The Markov-Switching Multifractal as a Generative Model of Musical Intensity: An Instrument, Measurements, and a Research Programme

Evan Tabak Atlas

Independent · ORCID 0009-0007-7374-2338

Evanatlas@gmail.com

Preprint. Intended for arXiv, cs.SD (Sound), cross-listed eess.AS (Audio and Speech Processing). Paper 3 in the Fenophone research-note series.

ABSTRACT

Musical intensity — the joint trajectory of onset density, dynamics, and harmonic tension — shares a distinctive phenomenology with financial volatility: bursts and lulls, clustered at every time scale, with no single characteristic period. We propose taking this correspondence literally. We adopt the Markov-Switching Multifractal (MSM) of Calvet and Fisher — a stationary, causal Markov construction of multifractal volatility — not as a metaphor or a texture source but as the generative core of a musical instrument: we discard the return process entirely and sonify the volatility state itself, read through a fixed band map into four control lanes (dynamics, density, timbre energy, harmonic tension) that drive an authored musical scaffold. We describe a complete reference implementation (the Fenophone), whose event layer is integer/fixed-point and bit-identical across platforms, whose player controls are exactly the MSM parameters `(m_hi, f_slow, S, k)`, and whose "conducting" gesture is clamping individual latent states. We report measured properties of the emitted note streams: with the authored bar template removed as a seasonal trend, per-scaffold median DFA exponents α(onset) of 0.695–0.770 and α(velocity) of 0.676–0.702, every seed exceeding its shuffled-surrogate null; MFDFA widths `h(−2) − h(2)` of 0.054–0.126, exceeding matched IAAFT surrogate nulls on every seed for three of the four measured packs (the fourth, the least intermittent by design, is consistent with linear long memory at this length and reported as such); and a monotone dose–response of α on the time-scale-span control, with a median gap ≥ 0.09 between span extremes. All measurements are CI-gated, surrogate-compared, estimator-anchored on known-Hurst ground truth, and reproducible from pinned seeds. We then lay out a six-experiment research programme — surrogate discrimination, a preference surface over the (α, width) plane, tension-lane validation, boundary perception against ground-truth latents, counterfactual intervention pairs, and corpus inversion — each with predictions and explicit falsification conditions. The corpus half of the inversion experiment now has first results (2026-07-12; companion analysis): 1,260 MAESTRO piano performances measure median α(onset) 0.717 and width 0.118 under the same protocol — inside the instrument's published band — with fitted parameters identifying composer above chance, while drum-kit onset counts are near-white with the cascade structure appearing on the velocity channel instead. The five perceptual

experiments (E1–E5) remain unrun, and the corpus inversion (E6) is complete only on its statistical half; this paper is a preregistration-grade programme, not a report of finished perceptual results.

1. Introduction: music lives in the volatility

A well-known fact about asset returns is that the returns themselves are nearly unpredictable while their *volatility* is highly structured: quiet periods and turbulent periods alternate, cluster, and nest, with correlations decaying slowly over horizons from minutes to years (Mandelbrot, Fisher & Calvet, 1997; Calvet & Fisher, 2002). The sign of the next price move carries almost no information; the *size* of recent moves carries a great deal.

We start from the observation that musical intensity has the same phenomenology. Consider not the pitches of a piece but its *activity*: how many onsets fall in each beat, how forcefully they are struck, how restless the harmony is. Music is not uniformly busy. It surges and recedes; the surges themselves contain smaller surges; a four-bar swell sits inside a section-length arc which sits inside the dramatic shape of the whole piece. This burst-and-lull structure, nested across scales, is exactly volatility clustering. The empirical literature supports the analogy at both ends of the time-scale range: composed rhythm exhibits 1/f-family spectra whose exponents vary by composer (Levitin, Chordia & Menon, 2012), and even the involuntary microtiming of human performers is long-range correlated, with listeners reportedly preferring such correlated fluctuations over white jitter (Hennig et al., 2011; Räsänen et al., 2015). Multifractal analyses of scores and performances find not just a single scaling exponent but a spectrum of them (Su & Wu, 2006; Jafari, Pedram & Hedayatifar, 2007).

The standard move in generative music, going back to Voss and Clarke (1978), is to take such measurements as a *filter specification*: generate 1/f noise, map it to notes. Our proposal is different in kind. Econometrics has spent five decades building *generative models* of volatility — models with latent states, estimators, forecasting theory, and a well-understood repair history — precisely because volatility is where the structure lives. Among these, the Markov-Switching Multifractal (MSM) of Calvet and Fisher (2001, 2004, 2008) is distinguished: it produces genuinely multifractal fluctuations (a whole spectrum of scaling exponents, not one), it is a low-dimensional parametric family (a handful of interpretable parameters), and — decisively for our purposes — it is *stationary, causal, and Markov*, so it can be iterated forward one tick at a time, forever, in real time.

The proposal of this paper is therefore: take MSM not as a metaphor but as the literal generative core of a musical instrument. Discard the return process — the part of the asset-price model that is unpredictable and musically uninteresting — and sonify the volatility state itself. Where the econometrician observes returns and must *filter* to recover the latent multipliers, the instrument simply *plays* the multipliers: their product becomes the intensity envelope, their band-wise sub-products become independent control lanes for dynamics, rhythmic density, timbre energy, and harmonic tension, and the notes themselves are supplied by a separate, deliberately conventional musical scaffold that the MSM state modulates.

This paper does three things. First, it states the theory-to-instrument mapping precisely (§4), including an honest account of which output dimensions are MSM-driven today and which are not. Second, it describes a complete, open reference implementation — the Fenophone, a web instrument whose event layer is integer/fixed-point and bit-identical across CPUs, browsers, and operating systems — and reports measured spectral and multifractal properties of its emitted note streams, all of which are enforced as continuous-integration gates and reproducible from pinned seeds (§5). Third, and mainly, it stakes out a research programme: six experiments, each with a concrete design, deterministic stimulus generation, a predicted outcome, and a statement of what a negative result would mean (§6–§7). The instrument exists and is playable; the perceptual claims are hypotheses. We try throughout to keep the two clearly separated.

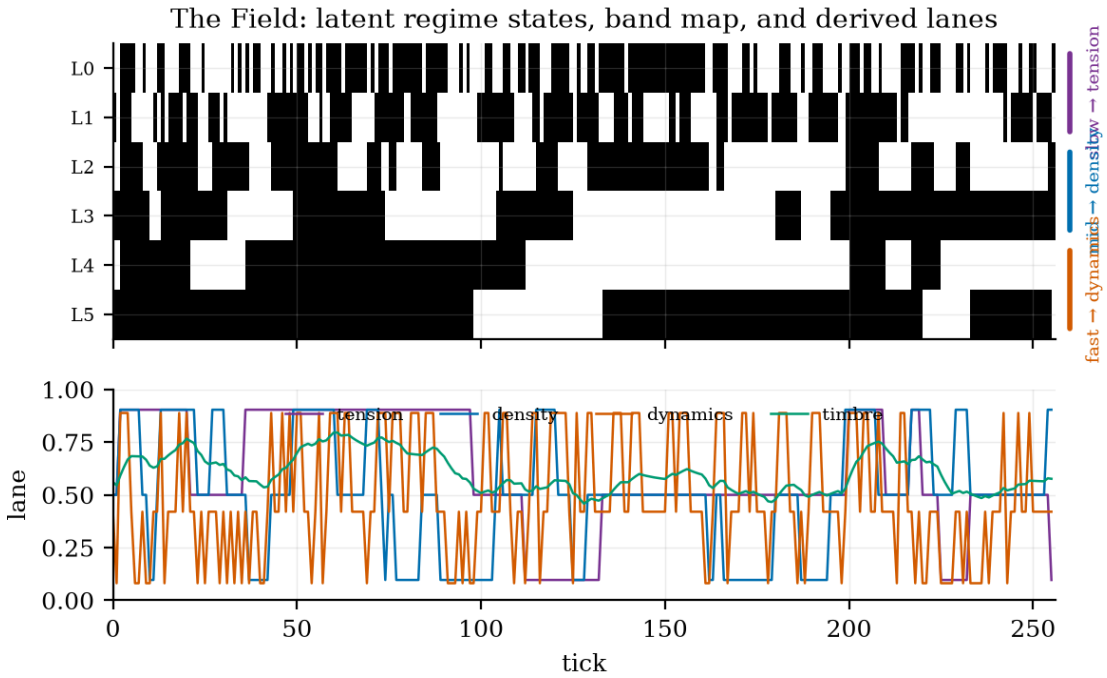

Figure 1. The instrument's central visual, "the Field": layer-state cells (one row per MSM multiplier, slowest at bottom) over 256 ticks, with the slow/mid/fast band partition and the four lanes they drive. Rendered from the engine's per-tick latent states (`foundry render` lanes.csv).

2. Background

2.1 1/f and multifractality in music

Voss and Clarke (1975, 1978) measured approximately 1/f power spectra in the audio power and pitch fluctuations of music and speech across genres, and showed in listening tests that melodies generated from 1/f noise were judged more musical than white-noise or brown-noise (random-walk) melodies. This pair of results — a measurement and a preference — founded a research thread that has been refined but never fully settled.

On the measurement side the modern evidence is reasonably strong. Levitin, Chordia and Menon (2012) analyzed the *rhythm* spectra of a large corpus of Western classical scores and found 1/f^β spectra with β varying systematically by composer — Beethoven and Mozart occupying different regions of the exponent axis — suggesting the scaling exponent is a stylistic degree of freedom rather than a universal constant. Hennig et al. (2011) showed that the timing errors of human performers (drumming and tapping) are long-range correlated over hundreds of events, i.e. that even involuntary human fluctuation is 1/f-family rather than white; Räsänen et al. (2015) confirmed long-range correlation in the microtiming of a professional studio

recording. Beyond single exponents, multifractal analyses — typically via multifractal detrended fluctuation analysis, MFDFA (Kantelhardt et al., 2002) — report nontrivial singularity spectra in scores and performances (Su & Wu, 2006; Jafari, Pedram & Hedayatifar, 2007): the small fluctuations and the large fluctuations scale with *different* exponents, which no Gaussian 1/f process reproduces. The same multifractal structure — read as "expressiveness," a fluctuating mix of predictable and unpredictable patterns across timescales — appears beyond human music, in thrush-nightingale song, where subtle note-timing and intensity deviations (not note arrangement alone) measurably carry the multifractality (Roeske, Kelty-Stephen & Wallot, 2018), suggesting a general engagement signature rather than a Western-repertoire artifact.

On the preference side the literature is more contested, and we lean on it only as motivation, not as settled fact. The Voss–Clarke listening results were informal by modern standards; the "1/f is preferred" generalization has been tested mainly on melodic pitch sequences and on microtiming humanization (Hennig et al., 2011, report preference for long-range-correlated jitter over white jitter), with mixed results across stimulus domains, and there are known statistical routes to *apparent* multifractality (finite size, heavy tails) that do not reflect genuine nonlinear structure (Bouchaud, Potters & Meyer, 2000). We therefore treat two questions as open and experimentally decidable: whether listeners can perceive multifractal structure *beyond* the power spectrum at all (E1, §6), and whether preference is systematically organized over the (exponent, multifractal width) plane (E2). The instrument described here was built, in part, to make these questions cleanly testable: it produces music whose scaling exponent and multifractal width are *knobs*.

2.2 MSM in econometrics — and why this construction specifically

Multiplicative cascades enter through Mandelbrot's work on turbulence (Mandelbrot, 1974): a coarse quantity is redistributed to finer and finer scales by multiplying independent random weights, producing intermittency — a multifractal measure whose moments scale with a nonlinear exponent function. The Multifractal Model of Asset Returns (MMAR; Mandelbrot, Fisher & Calvet, 1997) imported this into finance by subordinating a Brownian motion to a cascade-generated trading time, and matched the stylized facts of returns strikingly well.

But the MMAR's cascade has a structural defect that its own authors named: it is built recursively on a dyadic grid over a *bounded* interval `[0, T]`. The construction is tied to `T`; it is not stationary, and it is not causal — the weight at a coarse scale is fixed "at the start of time" and shared by an entire subtree of the future. It cannot be iterated forward. For econometrics this

frustrated estimation and forecasting; Calvet and Fisher's repair (2001, 2004; monograph 2008) was to replace the grid recursion with a *Markov chain*: `k` volatility components (multipliers), each a two-state (more generally, finitely supported) Markov process, switching at Poisson-like arrival rates spaced geometrically across scales, their product modulating instantaneous volatility. The cascade's scale hierarchy survives as a hierarchy of *switching frequencies*; grid recursion is replaced by time evolution. The resulting MSM is stationary, causal, Markov with `2^k` states, admits exact maximum-likelihood estimation by standard hidden-Markov filtering for moderate `k` (Calvet & Fisher, 2004; GMM alternatives for large state spaces in Lux, 2008), and reproduces multifractal moment scaling over a wide intermediate range of scales.

The point we want to stress is that *real-time music generation needed exactly the same repair, for exactly the same reason*. A grid-bound cascade can render a fixed-length texture offline, but an instrument must run indefinitely, respond to a parameter change at the next tick, and never privilege the moment the process was switched on. Those are the requirements "stationary, causal, Markov" formalize. Among the multifractal constructions available — grid-bound cascades, the multifractal random walk and its log-infinitely-divisible relatives (Bacry, Delour & Muzy, 2001), and MSM — the MSM is the one whose state is finite, whose per-tick update is a handful of independent Bernoulli switches, and whose parameters are few and independently meaningful. That is why this instrument is built on MSM specifically, and it is also why the *inverse* direction (fitting the model to musical corpora, E6 in §6) comes for free: the estimation theory exists because econometrics needed it.

3. Terminology and reading guide

Throughout, "the cascade" means the MSM latent process as implemented (a deliberate retention of Mandelbrot's term for the multiplicative hierarchy); "the scaffold" means the authored musical data pack (scales, chord-function graph, rhythmic grid and template, voice presets) that turns lane values into notes; "lane" means one of the four `[0,1]` control signals derived from the cascade each tick. `α` is the DFA-1 scaling exponent, `h(q)` the MFDFA generalized Hurst exponent, and "width" the multifractal width `h(−2) − h(2)`. All identifiers (`m_hi`, `f_slow`, `S`, `k`, `Δt`) match the reference implementation's source and configuration files, so every equation below can be checked against running code.

4. The model, precisely

4.1 The latent process: a mirror-balanced binary MSM

The core is a product of `k` switching multipliers. Multiplier `M_i` (layer `i = 1..k`, slowest to fastest) takes one of two states,

```
M_i ∈ { m_hi, m_lo },    m_hi > 1,    m_lo = 1/m_hi
```

mirror-balanced in log space so each multiplier's long-run geometric mean is exactly 1 (`E[log M_i] = 0`). The instantaneous intensity is the product

```
intensity_t = base · M_1,t · M_2,t · … · M_k,t
```

where `base` is an output-gain frame control applied downstream, never inside the model. Raising `m_hi` therefore widens the *contrast* between states without moving the long-run average — the design reason the player's primary control is honest ("Intensity raises drama, not loudness").

Switch rates are specified directly in Hz and are decoupled from `k`. With `f_slow` the slowest layer's rate and `S = f_fast / f_slow` the span of the scale hierarchy, layer `i` has target rate and per-tick switch probability

```
f_i = f_slow · S^((i−1)/(k−1))          // geometric spacing across the fixed span S
f_i ← min(f_i, 8 Hz, 0.9/Δt)            // clamp: stays musical, avoids p → 1 saturation
p_i = 1 − exp(−f_i · Δt),               // Δt = 60 / (BPM · ticks_per_beat) seconds
```

and at each scheduler tick layer `i` toggles its state with probability `p_i`, independently across layers. Because the spacing is parameterized by the span rather than a per-layer ratio, changing `k` refines the *same* slow-to-fast range: at `i = k` the exponent is 1, so `f_k = f_slow · S` for every `k`.

Correspondence with Calvet–Fisher. Writing `b = S^(1/(k−1))` for the per-layer rate ratio, the switch schedule is `p_i = 1 − exp(−f_slow · b^(i−1) · Δt) = 1 − (1 − p_1)^(b^(i−1))` — exactly the Calvet–Fisher (2004) transition-probability ladder `γ_i = 1 − (1 − γ_1)^(b^(i−1))` with `γ_1 = 1 − exp(−f_slow·Δt)`. Two deliberate deviations. First, the multiplier distribution: canonical binomial MSM draws `M ∈ {m_0, 2 − m_0}` with equal probability, normalizing the *arithmetic* mean (`E[M] = 1`); we use the log-symmetric pair `{m_hi, 1/m_hi}`, normalizing the *geometric* mean. For volatility forecasting the arithmetic normalization is the natural one; for an instrument the geometric one is — the log of the whole product is then `log2(m_hi)` times a bounded integer net state symmetric about zero, and "average loudness never drifts with the Intensity setting" holds exactly in log-amplitude, which is the perceptual domain. Second, the switch convention: at an arrival, canonical MSM draws a *fresh* state (which, for a symmetric two-point distribution, changes the value with probability 1/2), whereas our

implementation *toggles*. A toggle-with-probability-`p_i` chain is therefore the canonical fresh-draw chain with arrival probability `γ_i = 2·p_i`; we note this so that parameter values fitted under one convention transport to the other (relevant to E6).

4.2 The band map and the four lanes

Voices never read raw layers. The `k` layers partition into three bands — sized as equally as possible, remainder to mid: `slow = fast = floor(k/3)`, `mid = k − 2·floor(k/3)`, every band non-empty for `k ≥ 3` — and the music reads *band products*. Because all layers share `m_hi`, a band's product is `2^(log2(m_hi) · s)` where

```
s = (#high − #low) ∈ [−n, n]        // the band's integer net state, n = band size
```

Each lane is the band product mapped monotonically into `[0,1]`, affine in the log product, centered at 1/2 when the product is 1:

```
lane = clamp( 1/2 + log2(band_product) / (4·n), 0, 1 )
     = clamp( 1/2 + (log2(m_hi) / (4·n)) · s,   0, 1 )
```

where the reference scale `4·n = 2·n·log2(m_hi_max)` is pinned by the maximum `m_hi_max = 4.0`, so a fully aligned band spans the whole `[0,1]` range only at maximum Intensity. The four lanes and their routing:

lane	reads	musical effect
`dynamics`	fast band	moment-to-moment accent and swell
`density`	mid band	flurries vs. rests over phrases
`timbre`	smoothed whole product (all `k` layers; one-pole low-pass, default coefficient 0.125)	overall energy opens filters, adds drive
`tension`	slow band	high = unstable, leaning harmony; low = resolved

This routing is the instrument's central decoupling: different musical dimensions read different regions of the scale hierarchy, so they do not pump in lockstep — harmony breathes at the slow layers' rate while accent flickers at the fast layers' rate, all from one latent process.

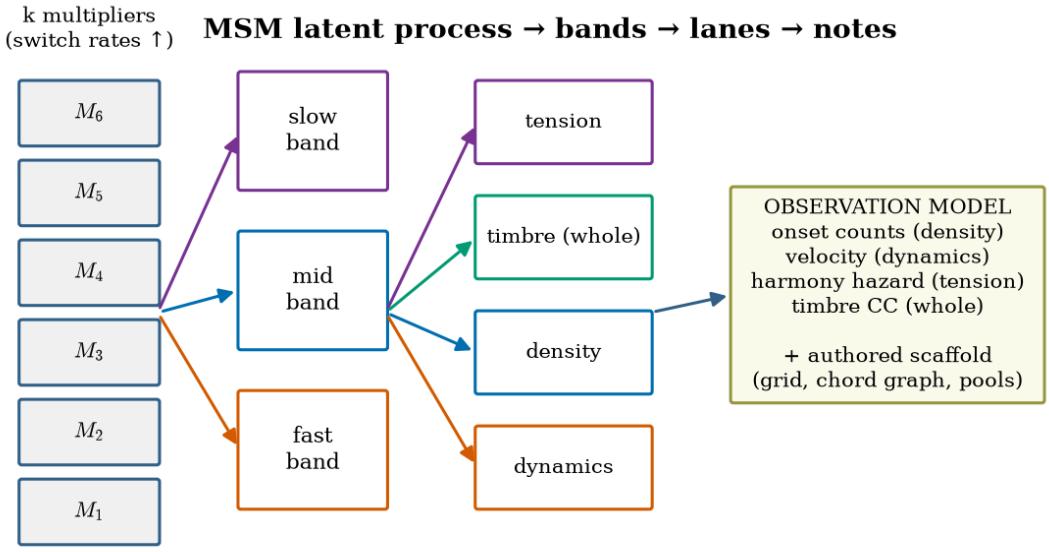

Figure 2. Model schematic: k two-state multipliers with geometrically spaced switch rates → band partition (slow/mid/fast) → four lanes → observation model (onset counts, velocity, harmonic-transition bias, timbre CC), with the authored scaffold shown as a separate, non-stochastic layer the lanes modulate.

4.3 The observation model: an MSM-modulated doubly stochastic point process

The emitted music is a marked point process whose conditional law is a function of the MSM state — in the terminology of point processes, a doubly stochastic (Cox-type) construction (Cox, 1955), with the driving intensity supplied by the MSM band products rather than a log-Gaussian field. Per scheduler tick (the scaffold's grid step), the pipeline is:

1. **Cascade → lanes.** One Bernoulli draw per free layer; the four lanes are computed as above and passed through the scaffold's monotone lane-mapping curves.
2. **Density → onsets.** At each bar downbeat, the `density` lane sets the bar's onset count through a monotone authored curve `onsets(density)`, bounded below and above by per-voice coherence bounds (never silent, never a wall); the onsets are placed on the bar's heaviest grid cells under the authored metric template. Placement is deterministic given the lane — the stochasticity of rhythm lives entirely in the MSM.
3. **Tension → harmony.** At the scaffold's harmonic-rhythm rate, a chord-function walk samples an outgoing edge of an authored chord graph with effective weight `base_weight · score(τ, T)`, where `τ ∈ [0,1]` is the edge's authored tension level, `T` is the live `tension` lane, and

$$\texttt{score}(\tau, T) = (1 - T)\cdot(1 - \tau) + T\cdot\tau + \varepsilon, \qquad \varepsilon = 1/16$$

so high `T` up-weights leaning progressions and low `T` up-weights resolution, with a positive floor guaranteeing the walk never stalls. The slow band thus modulates the *hazard structure* of harmonic transitions — a hidden-Markov-modulated jump process.

4. **Pitch → contour snap.** Each voice carries a continuous pitch-height process `h_t` (§4.4); at an onset, `h` snaps to the nearest pitch in the current chord's legal pool (chord tones plus tension-gated extensions, transposed by key/mode).

5. **Marks.** The note's velocity is a monotone quantization of the `dynamics` lane (`velocity = 1 + round(126 · lane)`), and the `timbre`/`tension` lanes are emitted as 14-bit controller values (`round(16383 · lane)`; CC 74+106 and CC 1+33 respectively), which drive filter/drive and are recorded in exports.

The per-tick onset-count series and the summed-velocity series analyzed in §5 are therefore, in model terms, bar-aggregated counts of a point process whose intensity is a monotone function of the mid-band product, and marks that are monotone functions of the fast-band product — both observed through an authored, deliberately periodic metric template, which every measurement below removes as a seasonal trend.

4.4 The pitch process (not MSM): an OU-type AR(2) walk with register gravity

For scope-honesty we state the pitch model precisely, because it is *not* multifractal. Each voice's pitch height evolves in semitone space as

```
h_{t+1} = h_t + φ·(h_t − h_{t−1})     // momentum ("Flow" memory)
              − γ·(h_t − center)       // register gravity (mean reversion)
              + σ·ε_t                   // bounded integer-Gaussian step noise
```

a discrete Ornstein–Uhlenbeck-style AR(2) on the deviation `d_t = h_t − center`: `d_{t+1} = (1+φ−γ)·d_t − φ·d_{t−1} + σ·ε_t`, with characteristic polynomial `z² − (1+φ−γ)·z + φ`. By the Jury criterion the walk is stationary iff `γ > 0`, `|φ| < 1`, and `2 + 2φ − γ > 0`; the implementation enforces all three (with `φ = 0.97·x` from the Flow control, clamped below 1), and its stationary standard deviation has the closed form

$$\sigma_d = \mathrm{sqrt}\left(\sigma^2 \cdot (1+\varphi) \,/\, [\ (1-\varphi)\cdot\gamma\cdot(2+2\varphi-\gamma)\ ] \right)$$

against which every voice's register band is validated (a mandatory "two-minute drift" CI gate runs each voice at parameter extremes and fails the pack if it exits `center ± span`). Defaults sit at `σ = 0.6–1.2` semitones/step, `γ = 0.05–0.15`, centers near MIDI 40 (bass) and 64 (lead), spans near 9 semitones.

4.5 Honest scope: what is MSM-driven today, and what is not

> **Scope statement.** The MSM claim of this paper covers the following output dimensions, each a memoryless monotone read-out of the latent state: **onset density** (mid band, through the bounded `onsets(density)` curve), **dynamics** (fast band → velocity marks), **timbre energy** (smoothed whole product → filter/drive and CC 74), and the **harmonic-transition hazard** (slow band → the edge-bias `score(τ, T)` and CC 1). It does **not** cover: **pitch increments**, which follow the stationary AR(2) walk of §4.4 with constant step variance on the packs measured here — making σ itself a function of a band product ("pitch-increment volatility") has since been implemented engine-side as a per-voice `pitch_volatility` field that scales the step noise by `2^(x·s/k)` off the cascade net state, but it ships disabled (`= 0`) on every launch scaffold, so those voices remain bit-identical to constant variance and nothing below assumes the coupling; the **authored periodic metric template** (grid weights, bar structure), which is deliberately outside the stochastic claim — it is composed data, the direct analogue of intraday seasonality in volatility data, and it is removed as a seasonal trend in every measurement in §5; and the **harmony graph, pitch pools, and voice timbres**, which are authored scaffold data the MSM modulates but does not generate. (One further layer landed after this draft: **multifractal microtiming**, a dedicated long-range-correlated timing sub-cascade adding sub-tick onset offsets in the spirit of Hennig et al. (2011). It too ships disabled by default on the measured packs, and because it is purely additive to onset *timing* and never alters onset *counts*, it leaves every §5 onset-series measurement unchanged.)

---

5. Reference implementation and measured properties

5.1 The instrument

The reference implementation is the Fenophone, a generative instrument that runs in the browser (Rust core compiled to WebAssembly; the same core ships as a DAW-plugin runtime). Two engineering properties matter for research use. First, the entire event layer — cascade switches, harmony walk, contour walks, scheduling, and quantization to MIDI-domain integers — is integer/fixed-point (Q16.16) with a single specified PRNG (xoshiro256**) and a fixed per-tick draw order; the transcendentals in `p_i = 1 − exp(−f_i·Δt)` are evaluated in fixed point only when a control changes, and the per-tick decision is one integer comparison against a precomputed 64-bit threshold. Given a state vector `(master_seed, scaffold_id,`

`scaffold_version, tempo_bpm, key_offset, mode_id, parameter_automation, routing, k, voice_seed_offsets)`, the emitted event stream is **bit-identical on every CPU, browser, and operating system** — a float ban enforced by lint and by a source-scanning test. Second, every stochastic component owns an independent seed substream (`master_seed ⊕ offset`): each cascade layer's switch sequence depends only on `(master_seed, i)`, never on `k` or on any other layer, so adding, removing, or intervening on one layer provably never perturbs another's draw sequence. These two properties make the experiments of §6 unusually clean: a stimulus is a short state vector, any laboratory can regenerate it exactly, and matched counterfactual stimuli are constructible by design rather than by post hoc editing.

The player-facing controls *are* the MSM parameters, through pinned monotone transfers (all ranges live in a versioned configuration file): **Intensity** → `m_hi = 1.05 · exp(1.3376295·x)`, range 1.05–4.0; **Change Rate** → `f_slow = 0.01 · 50^x` Hz, range 0.01–0.5 Hz; **Spread** → `S = 4 · 50^x`, range 4–200; **Layers** → `k = round(3 + 9x)`, range 3–12. Tempo, key, and level are explicitly framed as stage controls, not model parameters. The instrument's direct-manipulation gesture — "conducting" — is **pinning**: the player grabs a layer and clamps its latent state high or low. A pinned layer takes no PRNG draw (its substream freezes and resumes on release, exactly as if deactivated), and every pin/release is recorded as a keyframe `{t, param: "pin.<layer>", value}` in the state vector's automation track, so a conducted performance and its replay are bit-identical. Pinning a slow layer high is the instrument's signature build gesture; in model terms it is a *clamp intervention on a latent volatility component*, which §6 (E5) turns into an experimental method.

5.2 Measurement protocol

All spectral measurements below follow a single published protocol, implemented twice (once in the engine's test suite, once in the pack-validation crate; both copies are pinned by analytic anchors). Operating point: the instrument's documented first-run positions — Intensity 0.8, Spread 0.65, Change Rate 0.59, Layers 1/3 — through the pinned transfers (approximately `m_hi ≈ 3.06`, `S ≈ 51`, `f_slow ≈ 0.1` Hz, `k = 6`), at each scaffold's preferred tempo (72–124 BPM across the corpus; all four packs use a 16-tick bar, 4 ticks per beat). Series: from 8192-tick runs, (a) per-tick onset counts and (b) per-tick summed velocities. Both are **bar-template adjusted**: the bar-phase mean — the scaffold's authored, deliberately periodic metric template — is subtracted, exactly as DFA practice removes a known seasonal trend, so the estimator sees the model's fluctuations rather than the authored grid. (Unadjusted, the template dominates: raw series measure α ≈

0.39–0.61 across the corpus.) Estimator: DFA-1 (Peng et al., 1994) over a dyadic scale ladder of 16–1024 ticks for onsets (the suite's own conversion: 2–124 s of music at 124 BPM, 3.3–213 s at 72 BPM) and 16–256 ticks for velocity; MFDFA width `h(−2) − h(2)` (Kantelhardt et al., 2002) with near-zero-fluctuation windows (below 1e−9) excluded from the negative moment. Aggregation: median over a pinned 8-seed panel (`seed_i = 0x5FEC7A110BEA75D0 ⊕ (i · 0x9E3779B97F4A7C15)`). Analytic anchors pin the estimator itself, not the music: the classic endpoints (a deterministic white-noise series must measure α ∈ [0.4, 0.6] and its cumulative sum α ∈ [1.3, 1.7]; `h(2)` must equal DFA α to within 1e−12) plus an interior known-Hurst battery on exact fractional Gaussian noise (Hosking's method, deterministic innovations) in the protocol's own regime: dense fGn recovers H to ≤ 0.01 at the panel median across H ∈ {0.55, 0.70, 0.85}; an onset-like sparse binary observation of the H = 0.70 latent reads α ≈ 0.632 — a measured attenuation of ≈ 0.07, strictly downward — with the bar-template-removal step contributing ≈ 0.001 on top of that channel and shuffling collapsing α to ≈ 0.5. A companion null-model comparison (`foundry nulls`) rebuilds each measured series' IAAFT and shuffle surrogate families (99 each per seed, from the same bar-adjusted series the gates measure) and stores per-seed one-sided p95 verdicts (`research/data/nulls-v1.csv`).

5.3 The measured landscape

The per-scaffold medians at the default operating point (8 seeds × 8192 ticks, bar-template adjusted), verbatim from the suite:

scaffold	α(onset), scales 16–1024	α(velocity), scales 16–256	width(onset), 16–1024
launch-beat-driven	0.768	0.701	0.126
launch-cinematic	0.770	0.676	0.108
launch-warm-ambient	0.695	0.702	0.054
prototype-beat-forward	0.727	0.690	0.102

Two headline facts. **(i)** With the authored bar template removed, the emitted onset streams sit at DFA α ≈ 0.70–0.78: squarely in the persistent, long-range-correlated 1/f family (white noise α = 0.5; an exactly bar-periodic stream has no exponent at all once adjusted), and this is a property of the *final quantized note stream*, after the scaffold's bounds, masks, and snapping — not of the latent process alone. The claim is null-controlled: every pack's measured α exceeds its shuffled null's per-seed 95th percentile on 8/8 seeds (the shuffle family collapses to α ≈ 0.50 median, p95 ≈ 0.53–0.55), so the exponent lives in temporal order, not in the marginals; and the interior estimator anchors (§5.2) show the sparse observation channel attenuates rather than inflates latent persistence, so these values cannot be estimator

inflation. The velocity streams, which mix in the fast band, measure α ≈ 0.676–0.702 over the mid-scale decade (their long-range tail is weaker, ≈ 0.63 over 16–1024). **(ii)** Three of the four onset streams are multifractal by the surrogate criterion — measured widths 0.102–0.126 exceed the 95th percentile of matched IAAFT nulls on 8/8 seeds (beat-driven 0.126 vs. null p95 ≈ 0.080; cinematic 0.108 vs. ≈ 0.069; prototype 0.102 vs. ≈ 0.080) — structure beyond linear correlation plus marginals. The ambient pack's width (0.054 vs. its null p95 ≈ 0.092) does not exceed its null on any seed: at this length it is consistent with a linear long-memory process with these marginals (the finite-size "apparent multifractality" of Bouchaud, Potters & Meyer, 2000), and we claim no more for it. Indeed we claim strictly less than that sentence suggests: the two-point width statistic has since been measured, on identical IAAFT-null designs, to have near-zero detection power against canonical cascade alternatives (the scattering-moment audits of Fenocosm record v22 and Fenosoma record v3 — its x7 scattering audit), so a null width-vs-IAAFT outcome is a statement about the instrument's power, not about the process — "no verdict at this length and power," never "linear." The same measurement cuts the other way for the three positive packs: a rejection delivered by a near-blind instrument requires a large effect, so those excesses are stronger evidence than the bare thresholds suggest. The width ordering is musically sensible — the beat-driven pack is the most intermittent, the ambient the least — and the one negative fits that ordering: the pack authored to be least intermittent is the one whose multifractality is not statistically separable from its linear null. Moreover the **Spread control is a dose–response knob for long-range correlation**: with the other dials fixed at the default point, median α(onset) falls monotonically as Spread rises on every scaffold — Spread = 0 measures 0.913 / 0.865 / 0.747 / 0.851 against Spread = 1's 0.747 / 0.751 / 0.657 / 0.735 (beat-driven / cinematic / ambient / prototype), a gap ≥ 0.09 everywhere with no per-seed overlap. Physically: widening the span pushes the mid-band (density-lane) rates higher, so the onset stream decorrelates in fewer ticks. This is the manipulation E2 (§6) is built on.

All of these numbers are **enforced, not archival**. The engine's CI suite freezes per-scaffold bands around the measured medians (α(onset) ∈ [0.62, 1.10], α(velocity) ∈ [0.60, 1.05], width ≥ 0.025 — the width floor being a degeneracy guard, per the null comparison above, with the IAAFT excess carrying the multifractality evidence) and asserts the Spread ordering per scaffold; a separate, deliberately looser pack-level invariant — #10 of ten in the pack-validation gate `check_pack` — requires every shipped or third-party pack to keep median α(onset) ≥ 0.60 over 4 pinned seeds × 4096 ticks, a floor that sits above the shuffled null's measured p95 (α ≈ 0.53–0.55) with

margin. A future engine or content change that whitens, trivializes, or over-smooths the output fails CI before any listener hears it. Every figure in this section regenerates from the repository's pinned seeds with one test invocation.

Two **descriptive** measurement surfaces extend this landscape beyond the two gated moments (2026-07-19, `research/analysis/spectrum-ensemble-v1/`), each cut at the published protocol over the whole catalog and each explicitly not a gate. First, the **full MF-DFA spectrum** (`research/data/spectrum-v1.csv`): the generalized-Hurst ladder h(q) over $q \in \{-4 \ldots 4\}$ with an order-destroyed shuffle control curve. Every one of the twelve packs' ladders separates from its shuffle (measured two-point width > shuffle on 12/12; warm-ambient the tightest, consistent with its IAAFT honest-exception status), and the h(2) column reproduces the α(onset) bands above — a cross-validation from an independent estimator entry point. Second, the **ensemble path** the product actually renders (`research/data/ensemble-v1.csv`): measured per voice, in aggregate, and over a consonance-rank series. The aggregate two-point width regularizes below the per-voice widths on 10/12 packs while α transports to the aggregate — the per-voice-vs-superposition finding of Telesca & Lovallo on Bach's Sinfonias, in this instrument. The consonance-rank α spans 0.61–0.95 (catalog median ≈ 0.73), mostly below the consonance-fluctuation literature's composer band (≈ 0.86–0.90); under the metric-comparability caveat (per-tick sounding-set CR vs the literature's note-indexed scored-corpus CR), no claim is drawn from that channel here. The claim-bearing multifractality statistic remains the two-point width's IAAFT excess; these ladders are read against the weaker shuffle control only.

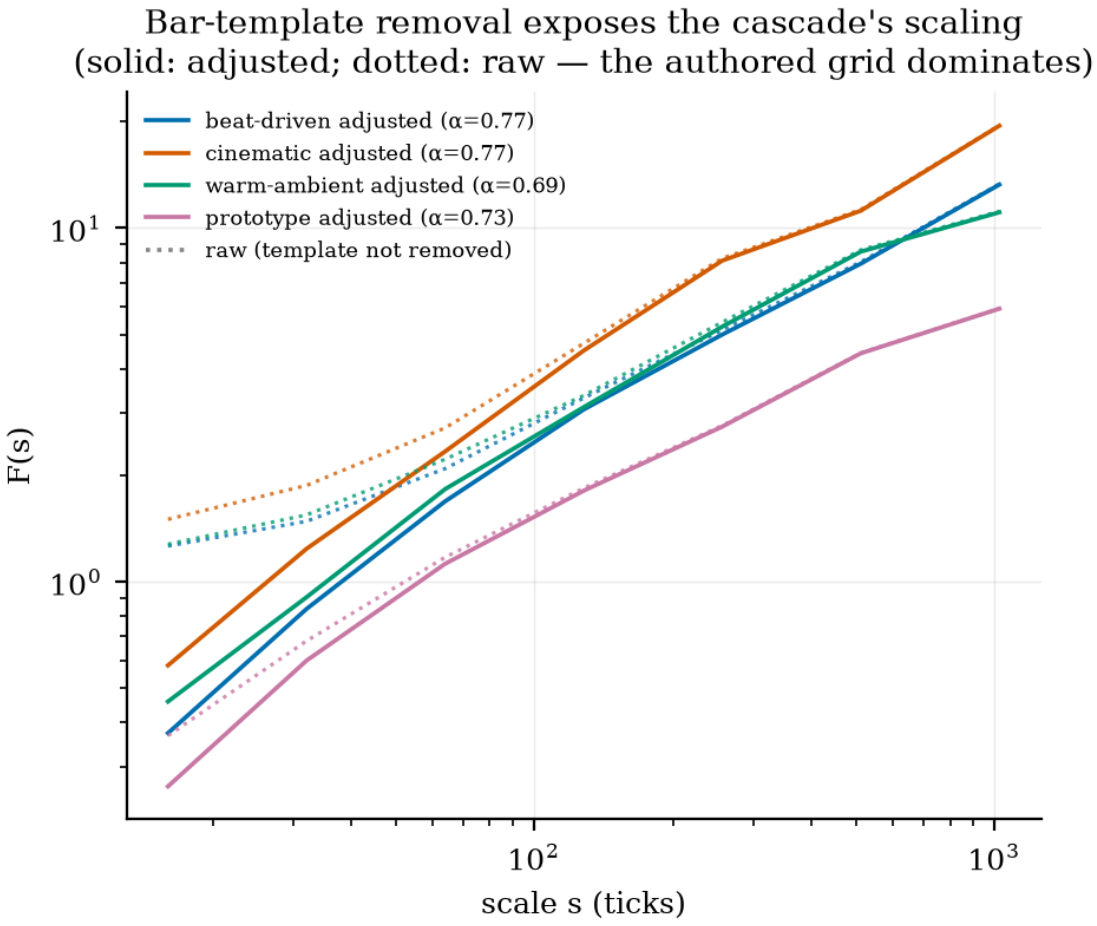

Figure 3. DFA log–log F(s) vs. s (scales 16–1024) for the four scaffolds' onset series at the default operating point, adjusted (solid) vs. unadjusted (dotted), with white-noise and 1/f reference slopes. The labelled α is the CI-gated median. Rendered from `foundry render` (seed 0).

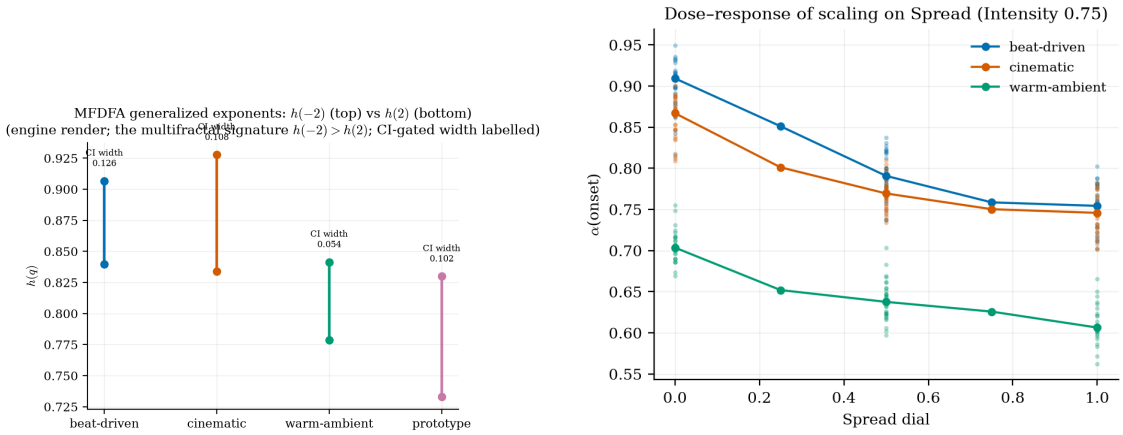

Figure 4. Left: MFDFA generalized exponents h(−2) (top) and h(2) (bottom) per scaffold from a rendered series — the multifractal signature h(−2) > h(2); the CI-gated width is labelled. Right: median α(onset) vs. Spread with per-seed points (Intensity 0.75) — the dose–response ordering.

6. The research programme: six experiments

The instrument turns several long-standing observational questions into intervention questions. Common methods: all stimuli are generated from pinned state vectors (deterministic seeds; stimulus identifiers are the vectors themselves, published with the data); audio is rendered offline through the engine's pinned render path so that stimulus audio is identical across sites within a published tolerance; where two stimuli must differ *only* in a designated cause, we use the counterfactual-pin method — because pinned layers take no draws and all other substreams are independent by construction, a pin-high/pin-low pair is bit-exactly matched everywhere except the intervened layer. Sample sizes below are indicative minimums for the stated effect thresholds; full power analyses belong to the individual preregistrations. (The product's own release process already runs a versioned human listening panel — median ratings over ≥ 30 random patches, ≥ 8 listeners — so the recruitment and blinding infrastructure exists; the experiments here are separate, hypothesis-driven studies.)

E1 — Beyond-the-spectrum discrimination

Design. Do listeners perceive multifractal structure *at all*, beyond the power spectrum? Take the per-tick onset-count and velocity series of a generated passage; construct IAAFT surrogates (Schreiber & Schmitz, 1996; Theiler et al., 1992) that preserve the amplitude distribution and power spectrum while destroying the higher-order (multiplicative) dependence that carries multifractality; re-quantize the surrogate series to valid event streams through the same scaffold machinery (identical template, pools, and voices — the surrogate constraint is matched marginal and spectrum within quantization tolerance, which we verify by re-measuring both); render both through identical synthesis. Two-alternative forced choice ("which is the original?" after familiarization, and separately "which is more musical?"), ≈ 40 listeners, passages of 30–120 s across the Spread range (longer passages give the slow scales room to matter). **Prediction.** Above-chance discrimination, increasing with passage length and with the pack's measured width; the surrogate's measured MFDFA width collapses toward the linear-Gaussian baseline while α is unchanged, so any reliable discrimination is attributable to structure beyond the spectrum. **A negative result would mean** spectrum-matched surrogates are perceptually sufficient at these widths (0.054–0.126): multifractality in this range is measurable but not audible. The instrument would stand — its 1/f claim and its CI gates are spectrum-level — but the specifically *multifractal* perceptual claim would be withdrawn, and E2's width axis would be expected flat. **Sharpening from the statistical nulls (§5.3).** The estimator-side comparison already run

gives E1 a measured zero point: the ambient pack's width does not exceed its IAAFT null, so its structure is spectrum-level by measurement — listeners should be at chance there by construction, making it the natural built-in control pack; conversely, any reliable discrimination on the three null-exceeding packs is attributable to exactly the structure the nulls isolate.

E2 — The (α, width) preference surface

Design. Extend Voss–Clarke from a line to a plane. Generate a Spread × Intensity grid (e.g. 5 × 4 positions, all else at the default point) per scaffold; each cell's stimuli are measured post hoc on the (α(onset), width) plane using the §5.2 protocol, so the design variables are controls but the analysis variables are measured statistics. Participants give pairwise preferences or scale ratings over 60–90 s excerpts; fit a preference surface over (α, width) with scaffold as a random effect. **Prediction.** A preference peak *interior* to the width axis — an inverted-U, echoing the Voss–Clarke finding that 1/f sits between "boring" (over-correlated) and "surprising" (white) — with the α ridge in the persistent band our packs occupy (0.65–0.95) rather than at either extreme. Secondarily, the peak location should shift with genre (scaffold), since the packs' measured widths already order sensibly by style. **A negative result** (flat or corner-maximal surface) would show preference in this domain is governed by the exponent alone, or by neither — directly constraining the "1/f preference" literature with stimuli whose scaling properties are knob-controlled rather than corpus-sampled. Note E2 is informative even if E1 is null: preference and discrimination can dissociate.

E3 — Tension-lane validation

Design. The architecture makes a falsifiable psychological claim by construction: the slow band *is* the tension lane — it biases harmonic transitions (§4.3) and is exported as CC 1. Test it with the continuous-response paradigm of the tonal-tension literature (Krumhansl, 1996; Lerdahl & Krumhansl, 2007): listeners move a slider tracking felt tension while 2–4-minute generated passages play; the ground-truth latent series — the slow band's net state `s_slow(t)`, and each lane — is exported bit-exactly from the same run. Analyze lagged cross-correlations between ratings and `s_slow(t)`, with the `dynamics` lane (fast band) and the raw onset-count series as within-stimulus controls; include passages where the slow band moves while density is held statistically flat (selectable by seed screening, or enforced with mid-band pins). **Prediction.** Tension ratings track the slow-band state above both controls (pre-specified threshold: median optimal-lag correlation exceeding the dynamics-lane control by $\Delta r \geq 0.1$), with the positive lag characteristic of continuous ratings. **A negative result** would not break the instrument but would falsify its semantics: if ratings track only

loudness/density, the “tension” label and the CC 1 mapping misdescribe what the slow band contributes, and the harmony-bias design (the `score(τ, T)` coupling) needs strengthening or renaming. This is the cheapest experiment and the one the architecture most directly exposes.

E4 — Boundary perception with ground-truth latents

Design. Musical segmentation studies must infer structural boundaries from annotations; here the generative regime shifts are *known*. Listeners press a key at perceived section boundaries during 3–5-minute passages; the engine’s per-tick state log gives every layer’s flip times exactly. Score responses against flips per band (slow/mid/fast) by signal-detection analysis with a tolerance window (d' per band, guessing-corrected), across `k` and Change Rate settings. **Prediction.** A d' hierarchy ordered by band: slow-layer flips attract boundary marks ($d' \geq 1$ at the default point), mid flips weakly, fast flips at floor ($d' \approx 0$, confidence interval covering 0). Effectively this measures the perceptual transfer function from latent switching scale to experienced form — which layers are “form”, which are “texture”. **A negative result** (no hierarchy, or boundaries driven purely by surface features like register or instrumentation with no latent alignment) would mean the cascade’s scale hierarchy does not project into experienced musical form at these parameter values — challenging the instrument’s teaching claim that watching slow layers is watching structure, and pointing to lane-mapping gain as the lever to revisit.

E5 — Counterfactual pairs via pins

Design. The pin gesture is a clean intervention: pin layer `i` high at bar `B` in condition A, low in condition B, all seeds identical — the two event streams are bit-identical up to `B` and differ afterwards only through the intervened layer (both pinned substreams freeze, all other substreams are untouched by construction). Render A/B/X triads and measure regime-shift audibility as a function of band and of the probe’s temporal distance from `B` (probe windows excerpted 0–120 s after the intervention). **Prediction.** Audibility persists at long horizons for slow-layer pins (above-chance ABX at ≥ 60 s after the intervention, since a slow layer holds its state for minutes at the default `f_slow ≈ 0.1` Hz and biases harmony throughout) but decays within seconds for fast-layer pins (chance within ≈ 4 s, a few switch periods of a layer near 5 Hz) — a psychophysical measurement of each band’s *perceptual* time constant against its known generative rate `f_i`. **A negative result** for slow layers — inaudible interventions — would mean the instrument’s signature gesture is theatrically satisfying but musically inert, i.e. the slow band’s routing gain is too subtle; that is directly actionable (and re-testable, since the whole design is two automation keyframes). More broadly E5 calibrates

every other experiment: it tells us which latent scales are within perceptual reach at all.

E6 — Corpus inversion and a statistical Turing test

Design. Close the loop with real music. Extract per-tick onset-count and velocity series (grid-quantized, bar-template adjusted, exactly as §5.2) from symbolic corpora — e.g. the composer corpora of Levitin et al. (2012) and performance sets such as MAESTRO (Hawthorne et al., 2019) — and fit the MSM by moment/spectral estimation in the econometric tradition (GMM on log-moment and autocovariance conditions in the style of Lux, 2008, or a Whittle-type spectral objective; toggle-vs-fresh-draw rescaling per §4.1), modeling counts as conditionally Poisson in the band product and velocity residuals as MSM-modulated; select `k` by information criterion; obtain per-composer/per-genre maps in `(m_hi, f_slow, S, k)` — the instrument's own control space. The estimator prescription is a committed methodological finding, not a taste — and the first external-corpus run (2026-07-12; `research/analysis/msm-corpus-fit/out/RESULTS-corpus-v1.md`) sharpened it into a **channel-conditional** one. The companion corpus-fitting analysis implemented the textbook alternative — exact maximum likelihood via the `2^k`-state hidden-Markov filter (Calvet & Fisher, 2004) — which passes a synthetic parameter-recovery gate (panel-median errors within ~10% at 10^4 ticks; a positive control on true MSM-Cox data leaves the discriminating classifier at chance). Fitted to the instrument's own onset streams, the per-tick likelihood under-weights the long memory: the MLE flattens the cascade to `m_hi^κ ≈ 1.0`, its resynthesis drops α from ≈ 0.80 to ≈ 0.54, and a held-out classifier separates resynthesis from original at 0.987 accuracy. Fitted to **real music**, the same estimator does not collapse: on 1,260 MAESTRO performances (beat-time grid, bar covariate) the fitted cascade is genuinely non-flat (`m_hi^κ` median 1.619, 92.6% of pieces above 1.2) and the resynthesis retains the persistence (α 0.691 vs. real 0.717). The collapse is therefore a property of the engine's near-Poisson emission marginals masking their own long memory, where real piano counts are overdispersed enough (Fano ≈ 1.55) for the likelihood to reward the structured cascade — the channel decides (the companion channel paper's thesis, paper 04). The registration is unchanged in substance: the ML filter's behavior is channel-dependent, its `κ·log m_hi` and rate-ladder ridges remain per-piece ill-conditioned, and its resynthesis still fails the statistical Turing test on both corpora (held-out classifier 0.758 piano / 0.932 drums against ~0.50 permutation nulls) — so GMM/spectral estimation remains the estimator of record, with the ML filter retained as the diagnostic and positive-control machinery it has now twice proven to be. Then resynthesize: run the fitted parameters through a matched scaffold and test both

statistically (do held-out corpus statistics — α, width, burst-length distributions — fall inside the resynthesis ensemble?) and perceptually (discrimination of resynthesis vs. spectrum-matched Poisson and shuffled controls). **A competitive Turing test against neural generators.** The sufficiency test above asks whether *our* fitted model can be told from real music; a stronger framing, now available externally, asks how the Fenophone compares to the current state of the art in generative music. Zhang, Sun & Xiong (2025) ran essentially this paper's descriptor battery — DFA α, singularity-spectrum width, skewness, the generalized-Hurst ladder h(q), and the Jensen–Shannon divergence between f(α) spectra — over 381 Billboard hits (1950–2024) against matched outputs of three state-of-the-art neural systems (Suno, DiffRhythm, YuE), and found the generators systematically miss the human multifractal signature (reduced fine-scale variability; narrower or more erratic spectra; the gap widest in the most complex modern eras), proposing fractal-aware training objectives as the remedy. This reframes E6's closing loop as a *competitive* statistical Turing test: score the Fenophone's emitted (and MSM-resynthesized) streams on the same battery, alongside those neural systems and against the human reference. The instrument's prediction is asymmetric — because its generative core *is* a multifractal process whose α and width are CI-gated (§5), it should sit inside the human band on precisely the descriptors the learned systems miss, and do so by construction rather than through the post-hoc fractal-aware loss those authors call for. One caveat is domain: Zhang et al. measure audio amplitude envelopes over a wide moment range ($q \in [-10, 10]$), whereas our claim-bearing statistic is the symbolic onset/velocity stream at $q \in \{-2, 2\}$; a clean comparison must either render the Fenophone through its pinned audio path to matched envelopes or reduce their audio to onset streams, on a shared q-grid. Their code and generated corpus are public, so the neural baselines are reusable rather than re-generated. **Prediction.** Fitted parameters separate composers/genres (a classifier on fitted parameters identifies composer above chance, extending Levitin et al.'s composer-specific β to a multi-parameter fingerprint), and under the moment/spectral estimator MSM resynthesis is measurably harder to reject than the Poisson control on both tests — noting that the statistical half is partly built in for the moments the estimator matches, so the width, burst-length, and perceptual tests carry the evidential weight. **A negative result** is the most theoretically informative in the programme: systematic misfit — e.g. real music demanding a continuum of scales where MSM has k discrete ones, or non-stationarity beyond regime switching — would locate exactly where musical intensity *departs* from the volatility analogy, and would bound the instrument's claim to "a good generator" rather than "a good model".

7. Predictions and falsifiability

For concreteness, the programme's registered predictions, stated as decision rules:

1. **(E1)** Group-level 2AFC discrimination of originals from IAAFT surrogates exceeds chance (pre-specified: > 55% correct, $p < .05$) for at least the two widest packs (widths 0.126, 0.108) at passage lengths ≥ 60 s.
2. **(E1)** Discrimination accuracy increases with the pack's measured MFDFA width across the corpus (widths 0.054–0.126); ordering violated → the width measurement is not the perceptually operative variable.
3. **(E2)** The preference surface has an interior maximum along the width axis: the mid-Spread cell is preferred over both Spread endpoints at fixed Intensity 0.8, per scaffold.
4. **(E3)** Continuous tension ratings correlate with the slow-band net state above the dynamics-lane control by $\Delta r \geq 0.1$ at the optimal lag, per passage median.
5. **(E4)** Boundary d′ is ordered slow > mid > fast, with d′(slow) ≥ 1.0 and d′(fast) statistically indistinguishable from 0 at the default point.
6. **(E5)** ABX audibility of slow-layer pin pairs remains above chance at probe distances ≥ 60 s; fast-layer pin pairs fall to chance within 4 s.
7. **(E6)** A classifier over fitted `(m_hi, f_slow, S, k)` identifies composer above chance on held-out pieces; MSM resynthesis under the prescribed moment/spectral estimator is rejected less often than a spectrum-matched Poisson control on both statistical and perceptual tests. As a competitive corollary, the Fenophone's own emitted streams are predicted to fall within the human multifractal band on the DFA/MFDFA descriptor battery of Zhang et al. (2025), where current neural generators (Suno, DiffRhythm, YuE) measurably do not. *Status after the first corpus run (2026-07-12):* the composer half has its first confirmation — even under plain ML fitting, held-out one-vs-rest classification over the fitted parameters identifies the top-8 MAESTRO composers at 0.276 against a 0.190 permutation chance (0.374 jointly with the spectral pair (α, width), which alone reaches 0.337 — the multi-parameter fingerprint adds to, but does not yet dominate, the exponents). The resynthesis half remains open: ML resynthesis fails the statistical Turing test on both corpora (§6, E6), which is why the moment/spectral estimator remains part of the registration.
8. **(cross-cutting)** Every stimulus statistic re-measured by an independent laboratory from the published state vectors matches ours exactly for

event-level quantities (bit-identical streams) and within the published tolerance for audio.

Equally important is what does *not* hang on these predictions. The following product-level claims are established by construction and by CI measurement, independent of every perceptual outcome above: the instrument is deterministic and bit-exactly replayable; its controls are genuine MSM parameters with monotone perceived-trend gates; its emitted streams are long-range correlated ($\alpha \approx 0.70$–0.78, template-adjusted, every seed beyond its shuffled null) and — for three of the four measured packs — multifractal beyond matched IAAFT nulls, with the Spread dose–response gap ≥ 0.09; and these properties are regression-gated for every future pack. If the perceptual programme fails wholesale, the Fenophone remains a well-characterized instrument with an unusually clean measurement story — but the claims that multifractality is *heard*, that the slow band is *felt as tension*, and that latent regime flips are *experienced as form* would fall, and this paper's title thesis would reduce from "generative model of musical intensity" to "generative mechanism for 1/f musical texture". We consider that a meaningful, checkable difference, which is the point of stating it now.

8. Limitations

Estimator finite-size bias. MFDFA on 8192-point series is known to produce nonzero widths for linear long-memory processes (finite-size and distributional effects; Bouchaud, Potters & Meyer, 2000; Kantelhardt et al., 2002). The statistical half of that concern is now addressed directly rather than deferred: the committed null comparison (§5.2, §5.3) shows the shuffled null alone measures widths of 0.04–0.09 on these series — so no fixed width threshold certifies nonlinearity — and the per-pack IAAFT verdicts are what we claim: three packs' widths exceed their nulls on every seed; the ambient pack's does not, and we explicitly do not claim measured nonlinear multifractality for it. A power caveat now attaches to every width verdict above: the width statistic's measured near-zero power against cascade alternatives (Fenocosm record v22 / Fenosoma record v3 — its x7 scattering audit) means the ambient pack's null outcome is a no-verdict, and the three positive excesses are underclaims. What remains for E1 is the *perceptual* question — whether the structure that survives the nulls is audible — plus the sensitivity of the negative-moment exclusion of near-zero-fluctuation windows (silent stretches), which we have not fully mapped.

DFA on count series. Our onset series are non-negative, integer-valued, and zero-inflated; DFA's behavior on such series (and under residual trends

and nonstationarities) has documented caveats (Hu et al., 2001; Bryce & Sprague, 2012). We mitigate with analytic anchors, a template-adjustment no-op check on aperiodic input, dyadic scale ladders averaging ≥ 8 windows at the largest scale, and — closing the loop the caveats open — an interior known-Hurst battery in exactly this regime (§5.2): dense fGn recovers H to ≤ 0.01; a sparse binary observation at the packs' fill reads latent H = 0.70 as α ≈ 0.632 (attenuation ≈ 0.07, strictly downward, with template removal contributing ≈ 0.001). The exponents should therefore be compared across conditions of this pipeline, not treated as absolute latent readouts — and where they err, they err conservative. Relatedly, α ≈ 0.70–0.77 corresponds to a spectral exponent β = 2α − 1 ≈ 0.4–0.5 — flatter than the β ≈ 1 often reported for musical rhythm spectra — and our bar-template adjustment removes metric-scale power that corpus studies retain, so direct numerical comparison to that literature is not licensed.

The preference literature. As stated in §2.1, the "1/f is preferred" claim is motivation, not foundation; E1/E2 are designed to stand or fall on their own data.

Corpus and cultural scope. The measured landscape covers four scaffolds — one authored pool, Western tonal harmony, 4/4 meter, a 16-tick bar grid, tempi 72–124 BPM. The bar-template-as-seasonality move presumes an isochronous, cyclic meter; the grid presumes quantized onsets. Non-Western meters, unmetered music, and expressive (off-grid) timing are outside the current observation model, and every perceptual result will initially be a result about this pool and (in expectation) Western-enculturated listeners.

Model scope. Pitch is not MSM-driven (§4.5); onset *placement* within a bar is deterministic given the lane; the harmonic graph is authored. The MSM claim is about intensity, and the paper should not be read as claiming more.

9. Related work

Stochastic and Markovian composition is old and rich — Hiller and Isaacson's Illiac Suite, Xenakis's stochastic formalisms (Xenakis, 1971), and the Markov-chain tradition surveyed by Ames (1989) — but these use Markov structure over *symbols* (notes, chords), whereas here the Markov chain lives in a latent volatility hierarchy and the symbols come from a separate authored layer. The 1/f-generation thread (Voss & Clarke, 1978; fractal-music constructions, e.g. Hsu & Hsu, 1990; the power-law aesthetics program of Manaris et al., 2005) specifies target spectra or target statistics; MSM instead supplies a *mechanism* whose spectrum, multifractality, and

regime structure follow from interpretable parameters, and which supports filtering, estimation, and intervention. Long-range-correlated humanization of microtiming (Hennig et al., 2011; Räsänen et al., 2015) modulates timing residuals of otherwise fixed music; we modulate the generative intensity itself. Methodologically the closest family is model-based sonification (Hermann & Ritter, 1999), where sound is produced by a model's dynamics rather than by direct parameter mapping; we share the model-first commitment but sonify no data at all — the model runs forward as a free generator with its observation equation discarded. On the interaction side, the clamping gesture relates to the emerging NIME thread on performing learned latent spaces (Scurto & Postel, 2023; Shepardson & Magnusson, 2023); the substrate differs in kind — a finite, enumerable econometric regime vector with per-component meaning, rather than a neural embedding — which is what makes exact, glitch-free clamp interventions and counterfactual stimulus pairs (E5) possible. On the analysis side we inherit DFA/MFDFA methodology from the physiological and econophysics literatures (Peng et al., 1994; Kantelhardt et al., 2002) and the multifractal-music measurements of Su and Wu (2006) and Jafari et al. (2007). Within econometrics, our construction is the Calvet–Fisher binomial MSM (2001, 2004, 2008) with a log-symmetric multiplier and a span-parameterized rate ladder (§4.1); the multifractal random walk (Bacry, Delour & Muzy, 2001) is the principal alternative stationary multifractal generator, and adapting it to the same instrument interface would be a natural comparison. We are not aware of prior work using MSM — or any estimable econometric volatility model — as the generative core of a musical instrument; the closest conceptual relatives are doubly stochastic point-process models of event streams (Cox, 1955), of which our observation model is a musically constrained instance.

10. Conclusion

The Markov-Switching Multifractal was built to model a quantity — volatility — whose structure is bursts within bursts, and it was built causal, stationary, and Markov because the field needed to run it forward and to invert it. Musical intensity has the same structure and imposes the same requirements, and the same construction serves. We have taken the mapping literally: an instrument whose controls are `(m_hi, f_slow, S, k)`, whose conducting gesture clamps latent states, whose emitted note streams measurably carry long-range correlation (α(onset) medians 0.695–0.770, template-adjusted, null-controlled) and — beyond matched IAAFT nulls on

three of four packs — multifractal width, with a knob-controlled dose–response on α; all bit-reproducible from a short state vector and enforced in CI. What remains is the science: whether the structure is heard beyond the spectrum, whether the latent hierarchy projects into felt tension and experienced form, where preference lives on the (α, width) plane, and whether real corpora, inverted through the model's own estimation theory, land in distinct and meaningful regions of its parameter space. The six experiments above are specified to decide these questions, and the instrument was engineered so that any laboratory can run them from the same bits. We intend this paper as the fixed point the programme can be checked against — including the parts of it that may fail.

Reproducibility note

Every measured number in §5 is regenerated by the repository's pinned test suites (the engine spectral suite and pack-validation invariant #10, entry point `check_pack`) from the published seed panel; the measurement reporter that derived the landscape table is shipped alongside the gates. State vectors for all figures will be published with the paper.

References

All citations below were verified against the published record during the pre-submission citation pass; the machine-readable bibliography with DOIs is `references.bib` (shared across papers 01–04).

1. Ames, C. (1989). The Markov process as a compositional model: a survey and tutorial. *Leonardo*, 22(2).
2. Bacry, E., Delour, J., & Muzy, J.-F. (2001). Multifractal random walk. *Physical Review E*, 64, 026103.
3. Bouchaud, J.-P., Potters, M., & Meyer, M. (2000). Apparent multifractality in financial time series. *European Physical Journal B*, 13.
4. Bryce, R. M., & Sprague, K. B. (2012). Revisiting detrended fluctuation analysis. *Scientific Reports*, 2, 315.
5. Calvet, L. E., & Fisher, A. J. (2001). Forecasting multifractal volatility. *Journal of Econometrics*, 105.
6. Calvet, L. E., & Fisher, A. J. (2002). Multifractality in asset returns: theory and evidence. *Review of Economics and Statistics*, 84.
7. Calvet, L. E., & Fisher, A. J. (2004). How to forecast long-run volatility: regime switching and the estimation of multifractal processes. *Journal of Financial Econometrics*, 2.
8. Calvet, L. E., & Fisher, A. J. (2008). *Multifractal Volatility: Theory, Forecasting, and Pricing*. Academic Press.

9. Cox, D. R. (1955). Some statistical methods connected with series of events. *Journal of the Royal Statistical Society, Series B*, 17.

10. Hawthorne, C., et al. (2019). Enabling factorized piano music modeling and generation with the MAESTRO dataset. *International Conference on Learning Representations (ICLR)*.

11. Hennig, H., Fleischmann, R., Fredebohm, A., Hagmayer, Y., Nagler, J., Witt, A., Theis, F. J., & Geisel, T. (2011). The nature and perception of fluctuations in human musical rhythms. *PLoS ONE*, 6(10), e26457.

12. Hermann, T., & Ritter, H. (1999). Listen to your data: model-based sonification for data analysis. In G. E. Lasker (Ed.), *Advances in Intelligent Computing and Multimedia Systems* (pp. 189–194). IIAS.

13. Hiller, L., & Isaacson, L. (1959). *Experimental Music: Composition with an Electronic Computer*. McGraw-Hill.

14. Hsu, K. J., & Hsu, A. J. (1990). Fractal geometry of music. *Proceedings of the National Academy of Sciences*, 87.

15. Hu, K., Ivanov, P. Ch., Chen, Z., Carpena, P., & Stanley, H. E. (2001). Effect of trends on detrended fluctuation analysis. *Physical Review E*, 64, 011114.

16. Jafari, G. R., Pedram, P., & Hedayatifar, L. (2007). Long-range correlation and multifractality in Bach's Inventions pitches and lengths. *Journal of Statistical Mechanics: Theory and Experiment*.

17. Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Havlin, S., Bunde, A., & Stanley, H. E. (2002). Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A*, 316.

18. Krumhansl, C. L. (1996). A perceptual analysis of Mozart's Piano Sonata K. 282: segmentation, tension, and musical ideas. *Music Perception*, 13.

19. Lerdahl, F., & Krumhansl, C. L. (2007). Modeling tonal tension. *Music Perception*, 24.

20. Levitin, D. J., Chordia, P., & Menon, V. (2012). Musical rhythm spectra from Bach to Joplin obey a 1/f power law. *Proceedings of the National Academy of Sciences*, 109(10).

21. Lux, T. (2008). The Markov-switching multifractal model of asset returns: GMM estimation and linear forecasting of volatility. *Journal of Business & Economic Statistics*, 26.

22. Manaris, B., Romero, J., Machado, P., Krehbiel, D., Hirzel, T., Pharr, W., & Davis, R. B. (2005). Zipf's law, music classification, and aesthetics. *Computer Music Journal*, 29(1), 55–69.

23. Mandelbrot, B. B. (1974). Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier. *Journal of Fluid Mechanics*, 62.

24. Mandelbrot, B. B., Fisher, A. J., & Calvet, L. E. (1997). A multifractal model of asset returns. *Cowles Foundation Discussion Paper No. 1164*.

25. Peng, C.-K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., & Goldberger, A. L. (1994). Mosaic organization of DNA nucleotides. *Physical Review E*, 49.

26. Räsänen, E., Pulkkinen, O., Virtanen, T., Zollner, M., & Hennig, H. (2015). Fluctuations of hi-hat timing and dynamics in a virtuoso drum track of a popular music recording. *PLoS ONE*, 10(6).

27. Roeske, T. C., Kelty-Stephen, D., & Wallot, S. (2018). Multifractal analysis reveals music-like dynamic structure in songbird rhythms. *Scientific Reports*, 8, 4570.

28. Schreiber, T., & Schmitz, A. (1996). Improved surrogate data for nonlinearity tests. *Physical Review Letters*, 77.

29. Scurto, H., & Postel, L. (2023). Soundwalking deep latent spaces. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*.
30. Shepardson, V., & Magnusson, T. (2023). The Living Looper: rethinking the musical loop as a machine action-perception loop. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*.
31. Su, Z.-Y., & Wu, T. (2006). Multifractal analyses of music sequences. *Physica D*, 221.
32. Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., & Farmer, J. D. (1992). Testing for nonlinearity in time series: the method of surrogate data. *Physica D*, 58.
33. Voss, R. F., & Clarke, J. (1975). "1/f noise" in music and speech. *Nature*, 258.
34. Voss, R. F., & Clarke, J. (1978). "1/f noise" in music: music from 1/f noise. *Journal of the Acoustical Society of America*, 63.
35. Xenakis, I. (1971). *Formalized Music: Thought and Mathematics in Composition*. Indiana University Press.
36. Zhang, K. K., Sun, Y., & Xiong, H. (2025). Multifractal comparison of Billboard and AI-generated music. In *Proceedings of the 33rd ACM International Conference on Multimedia (MM '25)*, 12419–12427.

Companion records of the Feno program (DOIs published 2026-08-31)

*Every record below is deposited on Zenodo — staged as a private draft with a **reserved** DOI 2026-08-30, **published 2026-08-31**: every DOI below is live and resolves. The deposit map is `research/zenodo/records.json`, and the source repositories are private, so each record carries its own reproduction materials. Cross-citations follow that map's `related_identifiers`.*

- Atlas, E. T. 2026. "The Fenophone: A Musical Instrument Built on the Markov-Switching Multifractal, with CI-Verified Spectral Invariants" (Fenophone research series, paper 01). Zenodo DOI 10.5281/zenodo.22180423 (published 2026-08-31).
- Atlas, E. T. 2026. "Theorems as Instrument Design: Formal Guarantees and Statistical Contracts in a Generative Musical Instrument" (Fenophone research series, paper 02). Zenodo DOI 10.5281/zenodo.22180427 (published 2026-08-31).
- Atlas, E. T. 2026. "From Latent Cascade to Note Stream: How Scaling Exponents Survive Musical Quantization" (Fenophone research series, paper 04). Zenodo DOI 10.5281/zenodo.22180431 (published 2026-08-31).
- Atlas, E. T. 2026. "Does the Markov-Switching Multifractal Describe Real Music? A Fitting Pipeline and a Statistical Turing Test" (companion analysis paper, `analysis/msm-corpus-fit/PAPER.md`). Zenodo DOI 10.5281/zenodo.22180435 (published 2026-08-31).