

Supporting Online Material for: Rhythmicity Index: Signifying Shape Consistency in Addition to Phase Consistency

Italicized section names in this document refer to sections of the main text.

Supplementary Figures

Blinded visual audit of 36 sweep runs stratified by RI category $\times$ Q tercile: appearance tracks RI (ordering concordance 0.71, ties excluded), not Q (0.44, below the 0.5 chance level)
each trace: 300-sample segment of the measurement window · ratings are the median of 3 blinded passes, 0–10 scale

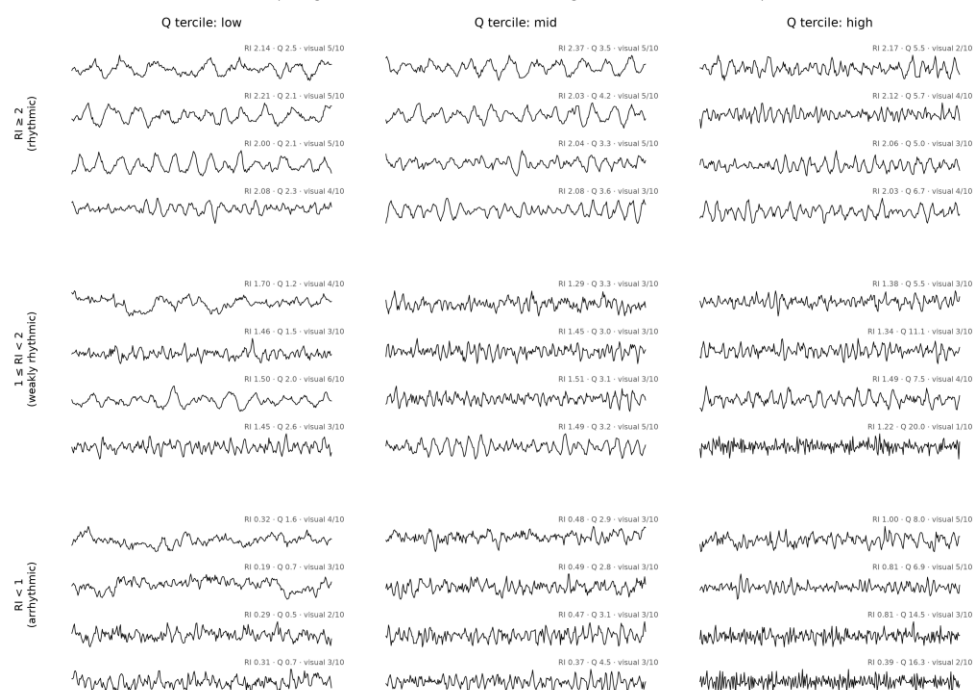


Figure S1 | Blinded visual appearance tracks the rhythmicity index, not the Q-factor. Of 36 admissible sweep runs split by RI category and Q tercile cells (four runs per cell with varying α ; all fully deterministically reproducible based on the sweep's deterministic seeds): each of the tiles depicts 300 samples (the zoom section; raters also assessed the whole measurement window); all traces were plotted using the same rendering method and assessed blindly for visual rhythmicity (rating 0–10 scale; each trace was independently rated three times; the median of these ratings is shown here). Visual rhythmicity assessment correlates with RI but not Q across all admissible runs: ratings show a positive correlation with RI (pairwise ordering concordance 0.71, ties discarded; $p = 0.005$, Spearman $\rho = 0.46$). However, they do not show a correlation with Q (concordance 0.44, below the 0.5 chance level; $p = 0.44$, $\rho = -0.13$). In addition, the highest-Q tercile received the lowest mean rating.

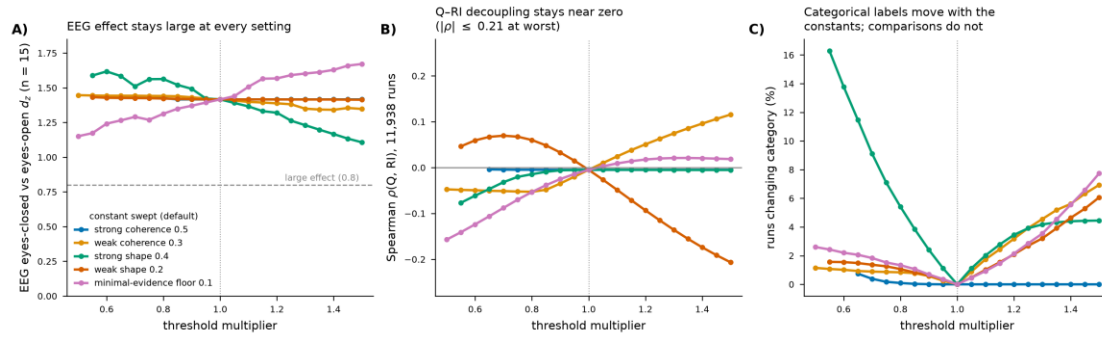


Figure S2 | The main-text results hold regardless of the precise values of the index’s constants. Each of the five independent constants (legend; the other two constants of *The rhythmicity index* are derived from the weak coherence threshold) was changed individually from 50% to 150% of its default with the others held constant; settings that violate the threshold ordering are not shown. **A)** The subject-level eyes-closed minus eyes-open effect size (paired Cohen’s d_z ; $n = 15$ subjects; *EEG validation* pipeline) is large relative to the conventional large-effect cut-off (dashed line) at each setting, and all 15 subjects increased. **B)** The rank correlation between Q and RI for the 11,938 admissible sweep runs remained essentially equal to zero (-0.005) and did not reach a substantial positive correlation (ρ from -0.21 to $+0.12$ across the constant settings; it exceeds $+0.10$ only at the highest values of the weak coherence threshold). **C)** The fraction of the 11,938 admissible runs whose categorical RI label (“arrhythmic”, “weakly rhythmic” or “rhythmic”) shifts: labels do move with the constants, since the bands are reporting conventions, while the orderings in A) and B) do not.

Supplementary Tables

Table S1 | Design and outcome of the blinded visual audit, by stratification cell.

RI category	Q tercile	n	RI range	Q range	Mean blinded rating (0–10)
Arrhythmic ($RI < 1$)	Low	4	0.19–0.32	0.54–1.56	3.00
Arrhythmic ($RI < 1$)	Mid	4	0.37–0.49	2.82–4.48	3.00
Arrhythmic ($RI < 1$)	High	4	0.39–0.998	6.95–16.33	3.75
Weakly rhythmic ($1 \leq RI < 2$)	Low	4	1.45–1.70	1.18–2.62	4.00
Weakly rhythmic ($1 \leq RI < 2$)	Mid	4	1.29–1.51	2.97–3.28	3.50
Weakly rhythmic ($1 \leq RI < 2$)	High	4	1.22–1.49	5.45–19.99	2.75
Rhythmic ($RI \geq 2$)	Low	4	2.00–2.21	2.11–2.46	4.75
Rhythmic ($RI \geq 2$)	Mid	4	2.03–2.37	3.28–4.19	4.50
Rhythmic ($RI \geq 2$)	High	4	2.03–2.17	4.99–6.68	3.25

Mean blinded ratings rise monotonically across RI categories ($3.25 \rightarrow 3.42 \rightarrow 4.17$) but fall monotonically across Q terciles ($3.92 \rightarrow 3.67 \rightarrow 3.25$).

Table S2 | Per-family discrimination of the benchmark battery: stationary rhythms versus each negative family.

Negative family	n	AUC (RI)	AUC (Q)	Instances with RI ≥ 2
White noise	100	0.95	0.15	0%
Pink noise	100	0.90	0.97	0%
Brown noise	100	0.80	0.98	11%
AR(2) narrowband (pooled over pole radii)	200	0.53	0.83	66%
AR(2), pole radius 0.9	50	0.77	0.97	24%
AR(2), pole radius 0.95	50	0.62	0.91	58%
AR(2), pole radius 0.98	50	0.44	0.78	88%
AR(2), pole radius 0.99	50	0.30	0.66	96%
Spectrum-matched surrogates of bursty rhythms	100	0.77	0.56	22%
Spectrum-matched surrogates of jittered rhythms	100	0.62	0.53	55%
Shape-randomized (regular clock, fresh waveform each cycle)	200	0.84	0.92	2%

Each row reports the ROC area for discriminating the 500 stationary positives (sustained, amplitude-modulated, and lightly jittered rhythms) from one negative family, for the rhythmicity index (inadmissible signals counted as non-detections) and for the Q-factor; AR(2) noise is given pooled and by pole radius. Pooled areas, threshold sensitivities and specificities, and the excluded diagnostic arm of phase-randomized surrogates are reported in the main text. Surrogates of lightly jittered rhythms often remain above RI = 2 for the same reason as the diagnostic arm: phase-randomizing a nearly periodic signal yields another nearly periodic signal.

Table S3 | One-at-a-time sensitivity of the paper’s conclusions to the five independent constants.

Constant	Default	Admissible settings	d_z range	$\rho(Q, RI)$ range	Min top-1%-Q fraction below RI = 2	Max fraction of labels changed
Strong coherence threshold (t_c)	0.5	18 of 21	1.42–1.43	−0.005 to −0.004	100.0%	0.7%
Weak coherence threshold (t_{cw})	0.3	21 of 21	1.34–1.45	−0.053 to +0.116	100.0%	6.9%
Strong shape threshold (t_r)	0.4	20 of 21	1.11–1.62	−0.077 to −0.005	97.5%	16.3%

Constant	Default	Admissible settings	d_z range	$\rho(Q, RI)$ range	Min top-1%-Q fraction below RI = 2	Max fraction of labels changed
Weak shape threshold (t_{rw})	0.2	20 of 21	1.41–1.43	−0.207 to +0.070	100.0%	6.1%
Minimal-evidence floor (t_{min})	0.1	21 of 21	1.15–1.67	−0.157 to +0.021	100.0%	7.7%

Each constant was varied from 50% to 150% of its default in 5% steps (21 settings each; settings violating the threshold ordering are excluded as inadmissible). EEG effect sizes are large at all admissible constant settings ($d_z = 1.11$ – 1.67 , $p \leq 7.5 \times 10^{-4}$; 15 out of 15 subjects had higher eyes-closed than eyes-open RI); eyes-closed epochs were labelled rhythmic in 22% to 85% of cases, whereas only 0% to 49% of eyes-open epochs were, and the two conditions differed at every setting.

Table S4 | Sensitivity along the two derived coherence-threshold axes, freed and swept independently.

Constant	Default	Admissible settings	d_z range	$\rho(Q, RI)$ range	Min top-1%-Q fraction below RI = 2	Max fraction of labels changed
Strong mean-coherence threshold (derived) ($t_{\bar{c}}$)	0.35	16 of 21	1.34–1.45	−0.006 to +0.004	100.0%	6.9%
Weak mean-coherence threshold (derived) ($t_{\bar{c}w}$)	0.25	18 of 21	1.40–1.45	−0.075 to +0.066	100.0%	1.6%

The two derived constants — the mean-coherence thresholds, fixed in the main text as $\max(0.35, 0.3 + 0.05)$ and $\max(0.20, 0.3 - 0.05)$ — were additionally freed and swept as independent axes over the same 50–150% grid. The exemplar run of *Spectral sharpness and genuine repetition can disagree* outscored its comparison run at every admissible setting of every axis in Tables S3 and S4.

Table S5 | Contiguous tolerance intervals around each default constant, with the factor that limits each edge.

Constant	Symbol	Default	Tolerance interval	Span	Limiting factor at lower edge	Limiting factor at upper edge
strong coherence threshold	t_c	0.5	[0.305, 0.750]	−39% to +50%	ordering limit (must exceed weak coherence 0.3)	sweep limit (+50%), no failure found
weak coherence threshold	t_{cw}	0.3	[0.150, 0.423]	−50% to +41%	sweep limit (−50%), no failure found	$\rho(Q, RI)$ crosses +0.10
strong shape threshold	t_r	0.4	[0.240, 0.548]	−40% to +37%	top-1%-Q fraction below $RI = 2$ drops to 0.975	exemplar RI falls below 2 (to 1.64; ordering vs comparison run preserved)
weak shape threshold	t_{rw}	0.2	[0.102, 0.242]	−49% to +21%	ordering limit (must exceed floor 0.1)	$\rho(Q, RI)$ crosses −0.10
minimal-evidence floor	t_{min}	0.1	[0.067, 0.150]	−33% to +50%	$\rho(Q, RI)$ crosses −0.10	sweep limit (+50%), no failure found

Table S6 | Subject-level Berger contrast on identical epochs ($n = 15$ paired subjects).

Measure (per-epoch, averaged per subject)	Eyes open, mean (SD)	Eyes closed, mean (SD)	d_z	Pooled d	p (paired t)
Rhythmicity index (RI)	1.178 (0.440)	1.904 (0.453)	1.42	1.63	8×10^{-5}
Q-factor	2.685 (1.408)	6.241 (1.921)	1.68	2.11	1×10^{-5}
Alpha-band $\log_{10}$ power	1.663 (0.473)	2.380 (0.612)	1.65	1.31	2×10^{-5}
Cycle correlation r alone	0.195 (0.089)	0.416 (0.147)	1.65	1.83	2×10^{-5}
specparam alpha peak power	0.498 (0.237)	1.270 (0.467)	1.65	2.08	2×10^{-5}

Pooled d is the difference of condition means divided by the pooled standard deviation of the per-subject means, $\sqrt{\frac{SD^2(\text{eyes open}) + SD^2(\text{eyes closed})}{2}}$; d_z divides the same difference by the

standard deviation of the within-subject differences. Neither is the epoch-level $d = 1.08$ of the original submission, which treated all 600 epochs as independent.

Supplementary Methods

The requirement of eight samples per cycle rests on a degrees-of-freedom count. A mean-removed cycle of n samples spans $n - 1$ dimensions, of which the fundamental occupies two (its cosine and sine components); the waveform therefore has $n - 3$ residual degrees of freedom in which it can differ from a pure sinusoid, and we confirmed this count numerically for $n = 3 \dots 12$. At $n = 3$ the residual dimension is exactly zero: any two mean-removed three-sample cycles that share a fundamental are perfectly correlated by construction, and two independent random three-sample cycles have a mean absolute correlation of $2/\pi \approx 0.64$, exceeding the index's shape threshold of 0.4 in magnitude by chance in 74% of draws. More generally, the null distribution of the sample correlation between two independent mean-removed cycles of length n is a symmetric Beta on $[-1, 1]$ with both parameters $(n - 2)/2$, so $P(r \geq 0.4)$ is 0.37 at $n = 3$, 0.22 at $n = 6$ and 0.16 at $n = 8$.

A worked example carries the trace of Figure 1B of the main text through every step of the index. The refined dominant frequency is $f_0 = 0.008336$ cycles per sample, and no subharmonic passes the autocorrelation test ($k = 1$), so the repetition period is 119.96 samples: 119.96 samples per cycle and 400.1 cycles in the record, both above their floors, so the index is admissible. The component measures are $c = 1.000$, $\bar{c} = 1.000$ and $r = 0.737$. These give $\Phi = \max(1.000/0.5, 1.000/0.35) = \max(2.000, 2.857) = 2.857$ and $\Psi = 0.737/0.4 = 1.842$, so the strong gate is $g_s = \min(2.857, 1.842) = 1.842$. Since $g_s \geq 1$ the first branch applies (the weak gate g_w is not needed) and $RI = 1 + 1.842 = 2.842$, which is ≥ 2 and therefore rhythmic. The shape term is the binding one: the waveform repeats well enough to pass, but it, not the timing, sets the score.

In the EEG analysis, the admissibility floors of the index correspond to the number of cycles and windows per epoch in the following way: an eyes-closed epoch at the group-median alpha frequency of 10.2 Hz contains about 41 cycles, while the multi-window coherence term averages over windows 1–10 cycles wide, corresponding to 40 down to 4 windows at this frequency.

Visual audit methodology

Every one of the 36 traces was plotted in the same way, namely: the full 1,500-sample measurement window together with a 300-sample magnification and without any labels, annotations or metadata (all traces in random order).

Note that the visual audit was performed by AI, which means we do not have access to the actual human raters used by other research projects. Specifically, the AI model used for this task was the version of Claude Sonnet 5 by Anthropic, accessed in August 2026. We queried this model with the default temperature of 1.0. However, because the task was relatively simple and we could not afford an alternative test setup (like recruiting an external human research assistant who could perform the visual audits in their own time), our methodology could still be informative, as the blinding setup, the randomization of the tested samples, and the multi-pass design follow the traditional design of a human observer study.