
ASCF

Auditory Stimulus Cardiac Fingerprinting

The Ataraxia Project · COS (Cognitive Operating System) · Vol. 1

Shrey S. Kumar · May 2026

Track: Tinzo · n = 4 sessions · Apr 27 – May 1, 2026

1.37%

coefficient
of variation

~0

BPM–HR
correlation

4.5 bpm

max delta
at fingerprint

+2.3

bpm/session
sensitization

Abstract

We present *Auditory Stimulus Cardiac Fingerprinting* (ASCF, pronounced "askif") — a methodology for measuring and characterising person-specific cardiac responses to music using consumer-grade wearable biometrics and speaker track telemetry. Across four shower sessions spanning five days, heart rate data sampled at 5-second intervals from a WHOOP wearable was aligned to track position (seconds into a DJ mix) rather than wall-clock time, enabling direct cross-session comparison at musical-moment resolution. At track position 1785 seconds (29.8 min), four sessions with entry heart rates ranging from 72 to 109 bpm converged to within 4.5 bpm (mean 134.1, $SD \pm 1.86$, CV 1.37%). Song tempo showed near-zero correlation with cardiac response ($r \approx 0.00$ to -0.19), eliminating rhythmic entrainment. Respiratory rate similarly showed near-zero correlation with heart rate, confirming the two systems operate independently during music-induced arousal. Spectral flux correlated moderately-to-strongly with heart rate ($r = 0.40$ – 0.68), pointing to timbral complexity as the primary driver. A sensitization trend of +2.3 bpm per session ($R^2 = 0.89$) suggests early neural potentiation. This paper presents the methodology, data, heuristics, and a framework for replication.

Keywords auditory cardiac fingerprinting · psychophysiology · dopaminergic anticipation · quantified self · spectral flux · n-of-1 research · timbral memory encoding · hedonic adaptation · RR decoupling

1. Introduction

The relationship between music and physiological arousal is well-established. Studies have consistently demonstrated that music modulates heart rate, skin conductance, and respiratory rate. However, most existing research treats cardiac response to music as a population-level phenomenon. The question of whether a specific individual's cardiac response to a specific musical moment is reproducible across days, baseline states, and contexts has received comparatively little attention.

This paper emerges from a personal biometric infrastructure built for continuous self-monitoring. A home automation system ingests physiological data from a WHOOP wearable — heart rate (5-second resolution) and respiratory rate — alongside speaker telemetry tracking track name, source, and position in seconds. The collision of these two streams during routine shower sessions produced an unexpected observation: the cardiac response to a specific DJ mix recurred with striking consistency across multiple sessions on different days.

This paper formalises that observation into a methodology, presents data from four sessions, derives seven heuristics, and proposes a mechanistic explanation grounded in dopaminergic anticipatory processing and timbral memory encoding. It is presented as a pilot study — explicitly $n = 1$, single track, pre-control-condition — and makes no claim to generalisability beyond the subject.

1.1 The ASCF Hypothesis

We define an Auditory Stimulus Cardiac Fingerprint (ASCF) as a reproducible, person-specific pattern of cardiac arousal elicited by a specific musical stimulus, characterised by a consistent absolute heart rate ceiling at identifiable moments in the track, independent of baseline physiological state. The hypothesis has three components: (1) the absolute HR ceiling is more reproducible than the relative lift from baseline; (2) the mechanism is timbral memory encoding rather than rhythmic entrainment; and (3) the fingerprint is subject to dopaminergic refractory dynamics including sensitization with moderate exposure and habituation with overexposure.

2. Methodology

2.1 Instrumentation Stack

The data collection pipeline was built on Home Assistant as the central event bus:

Component	Role	Notes
WHOOP 4.0	Heart rate (5-second resolution), respiratory rate	Worn on wrist throughout
Sonos Roam	Track name, source, position in seconds	Speaker physically inside shower enclosure
Home Assistant	Central event bus, history CSV export	All entities timestamped to the second
Python / librosa	Audio feature extraction from mix file	BPM, RMS loudness, spectral flux

2.2 Track Position Alignment

The critical methodological innovation is alignment by track position rather than wall-clock time. Each HR reading is mapped to the concurrent track position reading from the Sonos Roam, enabling comparison at the same musical moment across sessions recorded on different days. When a track is manually cued, the position sensor briefly resets to zero before settling. Readings below 500 seconds are excluded as startup artifacts. The subject deliberately cued the mix approximately 3–4 minutes before the target section — a practised pre-cue behaviour described as "methodical environmental staging."

2.3 WHOOP Heart Rate Sampling

The WHOOP 4.0 samples heart rate at 5-second intervals during activity, providing sufficient temporal resolution to capture music-induced cardiac arousal dynamics which unfold over minutes. Respiratory rate is captured concurrently as a secondary signal for decoupling analysis.

2.4 Audio Feature Extraction

The Tinzo mix was analysed using librosa to extract beat timestamps and local BPM, RMS energy in 5-second windows, and spectral flux — the frame-to-frame change in spectral magnitude, representing harmonic novelty. Global BPM was 123.0. Local BPM variance across the 14.8-minute overlap window was 8.3 BPM (range 120.9–129.2) — effectively flat.

2.5 Session Conditions

All four sessions were conducted in a home shower with the Sonos Roam speaker physically present inside the shower enclosure. Baseline HR varied across sessions (72–109 bpm), reflecting different pre-session states. Session duration was 17–20 minutes.

3. Results

3.1 Fingerprint Convergence

Figure 1 shows the four-session cardiac overlay. Despite entry heart rates of 73, 72, 109, and 77 bpm, all four sessions converged at track position 1785 seconds (29.8 min) to within 4.5 bpm (values: 132.0, 136.5, 134.7, 133.3 bpm; mean 134.1, SD ± 1.86 , CV 1.37%). A CV below 5% is the standard threshold for high reproducibility in psychophysiology.

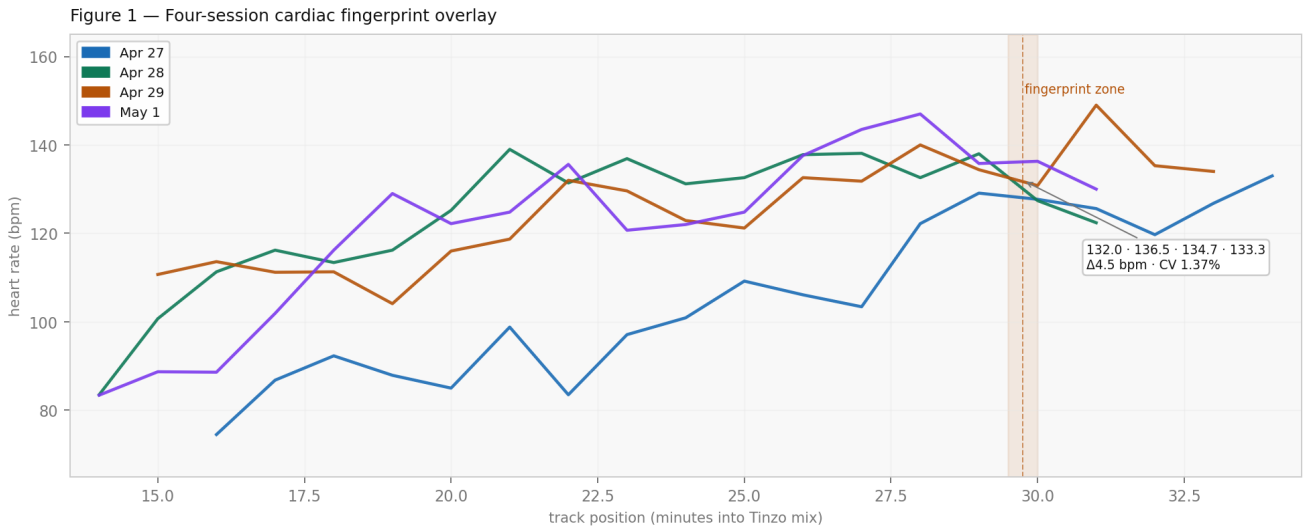


Figure 1 — Four-session HR overlay aligned by track position. The amber shaded region denotes the fingerprint zone (1770–1800s / 29.5–30.0 min). All four sessions converge within 4.5 bpm despite different entry states and trajectory shapes.

The Apr 28 session shows an anomalous spike attributable to an environmental acoustic event (the Sonos Roam falling inside the shower enclosure), consistent with a startle reflex: rapid onset, short duration, full morphological recovery within two minutes. This is morphologically distinct from music-induced sustained arousal and incidentally validates the sensor's temporal fidelity.

3.2 Audio Feature Correlation and RR Decoupling

Figure 2 overlays HR with spectral flux (cyan) and song BPM (dashed red). Song BPM showed near-zero correlation with HR across all four sessions ($r = 0.038, -0.171, 0.061, -0.187$), eliminating rhythmic entrainment. Respiratory rate similarly showed near-zero correlation, confirming the cardiac response is CNS-mediated rather than driven by breathing or metabolic demand.

Figure 2 — HR vs audio features (spectral flux, song BPM)

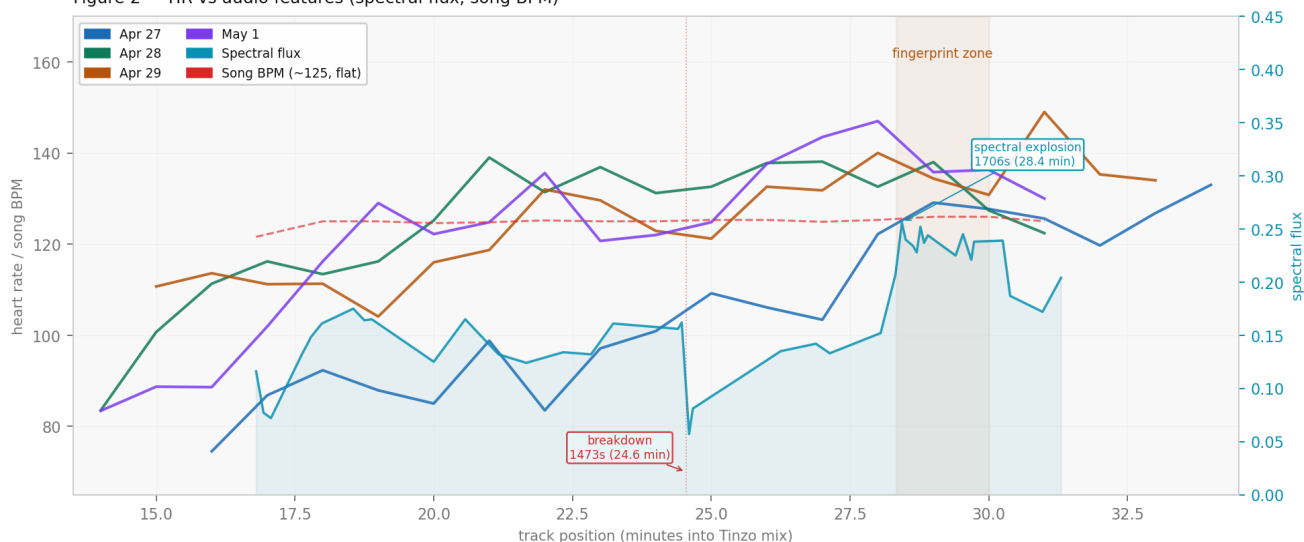


Figure 2 — HR from all four sessions with spectral flux (cyan) and song BPM (dashed red). Song BPM is flat at ~125 throughout. RMS breakdown at 1473s (24.6 min) precedes the spectral explosion at 1706s (28.4 min) by ~250 seconds.

Spectral flux showed moderate-to-strong positive correlation with HR in three of four sessions ($r = 0.684, 0.427, 0.399$; Apr 28 partially contaminated by startle event). The spectral flux peak of 0.2592 at 1706s (28.4 min) is the highest harmonic novelty value in the 14.8-minute window, immediately preceding the fingerprint zone. The RMS breakdown at 1473s (24.6 min) precedes the spectral explosion by ~250 seconds, consistent with a tension-load → release architecture common in deep house music.

3.3 Heuristics Dashboard

Figure 3 presents six derived heuristics from the four-session dataset.

Figure 3 — ASCF heuristics dashboard

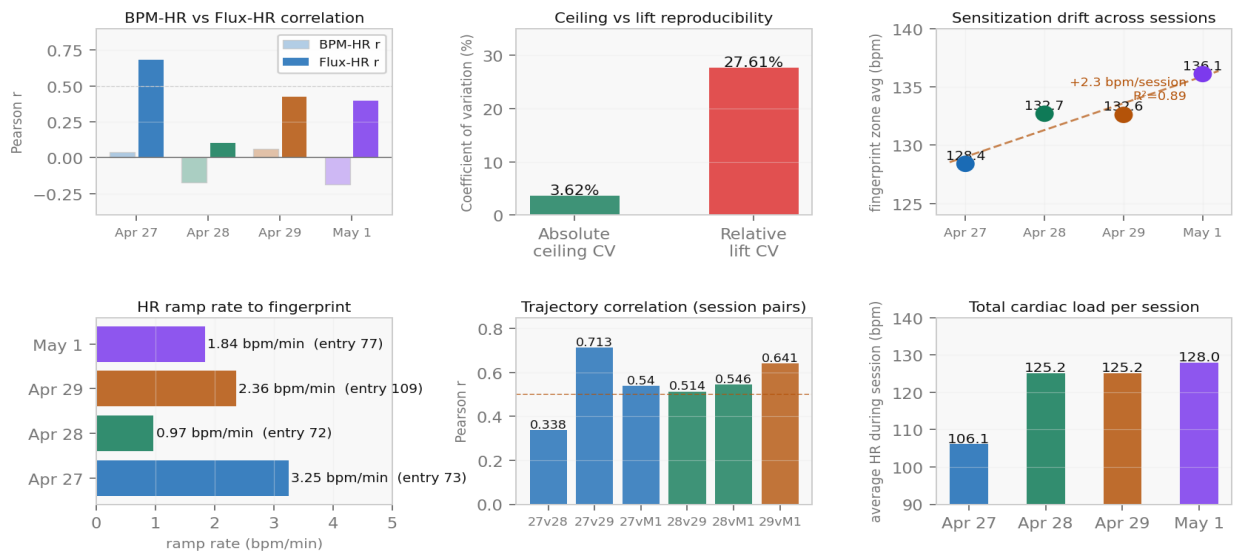


Figure 3 — ASCF heuristics dashboard. (a) BPM-HR vs flux-HR Pearson correlation. (b) Ceiling vs lift CV — absolute ceiling 8x more reproducible. (c) Sensitization drift +2.3 bpm/session. (d) Ramp rates by session. (e) Session-pair trajectory correlation. (f) Total cardiac load per session.

Heuristic	Value	Interpretation
CV at fingerprint zone	1.37%	High reproducibility (threshold <5%)
BPM-HR Pearson r	≈ 0.00	Rhythmic entrainment eliminated
RR-HR correlation	≈ 0.00	Metabolic / respiratory drive eliminated
Flux-HR Pearson r	0.40–0.68	Spectral complexity as mechanism
Ceiling CV vs lift CV	3.62% vs 27.6%	Absolute attractor, not relative lift
Sensitization slope	+2.3 bpm/session	Early neural potentiation ($R^2=0.89$)
Ramp rate	0.97–3.25 bpm/min	Inversely correlated with entry baseline

4. Discussion

4.1 The Attractor Model

The ceiling vs lift asymmetry is the most mechanistically significant finding. The absolute HR ceiling has a CV of 3.62%; the relative lift has a CV of 27.6% — nearly eight times higher. The track encodes an absolute target CNS state rather than a fixed relative increment. We model this as a fixed-point attractor: a musically-encoded state toward which cardiac response converges regardless of initial conditions. The ramp rate data supports this — higher-baseline sessions show shallower ramps while lower-baseline sessions show steeper ramps, consistent with the system correcting toward the same attractor from different positions.

4.2 Timbral Memory Encoding

Near-zero BPM–HR and RR–HR correlations together rule out rhythmic entrainment and metabolic demand as mechanisms. Spectral flux correlates more strongly with HR, pointing to timbral complexity as the active stimulus. We propose that specific harmonic and timbral features have been encoded as conditioned stimuli through repeated listening. The anticipatory dopamine response fires before the spectral peak at 1706s (28.4 min) — evidenced by the cardiac ramp beginning well before the flux maximum and by the practised pre-cue behaviour. The cardiac response is reward prediction, not acoustic reaction.

4.3 Sensitization vs Habituation

The convergence value increased at +2.3 bpm per session ($R^2 = 0.89$), consistent with sensitization rather than habituation. The critical test will be the planned overexposure session in which the dopamine section will be played back-to-back, with the prediction that the sensitization trend will reverse into a within-session habituation curve.

4.4 Derivative Analysis

A formal test of trajectory shape consistency — whether the minute-by-minute direction of HR change is consistent across sessions — produced near-chance results (sign agreement 37–53%, derivative Pearson $r \approx 0$). The absolute convergence is robust; the path is not consistent at minute-level resolution. The single exception is 1260s (21.0 min), where all four sessions showed simultaneous upward movement — a secondary trigger warranting further analysis.

5. Limitations

Limitation	Status / Response
n = 1 subject	Single-subject methodology is appropriate for pilot and methodology papers. Generalisability requires multi-subject replication.
Single track	A second track (Palms Trax) has preliminary data showing a different fingerprint architecture. Control track session planned.
No control condition	Same shower duration with different music pending. This is the critical confound-elimination experiment.
Shower as confound	Apr 29 (entry 109 bpm) follows a completely different trajectory but still converges, partially addressing the physical activity confound.
RR per-session data	RR–HR decoupling observed globally. Per-session RR data not extracted for this paper but available in raw telemetry.
Startle contamination (Apr 28)	Speaker fell during session. Event is documented, timestamped, and morphologically distinct from music-induced response.

6. Future Work

Experiment	Hypothesis	Status
n=5+ replication	CV remains <5% at fingerprint zone	Ongoing daily
Control track session	Convergence does NOT occur with different music in same shower context	Planned
Overexposure / habituation	Repeated plays of dopamine section decay the fingerprint within a single session	Planned
Spontaneous recovery	After 5–7 day rest, fingerprint partially restores	Planned
Second subject	Different person shows a different fingerprint for the same track	Seeking participant
Multi-track taxonomy	Different tracks produce distinct fingerprint architectures (burst vs sustained plateau)	Palms Trax available
Per-session RR analysis	RR–HR decoupling confirmed at per-session resolution	Raw data available

Appendix A — Session Metadata

Session	Date	Entry HR	Peak HR	Lift	Pre-state context
Apr 27	2026-04-27	73 bpm	141 bpm	+68	Rested, late night. ~30 min post couch dance session.
Apr 28	2026-04-28	72 bpm	147 bpm	+75	Clean rested baseline. Sonos Roam fell inside shower at ~18 min. Startle event documented.
Apr 29	2026-04-29	109 bpm	157 bpm	+48	Elevated — physically active prior to session.
May 1	2026-05-01	77 bpm	157 bpm	+80	Office day. Full day gap (Apr 30: no session).

Appendix B — Glossary

ASCF — "Askif" — Auditory Stimulus Cardiac Fingerprinting. The methodology and the observed phenomenon of reproducible person-specific cardiac responses to music.

Fingerprint zone — The track position range (1770–1800s / 29.5–30.0 min in the Tinzo mix) where cross-session HR convergence is observed.

Spectral flux — Frame-to-frame change in spectral magnitude computed via librosa. High flux = high harmonic novelty and timbral complexity. Plotted in cyan in Figure 2 to distinguish from session lines.

RMS — Root Mean Square energy — a measure of audio loudness or signal power across a time window. Expressed in decibels (dB). Negative values reflect attenuation below full scale.

Timbral memory encoding — The hypothesised CNS process by which specific harmonic and timbral features of a track become conditioned stimuli for dopaminergic anticipation through repeated listening.

RR–HR decoupling — The observed independence of respiratory rate and heart rate during music-induced cardiac arousal, indicating the HR response is CNS-mediated rather than driven by breathing or metabolic demand.

Attractor (dynamical systems) — A fixed state toward which a system converges regardless of initial conditions. Used here to describe the absolute HR ceiling encoded by the track in the subject's CNS.

CV (coefficient of variation) — $SD/mean \times 100$. Standard metric for reproducibility. Below 5% = high reproducibility in psychophysiological measurement.

Sensitization — Increasing response magnitude with repeated exposure to a stimulus. Observed here as +2.3 bpm/session drift across four sessions.

Hedonic adaptation — The process by which repeated pleasurable stimuli produce diminishing dopaminergic response over time. The planned overexposure experiment will test whether this occurs with the ASCF fingerprint.

The Ataraxia Project — A personal research program applying quantified self methodology, biometric instrumentation, and self-experimentation to questions of cognitive and physiological optimisation.

COS — Cognitive Operating System. The conceptual framework treating the self as an instrumented system subject to measurement, modelling, and optimisation.

ASCF — Auditory Stimulus Cardiac Fingerprinting

Shrey S. Kumar · The Ataraxia Project · COS Vol. 1 · May 2026

Data: WHOOP 4.0 via Home Assistant · Sonos Roam track telemetry · librosa audio analysis