The Fenophone: A Musical Instrument Built on the Markov-Switching Multifractal, with CI-Verified Spectral Invariants

Evan Tabak Atlas
Independent · ORCID 0009-0007-7374-2338
Evanatlas@gmail.com

Preprint draft. Target venue: NIME (alternates: ICMC, SMC).

ABSTRACT

We present the Fenophone, a browser-based generative music instrument whose intensity core is, by construction, the Markov-switching multifractal (MSM) of Calvet and Fisher: a product of k ∈ [3, 12] two-state switching multipliers whose Hz-defined switch rates are spaced geometrically across a player-controlled span. Every primary control is a genuine parameter of that model — Intensity sets the multiplier magnitude m_hi, Change Rate the slowest switch rate f_slow, Spread the time-scale span S, Layers the multiplier count k — and a fixed band map routes the cascade's slow, mid, and fast bands to harmonic tension, rhythmic density, and dynamics. Pitch comes from per-voice stochastic contour walks with provably bounded register drift, sampled at cascade-determined onsets. The event layer is integer/fixed-point and bit-identical across platforms, which makes the instrument learnable, performances exactly reproducible, and — unusually — statistical properties of the output enforceable in continuous integration: a DFA/MFDFA suite measures every shipped scaffold's emitted note stream (per-scaffold onset-stream DFA α medians 0.695–0.770, every seed beating its shuffled-surrogate null; multifractal widths 0.054–0.126, exceeding matched IAAFT surrogate nulls on every seed for three of the four measured packs, with the fourth reported as consistent with linear long memory at this length — a no-verdict at the width statistic's measured power, not evidence of linearity) and CI rejects any content or engine change whose median α falls below a published floor of 0.60 after bar-template removal. We describe the model, the instrument built around it, the playability theorems and mechanized gates, and the measured spectral landscape, and we reflect on what a strictly mathematical instrument buys and what it costs.

1. Introduction

Generative instruments usually stand in a loose relationship to the mathematics that inspires them. A composer reads about 1/f noise or chaotic maps, extracts a flavor — long memory, self-similarity, sensitive dependence — and builds a mapping in which the mathematics is a source of numbers and the musicality lives elsewhere. The result can be excellent music, but

the mathematical claim attached to it (“fractal music”, “1/f melody”) is typically neither precise about the generator nor verified on the output.

The Fenophone is an experiment in taking the opposite stance. Its design goal, stated before any code was written, is that *the controls and the outputs are the mathematics, not inspired by it*. Operationally this means three commitments, which are also the three claims this paper defends:

1. **A strict model core.** The process that decides when notes happen, how forceful they are, how dense the texture is, and where the harmony leans is exactly a discrete-time Markov-switching multifractal — the volatility model of Calvet and Fisher [6, 7, 8], descended from Mandelbrot’s multiplicative cascades [4] via the Multifractal Model of Asset Returns [5] — with two deliberate, named departures described in Section 3.4. The instrument’s primary controls (Intensity, Spread, Change Rate, Layers) are the model’s own parameters (m_hi, S, f_slow, k), reached through monotonic transfer functions; nothing about the core is cosmetic.
2. **Playability guaranteed by theorems and gates, not folklore.** Multiplicative processes burst and cluster; unconstrained random walks drift off the playable register within minutes. Rather than tuning these hazards away by ear and hoping, the Fenophone makes each one the subject of either a small theorem (a Jury stability analysis of the pitch walk, with a closed form for its stationary deviation; Section 5.2) or a mechanized release gate (a “two-minute drift” test, monotonic-trend controllability sweeps, rhythmic-coherence bounds; Sections 5.2–5.3). A control whose perceived effect is not monotonic, or a voice that can leave its register, fails the build.
3. **Multi-scale structure measured, never assumed.** The point of building on a multifractal is structure at every time scale. The Fenophone does not assert this property; it measures it, on the actual emitted note stream of every shipped content pack, with detrended fluctuation analysis (DFA) and its multifractal extension (MFDFA), and enforces published bounds in continuous integration (CI). A future engine or content change that whitens the output ($\alpha \to 0.5$), collapses it to a bar-periodic loop, or flattens its multifractal width fails CI before any listener hears it (Sections 5.4, 6).

The substrate that makes all three claims practical is determinism. The Fenophone’s event layer — everything that decides which notes happen, with what pitch, timing, velocity, and control data — is computed in integer/fixed-point arithmetic with a single specified pseudorandom generator and a fixed draw order, and is therefore bit-identical on every CPU, browser, and operating system. The same gesture from the same state always produces

the same result, which is what makes the instrument learnable rather than a slot machine; a captured performance is a small state vector that replays exactly; and, methodologically, statistical CI gates cannot flake (Section 5.1).

The engine is written in Rust and compiled to WebAssembly; the instrument runs in the browser (and as a plugin sharing the same core), with a four-voice ensemble, a routing matrix, and a direct-manipulation surface in which the visualization of the cascade *is* the control surface, including a "pin" gesture that holds a chosen multiplier in a chosen state — a musical intervention that falls directly out of the model's latent state (Section 4.4). The full invariant suite and the measurement code behind every number in this paper are published under the research tree's CC BY 4.0 license, whose engine reproduction grant covers the measurement crates (core, contracts, and the stimulus foundry) archived alongside this paper's deposit; all measurements are `cargo test` targets.

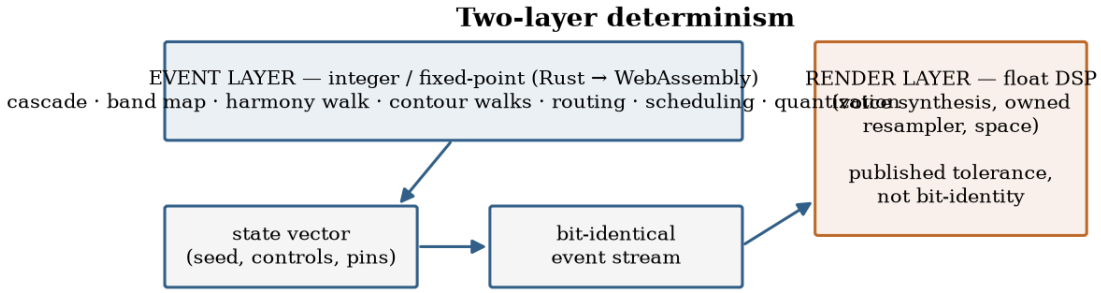

Figure 1. System overview: the two-layer determinism architecture. Left, the integer/fixed-point event layer (Rust → WebAssembly) — cascade, band map, harmony walk, per-voice contour walks, routing, scheduling, quantization — emitting a bit-identical event stream from a state vector. Right, the floating-point render layer (voice DSP, owned resampler), held to a published tolerance.

2. Related Work

1/f noise and music. Voss and Clarke measured 1/f spectra in music and speech [1] and then inverted the observation, generating melodies from 1/f noise and reporting that listeners preferred them over white- and brown-noise controls [2]; Gardner's column made the idea widely known [3]. A line of fractal and power-law composition followed: Bolognesi's self-similar dynamics [31], Dodge's explicitly self-similar pieces [23], Pressing's

nonlinear-map generators [32], Díaz-Jerez's FractMus toolbox of deterministic fractal algorithms [33], analyses claiming fractal geometry in existing repertoire [12], and Manaris et al.'s Zipf/power-law metrics as aesthetic classifiers [34]. This tradition established the target statistics but generally used fractal noise, chaotic maps, or deterministic self-similar procedures as a *source* to be mapped, rather than as a stationary, parametric, playable stochastic model with a watchable latent state.

Empirical spectral analyses of music. Levitin, Chordia, and Menon showed that rhythm spectra across a large notated corpus follow 1/f-family power laws with genre- and composer-dependent exponents [13]. Multifractal analyses — mostly via the MFDFA estimator of Kantelhardt et al. [11], building on the DFA of Peng et al. [10] — have repeatedly found not just long-range correlation but genuine multifractality in music: in note sequences [14], in Bach's Inventions pitches [15], and in Bach's Sinfonias as interacting voices [16]. The same MFDFA spectral width discriminates instrument families by their mode of playing [41] and tracks a soloist's performance-to-performance improvisation [42], while multiscale fractal descriptors distinguish instrument timbres in recognition tasks [43] — the multifractal signature carries musically salient structure, not merely a global exponent. These results set our measured targets — 1/f-family exponents and a nonzero multifractal width — verified on the emitted note streams in Section 6. That the targets are worth *enforcing* rather than assuming is underscored by recent work applying the same DFA/MFDFA battery to state-of-the-art neural music generators (Suno, DiffRhythm, YuE): their output departs systematically from the human multifractal signature — reduced fine-scale variability, narrower or more erratic spectra — even while passing surface-fidelity metrics [44]. (These are analyses of audio amplitude envelopes rather than symbolic streams, so we read them for the qualitative gap, not for numerical comparison to our onset-stream widths.)

Timing fluctuations and humanization. Hennig et al. found long-range correlated timing fluctuations in human rhythmic performance and showed listeners prefer 1/f-correlated "humanization" over conventional white-noise jitter [17], with follow-up work on interacting musicians [18] and on professional recordings [19]. The same expressiveness — a mixture of predictable and unpredictable structure across timescales — proves multifractal beyond human music: in thrush-nightingale song, subtle note-timing and intensity deviations (the analogues of accelerando and crescendo) measurably drive the multifractality [45], a cross-species echo of the two structure-generating mechanisms the Fenophone keeps separate (the cascade's arrangement and a dedicated long-range-correlated microtiming sub-cascade; Section 8). This strand supports the claim that long-memory

fluctuation is not an academic nicety but an audible property listeners respond to.

The MSM lineage. Multiplicative cascades originate in Mandelbrot's turbulence work [4]. The Multifractal Model of Asset Returns [5] composed a Brownian motion with a multifractal time deformation; Calvet and Fisher then recast the cascade as a *causal, stationary, discrete-time Markov process* — the Markov-switching multifractal — in which volatility is a product of k Markov-switching multipliers with geometrically spaced switching frequencies [6, 7, 8], a formulation with a substantial estimation literature [9]. MSM is, to our knowledge, an unusual choice for a musical instrument, but it is arguably the *instrumentable* multifractal: it runs forward one tick at a time, it is stationary, its parameters are few and each has a plain meaning (how strong the bursts are, how fast the fastest and slowest components move, how many scales participate), and its latent state is a finite vector a player can watch and grab.

Algorithmic composition and Markov models. Markov processes have been compositional material since Hiller and Isaacson [20] and Xenakis [21]; Ames surveys the tradition [22]. The Fenophone's harmony layer is in this lineage (a chord-function graph walked with biased edge sampling), but its distinctive commitment is that the *bias* comes from a slow band of the cascade, so harmonic restlessness is one more reading of the same multi-scale process.

Generative instruments and NIME. From Chadabe's interactive composing [24] and Eno's generative music [25] to the modern crop of generative sequencers inside commercial DAWs, systems that generate material under performer influence are well established; design principles for the control surfaces of such systems are equally established [26, 27], and the tradition of making the display and the control surface one object runs through instruments like the reacTable [28]; Drummond surveys the interactive-system design space [29]. The pin gesture (Section 4.4) sits in a more recent NIME thread that performs the *latent state* of a generative model: Scurto and Postel walk listeners through deep latent audio spaces [35], and Shepardson and Magnusson's Living Looper folds RAVE-style neural synthesis into a looping pedal's action–perception loop [36]. Those systems navigate a learned, high-dimensional, opaque embedding continuously; the Fenophone's latent state is instead the finite regime vector of an econometric Markov model — enumerable, renderable cell by cell, freezable without a glitch, and meaningful parameter by parameter — which is what lets the intervention be exact (Section 5.1) rather than exploratory. Against this background the Fenophone's contribution is not a novel

mapping strategy. It is, first, the model choice — a strict, parametric multifractal rather than an inspired-by texture — and second, a verification stance we have not seen elsewhere in the instrument literature: treating the *statistical* properties of an instrument's musical output the way engineering treats correctness, as published, versioned, CI-enforced invariants.

Model-based sonification. Hermann and Ritter's model-based sonification is the nearest methodological family: sound is produced by the dynamics of a model, not by directly mapping data to acoustic parameters [37]. The Fenophone shares the commitment to *sounding a model* but inverts the data flow — nothing is sonified from data at all; a volatility model that econometrics fits to markets is instead run forward as a free generator, its observation equation discarded and its latent state promoted to the performance surface. To our knowledge no prior instrument or sonification system uses an estimable econometric volatility model as its generative core.

3. The Model

3.1 The cascade

The Fenophone's intensity process is a product of k switching multipliers:

$$\text{intensity}_t = \text{base} \cdot M_{1,t} \cdot M_{2,t} \cdot \ldots \cdot M_{k,t}$$

Each multiplier M_i switches between a high state m_hi > 1 and a low state m_lo = 1/m_hi, *mirror-balanced in log space* so that each multiplier's long-run geometric mean is exactly 1. Raising Intensity raises m_hi, which widens the contrast between states without moving the long-run average: more Intensity means burstier, not louder (output gain is a separate frame control).

Switch rates are defined directly in hertz and are decoupled from k. Layer i = 1…k (i = 1 slowest) has target rate

$$f_i = f_{slow} \cdot S^{((i-1)/(k-1))},$$

geometric spacing across the span S = f_fast / f_slow. Because the spacing is parameterized by the *span* rather than a per-layer ratio, the exponent endpoints are 0 and 1: the top layer runs at exactly f_slow · S for every k, so adding layers fills the same slow-to-fast range with finer resolution and never creates runaway fast motion. A clamp $f_i \le \min(8\text{ Hz}, 0.9/\Delta t)$ keeps the fastest layer musical and keeps the per-tick switch probability

$$p_i = 1 - \exp(-f_i \cdot \Delta t), \quad \Delta t = 60 / (\text{BPM} \cdot \text{ticks\_per\_beat})$$

safely below saturation (the 0.9/Δt arm bounds p_i below 1 − e^(−0.9) ≈ 0.59). The exponential is evaluated only when a control changes, in fixed-point, and each p_i is quantized to an integer threshold (p_i · 2^64 at Q16.16 resolution); the per-tick decision for each layer is then a single 64-bit integer comparison against a draw from that layer's own PRNG substream. There is no floating point and no transcendental in the per-tick path.

Each multiplier is thus a symmetric two-state Markov chain: per tick, layer i toggles with probability p_i. Section 5.1 explains why each layer draws from its own seed substream.

3.2 The band map and the four lanes

The k layers partition into three bands — slow, mid, fast — for every k: slow = fast = ⌊k/3⌋, with the remainder to mid, so every band has at least one layer for all k ≥ 3 (at the default k = 6 the partition is 2/2/2; at k = 12 it is 4/4/4). Musical dimensions read *bands*, never individual layers, which is what decouples the Layers control from the audible ensemble: lowering k makes each band's motion coarser; it never deletes an instrument.

Each tick the cascade emits four control lanes in [0, 1]:

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	overall energy opens filters, adds drive
tension	slow band	high = unstable, leaning harmony; low = resolved

A band's product is the product of its multipliers. With a shared m_hi, the log of the band product is log2(m_hi) · s, where s = (#high − #low) is the band's integer net state in [−n, n] for band size n. The lane is a monotonic map of the band product, centered at 0.5 when the product is 1:

$$\begin{aligned} \text{lane} &= \text{clamp}\left(1/2 + \log_2(\text{band\_product}) \,/\, (4 \cdot n),\ 0,\ 1 \right) \\ &= \text{clamp}\left(1/2 + (\log_2(m_{hi}) \,/\, (4 \cdot n)) \cdot s,\ 0,\ 1 \right) \end{aligned}$$

The reference scale 4·n equals 2 · n · log2(m_hi_max) with the pinned maximum m_hi_max = 4.0, so at maximum Intensity a fully aligned band spans the whole [0, 1] range, while at lower Intensity the lane compresses toward 0.5 — the same average, less contrast. The per-tick cost is one fixed-point multiply. The timbre lane reads the whole product (all k layers) the same way and low-passes it with a one-pole filter (default coefficient 1/8).

This decoupled routing — different musical dimensions reading different time scales of one process — is what makes the output sound *composed*

rather than pumped: harmonic mood moves slowly while loudness flickers fast, because those literally are the slow and fast bands of one cascade.

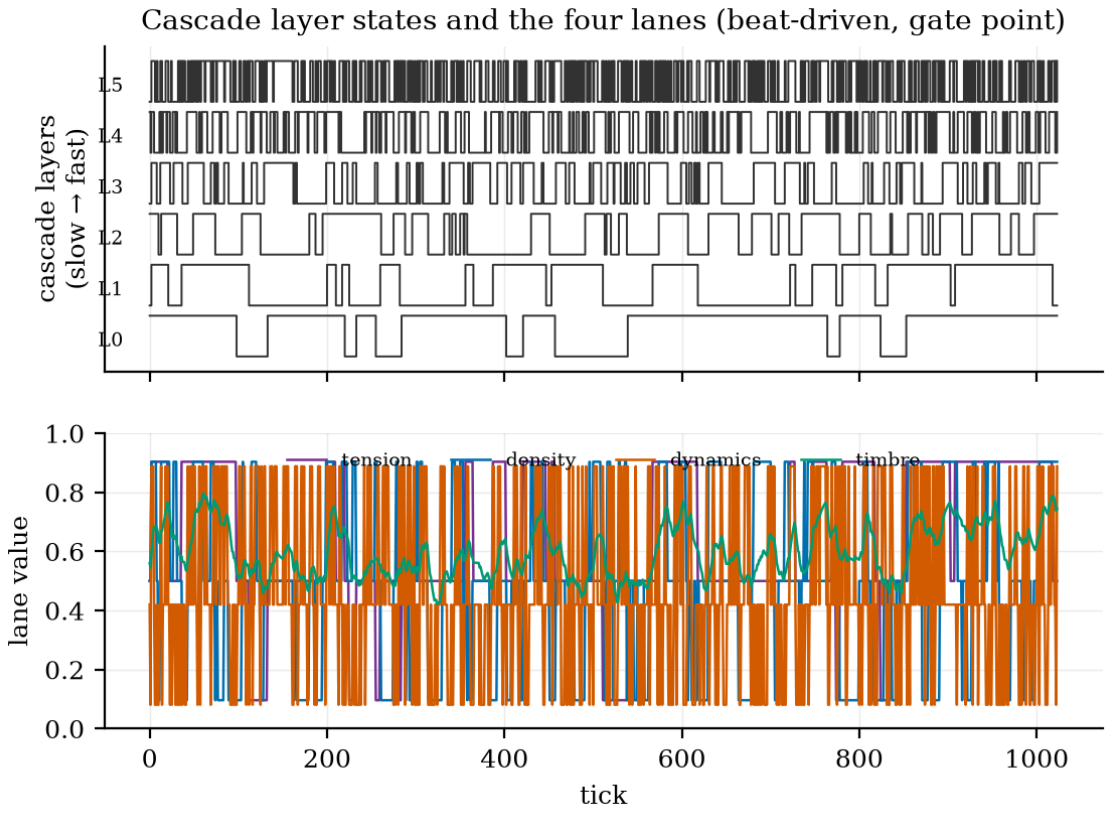

Figure 2. Cascade mechanics at the gate operating point (beat-driven): k = 6 layer state traces (slowest at bottom) over ~1024 ticks, with the four derived lane series (dynamics, density, timbre, tension) below. Rendered from a `foundry render` run at seed 0 of the published panel.

3.3 The control mapping

Each player control maps a normalized position x ∈ [0, 1] through an explicit, monotonic transfer function, evaluated only on control change and quantized to fixed-point. The model controls are, verbatim from the implementation:

Control	Model parameter	Transfer	Range
Intensity	high-state magnitude m_hi	exponential	≈ 1.05 (x = 0) → ≈ 4.0 (x = 1)
Spread	time-scale span S = f_fast/f_slow	S = 4 · 50^x	≈ 4× → 200×
Change Rate	slowest switch rate f_slow	f_slow = 0.01 · 50^x Hz	one switch per ~100 s → per ~2 s
Layers	multiplier count k	k = round(3 + 9x)	3 → 12
Flow	contour-walk memory φ (per voice)	φ = 0.97 · x, clamped < 1	0 → 0.97

Flow belongs to the pitch walk (Section 4.3) rather than the cascade; it is grouped with the model controls because it, too, is a genuine parameter of the generative process rather than a post-hoc effect. Three *frame controls* — Tempo (40–200 BPM; rescales Δt and recomputes switch probabilities at change time), Key (root over the pack's offered modes, applied as a global transpose), and Level (output gain, −60 dB + 60·x) — set the stage the model performs on. They are deliberately presented as stage settings, never dressed up as model parameters: the honesty of "every model knob is real" depends on also saying which knobs are not the model.

3.4 Relation to canonical MSM

We claim the core is *strictly* MSM, so we state exactly where the implementation departs from the canonical binomial MSM of Calvet and Fisher [7], and why.

1. **Normalization.** Canonical MSM draws multiplier states with arithmetic mean 1 (binomially, {m_0, 2 − m_0}). The Fenophone uses log-symmetric states {m_hi, 1/m_hi} with *geometric* mean exactly 1. Music reads intensity logarithmically (velocity, decibels), so neutrality in the log domain is the musically meaningful invariance: driving Intensity changes contrast while the long-run average level stays put.
2. **Rate parameterization.** Canonical MSM spaces switching probabilities as γ_i ≈ γ_1 · b^(i−1) for a fixed base b, so adding components extends the frequency range. The Fenophone fixes the *span* S and lets the effective base b = S^(1/(k−1)) adapt to k. Both are geometric ladders; the difference is ergonomic and essential for an instrument: Layers must be a resolution control, not a range control, or turning it up would manufacture uncontrolled fast motion.
3. **Toggle vs. redraw.** Canonical MSM redraws a switching multiplier from its state distribution (staying put half the time, in the binomial case); the Fenophone toggles deterministically on a switch event. For a symmetric two-state chain these coincide up to a factor of two in rate — a reparameterization, not a structural change.

With those departures named, the structure is the MSM structure: k first-order Markov chains at geometrically spaced frequencies whose product modulates activity. Strict-sense multifractality holds for the k → ∞ cascade; at finite k the process is multifractal over a bounded range of scales, which is precisely the regime Section 6 measures. Finally, where the MMAR composes a Brownian motion with a multifractal time deformation [5], the Fenophone composes a stationary second-order pitch walk with cascade-

determined *event times* — melody sampled in multifractal musical time. The parallel is structural, and we return to its honest limits in Section 7.

4. The Instrument

4.1 From lanes to notes

Each scheduler tick (the tick is the content pack's grid step; all shipped packs use four ticks per beat in 4/4, i.e. 16-tick bars):

1. The cascade advances and emits the four lanes.
2. **density → rhythm.** The density lane sets expected onsets in the current bar through the pack's monotonic onsets(density) curve plus a clumping factor, both bounded (Section 5.3); candidate onsets snap to the grid.
3. **tension → harmony.** At the pack's harmonic-rhythm rate, the next chord function is chosen by sampling the outgoing edges of a directed chord-function graph, *biased by the tension lane*: high tension up-weights leaning functions, low tension up-weights resolution toward the tonic. The chosen node defines the current legal pitch pool — chord tones plus extensions that join the pool only as tension rises. A held-note rule keeps pool changes from retuning sounding notes: transitions are voice-led so a tone held across a chord change is a common tone or a prepared suspension.
4. **Pitch.** Each voice's contour walk (Section 4.3) supplies a continuous pitch height; at each onset the height snaps to the nearest member of that voice's current legal pool, transposed by the global key.
5. The result is a stream of quantized note events (pitch, onset, duration, velocity, control data), every value traceable to the model core or to authored pack data, and every value computed in integer/fixed-point arithmetic.

4.2 The ensemble and the routing matrix

The default ensemble has four voices — bass, pad/harmony, lead, percussion — with the pad polyphonic up to 4 notes, the others monophonic, a global 16-note cap, and oldest-within-voice stealing. A band × voice routing matrix (3 × 8; up to eight addressable voices) mixes the three band activities into a per-voice *drive*:

```
drive(v) = ( Σ_band gain[band][v] · band_value[band] ) / ( Σ_band gain[band][v] )
```

The normalization makes the drive a weighted average, so it lies in [0, 1] regardless of how many bands feed a voice and — because the weights are non-negative — remains monotonic in each band value, preserving the

controllability guarantee through the routing layer. A voice with an all-zero column is silent. The drive sets each voice's onset count at the bar downbeat (through the pack's onset model) and each note's velocity at onset. The default wiring makes the multi-scale structure audible as arrangement: slow band → bass and pad (rare, deep moves), mid band → lead (phrasing, flurries), fast band → percussion (jittery, dense). Rewiring the matrix is where expert play lives; because the drive stays a monotone weighted average, no rewiring can make a control's effect reverse.

4.3 The contour walk

Each voice carries a continuous pitch-height signal h in semitone (MIDI) space, advanced once per tick by a discrete-time, fixed-point process:

```
h_{t+1} = h_t + φ · (h_t − h_{t−1})        (momentum: Flow)
              − γ · (h_t − center_voice)    (register gravity)
              + σ · ε_t                     (step noise)
```

φ is the Flow control's memory term (high φ gives long, correlated, singable runs; φ near 0 gives restless zig-zag); $\gamma > 0$ is a mandatory mean reversion toward the voice's register center (defaults 0.05–0.15 per step); σ is the step-noise scale in semitones per step (defaults 0.6–1.2); centers are authored per voice (bass ≈ 40, lead ≈ 64 in MIDI note numbers) with a register band of center ± span (span ≈ 9 by default). The noise ε_t is an integer-exact approximate Gaussian (Section 5.1). At each onset the continuous height snaps to the nearest legal pool member, with exact ties resolved to the lower pitch — a fixed rule, so the snap is deterministic. Each voice owns its own seed substream, so voices are statistically independent lines of one process family rather than copies, and adding or removing a voice never perturbs another voice's sequence.

4.4 The pin gesture: performing the latent state

The Fenophone's visualization — stacked horizontal rows of cells, the slowest layer at the bottom, scrolling as time advances — renders the cascade's latent state directly, and it is also the control surface. Beyond the dials, the player can *grab a layer and hold its state*: pinning a slow layer into its high state forces a sustained build, because the held layer lifts its band's net state (and the whole product), so the elevated tension and timbre lanes carry the build upward through the faster layers and into the music. This is the instrument's signature "conducting" gesture: the player changes the structure and watches music grow from it, never placing individual notes.

Mechanically, a pin is a held intervention on the latent Markov state. A pinned layer takes no PRNG draw and never toggles — its substream

freezes, exactly as a deactivated layer's does — and on release the layer returns to model-driven switching at the next tick, resuming its substream from where the pin paused it. Each pin and release is recorded as a keyframe in the state vector's automation track, so a live performance and its replay are bit-identical, and a pin survives capture, export, and sharing. A latent state you can render, you can grab; a substream you can freeze, you can release without a glitch (Section 7 returns to this point).

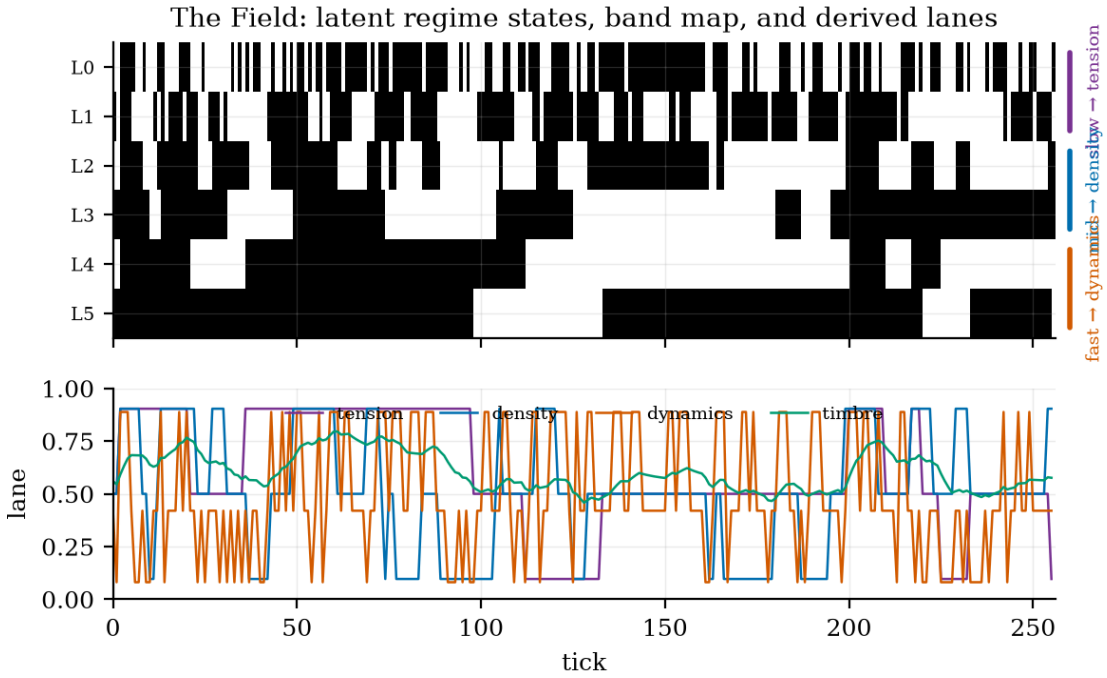

Figure 3. The Field: the instrument's latent surface. Layer-state cells (slow at bottom) over 256 ticks with the slow/mid/fast band partition annotated at the right edge and the four lanes they drive below. Rendered from the engine's per-tick latent states (`foundry render` lanes.csv).

4.5 Scaffolds: authored taste, played live

Genre lives entirely in a *scaffold*: a versioned, signed data pack (never code) containing the tuning and pitch pools, the chord-function graph and offered modes, the rhythmic grid and onset model with coherence bounds, the lane-mapping curves, the per-voice contour parameters, and the voice/timbre presets. The engine is genre-neutral; switching scaffolds (the launch set is beat-driven at 124 BPM preferred tempo, warm/ambient at 72, and cinematic at 90, plus a beat-forward prototype pack at 120) swaps only data. Because the cascade is genre-neutral and seeded independently of the pack, the same master seed and control positions produce a bit-identical cascade motion across genres while the emitted notes differ — the same structural weather heard through different instruments, which the instrument's teaching design uses as its one engineered revelation.

The scaffold is where taste comes from, and we are explicit about that division of labor: the model supplies motion with structure at every scale; the scaffold guarantees that the motion lands as music (consonance floors, voice-led transitions, bounded rhythm). The player's live surface is the model controls, the frame controls, the routing matrix, and the pin gesture; the scaffold is chosen, not performed.

5. Guarantees

5.1 Determinism: the event layer is bit-identical everywhere

The Fenophone splits its arithmetic into two layers with different, honestly stated guarantees. The *event layer* — cascade switching, band products, lanes, the harmony walk, contour walks, routing, scheduling, and quantization to note events — uses integer and fixed-point (Q16.16 or wider) arithmetic only, with no floating point and no platform-divergent transcendental in the per-tick path; transfer functions involving exp are evaluated only on control change and quantized before entering the event layer. Given the same state vector and engine version, the event stream is bit-identical on any CPU, browser, or OS. The *render layer* (synthesis and resampling, all owned in the instrument's own DSP rather than delegated to browser nodes) is floating-point and is guaranteed identical within a published tolerance: no output sample differs by more than 2^{-20} ($\approx 10^{-6}$) in normalized amplitude, with full-window RMS error below ≈ -120 dBFS; exports render offline at a pinned 48 kHz / 24-bit. MIDI export is therefore the bit-identical event stream; audio export reproduces the captured passage within the published, sub-audible tolerance.

A performance is fully determined by a small state vector: master seed, scaffold identity and version, tempo, key and mode, the automation track (including pins), the routing matrix, k, and per-voice seed offsets. Randomness happens once, at seeding. A single named integer PRNG (xoshiro256**) drives everything through *substreams*: each cascade layer and each voice draws from its own stream seeded master_seed ⊕ offset, and the per-tick draw order is fixed (cascade layers in order 1…k, then the harmony edge on harmonic steps, then per-voice walk noise in voice-index order). Two consequences matter musically. First, *k-invariance*: a layer's substream is a pure function of (master seed, layer index) — never of k — and all twelve multiplier slots are seeded up front, so changing Layers mid-performance keeps every retained multiplier's state and stream untouched; the arrangement re-resolves rather than jumping, and a deactivated layer later reactivated resumes rather than restarts. Second, voice independence: adding or removing a voice or layer never perturbs any other voice's

sequence, which is itself a tested invariant. Gaussian noise for the walks is produced without sin/log: twelve uniform draws (the top 16 bits of one PRNG word each) are summed and centered — the Irwin–Hall construction — giving unit variance exactly, support (−6, 6), and bit-identical noise on every platform.

Determinism is usually sold as an engineering convenience. Here it is load-bearing three times over: it is why the instrument is learnable (the same gesture from the same state always does the same thing); it is why capture, sharing, and honest export are true claims (a shared "exact take" replays the identical event stream); and it is why the statistical gates of Sections 5.4 and 6 are CI-admissible at all — the measured series are bit-identical everywhere, so a gate with margins of order 0.1 in α cannot flake on platform arithmetic that wiggles by a few ulps.

5.2 The register-gravity theorem and the two-minute gate

An unconstrained contour walk drifts without bound and floats off the usable register within minutes — in our experience the single most common silent failure of instruments of this kind. The Fenophone makes the fix a theorem plus a gate rather than a habit.

Let $d_t = h_t - \text{center}$. The walk of Section 4.3 gives the linear recurrence

$$d_{t+1} = (1 + \varphi - \gamma) \cdot d_t - \varphi \cdot d_{t-1} + \sigma \cdot \varepsilon_t,$$

a damped, noise-driven oscillator with characteristic polynomial $P(z) = z^2 - (1 + \varphi - \gamma)z + \varphi$. By the Jury stability test [30], the roots lie strictly inside the unit circle — i.e. the walk has a bounded stationary range — iff

$$P(1) = \gamma > 0 \quad \text{and} \quad |\varphi| < 1 \quad \text{and} \quad P(-1) = 2 + 2\varphi - \gamma > 0.$$

With $\gamma = 0$ the polynomial has a root at $z = 1$ and the process degenerates to an undamped random walk. The first condition is therefore a standing rule of the codebase: $\gamma > 0$ is mandatory, and a voice configuration with $\gamma \le 0$ is rejected at validation. Stability alone does not size the register band, so we also use the closed form for the walk's stationary standard deviation (the noise being unit-variance by construction):

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

with long-run maximum excursions settling at roughly 5–6 σ_d. A register band center ± span is safe only when the span is chosen consistently with (φ, γ, σ); the engine's default voices declare spans with margin over 6 σ_d (e.g. the bass preset's span of 9 semitones against $6\,\sigma_d \approx 8.0$; the lead's 10 against ≈ 8.7), and the launch packs author spans above 8 σ_d.

The mechanized half is the *two-minute drift gate*: CI runs every voice of every pack for at least two minutes of ticks at parameter extremes and fails the build if any voice leaves its declared band. Deliberately broken fixtures (a drifting voice among them) prove the gate has teeth. The theorem says why the instrument cannot drift; the gate makes sure no future edit quietly removes the reason.

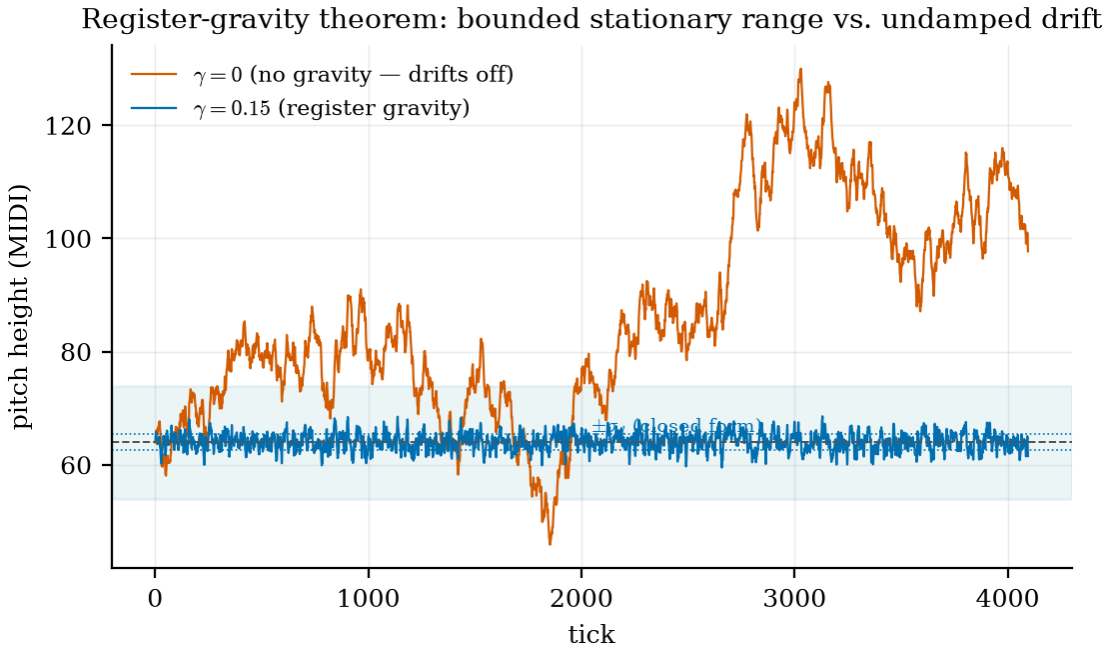

Figure 4. Register gravity: two contour-walk paths at identical seed, γ = 0 (undamped — drifts off the shaded register band) vs. γ = 0.15 (stationary within the band), with the ±σ_d closed-form envelope marked. Illustration of the documented §4.4 walk equation.

5.3 The invariant suite

Every scaffold — shipped or, later, community-authored — must pass ten automated invariants before it can be signed and distributed; the CI corpus gate runs the suite over every pack in the repository plus deliberately broken fixtures that prove each check can fail. Compactly:

1. **Schema-valid** against the versioned scaffold schema.
2. **Voice-leading:** every harmony-graph edge preserves a common tone or stepwise resolution for every chord tone, under every offered mode.
3. **Non-empty legal pool** at every node, including a consonance floor that stays legal at maximum intensity, under every mode and root.
4. **Reachability:** the harmony graph can always progress and always return toward the tonic.
5. **Register safety:** every voice stays in-band over the two-minute-drift run (Section 5.2).

6. **Rhythmic coherence:** every voice respects its authored minimum and maximum onset density and maximum clumping across the full parameter sweep — never silent, never a wall of noise.
7. **Lane monotonicity:** all lane-mapping curves are monotonic.
8. **Clip-free / no-NaN** rendering across a parameter sweep.
9. **Provenance complete** for all sampled assets.
0. **Spectral structure:** the emitted onset stream keeps 1/f-family long-range correlation at the published default controls (Section 5.4).

Alongside the pack suite, an engine-side controllability suite sweeps every control and asserts that its owned perceived feature trends monotonically (more Intensity ⇒ non-decreasing energy contrast; more Spread ⇒ non-decreasing time-scale separation; and so on), mechanizing the instrument's tameability guarantee. Two human gates remain deliberately human: a listening panel scores random patches each release on musicality and steerability (release threshold: median ≥ 3.5/5 on each dimension over ≥ 30 patches spanning all packs and extremes, panel of ≥ 8 listeners, blind to build), because no invariant proves that output is worth keeping.

5.4 Invariant #10: the spectral floor

Invariants 2–9 check bounds and properties; none of them checks *long-range correlation of what is actually emitted*, so a pack could satisfy all of them while pinning its rhythm into a stream that destroys the multi-scale structure the cascade exists to provide. Invariant #10 closes that gap, and we specify it fully because it is, to our knowledge, the first CI-enforced spectral invariant on a musical instrument's output.

Protocol. Run the pack's own pipeline at a pinned default operating point — the instrument's documented first-run positions, Intensity 0.8, Spread 0.65, Change Rate 0.59, Layers 1/3 (k = 6), through the pinned transfer functions, at the pack's preferred tempo, default mode, root 0 — for 4096 ticks on each of 4 pinned seeds. Extract the per-tick onset-count series. Require non-degeneracy (neither silent nor an unbroken wall). Subtract the *bar-phase mean* — the series' mean value at each position modulo the bar length (16 ticks on the shipped grids) — and estimate the DFA-1 scaling exponent α over dyadic scales 16–512 ticks (cumulative-sum profile, non-overlapping windows, per-window least-squares detrend, RMS fluctuation, least-squares log–log fit). The invariant passes iff every run has an exponent at all and the median α over the seed panel is ≥ 0.60.

Why bar-template removal. A scaffold's metric grid is an *authored, deliberately periodic* template; the estimator must see the model's

fluctuations, not the grid. Subtracting the bar-phase mean is exactly the seasonal-detrending step standard in DFA practice for known periodic trends. It matters: on the raw series the periodic template dominates and the corpus measures α ≈ 0.39–0.61, while the adjusted series expose the cascade's scaling cleanly. The step is validated in both directions — it zeroes an exactly bar-periodic series (which then correctly has no exponent, the failure mode the invariant names), and it is a statistical no-op on aperiodic input (white noise stays α ≈ 0.5 after adjustment).

Anchors. The estimator itself, not just the music, is pinned by analytic anchors that run in the same suites: a deterministic white-noise series must measure α ≈ 0.5 (accepted band [0.4, 0.6]) and its cumulative sum ≈ 1.5 ([1.3, 1.7]) — the classic uncorrelated-noise and random-walk fixed points; and the MFDFA generalized exponent h(2) must coincide with plain DFA α to within 10^−12 on a series with no zero-fluctuation windows.

Interior anchors. Endpoint anchors do not validate the interior, and the landscape lives in the interior (α ≈ 0.70–0.77) on sparse, integer, zero-inflated series seen through an authored periodic template — the regime where DFA has documented caveats. A known-Hurst battery therefore anchors the estimator in its own operating regime, with exact fractional Gaussian noise (Hosking's recursion, deterministic innovations) as ground truth. Measured at the published length and ladder, panel medians: dense fGn recovers H to within 0.01 across the interior (0.548 / 0.691 / 0.855 at H = 0.55 / 0.70 / 0.85); an onset-like sparse binary observation of the same latent process (indicator above the 75% quantile, ≈ the packs' fill) stays clearly persistent but reads *low* — 0.632 at latent H = 0.70, a frozen attenuation of ≈ 0.07, downward, never upward; the bar-template-removal step reproduces the constant-threshold reading to ≈ 0.001 in that sparse regime and remains a no-op on sparse white onsets; and shuffling the fGn collapses α to ≈ 0.5, so the measured exponent lives in temporal order, not in marginals. Two readings the rest of this paper depends on: the emitted-stream exponents cannot be an estimator inflation (the channel's bias is conservative), and emitted α compares across conditions of this one pipeline rather than reading out latent H absolutely.

Why a loose floor. White noise sits at 0.5; an exactly bar-periodic stream has no exponent at all; the shipped packs measure ≈ 0.70–0.78 under this protocol — and the order-destroyed (shuffled) null of those very streams reaches a per-seed 95th percentile of only α ≈ 0.53–0.55 (Section 6's null comparison). The 0.60 floor therefore sits above everything the marginal distribution alone can produce, rejects whitening and periodic collapse with margin, and deliberately does not legislate style: the invariant guards

against *destroying* multi-scale structure in any pack the pipeline will accept (including future community packs), and the tighter per-scaffold bands of Section 6 sit on top of it for the shipped set.

6. The Measured Spectral Landscape

The per-scaffold companion suite measures the shipped corpus under a stronger protocol and freezes the results as published bands. Runs are 8192 ticks (2048 beats on the shared four-ticks-per-beat grid — roughly 17 to 28 minutes of music at the launch tempi of 124 down to 72 BPM) over a pinned 8-seed panel that shares its base with the invariant's 4-seed panel. Two series are extracted from the emitted notes — per-tick onset counts, and per-tick summed velocities (0 when silent) — and bar-template-adjusted as above. DFA α(onset) is read over dyadic scales 16–1024 ticks (2–124 s of music at 124 BPM; 3.3–213 s at 72 BPM); α(velocity) over 16–256; the multifractal width is MFDFA's $h(-2) - h(2)$ over the onset ladder, with (near-)zero-fluctuation windows excluded from the negative moment. Aggregation is the median over the seed panel, per scaffold. The measurement corpus is fixed at the four packs below — the three launch scaffolds plus one beat-forward prototype — and stays the frozen gate corpus for provenance even though the shipped library has since grown past them (additional base and premium packs pass the same `check_pack` spectral gate before shipping, but the published per-scaffold bands are derived from these four).

Measured medians at the default operating point (8 seeds × 8192 ticks, bar-template adjusted), reproduced exactly from the measurement suite:

Scaffold	α(onset), 16–1024	α(velocity), 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

From these measurements, frozen bands now gate every release: median α(onset) must lie in [0.62, 1.10] (the floor sits 0.075 under the weakest pack and excludes white noise at 0.5; the ceiling excludes a drift-dominated stream while leaving ≈ 0.33 headroom), median α(velocity) in [0.60, 1.05], and the median multifractal width must reach the published floor of 0.025 (≥ 2× margin under the weakest pack). One honesty note the null comparison below forced: the width floor is a *degeneracy* guard, not a multifractality certificate — at this length the width estimator reads 0.04–0.09 even on

surrogates with no temporal structure at all, so no fixed width threshold can certify nonlinearity; the certificate is the per-pack surrogate excess reported next.

The null comparison. Because finite length and non-Gaussian marginals alone produce nonzero measured widths [38], every per-scaffold measurement is compared against two families of 99 surrogates per seed, built from the same bar-adjusted onset series (`foundry nulls`; committed as `research/data/nulls-v1.csv`): IAAFT surrogates [39, 40], which keep the amplitude distribution exactly and the power spectrum iteratively while destroying higher-order structure, and Fisher–Yates shuffles, which keep only the amplitudes. The verdict convention is one-sided: the measured statistic must exceed the family's 95th-percentile order statistic, per seed. Results: **α(onset) is order-borne everywhere** — every pack's measured α beats its shuffle null's p95 on 8/8 seeds, and the shuffle null collapses to α ≈ 0.50 (median) with p95 ≈ 0.53–0.55; **the width claim splits the corpus** — beat-driven (0.126 vs. IAAFT p95 ≈ 0.080), cinematic (0.108 vs. ≈ 0.069), and prototype (0.102 vs. ≈ 0.080) exceed their IAAFT nulls on 8/8 seeds, evidence of multifractal structure beyond linear correlation plus marginals, while warm-ambient (0.054 vs. IAAFT p95 ≈ 0.092) does not on any seed, so its measured width is consistent with a linear long-memory process observed at this length, and we do not claim otherwise — and, since the two-point width statistic has measured near-zero power against genuine cascade structure under this very null design (the scattering-moment audits of Fenocosm v22 and Fenosoma x7), the honest reading is "no verdict at this power," while the three positive packs' excesses, delivered by a near-blind instrument, are correspondingly stronger evidence than their bare thresholds suggest. The measured landscape's multifractality claim is therefore per-pack, surrogate-controlled, and honest about its one negative. The velocity exponent is gated over the mid-scale 16–256 ladder because the velocity stream mixes in the fast dynamics band and its long-range tail is weaker — ≈ 0.63 over 16–1024, too close to the 0.60 floor to gate honestly; the mid-scale decade is where the check has teeth.

The Spread → α ordering. The one interaction between a dial and the output spectrum that we publish as a gate is monotone and fully consistent across the corpus: with the other dials at the default point, median α(onset) *falls* as Spread rises. At Spread = 0 the four packs measure 0.913 / 0.865 / 0.747 / 0.851 (beat-driven / cinematic / ambient / prototype), against 0.747 / 0.751 / 0.657 / 0.735 at Spread = 1 — a median gap ≥ 0.09 on every scaffold, with no per-seed overlap anywhere. The mechanism is physical: widening the span pushes the mid-band rates — the corners of the density lane — higher, so the onset stream decorrelates in fewer ticks. A CI sub-gate asserts the

direction on every pack. We note the reading for players: Spread is audible not only as arrangement stratification but as a controlled tilt of the output's long-range memory.

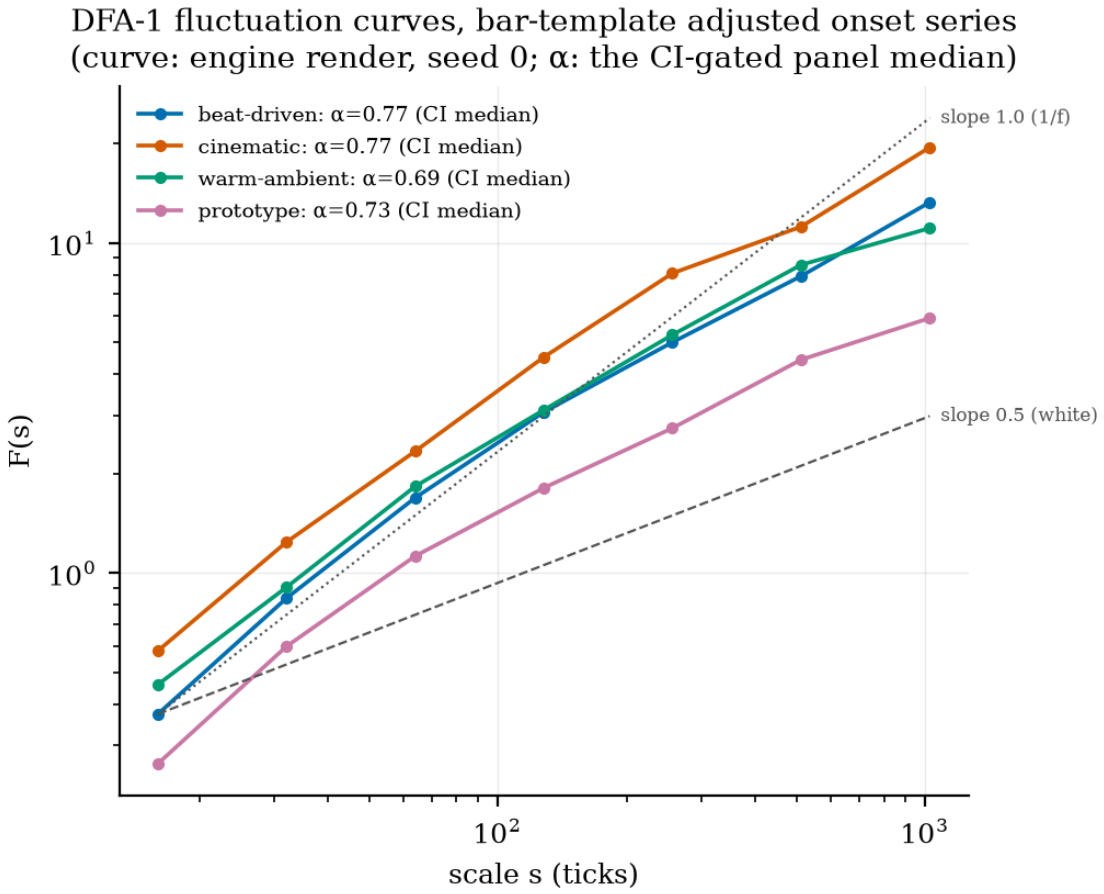

Figure 5. DFA-1 fluctuation curves F(s) vs. scale s (16–1024 ticks) for the four measured packs' bar-template-adjusted onset series, with reference slopes 0.5 (white) and 1.0 (1/f). Curves rendered from `foundry render` (seed 0); the labelled α is the CI-gated panel median (Section 6). (Shared figure, also used in the companion paper 03.)

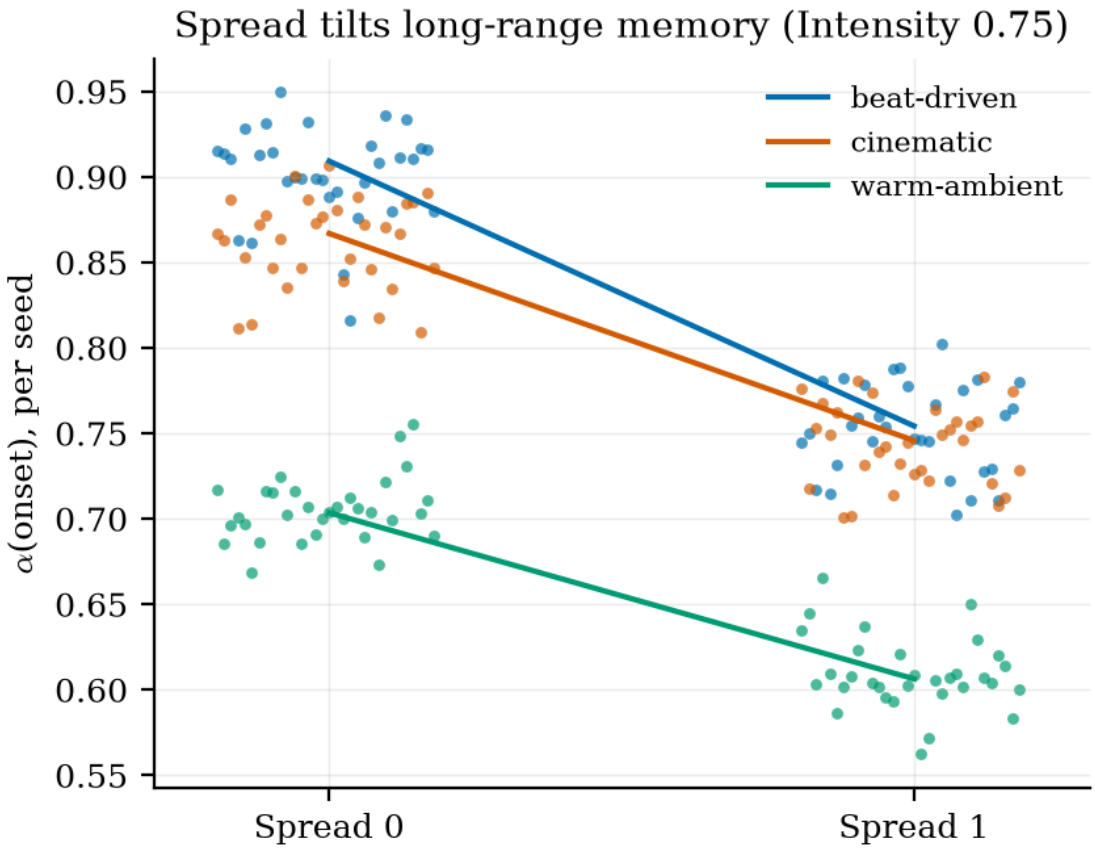

Figure 6. The Spread ordering: per-seed α(onset) at Spread = 0 vs. Spread = 1 for each launch scaffold, from `sweep-v2.csv` (Intensity 0.75, the nearest grid point to the 0.8 gate). The medians drop with Spread; per-seed ranges are disjoint.

Interpretation. Two properties were the point of choosing this model, and both are now measured, null-controlled facts about the shipped instrument rather than intentions. First, the emitted onset and velocity streams sit in the 1/f family — α medians of 0.695–0.770 and 0.676–0.702 respectively, comfortably away from both the uncorrelated (0.5) and random-walk (1.5) fixed points, with every seed beating its shuffled null — the same family reported for rhythm and fluctuation spectra of composed and performed music [1, 13, 17]; the interior anchors of Section 5.4 additionally bound the estimator's role, since the sparse observation channel attenuates rather than inflates latent persistence. Second, three of the four packs' onset streams are multifractal by the surrogate criterion — widths 0.102–0.126, beyond their IAAFT nulls on every seed — consistent with the music-MFDFA literature's finding that real music is multifractal rather than monofractal [14, 15, 16]; the fourth (warm-ambient, width 0.054) is not separable from a linear long-memory null at this length, a finding we report as-is and which fits its musical character (the least intermittent pack by design). We are careful about the claim's strength: the exponents were not tuned to match

any particular corpus (they vary by genre pack, as they do across composers in the literature), and the measurement is of a generative note stream under a fixed protocol, not of audio recordings. The claim is family membership plus per-pack surrogate-controlled multifractality, measured on what the instrument actually emits, with published estimators, scales, seeds, bands, and nulls.

The corpus comparison. Because the band was fixed before any corpus was measured, the first external-corpus run (2026-07-12; the companion corpus-fitting analysis, `research/analysis/msm-corpus-fit/`) is an informative post-hoc test rather than a fit: the 1,260 analyzed MAESTRO piano performances (of 1,264 kept after the ingest gate), processed with the same estimator lineage (DFA-1 with bar-template removal, MFDFA $h(-2) - h(2)$, dyadic ladders, shuffle nulls) on a 12-ticks-per-beat beat-time grid, measure median α(onset) 0.717 (shuffled 0.500; 83% of pieces above 0.6) and width 0.118 — inside the instrument's published band (α 0.695–0.770, width 0.054–0.126; these are the 8-seed protocol medians — a 32-seed panel reproduces them within ≈ ±0.012, so where the band is read as a fixed target across the program it should be read as ≈ 0.70 ± 0.04 at panel scale; companion channel paper, §4.1). The agreement is band-level, not protocol-identical (finer grid, one idiom), and it does not extend to every idiom: the same run finds drum-kit onset counts near-white at the beat grid, with the multiplicative structure appearing on the velocity channel instead. Within those bounds, the property this section gates — the 1/f-family, positive-width statistical texture — is the measured texture of real performance in the instrument's nearest idiom, not only a designed target.

Reproducibility. The research tree (papers, datasets, figures) is published under CC BY 4.0 with a reproduction grant covering the engine's measurement crates, which are archived alongside this paper's deposit, and every number above is the output of a `cargo test` target: the hard gates run in CI on every commit, and a measurement reporter regenerates the full per-seed landscape (medians, per-seed values, and the Spread comparison) from the pinned protocol. Because the event layer is bit-identical everywhere, these are not "results on our machine": any reader will measure the same series to the bit, with only ulp-level variation in the floating-point estimation on top.

7. Design Reflections

What strictness bought. More than we expected. *Reproducibility* is the headline: a performance is a small state vector; a bug report replays exactly; a shared take is the identical event stream; and the instrument is learnable

in the specific sense that the same gesture from the same state always does the same thing — the felt correlation that teaches the model without ever naming it. *Testability* followed: an integer-deterministic core makes exhaustive property and golden testing cheap, and — the methodological point we would most like other instrument builders to take — it makes *statistical* properties CI-enforceable, because deterministic series cannot flake. The invariant suite in turn changes the economics of content: a scaffold is data that can be authored at volume, with CI proving each pack musical, bounded, register-safe, and spectrally alive before a human ever curates it. Finally, the pin gesture — to us the most instrument-like thing about the Fenophone — was not designed; it was *found*, because a model with an explicit, renderable latent state invites intervention on that state, and a freezable substream gives the intervention exact, glitch-free semantics. Staying strictly inside the mathematics surfaced an interaction a looser mapping would likely never have exposed.

What it cost. The fixed-point discipline is a real tax: no floating point in the event layer means owned fixed-point transcendentals evaluated only at control-change time, saturating arithmetic, a fixed evaluation order so no two builds disagree on a rounding tie, and test oracles (like the σ_d closed form, which needs a square root) exiled to float-legal test code. Transfer functions frozen at control-change time are a subtler cost: a continuous gesture is, to the engine, a sequence of frozen re-evaluations, so response smoothing must live downstream, and any future feature wanting continuously varying model parameters must be designed around the quantization boundary. And the stance constrains scope by construction: a feature that cannot be made integer-deterministic and gate-testable gets redesigned or cut.

Honest limits. Three should be stated plainly. First, *taste is authored, not emergent*. The cascade supplies motion with structure at every scale; consonance, voice-leading, groove, and the register architecture are scaffold data written by a human. The invariant suite proves a pack safe and structurally sound; the listening panel judges whether it is good; nothing about the model conjures musicality on its own, and we regard claims to the contrary — for this or any generative system — with suspicion. Second, *the strict-MSM claim has a boundary*. It covers intensity, rhythm, dynamics, timbre modulation, and harmonic bias — everything the lanes drive. The pitch contour is an AR(2)/Ornstein–Uhlenbeck-style walk: short-memory by design (that is what makes register gravity provable), so melodic *pitch increments* are not multifractal, and we do not claim they are. Driving the walk's step-noise scale σ from the cascade — multifractal pitch-increment volatility, the direct analogue of MSM's original role — has since landed

engine-side behind a per-voice scaffold field (`pitch_volatility`, default 0, so every launch pack is bit-identical to before pending a listening pass), and its measured effect is instructive: at full depth the regime-conditional mean step size swings ×1.6 with the cascade state, while the event-indexed DFA α of the snapped interval series moves only ≈ +0.003 — the variance coupling is live, but the chord-snap's crossing noise dominates the increments' scaling exponent at these run lengths (the engine's interval-series reporter documents both numbers). Its interaction with the register-gravity theorem (σ_d depends on σ) is policed exactly as designed: the two-minute drift gate re-runs every coupled voice at σ_max = σ·2^x. Third, the published spectral bands are *point measurements at one operating point* (the documented first-run controls) plus one interaction (Spread); the controllability suite sweeps every control for monotone trends, but a full map of the output spectrum over the control space remains to be published rather than gated. MFDFA beyond q ∈ {−2, 2} is now *measured* as a descriptive research surface (the full `h(q)` ladder over q ∈ {−4…4}, `research/data/spectrum-v1.csv`), as is the multi-voice ensemble path the product actually renders (per-voice, aggregate, and consonance-rank, `research/data/ensemble-v1.csv`; the aggregate two-point width regularizes below the per-voice widths, the Sinfonia pattern of [16]); both remain descriptive and ungated — the claim-bearing multifractality statistic stays the two-point width's IAAFT excess.

8. Future Work

Determinism makes the instrument an unusually clean stimulus generator, and the studies we most want to run are perceptual: surrogate discrimination (can listeners distinguish cascade-driven streams from bar-template-preserving surrogates — phase-randomized or shuffled — at matched densities?); validation of the tension lane against continuous perceived-tension ratings; and boundary detection, testing whether listeners segment the music at slow-layer switches they have never been shown. On the modeling side: corpus fitting — estimating MSM parameters from performed-music onset and dynamics series (the estimation machinery exists in the econometrics literature [7, 9]) and shipping fitted parameter regions as scaffold data. The multifractal account of pitch and microtiming, by contrast, has since landed rather than remaining an intention: cascade-modulated σ for the contour walk (Section 7's `pitch_volatility`) and a dedicated long-range-correlated microtiming sub-cascade in the spirit of [17] both ship behind the same invariant suite that gates everything else, disabled by default on the launch packs so that they stay bit-identical pending a listening pass. The microtiming layer — a dedicated second Markov-switching sub-cascade of the same family as the intensity core, each

onset's deviation a bounded linear image of that sub-cascade's net state, decorrelated per voice on a disjoint salted substream and clamped strictly within ±7/16 of a grid tick so onsets provably never reorder (a CI invariant, not a tendency) — is already measured: over the published seed panel (`research/data/microtiming-dfa-v1.csv`) its emitted sub-tick deviation series has DFA α ≈ 0.90 against a shuffled-null α ≈ 0.49 — an LRC excess of ≈ 0.41 that is itself CI-gated, i.e. humanization that is long-range correlated rather than white, and that never loops. That humanization has since been benchmarked against measured human drummers (195 Groove MIDI performances; `research/analysis/microtiming-benchmark/`, 2026-07-19): human drum microtiming is long-range correlated (DFA α ≈ 0.81) and the engine's deviation series (catalog-median α ≈ 0.85) sits inside the human range, with multifractality weak in both (width ≈ 0.05, near-monofractal). On regime structure — the property that separates the switching mechanism from Hennig-style linear filtered noise — human timing shows a *weak* sign-persistence excess over a phase-randomized LRC null, and the engine shows the same kind, running somewhat stronger than the median human; we report the overshoot as-is.

References

1. R. F. Voss and J. Clarke, "'1/f noise' in music and speech," *Nature*, vol. 258, pp. 317–318, 1975.
2. R. F. Voss and J. Clarke, "'1/f noise' in music: Music from 1/f noise," *Journal of the Acoustical Society of America*, vol. 63, no. 1, pp. 258–263, 1978.
3. M. Gardner, "Mathematical games: White and brown music, fractal curves and one-over-f fluctuations," *Scientific American*, vol. 238, no. 4, 1978.
4. B. B. Mandelbrot, "Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier," *Journal of Fluid Mechanics*, vol. 62, pp. 331–358, 1974.
5. B. B. Mandelbrot, A. Fisher, and L. Calvet, "A multifractal model of asset returns," Cowles Foundation Discussion Paper No. 1164, Yale University, 1997.
6. L. Calvet and A. Fisher, "Forecasting multifractal volatility," *Journal of Econometrics*, vol. 105, pp. 27–58, 2001.
7. L. E. Calvet and A. J. Fisher, "How to forecast long-run volatility: Regime switching and the estimation of multifractal processes," *Journal of Financial Econometrics*, vol. 2, no. 1, pp. 49–83, 2004.
8. L. E. Calvet and A. J. Fisher, *Multifractal Volatility: Theory, Forecasting, and Pricing*. Academic Press, 2008.
9. T. Lux, "The Markov-switching multifractal model of asset returns: GMM estimation and linear forecasting of volatility," *Journal of Business & Economic Statistics*, vol. 26, no. 2, 2008.
10. C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, "Mosaic organization of DNA nucleotides," *Physical Review E*, vol. 49, pp. 1685–1689, 1994.

11. J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. E. Stanley, “Multifractal detrended fluctuation analysis of nonstationary time series,” *Physica A*, vol. 316, pp. 87–114, 2002.

12. K. J. Hsü and A. J. Hsü, “Fractal geometry of music,” *Proceedings of the National Academy of Sciences*, vol. 87, pp. 938–941, 1990.

13. D. J. Levitin, P. Chordia, and V. Menon, “Musical rhythm spectra from Bach to Joplin obey a 1/f power law,” *Proceedings of the National Academy of Sciences*, vol. 109, no. 10, pp. 3716–3720, 2012.

14. Z.-Y. Su and T. Wu, “Multifractal analyses of music sequences,” *Physica D*, vol. 221, pp. 188–194, 2006.

15. G. R. Jafari, P. Pedram, and L. Hedayatifar, “Long-range correlation and multifractality in Bach’s Inventions pitches,” *Journal of Statistical Mechanics: Theory and Experiment*, 2007.

16. L. Telesca and M. Lovallo, “Revealing competitive behaviours in music by means of the multifractal detrended fluctuation analysis: application to Bach’s Sinfonias,” *Proceedings of the Royal Society A*, vol. 467, 2011.

17. H. Hennig, R. Fleischmann, A. Fredebohm, Y. Hagmayer, J. Nagler, A. Witt, F. J. Theis, and T. Geisel, “The nature and perception of fluctuations in human musical rhythms,” *PLoS ONE*, vol. 6, no. 10, e26457, 2011.

18. H. Hennig, “Synchronization in human musical rhythms and mutually interacting complex systems,” *Proceedings of the National Academy of Sciences*, vol. 111, no. 36, 2014.

19. E. Räsänen, O. Pulkkinen, T. Virtanen, M. Zollner, and H. Hennig, “Fluctuations of hi-hat timing and dynamics in a virtuoso drum track of a popular music recording,” *PLoS ONE*, vol. 10, no. 6, e0127902, 2015.

20. L. A. Hiller and L. M. Isaacson, *Experimental Music: Composition with an Electronic Computer*. McGraw-Hill, 1959.

21. I. Xenakis, *Formalized Music: Thought and Mathematics in Composition*. Pendragon Press, 1992.

22. C. Ames, “The Markov process as a compositional model: A survey and tutorial,” *Leonardo*, vol. 22, no. 2, pp. 175–187, 1989.

23. C. Dodge, “Profile: A musical fractal,” *Computer Music Journal*, vol. 12, no. 3, pp. 10–14, 1988.

24. J. Chadabe, “Interactive composing: An overview,” *Computer Music Journal*, vol. 8, no. 1, pp. 22–27, 1984.

25. B. Eno, “Generative music,” talk transcript, *In Motion Magazine*, 1996.

26. P. R. Cook, “Principles for designing computer music controllers,” in *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, 2001.

27. D. Wessel and M. Wright, “Problems and prospects for intimate musical control of computers,” *Computer Music Journal*, vol. 26, no. 3, pp. 11–22, 2002.

28. S. Jordà, G. Geiger, M. Alonso, and M. Kaltenbrunner, “The reacTable: Exploring the synergy between live music performance and tabletop tangible interfaces,” in *Proceedings of the International Conference on Tangible and Embedded Interaction (TEI)*, 2007.

29. J. Drummond, “Understanding interactive systems,” *Organised Sound*, vol. 14, no. 2, pp. 124–133, 2009.

30. E. I. Jury, *Theory and Application of the z-Transform Method*. Wiley, 1964.

31. T. Bolognesi, “Automatic composition: Experiments with self-similar music,” *Computer Music Journal*, vol. 7, no. 1, pp. 25–36, 1983.

32. J. Pressing, “Nonlinear maps as generators of musical design,” *Computer Music Journal*, vol. 12, no. 2, pp. 35–46, 1988.

33. G. Díaz-Jerez, *Algorithmic Music: Using Mathematical Models in Music Composition* (the FractMus system), D.M.A. thesis, Manhattan School of Music, 2000.

34. B. Manaris, J. Romero, P. Machado, D. Krehbiel, T. Hirzel, W. Pharr, and R. B. Davis, “Zipf’s law, music classification, and aesthetics,” *Computer Music Journal*, vol. 29, no. 1, pp. 55–69, 2005.

35. H. Scurto and L. Postel, “Soundwalking deep latent spaces,” in *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, 2023.

36. V. Shepardson and T. Magnusson, “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)*, 2023.

37. T. Hermann and H. Ritter, “Listen to your data: Model-based sonification for data analysis,” in *Advances in Intelligent Computing and Multimedia Systems*, G. E. Lasker, Ed. Baden-Baden: IIAS, 1999, pp. 189–194.

38. J.-P. Bouchaud, M. Potters, and M. Meyer, “Apparent multifractality in financial time series,” *European Physical Journal B*, vol. 13, pp. 595–599, 2000.

39. T. Schreiber and A. Schmitz, “Improved surrogate data for nonlinearity tests,” *Physical Review Letters*, vol. 77, pp. 635–638, 1996.

40. J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, “Testing for nonlinearity in time series: The method of surrogate data,” *Physica D*, vol. 58, pp. 77–94, 1992.

41. A. Banerjee, S. Sanyal, T. Guhathakurata, R. Sengupta, and D. Ghosh, “Categorization of stringed instruments with multifractal detrended fluctuation analysis,” arXiv:1601.07709, 2016.

42. D. Ghosh, R. Sengupta, S. Sanyal, and A. Banerjee, “Improvisation—A new approach of characterization,” in *Musicality of Human Brain through Fractal Analytics*, ch. 9, pp. 185–212. Singapore: Springer, 2018.

43. A. Zlatintsi and P. Maragos, “Multiscale fractal analysis of musical instrument signals with application to recognition,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 4, pp. 737–748, 2013.

44. K. K. Zhang, Y. Sun, and H. Xiong, “Multifractal comparison of Billboard and AI-generated music,” in *Proceedings of the 33rd ACM International Conference on Multimedia (MM ’25)*, pp. 12419–12427, 2025.

45. T. C. Roeske, D. Kelty-Stephen, and S. Wallot, “Multifractal analysis reveals music-like dynamic structure in songbird rhythms,” *Scientific Reports*, vol. 8, 4570, 2018.

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

research series, paper 03). Zenodo DOI 10.5281/zenodo.22180429 (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).