

# Speech Understanding

## Souvik Maji

## Introduction

This document provides a comparative analysis of four songs, each belonging to a distinct genre. **Spectrograms** for these tracks can be found at: Spectrograms Repository. The corresponding **song files** are located here: Songs Repository.

Each spectrogram has been visually examined to understand the frequency-energy distribution, transitions, and unique characteristics that align with the typical features of their respective genres.

## 1 Classical Indian Track

**Song:** *Dar Lage Garje Badariya*

**Genre:** Indian Classical

### Overall Structure

- The spectrogram reveals multiple vertical striations in the mid-to-upper frequency range (around 500 Hz to 3–4 kHz), appearing as closely spaced, high-energy bands (yellow/orange columns).
- Lower frequencies (below  $\sim 150$ –200 Hz) display a more diffuse energy distribution, indicating a lesser focus on strong, repetitive bass patterns typical of pop or electronic music.

### Mid-Range & Vocals

- Indian classical vocals predominantly occupy a broad midrange (200 Hz to 2–3 kHz). Their melodic movement appears as shifting energy bands within this range.
- Microtonal ornamentations and slides (*gamakas*) are strongly present, visible as short, closely spaced horizontal lines or small frequency swoops between 500–2,000 Hz.

### High Frequencies

- The higher frequencies (above  $\sim 3$  kHz) feature bursts of energy, likely from vocal harmonics or accompanying instruments such as the harmonium or a high-pitched percussion instrument (e.g., manjira). However, these bursts are more intermittent compared to the electronic track.

## Dynamic Changes

- Cyclic shifts in overall amplitude (color variations from orange to purple) are visible, likely reflecting moments when the vocalist intensifies or softens their delivery.

## 2 Electronic/Dance Track

**Song:** *Matushka Ultrafunk*

**Genre:** Dance/Electronic

### Overall Structure

- The lower half of the spectrogram (below  $\sim 500$  Hz) remains consistently yellow to orange, indicating high-energy low frequencies—characteristic of an electronic kick and bassline that drive the track throughout.
- Large, uniform blocks of energy, especially after  $\sim 0:30$ , indicate the steady, pulsating rhythm of the “four-on-the-floor” beat.

### Low-Frequency/Bass Patterns

- Dance/electronic music typically features strong, continuous low-frequency energy. The 50–200 Hz range remains bright (yellow) for extended periods, representing a powerful, pumping bass.

### High Frequencies

- The higher frequencies ( $\sim 3$  kHz to  $\sim 16$  kHz) show a near-continuous wash of mid-intensity energy (purple/orange hues), commonly produced by hi-hats, cymbals, or synth pads. This sustained brightness suggests a consistent high-frequency presence, typical of dance music.

### Transitions/Breaks

- Notable drops or breakdowns are evident as sudden darkening in parts of the spectrogram, such as around 0:15–0:30 and near 1:00—features typical of EDM-style transitions.

## 3 Rock Track

**Song:** *Natural*

**Genre:** Rock

## Overall Structure

- The midrange (200 Hz–2 kHz) and upper midrange (2–5 kHz) show sustained energy, reflecting the prominence of electric guitars.
- The spectrogram displays distinct segments with varying textures, likely corresponding to different song sections (verse/chorus transitions, solos, breakdowns).

## Bass & Drums

- Bass frequencies ( $\sim 50$ –150 Hz) are bright but do not form an unbroken block like in electronic tracks. Instead, more “pulses” or narrow vertical columns are visible, reflecting the percussive nature of rock drum hits and short bass guitar sustains.

## Guitars & Vocals

- Electric guitars typically range from 100 Hz to about 6–7 kHz, with the most energy concentrated around 1–3 kHz (giving them their distinctive “bite” or “edge”). The blocky, vertically banded patterns in this range indicate chords and strumming.
- Vocals generally occupy the 200 Hz–3 kHz range, though in a dense rock mix, they may be visually masked by the guitar layers. Some darker horizontal or diagonal lines in this range could represent sustained vocal notes.

## Notable Break

- A significant shift in spectrogram color and texture occurs around 1:30–1:50 and again near 2:10–2:30 (dark bands or frequency drops). These may correspond to guitar solos, quieter bridge sections, or dynamic breakdowns.

# 4 Indian Pop Track

**Song:** *Tum hi ho*

**Genre:** Indian Pop

## Overall Structure

- Compared to the classical track, this one exhibits more sustained low-to-mid frequency energy, characteristic of a modern, studio-produced pop track with bass and drums.
- Consistent blocks of orange around  $\sim 100$ –300 Hz indicate a stronger, more contemporary pop-style bassline than in traditional classical music.

## Vocals & Ornamentation

- Vocals are prominent in the midrange (500–3 kHz) with fairly continuous presence. They appear as lighter-colored (higher-energy) lines that rise and fall throughout the track.
- Indian pop vocals often incorporate ornamentation (though less elaborate than in classical music), seen as pitch slides but within a more structured pop format (shorter runs, distinct phrases).

## High Frequencies

- Frequencies from  $\sim 3$  kHz to 8 kHz show moderate strength, likely due to cymbals, digital percussion, or synth pads. While not as sustained and bright as in EDM, they still contribute to the polished production typical of pop music.

## Transitions

- Verse-chorus shifts occur at regular intervals (e.g., around 0:45, 1:15, 1:45, 2:15), marked by changes in intensity across the frequency spectrum. The chorus sections tend to be denser or louder, which slightly brightens the spectrogram in multiple frequency bands.

# 5 Key Comparative Takeaways

## Low-Frequency Energy (Kick/Bass)

**Dance/Electronic:** Strong, sustained low-frequency blocks.

**Rock:** Pulsed low-end from drums and bass guitar.

**Indian Pop:** Moderate but steady pop-style bass presence.

**Classical:** Less continuous bass, with more intricate, nuanced patterns.

## Midrange (Vocals, Guitars, Lead Instruments)

**Indian Classical:** Dominated by melodic vocals (or traditional instruments) with prominent pitch ornamentation.

**Rock:** Guitars and vocals strongly fill the midrange, creating a dense, saturated texture.

**Dance/Electronic:** Vocals and lead synth lines exist but are often overshadowed by bass and percussion.

**Indian Pop:** Clear vocal lines with occasional ornamentation, layered over a modern pop mix.

## High Frequencies (Cymbals, Hi-Hats, Sibilance)

**Dance/Electronic:** Consistent high-frequency wash (hi-hats, synths).

**Rock:** High-frequency bursts from cymbals and distortion harmonics, with some sections dropping out.

**Indian Pop:** A polished, moderate high-end sheen.

**Classical:** Intermittent high-frequency bursts tied to vocal harmonics or high-pitched instruments.

## Transitions & Structure

**Classical:** More fluid, continuous movement (typical of raga-based compositions).

**Dance/Electronic:** Distinct breakdowns, drops, and repetitive sections.

**Rock:** Clearly defined verse, chorus, and solo transitions, creating visually distinct blocks.

**Indian Pop:** Verse-chorus transitions are evident but maintain a relatively uniform mix throughout.

These spectrogram patterns align with the fundamental characteristics of each genre: the driving beats and steady low-end energy of electronic music; the guitar-driven, midrange-intense texture of rock; the fluid ornamentation of Indian classical vocals; and the structured yet expressive nature of Indian pop.