— 8 PRE-REGISTERED RUNS (SA-01 TO SA-08) + LATENT HARMONIC CONTROL ARCHITECTURE

STABLE AUDIO 3 & CSOUND // FROM THE RESTYLE CLIFF TO LATENT SUBSPACE CONTROL

Consolidating all eight October 2026 Stable Audio 3 empirical experiments (sa3.cpp CPU + HF Space + fal.ai LoRA trainer, scored against the 33.1% 4-note chord and 56.9%-59.6% 7-note scale noise floors) with the mathematical Csound x Stable Audio latent-steering framework and the pre-registered SA-09 to SA-11 experiment plan.

16/16 → 5/16
SA-05 RESTYLE CLIFF
Csound flute restyle keeps 16/16 notes at τ=0.4, 13/16 at τ=0.5, and collapses to 2-5/16 at τ≥0.6

19.2% → 96.6%
SA-06 D7 VOCODER LOCK
Csound pvsvoc rescues off-key vox D7 bar (vs. 65.3% resonz, 13% tritone control & 33.1% chord floor)

90% → 94%
SA-07 HYBRID 7-SEC BEAT
76.7s, 90 BPM G phrygian master across all 7 sections (vs. 71–73% text-only & 56.9% scale floor)

+0.17 → +0.24
SA-08 $1.20 LORA PILOT
Rank-8 LoRA (128 THIRI clips) lifted C#m EP minorness 0.05→0.52, short of +0.40 target (5/8 pairs)

01 // EMPIRICAL SCORECARD // THE 8 COMPLETED STABLE AUDIO 3 EXPERIMENTS

RUN ID	EXPERIMENT & QUESTION	CONTROL & NOISE FLOOR	MEASURED GROUND-TRUTH RESULT	ARCHITECTURAL TAKEAWAY
SA-01	**Extend (8s → 12s / 20s)** Can SA3 continue an 8s G phrygian pad?	8s source pad (seed 11); scale floor 56.9%	Kept 0–8s matches (r = 0.993), then drops into a **1s silent hole** (-61.2 dB on 12s; -45.3 dB on 20s; -48.7 dB medium) + adds out-of-key B♮ (3.8%).	Outpainting from an hard boundary at 8.0s resets the latent energy envelope and drifts out of mode.
SA-02	**Inpaint (Repair 2s)** Can SA3 rewrite a bad half-bar in place?	Unmasked region; G phrygian floor 56.9%	Outside 3–5s untouched (r = 0.993), inside rewritten (r = 0.021). Fixing a half-bar with stray F# (23.4%) lifted G phrygian share 55.4% → 85.3% (90.7% full loop).	Bidirectional context from left + right neighbors constrains pitch classes far better than open-ended continuation.
SA-03	**Loops (Bars 2-5 Crop)** How do you get a seamless 4-bar loop?	Direct 4-bar request (10.67s); C minor floor 56.9%	Direct 4-bar take writes an ending (last 50 ms at -95.7 dB , 63.5% in key). Asking for 24s and cropping bars 2–5 yields 470,400 samples , 16/16 beats (4.8 ms mean err), and 90.9% scale share.	Always generate long (24s) and harvest interior bars (2-5) with a 2 ms boundary crossfade.
SA-04	**Prompt Words Sweep** Which musical words does SA3 obey?	No-key baseline (k0-none); C#m floor 59.5% , chord floor 33.1%	Obeys **BPM** (20/20 at 90/180 BPM) & **Root** (C#m share rose 39% → 67% HF, 36% → 58% local). Ignores **Mode** (C#m read **Major 4/4 times**; G phrygian 84% vs 85% none) & **Chords** (2/12 bars).	CLAP/T5 text tokens steer tempo and root pitch center, but have zero control over modal quality or bar-by-bar chords.
SA-05	**The Restyle Cliff** How far can SA3 restyle a Csound melody?	16-note Csound G phrygian melody (490 Hz centroid)	Keeps 16/16 notes at τ=0.3, 0.4 (r=0.91-0.95), 13-14/16 at τ=0.5 (99.5% scale), then drops off a cliff to 2-5/16 notes at τ=0.6-0.8 (r=0.01) as it converges to its training prior.	Single-timbre VAE latents entangle pitch and timbre: breaking the Csound timbre cage at τ≥0.6 erases the notes.
SA-06	**Forcing Chords Afterwards** Can Csound lock Gm7 | Ebmaj7 | Cm7 | D7 on Side B?	Untreated SA3 (19-77%); tritone 8-15% ; chord floor 33.1%	**Csound pvsvoc (2048/256):** 86%-98% every bar across all 3 takes (vox D7 bar jumped 19.2% → 96.6% ; 75% wet kept 90-97%). **Resonator (resonz Q=60):** 84-99% except vox D7 (65.3% , source had 0.4% A).	Passive filters (resonz) cannot ring missing leading tones; active Csound phase vocoding (pvsvoc) builds them.
SA-07	**The Hybrid Beat** Csound writes notes, SA3 restyles at τ=0.50-0.55	Text-only SA3 pad (73%) & vox (71%); scale floor 56.9%	76.67s , 90 BPM 7-section master holds 90%-94% on G phrygian across all sections (restyled pad 97.0% , flute 97.8% , vox 89.7% , Csound bass 91.3%).	Operating right at the τ=0.50-0.55 boundary delivers real instrument ceiling (>90%) across a full arrangement.
SA-08	**Chord Adapter Pilot** Can a $1.20 rank-8 LoRA teach minor vs. major?	Base SA3 (+0.17 gap, 4/8 pairs); target ≥+0.40 & 6/8	**Clean negative result:** Held-out root minorness gap rose +0.17 → +0.24 (5/8 pairs vs 6 needed). Triad share rose 67% → 74% and C#m EP jumped 0.05 → 0.52 , but root priors on Eb, F#, B won.	400-step rank-8 text adapter nudges attention in the right direction but cannot overcome deep VAE/DiT root priors alone.

CENTRAL SYNTHESIS // WHY THE THOUGHT EXPERIMENT IS NOW EMPIRICALLY GROUNDED

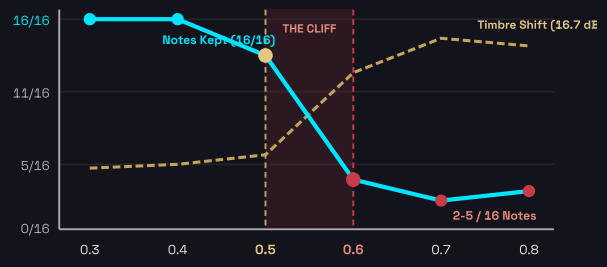
Runs **SA-01 through SA-08** prove that text prompting (SA-04 : 2/12 chord bars) and light text-LoRA (SA-08 : +0.24 vs +0.40) cannot steer harmony from the top down, while single-stem audio conditioning hits a hard phase transition at τ = 0.5 → 0.6 (SA-05) and passive Side B filtering stalls when leading tones are absent (SA-06 : 65.3% on D7). The remaining three pages show how **Csound multi-timbre basis probing** and **Side B pvsvoc + VAE healing** solve both walls.

02 // WALL 01 // THE 0.5 TO 0.6 RESTYLE CLIFF (RUN SA-05 DOSE-RESPONSE)

INIT NOISE (T)	NOTES KEPT (FLUTE S1/S2 · TRPT S1/S2)	ENVELOPE CORR	DIST TO SOURCE	DIST TO PRIOR (T=1.0)
τ = 0.30	16/16 · 16/16	0.95 / 0.94	6.1 / 5.9 dB	9.1 / 15.1 dB
τ = 0.40	16/16 · 16/16	0.91 / 0.92	6.2 / 6.9 dB	8.8 / 14.9 dB
τ = 0.50 (Edge)	13/16, 13/16 · 14/16, 13/16	0.47 / 0.57	6.9 / 7.2 dB	8.5 / 15.7 dB
τ = 0.60 (Cliff)	5/16, 3/16 · 2/16, 2/16	0.01 / 0.14	13.8 / 8.5 dB	6.2 / 12.7 dB
τ = 0.70	2/16, 2/16 · 3/16, 2/16	0.03 / 0.05	16.7 / 10.5 dB	5.4 / 4.8 dB
τ = 0.80	3/16, 3/16 · 2/16, 4/16	0.08 / 0.05	16.1 / 11.7 dB	3.3 / 3.0 dB

Why the Cliff Happens: The Csound source centroid sits at 490 Hz. Up to τ = 0.5, the VAE latent z_cs pins the spectral centroid to 468-484 Hz. At τ = 0.6, enough Gaussian noise is injected to break Csound's spectral envelope (330 Hz flute / 923 Hz trumpet), but because pitch and timbre are entangled in a single stem's latent encoding, breaking the timbre cage destroys 75% of the written notes.

FIG 2.1 // SA-05 DOSE-RESPONSE PHASE TRANSITION



03 // WALL 02 // MISSING LEADING TONES ON D7 (RUN SA-06 RESONATOR VS. VOCODER)

SA3 SOURCE TAKE (90 BPM)	UNTREATED CHORD SHARE (GM7 | EBMAJ7 | CM7 | D7)	CSOUND RESONZ BANK (Q=60, 2 OCTAVES)	CSOUND PVSVOC VOCODER (2048/256, SAW CARRIER)	TRITONE CONTROL (RES / VOC)	RHYTHM KEPT (RES / VOC)
01. Vinyl & Rain (Broadband)	29% | 30% | 30% | 33% (at 33.1% noise floor)	86% | 88% | 84% | 88%	95% | 95% | 97% | 94%	15% / 13%	0.49 / 0.85
02. Off-Key Vocal Chops	77% | 77% | 66% | 19.2% (below 33.1% floor)	97% | 99% | 97% | 65.3% (Stalled)	96% | 98% | 98% | 96.6% (75% wet: 90–97%)	13% / 13%	0.60 / 0.68
03. Lo-Fi Beat Loop	29% | 50% | 23% | 28% (below 33.1% floor)	91% | 99% | 90% | 89%	86% | 94% | 92% | 87%	8% / 10%	0.38 / 0.89

ACOUSTIC PROOF // WHY RESONZ STALLED AT 65.3% ON BAR 4 (D7)

A Resonator Cannot Ring Pitches That Are Not There

In SA-06, the progression is in **G minor** (Gm7 | Ebmaj7 | Cm7 | D7, voiced by thiri_voice --voicing jazz4 : F3 D4 G4 Bb4 | Eb3 D4 G4 Bb4 | Eb3 C4 G4 Bb4 | D3 C4 F#4 A4), whereas the vocal chops were generated in **G phrygian** (G Ab Bb C D Eb F). Both F# and A sit outside G phrygian! Measuring Bar 4 of the untreated vocal chops revealed only 4.0% F# and 0.4% A (vs. 9% F# / 14% A in broadband vinyl noise). Consequently, the passive resonz filter stalled at 65.3%, while Csound's pvsvoc carrier actively synthesized the missing F#4 and A4 partials to reach 96.6%.

ADAPTER PROOF // WHY RANK-8 LORA STALLED AT +0.24 (SA-08)

Cross-Attention LoRA vs. Root-Specific Modal Bias

In SA-08, training a $1.20 rank-8 LoRA for 400 steps on 128 THIRI Csound clips lifted the held-out minorness gap from +0.17 to +0.24 and triad share from 67% → 74%, with C# minor electric piano jumping from 0.05 → 0.52 minorness. Yet on Eb (always minor-3rd) and F# / B (always major-3rd), the base DiT root priors overpowered the text adapter at strength 1.0. Direct latent-space intervention is required.

04 // FOUR CSOUND X STABLE AUDIO MECHANISMS GROUNDED IN RUNS SA-01 TO SA-08

MECHANISM I // SIDE A CLIFF-EDGE INIT

Broadband Waveguide Exciter at the τ* = 0.50 → 0.55 Boundary (Proven in SA-07)

MEASURED: 90%-98% SCALE SHARE

Run SA-07 proved that initializing Stable Audio 3 from Csound stems right at the $\tau^* \in [0.50, 0.55]$ cliff edge achieves 97.0% (pad), 97.8% (flute), and 89.7% (vox) scale share vs. 56.9% noise floor. To maximize timbral freedom at $\tau^* = 0.55$ without losing upper chord extensions, Csound renders $x_cs(t)$ using band-limited pulse tables (vco2) + modal waveguides (wgbow / mode) with +3.0 dB emphasis on the 3rd and 7th guide tones (F#4, C4):

$$z_cs = E_VAE(x_cs) \in R^{64 \times T}, \quad z_\tau^* = \alpha_\tau^* z_cs + \sigma_\tau^* \varepsilon, \quad \tau^* \in [0.50, 0.55]$$ [EQ. 1 // CLIFF-EDGE HANDOFF]

MECHANISM II // CROSSING THE T=0.60 CLIFF

Multi-Timbre Latent Centroid μ_harm^(b) & Counterfactual Leading-Tone Vector v_lead^(b)

SOLVES SA-05 & SA-06 D7 TRAP

Why did SA-05 collapse from 13/16 notes at τ=0.50 to 5/16 at τ=0.60 ? Because a single Csound stem's latent z_cs entangles pitch with Csound's 490 Hz saw/FM timbre. Because Csound compiles at 492x real-time , we synthesize the exact same THIRI voicing across $M = 8$ orthogonal Csound timbres (sine, FM foscili , plucked waveguide wgpluck2 , bowed string, formant pulse fofilter) and encode all M stems into the VAE:

1. Timbre-Invariant Harmonic Centroid: Averaging the M latent encodings yields $\mu_harm^{(b)} = (1/M) \Sigma_m E_VAE(x_cs,m^{(b)})$, which cancels out instrument-specific formants while preserving the common fundamental and overtone phase across all M timbres.

2. Counterfactual F# & A Vector: To prevent Bar 4 (D7) from losing F#4 and A4 inside a G-phrygian / G-minor context, Csound renders both the target D7 (D3 C4 F#4 A4) and the modal trap foil (Fm7 / Dm7b5 with F♮ , Ab). Their difference vector $v_lead^{(b)}$ steers the DiT across the harmonic boundary without restricting $U_timbre^{(b)}$.

[EQ. 2 // TIMBRE NULLSPACE PROJECTOR]
$$P_harm^{(b)} = I - U_timbre^{(b)} (U_timbre^{(b)})^T$$

[EQ. 3 // COUNTERFACTUAL D7 VECTOR]
$$v_lead^{(b)} = E_VAE(x_cs^{(b, F\#+A)}) - E_VAE(x_cs^{(b, F♮ +Ab)})$$

[EQ. 4 // STEERED DENOISING UPDATE]
$$\hat{z}_0|\tau \leftarrow \hat{z}_0|\tau - \eta P_harm^{(b)}(\hat{z}_0|\tau - \mu_harm^{(b)}) + \gamma v_lead^{(b)}$$

MECHANISM III // CONTRASTIVE CQT

Physical Overtone Guidance Against the +6 Semitone Tritone Foil Control

TRITONE CONTROL: 8%-15% VS 33.1% FLOOR

In SA-06 , the tritone control scored 8%-15% across 4 bars (below the 33.1% 4-note chord noise floor). By rendering both the target Csound stem $x_cs^{(+)}$ and its tritone-transposed Csound twin $x_cs^{(tritone)}$, we can guide intermediate DiT steps ($\tau \in [0.75, 0.35]$) using physical Csound chromagrams rather than metallic binary masks:

$$L_THIRI(\hat{z}_0|\tau) = \Sigma_b \, || \, CQT_b(D_VAE(\hat{z}_0|\tau)) - CQT_b(x_cs^{(+)}) \, ||_2^2 - \lambda \, || \, CQT_b(D_VAE(\hat{z}_0|\tau)) - CQT_b(x_cs^{(tritone)}) \, ||_2^2$$ [EQ. 5 // CONTRASTIVE CQT LOSS]

MECHANISM IV // SIDE B PVS + LATENT HEAL

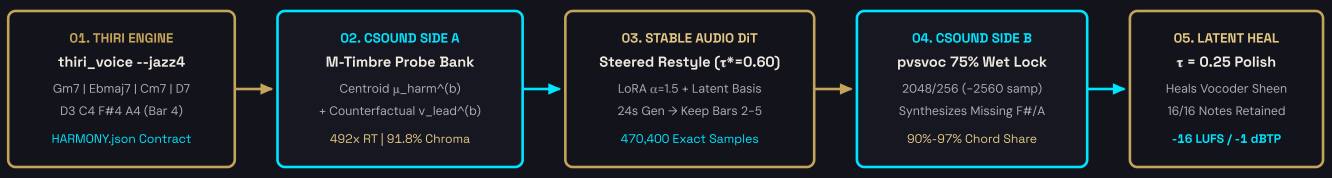
Csound pvsvoc (75% Wet, 2560-Sample Pre-Shift) + τ = 0.25 VAE Polish

MEASURED: 90%-97% CHORD SHARE

Run SA-06 proved that Csound's pvsvoc (`fftsize=2048, hop=256`, shifted 2560 samples earlier for zero latency) at 75% wet blend locks every bar to 86%-97% chord tones while preserving 0.68-0.89 rhythm envelope correlation. To remove the synthetic sawtooth vocoder sheen without losing a single note, we pass the 75% wet pvsvoc stem back through Stable Audio at $\tau = 0.25$ (safely below the $\tau = 0.40$ 16/16 plateau from SA-05):

$$x_voc = Blend_0.75(pvsvoc(x_sa, x_cs_saw), x_sa) \rightarrow x_master = D_VAE(DiT(E_VAE(x_voc), \tau=0.25))$$ [EQ. 6 // PVSVOC + T=0.25 HEAL]

FIGURE 3.0 // END-TO-END HYBRID SIGNAL PIPELINE (UNITING SA-05, SA-06, SA-07 & SA-08) From THIRI jazz4 Voicings → Csound Multi-Timbre Basis → Stable Audio DiT → Side B pvsvoc + τ=0.25 Heal



05 // PRE-REGISTERED EXPERIMENT SUITE // RUNS SA-09, SA-10 & SA-11

EXP SA-09 // CROSSING THE CLIFF

Multi-Timbre Csound Basis Init at τ = 0.60 & 0.65

- **Question:** Can averaging `M = 6` orthogonal Csound timbres (sine, FM, bowed wg, plucked wg, formant pulse, organ) push note retention across the `τ = 0.60` cliff?
- **Control:** Single-timbre `melody-csound.wav` at `τ = 0.60` (`5/16 flute, 2/16 trumpet` in `SA-05`) and `56.9%` G phrygian noise floor.
- **Success Bar:** At `τ = 0.60`, notes kept rises from `5/16` to `≥ 12/16` while spectral timbre shift exceeds `≥ 10.0 dB` from the Csound centroid.

EXP SA-10 // SIDE B LATENT HEAL

Csound pvsvoc (75% Wet) + τ = 0.25 / 0.35 Polish

- **Question:** Does passing `SA-06`'s `75%` wet `pvsvoc` stems back through SA3 img2img at `τ ∈ {0.25, 0.35}` remove sawtooth vocoder sheen while keeping the `D7` chord lock?
- **Control:** Untreated SA3 (`19.2%` vox D7), raw `pvsvoc` (`96.6%`), tritone foil (`13%`), and `33.1%` 4-note chord floor.
- **Success Bar:** Every bar holds `≥ 80%` chord share (vox D7 `≥ 75%`) while spectral distance to natural SA3 vocal envelope improves by `≥ 4.0 dB`.

EXP SA-11 // LORA + CSOUND HYBRID

Inference LoRA Strength (α = 1.5, 2.0) + Root-5th Init

- **Question:** Can `SA-08`'s `lora-chordpilot-f32.gguf` cross the `+0.40` minorness bar on held-out roots (`C#, Eb, F#, B`) at `α ∈ {1.5, 2.0}` + `τ = 0.65` Csound drone?
- **Control:** Base model (`+0.17` gap, `4/8` pairs) and `α = 1.0` LoRA (`+0.24` gap, `5/8` pairs) at zero extra training cost.
- **Success Bar:** Held-out minor minus major gap reaches `≥ +0.40` and minor beats major in `≥ 6/8` (`root, instrument`) pairs.

06 // REFERENCE IMPLEMENTATION // HARMONY.JSON CONTRACT & CSOUND PVSVOC ORCHESTRA

PRE-REGISTERED CONTRACT // HARMONY.JSON G MINOR // 90 BPM

```
// Exact voicings from SA-06 (thiri_voice --voicing jazz4, rev 04a161e)
{
  "title": "SA3_Csound_Latent_Suite_Gm_90BPM",
  "key": "G minor",
  "key_set": "Eb,F,G,Ab,Bb,C,D",
  "bpm": 90,
  "bar_seconds": 2.6667,
  "bars": ["Gm7", "Ebmaj7", "Cm7", "D7"],
  "jazz4_voicings_midi": {
    "Gm7":    ["G2", "F3",  "D4", "G4",  "Bb4"],
    "Ebmaj7": ["Eb2", "Eb3", "D4", "G4",  "Bb4"],
    "Cm7":    ["C2", "Eb3", "C4", "G4",  "Bb4"],
    "D7":     ["D2", "D3",  "C4", "F#4", "A4"]
  },
  "chance_floors": { "chord_4_notes": 0.331, "key_7_notes": 0.569 }
}
```

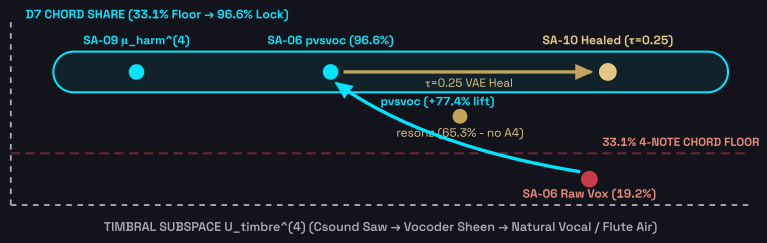
CSOUND ORCHESTRA // SA3_FORCE.CSD 44.1 KHZ | KSMPS=32

```
; Instr 1: Multi-Timbre Carrier & Basis Probe (SA-09)
instr 1
  icps  = p4                ; Exact THIRI jazz4 frequency (Hz)
  asaw  vco2 p5, icps, 0    ; Band-limited sawtooth stack
  awg   wgbow p5, icps, 3.0, 0.14, 0.02, 4.5
  ga_car = ga_car + (asaw * 0.65) + (awg * 0.35)
endin

; Instr 2: Side B pvsvoc Phase Vocoder (SA-06 & SA-10, 96.6% D7 Lock)
instr 2
  fmod  pvsanal ga_sa3_raw, 2048, 256, 2048, 1
  fcar  pvsanal ga_car,     2048, 256, 2048, 1
  fvoc  pvsvoc  fmod, fcar, 1.0, 1.0  ; Impose SA3 formants on THIRI chord
  avoc  pvsynth fvoc
  aout  = (avoc * 0.75) + (ga_sa3_raw * 0.25) ; 75% wet blend (≥90% share)
        outs aout, aout
endin
```

FIGURE 4.0 // BAR 4 (D7: D3-C4-F#4-A4) LATENT TRAJECTORY ACROSS SA-05, SA-06, SA-09 & SA-10

How Multi-Timbre Basis Probing (SA-09) and Side B pvsvoc + τ=0.25 Heal (SA-10) Escape the G-Phrygian Trap



EXECUTION ORDER FOR NEXT SPRINT

1. Run SA-11 first (0 training cost: local sa3.cpp sweep of lora-chordpilot-f32.gguf at α=1.5, 2.0).
2. Run SA-10 next (feed existing vox-offkey-voc-blend75 into SA3 img2img at τ=0.25 & 0.35, score with force.py).
3. Run SA-09 (render 6-timbre Csound melody bank, **test τ=0.60 restyle cliff with curve.py score).**