The Cello as a Complex Wave System

Part 2: From Signal to Feature Vector

Arina Abbiasova

Prerequisites. This part builds directly on Part 1. The reader should be familiar with the wave equation, normal modes, Helmholtz motion, and the Schelleng bounds. Here we ask: given the acoustic signal $x(t)$ recorded by a microphone, how do we extract a compact, physically meaningful description of the wave field that produced it?

1. The Problem: From Waveform to State

The microphone records a time series $x(t)$. This signal is a one-dimensional projection of a high-dimensional physical state: the superposition of all body modes of the instrument, the acoustic radiation pattern, and the room response. We cannot invert this projection exactly — the map from wave field to microphone signal is many-to-one.

What we *can* do is extract a compact vector

$$\Phi[x] = \big(F_1[x],\, F_2[x],\, \ldots,\, F_9[x]\big) \in \mathbb{R}^9 \tag{1}$$

whose components are physically motivated functionals of $x(t)$, chosen so that different playing regimes and different instruments occupy well-separated regions of $\mathbb{R}^9$.

This is not a new idea in signal processing. What is specific to our program is that each component of $\Phi[x]$ is *derived from the physics of Part 1*, not chosen arbitrarily. The link between the mathematics of the wave equation and the computational features is explicit and traceable.

2. Discretisation: From Continuous Signal to Samples

2.1. Sampling and the Nyquist Theorem

A real microphone produces a **discrete time series** $x[k] = x(k/f_s)$, $k = 0, 1, \ldots, N-1$, where f_s is the **sampling rate** (typically $f_s = 44\,100\,\mathrm{Hz}$ for audio). The Nyquist–Shannon theorem states that this discrete representation faithfully captures all frequency content up to the **Nyquist frequency**:

Nyquist Theorem. A signal band-limited to $f_{\max}$ is exactly recoverable from uniform samples at rate f_s, provided:

$$f_s > 2\, f_{\max}. \tag{2}$$

For $f_s = 44\,100\,\mathrm{Hz}$: maximum recoverable frequency $= 22\,050\,\mathrm{Hz}$, well above the range of cello overtones ($\lesssim 10\,\mathrm{kHz}$).

2.2. Quasi-Stationary Segments

A real cello tone is not perfectly stationary — the player's bow speed and pressure fluctuate slowly. We therefore apply the analysis to a **quasi-stationary segment**: a window of length N_w samples selected from the steady portion of the tone (after the attack, before the release).

For a segment of duration $T_w = N_w/f_s$, the **frequency resolution** of any spectral analysis is:

$$\Delta f = \frac{1}{T_w} = \frac{f_s}{N_w}. \tag{3}$$

We require $\Delta f \ll f_1$ to resolve individual harmonics. For the cello C2 string ($f_1 = 65.4\,\mathrm{Hz}$), a window of $T_w = 1\,\mathrm{s}$ gives $\Delta f = 1\,\mathrm{Hz}$ — more than adequate.

3. The Discrete Fourier Transform

3.1. Definition and Physical Meaning

The **Discrete Fourier Transform** (DFT) of a length-N sequence $x[k]$ is:

DFT

$$X[m] = \sum_{k=0}^{N-1} x[k]\, e^{-2\pi i\, mk/N}, \qquad m = 0, 1, \ldots, N-1. \tag{4}$$

The coefficient $X[m]$ gives the **complex amplitude** of the frequency component $f_m = m\, f_s/N$.

The DFT is the discrete analogue of the Fourier series expansion (**??**) from Part 1. The magnitude $|X[m]|$ corresponds to the amplitude of the m-th frequency component; the phase $\arg X[m]$ to its phase offset. In practice we compute the DFT via the **Fast Fourier Transform** (FFT) algorithm, which reduces the complexity from $\mathcal{O}(N^2)$ to $\mathcal{O}(N \log N)$.

Since $x[k]$ is real-valued, $X[N-m] = X^*[m]$ (conjugate symmetry), and we retain only the **one-sided spectrum**: $m = 0, 1, \ldots, \lfloor N/2 \rfloor$.

3.2. The Spectral Leakage Problem

The DFT implicitly assumes that the signal $x[k]$ is periodic with period N. When a sinusoidal component has a frequency that is *not* an exact integer multiple of Δf, the periodicity assumption creates a discontinuity at the segment boundaries. The Fourier transform of this discontinuity spreads energy across many frequency bins — this is **spectral leakage**.

Example. A pure tone at $f = 65.4\,\text{Hz}$ in a segment of $N = 44\,100$ samples gives exactly $65.4\,\text{Hz}\,/\Delta f = 65.4$ — not an integer. The nearest bins are $m = 65$ and $m = 66$. Without windowing, energy leaks from these two bins across the entire spectrum, masking weaker harmonics.

3.3. The Hann Window

To suppress spectral leakage, we multiply $x[k]$ by a **window function** $w[k]$ before computing the DFT. The most widely used choice for harmonic analysis is the **Hann window**:

Hann Window

$$w[k] = \frac{1}{2}\left[1 - \cos\left(\frac{2\pi k}{N-1}\right)\right], \qquad k = 0, \ldots, N-1. \tag{5}$$

The windowed DFT is computed on $\tilde{x}[k] = w[k] \cdot x[k]$.

The Hann window tapers the signal smoothly to zero at both ends, eliminating the boundary discontinuity. The cost is a reduction in frequency resolution by a factor of ≈ 2 (the main lobe width doubles). This is an acceptable trade-off: for the cello, harmonics are separated by $f_1 \geq 40\,\text{Hz}$, far wider than the broadened bin.

3.4. Power Spectrum and dB Scale

We define the **power spectrum** as:

$$S[m] = |X[m]|^2. \tag{6}$$

For display and analysis we convert to **decibels**:

$$S_{\text{dB}}[m] = 20\log_{10}\left(\frac{|X[m]|}{|X|_{\max}}\right), \tag{7}$$

so that the peak bin has $S_{\text{dB}} = 0$ and quieter components appear as negative values. The logarithmic scale compresses the dynamic range (typically 60–80 dB for a cello tone) into a visually interpretable range.

4. The Feature Vector: Nine Physically Grounded Quantities

We now define each component of $\Phi[x]$ and derive it from the physics of Part 1. All quantities are computed from the windowed power spectrum $S[m]$ of a quasi-stationary segment.

4.1. Features 1–7: Relative Harmonic Amplitudes $H_2, \ldots, H_8$

Physical origin. From the Helmholtz solution (Part 1, eq. (7)): $A_n \propto 1/n$. Real instruments deviate from this ideal, and the *pattern of deviations* is instrument-specific.

Definition. For harmonic n, we search for the spectral peak in a window of $\pm 7\%$ around the expected frequency nf_0:

$$H_n = \frac{\max_{|f-nf_0|\le 0.07\,nf_0} |X(f)|}{\max_{|f-f_0|\le 0.07\,f_0} |X(f)|}, \quad n = 2, \ldots, 8. \tag{8}$$

By construction, $H_1 = 1$. The ratio H_n measures how much energy the n-th harmonic carries *relative to the fundamental*. The $\pm 7\%$ window accommodates the inharmonicity shift Δf_n from eq. (**??**) of Part 1.

Physical interpretation.

- Violin: slow roll-off ($H_n \approx n^{-0.7}$) — bright, rich harmonics.
- Cello: moderate roll-off ($H_n \approx n^{-1.1}$) — warm, full tone.
- Double bass: steep roll-off ($H_n \approx n^{-2}$) — dominated by fundamental.

These exponents are not arbitrary: they reflect the frequency-dependent impedance of the instrument body, which attenuates higher modes more strongly in heavier, larger instruments.

4.2. Feature 8: Spectral Centroid ν_c

Physical origin. The spectral centroid is the energy-weighted mean frequency of the signal. It measures the *centre of mass* of the power spectrum, and correlates with perceived brightness.

Spectral Centroid

$$\nu_c = \frac{\sum_m f_m\, S[m]}{\sum_m S[m]}, \tag{9}$$

where the sum runs over all frequency bins m.

Connection to Part 1. For a pure Helmholtz sawtooth with $A_n = A_1/n$, the centroid can be computed analytically:

$$\nu_c^{\text{Helm}} = \frac{\sum_{n=1}^{\infty} nf_1 \cdot (1/n)^2}{\sum_{n=1}^{\infty}(1/n)^2} = \frac{f_1 \sum n^{-1}}{\sum n^{-2}}. \tag{10}$$

Both sums diverge/converge differently, so ν_c is sensitive to the high-frequency content: even a small boost of upper harmonics (as in the harsh regime) shifts ν_c upward substantially. This is why ν_c is one of the most discriminating features in our analysis.

4.3. Feature 9a: Spectral Bandwidth $\Delta\nu$

The bandwidth measures the *spread* of energy around the centroid:

Spectral Bandwidth

$$\Delta\nu = \sqrt{\frac{\sum_m (f_m - \nu_c)^2 \, S[m]}{\sum_m S[m]}}. \tag{11}$$

For a clean Helmholtz tone, energy is concentrated in narrow harmonic peaks and $\Delta\nu$ is small. In a noisy regime (near-breaking), broadband noise fills the gaps between harmonics and $\Delta\nu$ grows sharply. In our data, $\Delta\nu$ increases by a factor of 4.3 from stable to near-breaking playing — one of the clearest numerical signatures of the transition.

4.4. Feature 9b: Spectral Slope α

Spectral Slope

$$\alpha = \frac{d}{d\nu}\big[S_{\mathrm{dB}}(\nu)\big] \approx \text{slope of linear fit to } S_{\mathrm{dB}}(\nu) \text{ on } [20\,\mathrm{Hz}, 5\,\mathrm{kHz}]. \tag{12}$$

Units: dB/kHz.

Physical meaning. The Helmholtz amplitude law $A_n \propto 1/n$ implies $S_{\mathrm{dB}} \propto -20\log_{10}(f/f_1)$: a logarithmically decreasing spectrum. On a linear frequency axis this is a convex curve, well approximated locally by a negative slope. A steeper slope (α more negative) indicates faster attenuation of high modes — characteristic of larger, heavier instruments with more body damping.

4.5. Feature 9c: High-Frequency Energy Ratio ρ_{HF}

High-Frequency Energy Ratio

$$\rho_{\mathrm{HF}} = \frac{\sum_{f_m \geq 2\,\mathrm{kHz}} S[m]}{\sum_m S[m]}. \tag{13}$$

The threshold of 2 kHz is chosen to lie above the strongest harmonics of the cello ($f_1 = 65.4\,\mathrm{Hz}$; harmonic $n = 30$ falls at $\approx 2\,\mathrm{kHz}$) but within the range amplified by the instrument body's high-frequency resonances. ρ_{HF} is near zero for a clean tone and grows rapidly in harsh playing, when the bow excites incoherent high-frequency motion. In our data it increases by a factor of ≈ 20 from stable to near-breaking.

4.6. Feature 9d: Inharmonicity Index η

Physical origin. From Part 1 (eq. (**??**) there): real partials deviate from exact harmonics due to string stiffness.

Inharmonicity Index

$$\eta = \frac{1}{N_h - 1} \sum_{n=2}^{N_h} \frac{|f_n^{\text{actual}} - n f_0|}{n f_0}, \tag{14}$$

where f_n^{actual} is the measured peak frequency near nf_0, and $N_h = 8$.

η is computed from the same peak-search used for H_n. It is instrument-specific (double bass has largest η, violin smallest) and largely independent of playing regime, making it a stable "identity" feature of the string construction.

4.7. Summary: The Feature Vector

Summary. The complete feature vector is:

$$\Phi[x] = \big(\underbrace{H_2, H_3, H_4, H_5, H_6, H_7, H_8}_{\text{harmonic profile}}, \underbrace{\nu_c, \Delta\nu}_{\text{spectral shape}}, \underbrace{\alpha}_{\text{slope}}, \underbrace{\rho_{\text{HF}}}_{\text{HF energy}}, \underbrace{\eta}_{\text{inharmonicity}}\big). \tag{15}$$

All 12 components are dimensionless or carry units of Hz or dB/kHz. Each is derived from a physically motivated property of the wave field.

5. From Features to State Space

5.1. Why This Vector Is Sufficient (and Why It Is Not)

The feature vector $\Phi[x]$ is a *coarse* description. It captures the steady-state spectral structure but discards:

- **Transient behaviour** — the attack and release of each note, which carry information about the stick-slip initiation.
- **Phase relationships** between harmonics — destroyed by taking $|X[m]|^2$.
- **Temporal evolution** within the quasi-stationary segment — slow modulations of ν_c or $\Delta\nu$ within a single tone.

These omissions are deliberate at the pilot stage: we want the smallest set of features that still separates regimes. Parts 3 and 4 will assess whether this minimal set is sufficient, and where it fails.

5.2. Geometric Interpretation

Each recording — one instrument, one playing regime, one repetition — is a single point $\Phi[x] \in \mathbb{R}^{12}$. A dataset of N_{rec} recordings is a **point cloud** in this space. The central question of the research program is:

Does the point cloud have a physically meaningful geometric structure? Do different instruments and regimes form well-separated clusters? Are there smooth trajectories connecting them?

These questions are addressed computationally in Part 3 via PCA and k-means clustering, and analytically in Part 4 via connection to the Schelleng diagram.

5.3. Dimensional Analysis of the Features

Before any computation, we can reason about which features should separate instruments versus regimes, using only the physics of Part 1.

Instrument-discriminating features (vary across instruments, stable within one instrument across regimes): $H_2, \ldots, H_8$ (set by body impedance), η (set by string construction), α (set by body damping).

Regime-discriminating features (vary across regimes of one instrument): ν_c, $\Delta\nu$, ρ_{HF} (all sensitive to noise level and high-mode excitation, which change with bow force).

> **Key Prediction.** The feature space $\mathbb{R}^{12}$ should have a **block structure**: the first group $\{H_n, \eta, \alpha\}$ discriminates instruments; the second group $\{\nu_c, \Delta\nu, \rho_{\mathrm{HF}}\}$ discriminates regimes. This prediction is testable by examining the PCA loadings in Part 3.

6. Computational Implementation

6.1. Pipeline Architecture

The analysis is implemented as a two-language hybrid pipeline:

C++ layer (low-level, time-critical):

1. Read WAV file; extract 16-bit or 32-bit PCM samples.
2. Locate quasi-stationary segment (minimum RMS fluctuation over sliding 50 ms windows).
3. Apply Hann window (eq. 5).
4. Compute real FFT via Cooley–Tukey algorithm.
5. Extract peak frequencies and amplitudes for $n = 1, \ldots, 8$.
6. Compute H_n, η from peak data.
7. Write feature vector to CSV.

Python layer (statistical analysis):

1. Read feature CSV; compute ν_c, $\Delta\nu$, α, ρ_{HF} from FFT magnitudes.
2. Standardise: $\tilde{\Phi} = (\Phi - \mu)/\sigma$ per feature.
3. PCA, k-means, trajectory analysis (Part 3).

6.2. Standardisation

The components of $\Phi[x]$ have very different numerical scales: $H_n \in [0,1]$, $\nu_c \in [50, 2000]\,\mathrm{Hz}$, $\alpha \in [-5, 0]\,\mathrm{dB/kHz}$. Before any distance-based algorithm (PCA, k-means), we standardise each feature to zero mean and unit variance:

Standardisation

$$\tilde{F}_j = \frac{F_j - \mu_j}{\sigma_j}, \tag{16}$$

where μ_j and σ_j are the mean and standard deviation of feature j across all recordings in the dataset.

Without standardisation, features with large numerical range (like ν_c) would dominate the Euclidean distance metric used by PCA and k-means, regardless of their actual discriminating power.

7. Validation: Synthetic vs. Real Data

7.1. Pipeline A: Synthetic Signals

In the absence of a recording session, we first validate the pipeline on synthetic signals constructed directly from the physics:

$$x(t) = \sum_{n=1}^{N_h} A_n \sin(2\pi f_n t + \phi_n) + \sigma\,\xi(t), \tag{17}$$

where $f_n = n f_0(1 + \beta_{\mathrm{inh}} n^2)$, the amplitudes $A_n = A_1 n^{-\gamma}$ implement the instrument-specific roll-off, and $\xi(t)$ is white noise.

The key parameters are:

Instrument	f_0 (Hz)	γ	σ
Violin	196.0	0.70	0.04
Cello	65.4	1.10	0.06
Double Bass	41.2	2.00	0.05

For the cello playing regimes, σ is varied from 0.04 (stable) to 0.35 (near-breaking), and upper harmonics are boosted proportionally to model the increased high-mode excitation (Schelleng: excess bow force drives higher modes).

7.2. Pipeline B: Real Recordings

The identical feature-extraction code is applied to WAV files recorded from a real cello. Three playing regimes are recorded for the open C2 string: stable tone, intentionally harsh sound, and near-breaking bowing. Multiple repetitions ($N_{\mathrm{rep}} \geq 5$) are recorded under fixed microphone placement and room conditions.

Status. Pipeline A results are complete and reported in Part 3. Pipeline B recordings are in preparation. The comparison between Pipeline A and Pipeline B feature vectors will reveal which synthetic parameters correctly capture the physical system and which require adjustment.

8. Summary

We have established:

1. The **DFT** (eq. 4) as the discrete analogue of the normal-mode decomposition from Part 1, and the **Hann window** (eq. 5) as the standard remedy for spectral leakage.
2. **Seven harmonic amplitude features** $H_2, \ldots, H_8$ (eq. 8), derived from the Helmholtz amplitude law $A_n \propto 1/n$.
3. **Four spectral shape features**: centroid ν_c (eq. 9), bandwidth $\Delta\nu$ (eq. 11), slope α (eq. 12), HF ratio ρ_{HF} (eq. 13).
4. **Inharmonicity index** η (eq. 14), derived from the stiffness correction of Part 1.
5. A **block-structure prediction**: $\{H_n, \eta, \alpha\}$ should discriminate instruments; $\{\nu_c, \Delta\nu, \rho_{\mathrm{HF}}\}$ should discriminate regimes.

Part 3 tests these predictions computationally using PCA, k-means clustering, and continuous stability trajectories.

References

Cooley, J. W. & Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, **19**(90), 297–301.

Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, **66**(1), 51–83.

Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the IRE*, **37**(1), 10–21.

Next: Part 3 — Phase Space, Clustering, and Stability Transitions