FFT Spectral Analysis

Overview

The **FFT (Fast Fourier Transform) Spectral Analysis** transforms identity drift time-series into the frequency domain. This reveals oscillation patterns that are invisible in time-domain plots: how often does identity 'flicker'? Do some providers show high-frequency instability masked by low time-domain drift?

This analysis treats each experiment's drift trajectory as a signal and decomposes it into constituent frequencies using FFT. The resulting power spectral density (PSD) reveals the 'spectral signature' of each provider's identity dynamics.
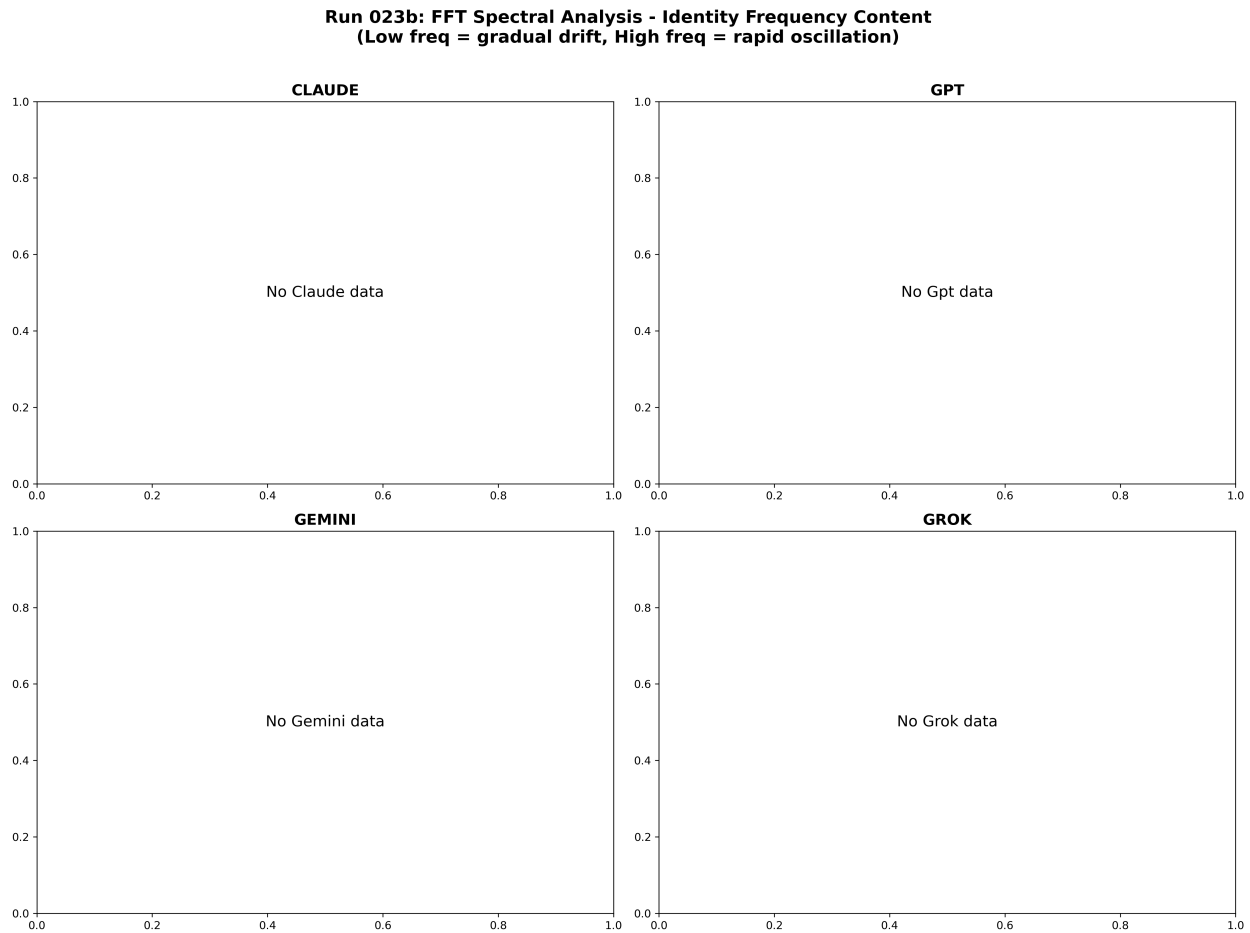
1. Provider Spectral Signatures

Run 023b: FFT Spectral Analysis - Identity Frequency Content
(Low freq = gradual drift, High freq = rapid oscillation)

Figure 1: FFT power spectral density by provider

What it shows: The power spectral density (PSD) for each provider family, showing how drift 'energy' is distributed across frequencies. The X-axis represents frequency (oscillations per iteration), and the Y-axis represents power (amplitude squared).

Reading the spectrum:
- **Low frequencies (left):** Slow, gradual drift - identity evolving smoothly
- **High frequencies (right):** Rapid 'flickering' - identity oscillating quickly
- **Peaks:** Dominant oscillation modes - characteristic 'resonances' in identity
- **Flat spectrum:** White noise - no preferred oscillation frequency

Key patterns to look for:
- Providers with **strong low-frequency components** show smooth, gradual drift
- Providers with **significant high-frequency content** exhibit rapid identity fluctuation
- **Sharp peaks** indicate resonant modes - identity 'rings' at specific frequencies
- **1/f patterns** (falling spectrum) are common in natural systems

2. The EEG Analogy

Human brain activity is characterized by spectral bands: delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), gamma (30+ Hz). Each band correlates with cognitive states (sleep, relaxation, focus, active thinking, high-level processing).

The hypothesis: If LLMs trained on human text capture human cognitive dynamics, they may exhibit analogous 'identity bands' - characteristic frequency regimes that correlate with different operational states (baseline, stressed, recovering).

Preliminary observations:
- Most providers show dominant low-frequency content (gradual drift)
- High-frequency components are generally smaller but provider-specific
- Gemini may show different spectral profile reflecting its 'transform' behavior
- Mistral's stability may manifest as very narrow, low-frequency spectrum

3. Future Analysis Directions

The FFT spectral view opens several analysis directions:

Spectrogram (Time-Frequency): How does the spectrum evolve over the course of an experiment? Does crossing the Event Horizon trigger spectral changes?

Cross-Spectral Analysis: Do different providers share spectral features? Coherence analysis could reveal shared frequency components across architectures.

Spectral Clustering: Can we cluster providers by spectral similarity rather than time-domain metrics? This might reveal hidden architectural relationships.

Band-Pass Filtering: Isolate specific frequency bands and analyze their contribution to total drift. Which frequencies carry the 'identity stress' signal?

Technical Notes

FFT Implementation: Standard NumPy FFT applied to drift time-series. Each experiment (30 iterations) provides 30 samples. Nyquist frequency = 0.5 oscillations per iteration. Zero-padding used for spectral resolution.

Windowing: Hanning window applied to reduce spectral leakage. This smooths the spectrum at the cost of some frequency resolution.

Power Spectral Density: Computed as |FFT|^2 normalized by sample length. Units are arbitrary but consistent across providers for comparison.

Interpretation Caution: With only 30 samples per trajectory, frequency resolution is limited. Higher-resolution spectral analysis would require longer experiments (more iterations) or concatenated trajectories.