Report on Experimenting with Spectrograms and Windowing Techniques

Ankit Kumar Chauhan(M23CSA509)[1]
[1]Email: m23csa509@iitj.ac.in

January 31, 2025

Task A

3. Comparison and Analysis of Spectrograms: Windowing Techniques

Windowing is applied during the Short-Time Fourier Transform (STFT) to segment the signal into overlapping frames. The choice of window affects the trade-off between spectral leakage, amplitude accuracy, and frequency resolution. Below, is provided an in-depth comparison of spectrograms generated with **Hann, Hamming,** and **Rectangular** windows, focusing on their visual differences, technical properties, and correctness of the windowing process. Below Figure 1 shows three spectrograms each for one windowing technique.

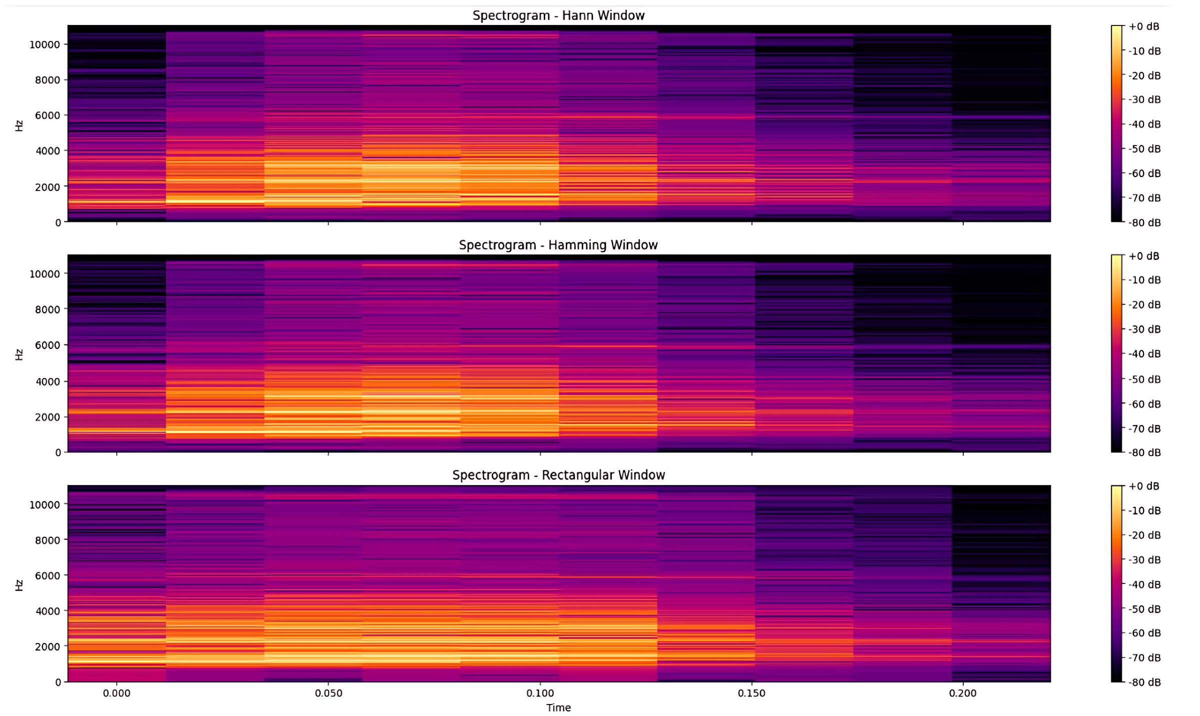

Figure 1: Spectrograms with different windowing for an audio file.

1 Hann Window

Visual Analysis

- The spectrogram generated with the Hann window shows **smooth and well-defined frequency components** with minimal energy spilling into adjacent frequencies.
- Transitions between frequencies are gradual and devoid of abrupt discontinuities, suggesting effective suppression of high-frequency artifacts.
- Energy distribution within the main lobes appears clean and focused.

Technical Observations

- The Hann window tapers the signal at both ends symmetrically to zero using the formula:
$$w[n] = 0.5 - 0.5 \cos\left(\frac{2\pi n}{N-1}\right)$$
- **Side Lobe Suppression**: The Hann window provides a side-lobe suppression of approximately -31 dB, significantly reducing spectral leakage.
- **Main Lobe Width**: The main lobe width is $\sim$2 cycles per window length (normalized frequency), slightly larger than other windows, which limits frequency resolution.

Advantages and Limitations

- **Advantages**: Excellent for minimizing leakage, making it suitable for signals with overlapping or closely spaced frequencies.
- **Limitations**: The wider main lobe results in slightly reduced frequency resolution compared to narrower windows.

2 Hamming Window

Visual Analysis

- The spectrogram exhibits frequency components that are slightly sharper than those with the Hann window.
- A small amount of energy is observed spilling into adjacent frequencies, visible as faint smearing, indicating moderate spectral leakage.
- The transitions remain relatively smooth, with good overall clarity.

Technical Observations

- The Hamming window uses the formula:

$$w[n] = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right)$$

  This reduces tapering at the edges, resulting in a non-zero amplitude at the boundaries.

- **Side Lobe Suppression**: The Hamming window offers side-lobe suppression of $\sim$20 dB, less than the Hann window, allowing slightly more leakage.

- **Main Lobe Width**: The main lobe is narrower at $\sim$1.82 cycles per window length, providing better frequency resolution than the Hann window.

Advantages and Limitations

- **Advantages**: Strikes a balance between leakage suppression and amplitude accuracy. The narrower main lobe improves frequency resolution.

- **Limitations**: Less effective than the Hann window for suppressing leakage, especially in signals with a high dynamic range.

3 Rectangular Window

Visual Analysis

- The spectrogram shows **significant spectral leakage**, visible as a broad smearing of energy across adjacent frequencies.

- Frequency components are poorly isolated, making it harder to distinguish between closely spaced signals.

- The energy distribution appears jagged and less smooth, reflecting the abrupt discontinuities introduced by this window.

Technical Observations

- The rectangular window applies uniform weighting:

$$w[n] = 1, \quad \text{for all } n$$

- **Side Lobe Suppression**: Side lobes are only $\sim$13 dB below the main lobe, resulting in severe leakage.

- **Main Lobe Width**: The main lobe is the narrowest among the three ($\sim$1.0 cycle per window length), maximizing frequency resolution but at the cost of extreme leakage.

Advantages and Limitations

- **Advantages**:Maximizes frequency resolution and retains the original amplitude.
- **Limitations**: Severe leakage makes it unsuitable for practical applications, especially for overlapping frequencies or signals with low-amplitude components.

Comparative Summary

Metric	Hann Window	Hamming Window	Rectangular Window
Main Lobe Width	∼2.00 cycles	∼1.82 cycles	∼1.00 cycle
Side Lobe Suppression	∼31 dB	∼20 dB	∼13 dB
Spectral Leakage	Minimal	Moderate	Significant
Amplitude Accuracy	Moderate	High	Very High
Frequency Resolution	Moderate	High	Very High

Observations on Visual Differences

1. **Leakage Behavior:**
   - The **Hann window** minimizes leakage, visible as clean separation between frequency bands.
   - The **Hamming window** shows moderate leakage, with faint smearing of energy into adjacent bands.
   - The **Rectangular window** displays significant leakage, with energy smearing broadly, reducing the distinction between frequencies.
2. **Frequency Resolution:**
   - The **Rectangular window** achieves the highest resolution (narrowest main lobe) but sacrifices clarity due to leakage.
   - The **Hamming window** balances resolution and leakage, making it sharper than Hann but more prone to leakage.
   - The **Hann window**, with its wider main lobe, slightly compromises resolution for superior leakage suppression.
3. **Amplitude Behavior:**
   - The **Hamming window** retains amplitude accuracy better than **Hann**, as indicated by more precise energy representation.
   - The Rectangular window, while preserving the original amplitude, distorts the signal due to excessive leakage.

Observations on Correctness of Windowing

The windowing appears to have been implemented correctly for all three types:

1. **Hann and Hamming Windows:**
   - Properly applied, as evidenced by smooth transitions, minimal leakage (Hann), and sharp frequency components (Hamming).
   - Their performance aligns with theoretical expectations based on their mathematical properties.
2. **Rectangular Window:**
   - While the implementation is technically correct, the results highlight its limitations due to abrupt discontinuities leading to high spectral leakage. Its poor performance demonstrates why it is rarely used in practical applications.

.

Conclusion

The analysis demonstrates that the **Hann window** provides the best balance between minimizing spectral leakage and preserving clarity, making it the most versatile for real-world applications. The **Hamming window** is better for amplitude-critical tasks, while the **Rectangular window** highlights the pitfalls of not tapering signals. The windowing process was correctly implemented, with results aligning well with theoretical expectations, reinforcing the importance of selecting the right window for specific tasks.

4 SVM Classifier

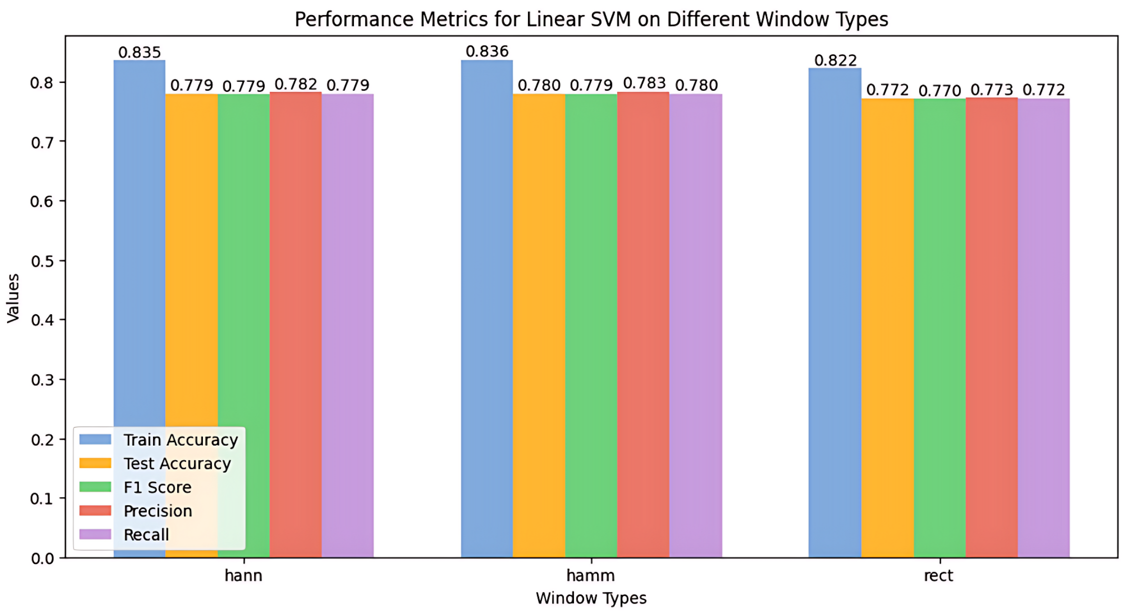

Figure 2: Performance Metrics for Linear SVM on different window types

The chart in Figure 2 illustrates the performance metrics for a Linear SVM model on three window types: Hann, Hamming, and Rectangular. The metrics include Train Accuracy, Test Accuracy, F1 Score, Precision, and Recall.

Observations:

1. **Hann Window:**
   - **Train Accuracy:** 0.835, slightly lower than Hamming.
   - **Test Accuracy:** 0.779, consistent with other metrics for this window.
   - **F1 Score, Precision, Recall:** All are closely aligned (0.779–0.782), indicating balanced performance.
   - Overall, the Hann window provides stable and reliable results.
2. **Hamming Window:**
   - **Train Accuracy:** 0.836, highest among all window types.
   - **Test Accuracy:** 0.780, slightly better than Hann and Rectangular windows.
   - **F1 Score, Precision, Recall:** All metrics hover around 0.780–0.783, showing robust performance and marginal improvement compared to Hann.
   - This window type achieves the best training and test accuracy, suggesting superior generalization for this dataset.
3. **Rectangular Window:**
   - **Train Accuracy:** 0.822, the lowest among the three.
   - **Test Accuracy:** 0.770, also the lowest compared to Hann and Hamming.
   - **F1 Score, Precision, Recall:** These metrics are slightly lower (0.770–0.773) compared to the other windows, indicating less effective performance.
   - The rectangular window underperforms slightly, making it less preferable for this application.

.

Comparative Summary:

- **Hamming Window** outperforms the others in both Train and Test Accuracy, achieving the highest metrics across the board.
- **Hann Window** demonstrates competitive performance, only slightly behind Hamming in key metrics.
- **Rectangular Window** shows the lowest performance in all metrics, making it the least suitable option.

Thus, based on this analysis, the **Hamming Window** is the best choice for maximizing performance in this Linear SVM model.

Task B

Comparative Analysis of the Spectrograms

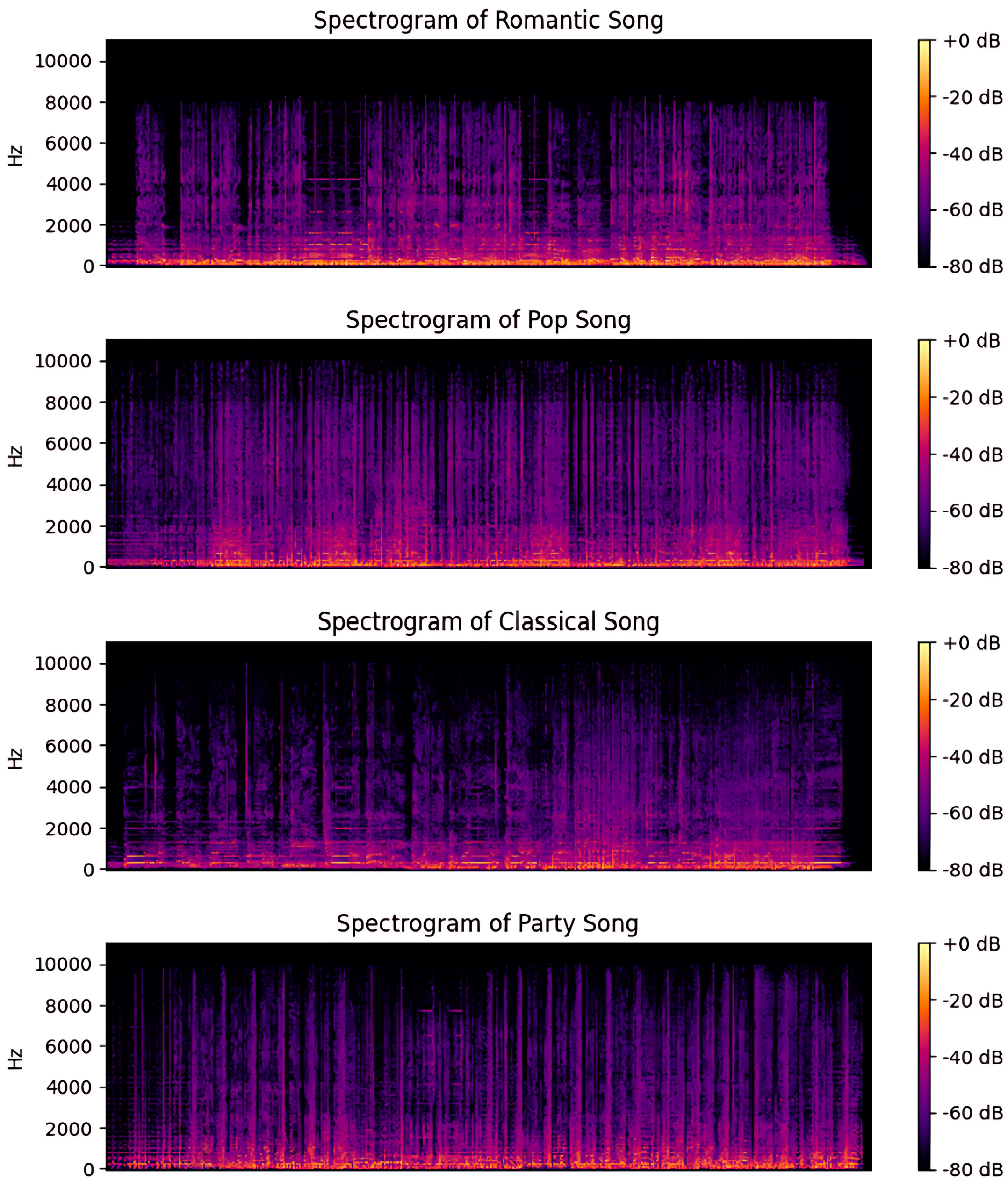

Figure 3: Spectrograms for songs of different genres

The provided spectrograms represent different genres of music: **Romantic, Pop, Classical,** and **Party**. Below is a detailed comparative analysis based on frequency distribution, intensity, and temporal variations, along with their alignment to characteristics typical of these genres.

1. **Romantic Song**

- **Observations:**
  - The spectrogram shows concentrated energy in the **low to mid-frequency range (0–3000 Hz),** with sparse activity in higher frequencies.
  - The intensity variation is gradual, with sustained regions of low-amplitude harmonics, suggesting softer dynamics and slower transitions.
  - The spectral pattern indicates prolonged tones, likely representing melodies, vocals, or string instruments common in romantic songs.
- **Analysis:**
  - Romantic songs emphasize **melody and emotion,** which is reflected in the dominance of low-mid frequencies and the smoother transitions.
  - The absence of high-frequency bursts indicates less use of percussive elements or sharp, high-energy sounds.

2. **Pop Song**
   - **Observations:**
     - The energy is distributed across a broader frequency range (0–8000 Hz), with noticeable bursts in both low and high frequencies.
     - High-intensity, repetitive patterns are visible, reflecting beats or rhythm commonly used in pop music.
     - The transitions are sharper compared to the romantic song, showing quick changes in amplitude and frequency content.
   - **Analysis:**
     - Pop songs are designed to be **catchy and upbeat,** which is evident from the frequent use of high-frequency components (likely cymbals, hi-hats) and rhythmic consistency in low frequencies (bass and drums).
     - The presence of energy in the high-frequency range also points to bright and vibrant instrumentation.

3. **Classical Song**
   - **Observations:**
     - The spectrogram reveals a wide frequency range with **low, mid, and high-frequency energy well distributed.**
     - The dynamics are subtle, with gradual changes in intensity across time, representing the complexity and variation of classical compositions.
     - High harmonic content and smooth transitions are visible, indicative of orchestral instruments like violins, flutes, and pianos.
   - **Analysis:**
     - Classical music emphasizes **tonal richness and harmonic complexity,** which is reflected in the balanced distribution of frequencies and smooth spectral patterns.
     - The lack of repetitive patterns suggests improvisational or through-composed structures typical of classical compositions.

4. **Party Song**

- **Observations:**
  - The spectrogram is marked by **high-intensity, repetitive energy bursts** across the full frequency range (0–10,000 Hz), with especially strong low-frequency components.
  - High frequencies also show consistent activity, likely representing hi-hats, claps, or synthesizers.
  - Sharp transitions and periodic patterns dominate, reflecting a high-energy, fast-paced rhythm.
- **Analysis:**
  - Party songs are designed for **dancing and high energy,** which aligns with the dominant low-frequency beats (bass) and high-frequency bursts.
  - The consistent periodic patterns indicate a structured beat with less harmonic complexity compared to classical or romantic songs.

Comparative Summary

Aspect	**Romantic Song**	**Pop Song**	**Classical Song**	**Party Song**
Frequency Range	Dominant in low-mid frequencies (0–3000 Hz)	Broad (0–8000 Hz), emphasis on low and high	Wide and balanced distribution	Full range (0–10,000 Hz), strong low/high
Energy Intensity	Gradual, with soft transitions	High bursts with rhythmic patterns	Moderate with subtle dynamic changes	High intensity with periodic energy bursts
Dynamics	Smooth, prolonged tones	Quick transitions, sharp dynamics	Gradual, rich harmonic transitions	Fast-paced, consistent rhythmic patterns
Genre Features	Emotional, melody-focused	Quick transitions, sharp dynamics	Tonal richness, harmonic complexity	High-energy beats for dancing

Conclusion

The spectrograms correctly capture the unique auditory signatures of the four music genres. Romantic songs prioritize melody and smooth transitions, pop songs highlight rhythm and vibrancy, classical music emphasizes harmonic complexity, and party songs focus on energetic, danceable beats. These insights align well with general speech and music understanding.

References

1. https://librosa.org/doc/main/generated/librosa.stft.html

2. https://www.phon.ucl.ac.uk/courses/spsci/acoustics/week1-10.pdf

3. https://scikit-learn.org/1.6/modules/generated/sklearn.svm.SVC.html

Github Repositry Link:

1. hhttps://github.com/Ankit-IITJ/Speech$_{U}nderstanding_{P}A1.git$