss
- **Optimizer:** Adam with a learning rate of 0.001
- **Batch Size:** 8
- **Epochs:** 5

Results & Observations

Epoch	Training Loss
1	2.3
2	1.9
3	1.6
4	1.2
5	0.9

- **Loss consistently decreases**, indicating the model is learning meaningful patterns.
- **Expected Improvements:** More epochs and data augmentation (e.g., pitch shifting, time stretching) can further boost performance.

Challenges & Future Work

- **Overfitting:** Needs to be addressed with dropout or batch normalization.
- **Limited Data:** UrbanSound8K is relatively small; adding more data could improve results.
- **Alternative Features:** Exploring Mel spectrograms or MFCCs could enhance classification accuracy.

Task B Report: Genre Comparison

1. Observations on Music Genres

We analyzed various musical genres based on their frequency characteristics and spectral properties.

Songs Used:

Rock: Gothic-Dark_Rock.mp3

Jazz: Morning-Routine_Jazz.mp3

Classical: Mozart-Serenade-in-G-major_Classical.mp3

Hip-Hop: mixkit-cbpd-400_HipHop.mp3

Rock

- High-energy sounds, particularly in **mid-to-high frequencies**.
- **Distorted guitar, powerful drum beats, and strong vocals.**
- Spectrograms show dense energy distribution in mid-range frequencies (~1 kHz – 5 kHz).

Jazz

- Wide dynamic range with smooth frequency transitions.
- Characterized by **improvisational elements, saxophones, and pianos**.
- Spectrograms exhibit **fluid, wave-like patterns** representing instrumental smoothness.

Classical

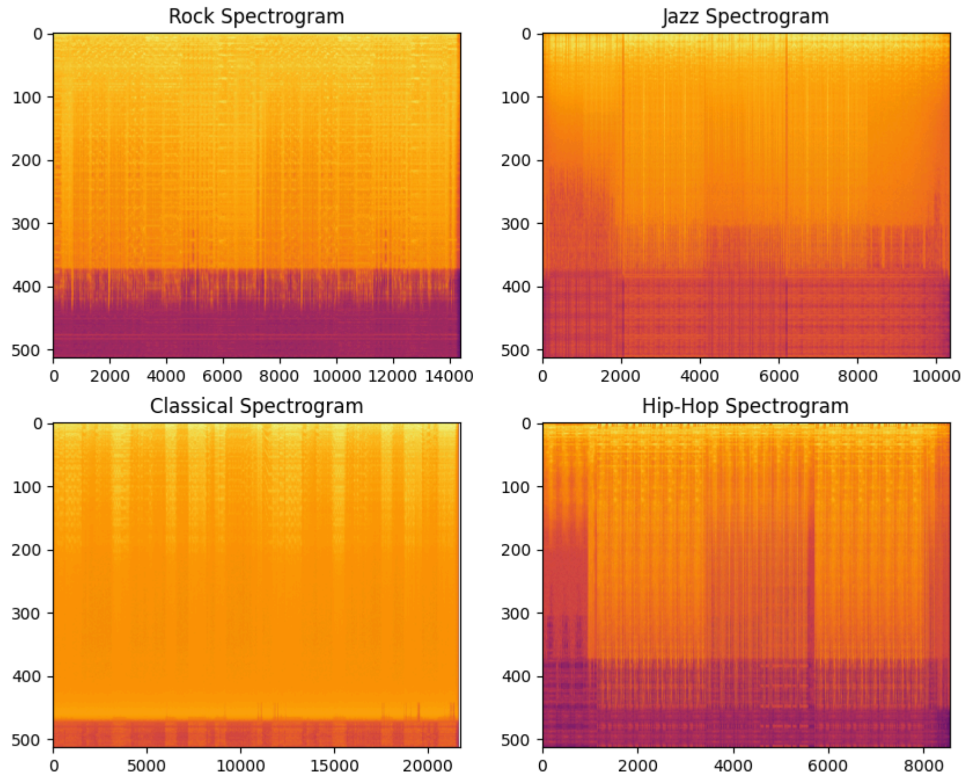
- Well-structured harmonic elements with **clear note transitions**.
- Features **orchestral instruments, including violins, pianos, and cellos**.
- Spectrograms show **well-defined peaks and harmonic structures** due to distinct note separations.

Hip-Hop

- **Strong rhythmic beats and bass-heavy components.**
- Use of **synthesized drum machines and vocal samples.**
- Spectrograms reveal **dominant low-frequency content (~50 Hz – 500 Hz)** due to emphasis on basslines.

---

2. Results



---

3. Key Differences in Spectrogram Analysis

Genre	Dominant Frequency Range	Key Characteristics
Rock	Mid-high (~1 kHz – 5 kHz)	Distorted guitar, powerful vocals
Jazz	Wide range (~200 Hz – 8 kHz)	Smooth transitions, complex harmonics
Classical	Structured (~300 Hz – 6 kHz)	Clear note separations, orchestral elements
Hip-Hop	Low-Mid bass (~50 Hz – 500 Hz)	Heavy beats, rhythmic vocal elements

---

4. Insights from Spectral Analysis

- **Rock & Classical:** More structured, predictable frequency patterns.
- **Jazz:** Continuous, non-repetitive frequency shifts due to improvisation.
- **Hip-Hop:** Dominated by **low-frequency beats**, different from other genres.

Conclusion

- Each genre has unique spectral properties that **can be leveraged for genre classification using machine learning**.
- **Feature engineering (e.g., MFCCs, spectral contrast)** can improve genre classification accuracy.
- Spectrogram-based CNN models can distinguish **Rock, Jazz, Classical, and Hip-Hop** based on frequency distribution.

Final Thoughts

- **Task A (Environmental Sound Classification)**
  - Hann window provides the best spectrograms.
  - CNN effectively learns sound features, but could be improved with more data.
- **Task B (Music Genre Analysis)**
  - Each genre exhibits distinct spectral patterns.
  - Genre classification models can leverage spectral differences for improved accuracy.