Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Bayesian tonal-state geometry in Western art music: Uncertainty and displacement jointly determine tonal continuity

Chen-Gia Tsai

cgtsai1971@gmail.com

Independent Researcher

Article

Keywords: computational musicology, tonality, Bayesian inference, chromaticism, corpus analysis, historical change

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11233232/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Bayesian tonal-state geometry in Western art music: Uncertainty and displacement jointly determine tonal continuity

Short title: Uncertainty, displacement, and tonal continuity in music

Chen-Gia Tsai
Independent researcher

Corresponding author:
Chen-Gia Tsai
Email: cgtsai1971@gmail.com

Abstract

Historical change in Western art music is often described as a gradual expansion beyond the pitch and harmonic resources of the prevailing key. Existing studies have traced corresponding developments in pitch organization and key-transition structure, but have not directly quantified tonal continuity across successive local contexts. This study develops a Bayesian framework in which each temporal window is represented as a posterior distribution over 24 major and minor keys. Summed posterior concentration S_{Σ} serves as the concentration counterpart to within-context uncertainty, whereas squared Euclidean distance D^2 provides a Bayesian-surprise-like measure of between-context displacement. Their joint geometry determines successive-state overlap and, through a monotonic transformation, the leakage threshold λ: the fraction of inherited prior weight that must be replaced by a uniform, key-neutral component before the two receive equal posterior support. Thus, λ indexes the tonal contribution to continuity across successive contexts, with lower values indicating more fragile tonal continuity. The evidence-strength parameters governing posterior sharpness were bracketed from two parallel-mode modulations in Mozart by requiring each modulation to register as a distinct tonal event while retaining continuity with the preceding tonal context. The framework was applied to 26 subcorpora of the Distant Listening Corpus, composed between 1736 and 1925, at a 4-bar integration span. Mean λ declined with subcorpus median composition year, indicating that tonal continuity between successive contexts weakened in later repertoire. Episodes with $\lambda < .5$ were uncommon but theoretically informative, spanning a continuum from high posterior concentration with large displacement to lower concentration with smaller displacement. Under such conditions, emerging tonal interpretations may gain support with weaker constraint from the preceding tonal context, although listener-side attenuation may also depend on other musical and cognitive factors.

Keywords: computational musicology; tonality; Bayesian inference; chromaticism; corpus analysis; historical change

1. Introduction

When listening to music, we do not normally abandon an established tonal center simply because an out-of-key tone or altered chord appears. Experimental and computational studies show that tonal information is integrated across successive musical events, allowing an established key context to continue shaping the interpretation of subsequent chords and pitches (Krumhansl and Kessler 1982; Huron and Parncutt 1993; Toiviainen and Krumhansl 2003). As long as new evidence remains compatible with the existing tonal context, carrying forward the current interpretation provides continuity across successive events. When accumulating pitch and harmonic evidence instead comes to support a remotely related tonal state, continued reliance on the inherited interpretation may increasingly constrain adaptation to the emerging context. In such circumstances, reducing the influence of the inherited tonal belief allows new evidence to support an alternative interpretation under weaker prior constraint.

This is a fundamental problem in sequential inference. In sequential Bayesian updating, the posterior belief formed at one time point can serve as the prior for subsequent inference and be revised as new observations arrive. If the latent state has changed, excessive reliance on the preceding belief may hinder adaptation to the emerging state. An observer in a dynamic environment therefore faces the question of how strongly accumulated information should continue to constrain new inference. Research on volatility and change-point learning shows that as environmental change becomes more likely, earlier evidence should exert less influence, while recent observations receive greater weight (Behrens et al. 2007; Nassar et al. 2010; Glaze et al. 2015). In dynamic Bayesian inference, this reduction in the influence of older evidence can be expressed as leakage or forgetting.

Tonal perception provides a natural musical instance of this problem, with key treated as a latent state and accumulating pitch-class content providing evidence about that state. Huron and Parncutt (1993) incorporated echoic-memory decay into a key-profile framework, allowing the influence of earlier pitches on the current tonal interpretation to diminish over time. Temperley (1999) formulated key finding as an inference problem across successive short segments and introduced a key-change penalty that favors persistence of an established key interpretation. Later probabilistic key-tracking models represented tonal persistence through transition probabilities among latent key states. López et al. (2019) assigned transition probabilities according to graded distances among neighboring keys, favoring persistence and transitions to closely related keys, whereas Weiß et al. (2020) used a diagonal-enhanced transition matrix that strongly favored self-transitions. These studies establish a computational tradition for modeling the persistence and updating of tonal interpretations over time.

The present study builds on this tradition by asking a complementary question. Conventional tonal Hidden Markov Models (HMMs) specify or learn a state-transition

structure and use it to infer sequential tonal states. Here, adjacent tonal beliefs are estimated first, and the question is instead how much the inherited belief must be attenuated before a uniform, key-neutral alternative becomes equally competitive under the current evidence. For this single-step comparison, the prior carried forward from the preceding context is represented as a mixture of the inherited belief and a uniform component. Bayesian comparison of these components defines the critical leakage weight λ, hereafter termed the *leakage threshold*. At this threshold, the inherited component has weight $1-\lambda$ and the uniform component has weight λ, with the two receiving equal posterior support. A trough in λ therefore marks a local reduction in tonal continuity: the preceding tonal context provides relatively weak support for what follows.

The leakage threshold does not, however, determine whether or when listeners actually attenuate that prior carry-over. Event Segmentation Theory proposes that transient increases in prediction error trigger updating of the current event model (Zacks et al. 2007), while information preceding an event boundary becomes less accessible to subsequent processing (Swallow et al. 2009). In music, such boundaries can be signaled by changes in harmony, texture, register, dynamics, motivic material, and sectional structure (Clarke and Krumhansl 1990; Bruderer et al. 2009). These cues may modulate whether and when listeners reduce reliance on the preceding tonal context. Thus, λ characterizes how strongly the current evidence supports continued reliance on the inherited tonal prior, rather than directly measuring listener-side prior carry-over or its attenuation.

Because harmonic vocabulary changes historically, the leakage threshold may itself exhibit systematic historical change. In tonal music, *chromaticism* broadly refers to the use of pitches or harmonies that fall outside the diatonic collection associated with the prevailing key. Recent large-scale corpus studies have quantified related historical changes from several perspectives. Moss et al. (2023) documented a historically increasing fifths range, that is, an expansion in the span of tonal pitch classes occupied along the line of fifths. Moss et al. (2024) modeled tonal interval distributions and found increasing complexity in interval relations, including a marked rise in the use of major and minor thirds during the nineteenth century. Guan et al. (2026) quantified chordal and non-chordal chromaticism and chordal dissonance, providing large-scale evidence for a gradual rise in both chromaticism and dissonance over three centuries. Using bar-by-bar local-key assignments and information-theoretic measures, Gonzalez-Espinoza and Plotkin (2026) examined historical changes in key uncertainty and diversity, as well as novelty in key transitions. These studies characterize historical change in tonal pitch-class usage, interval structure, chromaticism and dissonance, and transition patterns. What has not been directly quantified is how tonal continuity across successive local contexts has changed historically. Addressing this question requires measuring how far the full posterior over tonal states shifts from one context to the next and how that displacement combines with within-context uncertainty to determine λ.

These state-level quantities are analogous to, but distinct from, the uncertainty and surprise measures widely used in research on musical expectancy. In that literature, uncertainty and surprise are typically defined at the level of individual events: uncertainty as the entropy of a predictive distribution, and surprise as the information content of the event that occurs (Pearce and Wiggins 2012; Cheung et al. 2019; Gold et al. 2019; Pearce 2018). The present study shifts the representational level from events to latent tonal states. At this level, the relevant questions are which key or tonal region best explains accumulating pitch evidence, how uncertain that inferred state is, and how far it shifts between successive contexts. This state-level formulation is consistent with behavioral evidence for dynamically graded representations of key as musical context unfolds (Krumhansl and Kessler 1982; Toiviainen and Krumhansl 2003; Mizener and Dowling 2022), and with probabilistic accounts that treat key as a latent structure inferred from observed pitch patterns (Temperley 2004, 2008).

To operationalize these state-level quantities, the present framework represents each temporal window as a posterior distribution over 24 major and minor keys. Summed posterior concentration S_Σ serves as the concentration counterpart to within-context uncertainty, whereas squared Euclidean distance D^2 between adjacent posteriors provides a Bayesian-surprise-like measure of between-context displacement. Their joint geometry determines posterior overlap and, through a monotonic transformation, the leakage threshold λ. These quantities are evaluated at a 4-bar integration span, so that each endpoint compares two adjacent tonal contexts spanning 8 bars in total. This scale is comparable to a conventional phrase-length context while avoiding reliance on substantially longer-range tonal integration, consistent with behavioural evidence that memory for tonal centres weakens over extended temporal intervals (Woolhouse et al. 2016; Farbood 2016).

The framework is applied to 26 expert-annotated subcorpora from the Distant Listening Corpus, comprising repertoire composed between 1736 and 1925. The study addresses three questions. First, does D^2 correspond closely to Bayesian? Second, does mean λ change systematically with composition year? Third, how do λ, S_Σ, and D^2 relate to five established score-derived tonal descriptors adapted to the same endpoint resolution? Together, these analyses shift the historical focus from the prevalence of chromaticism and key-change events to changes in tonal continuity across successive contexts.

2. Bayesian tonal-state model and analytical relations

This section develops the model from posterior geometry to a sequential Bayesian interpretation. It first defines concentration, displacement, and overlap between successive tonal-state posteriors, then derives the leakage threshold λ and its probability-simplex constraints.

2.1 Bayesian tonal-state representation and the geometry of successive posteriors

Each temporal window is represented as a pitch-class distribution over the 12 pitch classes, weighted by sounding duration. The 24 candidate keys are represented by the Temperley–Kostka–Payne (TKP) profiles (Temperley 2007), estimated as the proportion of corpus segments in which each scale degree occurs. Being frequencies already, they require only normalisation to sum to one, and were transposed to all 12 tonics by cyclic rotation of the major and minor prototypes.

For candidate key i, let f_i denote its profile and x the observed pitch-class distribution in the window. Mismatch between the two is quantified by the Kullback–Leibler (KL) divergence $D_{\mathrm{KL}}(x \parallel f_i)$ and converted into likelihood-like support by exponentiation, following the general Bayesian, or Gibbs-posterior, principle by which a loss function supplies the evidence term of a posterior (Bissiri et al. 2016):

$$L_i = \exp[-N\, D_{\mathrm{KL}}(x \parallel f_i)],$$

where D_{KL} is evaluated with natural logarithms.

Under a uniform reference prior over the $J = 24$ major and minor keys, Bayes' rule gives the tonal-state posterior for the window (Murphy 2022):

$$P_i = \frac{L_i}{\sum_{j=1}^{J} L_j}.$$

Effective evidence strength N controls posterior sharpness. Its functional specification is described in Section 3. At each endpoint, let p denote the posterior from the preceding window and q the posterior from the current window. The two windows are adjacent and non-overlapping, and both posteriors are inferred under the same uniform reference prior.

For any tonal-state posterior r, define quadratic concentration as

$$S(r) = \sum_{i=1}^{J} r_i^2.$$

For J states, $S(r)$ ranges from $1/J$ for a uniform posterior to 1 for a point mass. Higher concentration therefore corresponds to lower within-context uncertainty and lower Rényi-2 entropy (Rényi 1961). For successive windows, define

$$S_\Sigma = S(p) + S(q).$$

Thus, S_Σ summarizes the combined concentration of the two adjacent tonal-state beliefs.

Between-context displacement is defined as squared Euclidean distance between successive tonal-state posteriors:

$$D^2 = \| q - p \|_2^2 = \sum_{i=1}^{J} (q_i - p_i)^2.$$

D^2 is symmetric and bounded. It is not identical to KL-divergence Bayesian surprise, conventionally defined through the change between prior and posterior beliefs (Itti and Baldi 2009); instead, it is used here as a Bayesian-surprise-like measure of how far the complete inferred tonal state moves from one context to the next. Its correspondence with KL-based posterior change is tested empirically in Section 4, while the information-theoretic relation to Jensen–Shannon divergence is discussed in Supplementary Note.

The quadratic form is especially useful because it has an exact relation to posterior concentration. Expanding the squared distance gives

$$D^2 = S(p) + S(q) - 2\langle p, q \rangle = S_\Sigma - 2\langle p, q \rangle,$$

and hence

$$\langle p, q \rangle = \frac{S_\Sigma - D^2}{2} \quad (1)$$

The inner product $\langle p, q \rangle$ quantifies posterior overlap. Lower concentration and greater displacement both reduce this overlap.

2.2 Uniform-leak transition family

To give posterior overlap a sequential-inference interpretation, the framework adopts the transition logic of an HMM but uses it as a single-step reference family rather than as recursive filtering over an entire piece. Standard HMMs propagate a posterior through a transition matrix to form the predictive prior for the next observation (Rabiner 1989), and tonal key-tracking models similarly encode persistence through key-change penalties or high self-transition probabilities (Temperley 1999; López et al. 2019). Here, however, p and q are estimated independently from adjacent windows. The transition family asks how strongly the inherited tonal belief would have to be attenuated before a uniform, key-neutral alternative becomes equally competitive.

Let u be the uniform distribution over J tonal states, so that $u_j = 1/J$. For leakage weight $\ell \in [0,1]$, define the symmetric transition kernel

$$T_{ij}(\ell) = (1 - \ell)\delta_{ij} + \frac{\ell}{J},$$

where δ_{ij} is the Kronecker delta. Propagating the inherited posterior p through this kernel gives

$$\pi(\ell) = pT(\ell) = (1 - \ell)p + \ell u.$$

Thus, weight $1 - \ell$ preserves inherited tonal-state structure, whereas ℓ shifts weight toward a uniform, key-neutral alternative. The uniform component is deliberately unstructured, providing a neutral benchmark that does not embed any particular theory of inter-key transition probabilities.

2.3 Critical equal-support leakage threshold

The following comparison uses the standard Bayesian decomposition of a mixture prior, in which posterior component odds are determined by prior mixture weights and component-wise marginal support (Murphy 2022).

Let L_i denote the likelihood-like support of the current window for tonal state i. The marginal support of the current evidence under the inherited posterior is

$$Z_p = \sum_{i=1}^{J} p_i L_i,$$

whereas its support under the uniform prior is

$$Z_u = \frac{1}{J}\sum_{i=1}^{J} L_i.$$

Because the current-window posterior q was itself formed under the uniform reference prior,

$$q_i = \frac{L_i}{\sum_{j=1}^{J} L_j},$$

and therefore

$$\frac{Z_p}{Z_u} = J\langle p, q\rangle$$

This evidence-side support ratio is determined entirely by the inherited posterior and the current evidence; it does not vary with the leakage weight ℓ. Under the mixture prior $\pi(\ell)$, the unnormalized posterior contributions of the inherited and uniform components are, respectively,

$$C_p(\ell) = (1-\ell)Z_p, \qquad C_u(\ell) = \ell Z_u.$$

Their posterior-support odds are therefore

$$\frac{C_p(\ell)}{C_u(\ell)} = \frac{1-\ell}{\ell}\, J\langle p, q\rangle.$$

Define the leakage threshold λ as the value of ℓ at which the two components receive equal posterior support. Setting $C_p(\lambda) = C_u(\lambda)$ gives

$$\lambda = \frac{Z_p}{Z_p + Z_u} = \frac{J\langle p, q\rangle}{1 + J\langle p, q\rangle}.$$

Substituting the overlap identity in equation (1) yields the central analytical relation in equation (2)

$$\lambda = \frac{J\left(S_\Sigma - D^2\right)}{2 + J\left(S_\Sigma - D^2\right)} \tag{2}$$

For $J = 24$,

$$\lambda = \frac{S_\Sigma - D^2}{S_\Sigma - D^2 + 1/12}$$

Thus, posterior concentration and between-context displacement jointly determine the leakage threshold. λ is a strictly monotonic reparameterization of posterior overlap and therefore does not introduce an additional informational dimension. Its added value is interpretive: it expresses tonal continuity as a break-even quantity—how much prior carry-over must be attenuated before a uniform, key-neutral component becomes equally competitive.

2.4 Probability-simplex constraints and the equal-mixture boundary

The feasible D^2–S_Σ geometry is constrained by the probability simplex. Let

$$m = \frac{p + q}{2}.$$

Because m is itself a probability distribution over J states,

$$S(m) \geq \frac{1}{J}.$$

By the parallelogram identity,

$$S_\Sigma = \frac{D^2}{2} + 2S(m),$$

and therefore

$$S_\Sigma \geq \frac{D^2}{2} + \frac{2}{J}$$

A second bound follows from nonnegative posterior overlap:

$$\langle p, q \rangle \geq 0 \quad \Leftrightarrow \quad S_\Sigma \geq D^2.$$

Equality corresponds to disjoint posterior support and therefore $\lambda = 0$. For $J = 24$, the simplex lower bound becomes

$$S_\Sigma \geq \frac{D^2}{2} + \frac{1}{12}.$$

The equal-mixture boundary follows from equation (2) at $\lambda = 0.5$. Equal prior mixture weights give equal posterior support when

$$J\langle p, q \rangle = 1,$$

which, using equation (1), is equivalent to

$$S_\Sigma = D^2 + \frac{2}{J}$$

For $J = 24$,

$$S_\Sigma = D^2 + \frac{1}{12}$$

This line is therefore not an arbitrary decision boundary. At equal prior mixture weights, points above it give greater posterior support to the inherited component, whereas points below it give greater support to the uniform component.

3. Methods

3.1 Corpus, score preprocessing, and integration spans

The analytical corpus was drawn from the Distant Listening Corpus (DLC; Hentschel et al. 2025) and comprised 26 subcorpora with note, measure, and expert harmonic annotations. The included repertoire spans 1736–1925, from Pergolesi's *Stabat Mater* to Medtner's *Two Tales*, Op. 48. For the historical analysis, each subcorpus was represented by the median of its work-level composition-year midpoints; these medians span 1736–1916.5.

Two movement-level inclusion criteria were applied: the notes, measures, and harmonies tables had to be present and readable, and at least 12 complete bars had to remain after preprocessing. Of 754 candidate movements, 35 were excluded for missing required tables and four for insufficient complete bars. The final corpus comprised 715 movements and 85,553 complete bars. Subcorpus composition is reported in Supplementary Table S1.

Scores were converted to duration-weighted pitch-class evidence on a complete-bar grid. The model was evaluated at a single integration span of $B = 4$ complete bars. At each endpoint, the current 4-bar window and the immediately preceding non-overlapping 4-bar window were compared, with a 1-bar hop. Full reconstruction, filtering, and evidence-validity rules are given in Supplementary Methods.

3.2 Evidence-strength specification

Effective evidence strength was specified as

$$N = k(h_b B)^{\alpha} \tag{3}$$

B denotes the length of the analysis window in complete bars (e.g., $B = 4$ for a 4-bar window), and h_b is the movement-level harmonic-event density in chords per bar, calculated as the number of nonempty chord annotations divided by the movement's number of notated bars. The product $h_b \cdot B$ therefore approximates the expected number of annotated harmonic events represented in the window. N is interpreted as an effective, rather than literal, sample size because temporally adjacent musical events are not statistically independent. The exponent α controls how effective evidence strength scales with the expected number of annotated harmonic events in a window, $h_b \cdot B$. Values below 1 implement sublinear pooling, limiting the increase in posterior confidence as the amount of harmonic evidence grows. The gain parameter k sets the overall evidence sensitivity. Because B is fixed at 4 in the present study, variation in $h_b \cdot B$ arises from differences in harmonic-event density across movements.

3.3 Bracketing the evidence-strength parameters

The evidence-strength parameters α and k are not determined by the model itself and were therefore constrained using parallel-mode modulation as an explicit musical criterion. Parallel major and minor keys occupy an intermediate position in the same TKP tonal space used by the model: they share a tonic but differ in mode, making them less similar than the closest relative and dominant/subdominant relations while remaining substantially more similar than more distant key relations. For example, relative to C major, the cosine similarity of the TKP profile is .879 for A minor and .875 for G major, .807 for C minor, .705 for D major, .632 for E-flat major, and .364 for F-sharp major; the corresponding KL divergences show the same broad ordering. This pattern is broadly consistent with established accounts of interkey relatedness, which place relative and dominant/subdominant relations among the most closely connected keys while distinguishing progressively less direct relations according to tonal-space proximity and common-tone structure (Krumhansl 1990; Kopp 2002). Parallel-mode modulation therefore provides a particularly useful intermediate case: it should produce a detectable shift in the inferred tonal state, but not the degree of tonal separation associated with substantially more distant key relations.

This calibration concerns posterior overlap rather than profile similarity per se. Tonal-state posteriors are determined jointly by the observed pitch-class evidence and effective evidence strength N, although the similarity structure of the candidate-key profiles constrains how strongly a given body of evidence can differentiate among keys. As N increases, these differences are amplified: posteriors become more concentrated within stable tonal contexts, while pre- and post-modulation posteriors become more clearly separated. Parallel-mode modulation therefore provides a useful basis for bracketing N: the modulation should produce a distinct local reduction in λ while preserving sufficient overlap for the inherited tonal context to remain more competitive than a uniform, key-neutral alternative.

Two Mozart movements were selected for this purpose: Piano Sonata K. 331/iii, with a mode change at m. 24, and K. 332/ii, with a mode change at m. 25 and a return at m. 29. Their harmonic densities differ by approximately a factor of four, providing leverage to distinguish the scaling parameter α from the gain parameter k. In K. 332/ii, the corpus encoding gives C♯ on the fourth beat of m. 24, whereas corrected editions give C♮; this pitch was emended before analysis.

Fig. 1 shows the λ trajectory through each modulation at a series of fixed values of N. Raising N has two effects simultaneously. First, the plateaus on either side of a modulation rise because greater evidence strength produces more concentrated posteriors within stable tonal contexts. Second, the modulation trough deepens because the same sharpening reduces overlap between the pre- and post-modulation posteriors. These effects motivate

two constraints. The minimum λ must remain above the equal-mixture boundary λ = .5 derived in Section 2.4: a parallel-mode change should not make a uniform, key-neutral alternative more strongly supported than the inherited tonal context. At the same time, the trough must lie at least 0.25 below the surrounding plateau, defined as the median λ over the analysed range excluding the trough and the bar on either side, so that the mode change registers as a distinct tonal event rather than being absorbed into its surroundings. Both criterion quantities vary monotonically with N over the relevant range; the first therefore supplies an upper bound and the second a lower bound, yielding an admissible interval of N for each movement.

Because the integration span is fixed at four bars, N varies across movements only with harmonic density h_b. An admissible interval of N at a given density therefore defines a band in the (α, k) plane. Because the two Mozart movements differ substantially in harmonic density, these bands have different slopes and intersect in a bounded region, which defines the admissible parameter set. A reference setting of (α, k) was specified within this region, and results are reported throughout for that setting. Robustness was assessed by repeating the D^2–KL correspondence, historical, and tonal-descriptor analyses at eight settings spanning the admissible region: values of α across the admissible range, each paired with the midpoint of its admissible k interval, together with the lower and upper admissible k values at the reference α.

3.4 Information-theoretic correspondence of displacement

Because D^2 is intended as a Bayesian-surprise-like measure of between-context tonal-state displacement, its agreement with a conventional information-theoretic measure of posterior change was assessed directly. For every valid endpoint, directional KL divergence from the previous posterior p to the current posterior q was computed as

$$D_{\mathrm{KL}}(\mathrm{q} \parallel \mathrm{p}) = \sum q_j \log(q_j / p_j)$$

Monotonic correspondence between D^2 and $D_{\mathrm{KL}}(\mathrm{q}\|\mathrm{p})$ was assessed by Spearman rank correlation within each movement, thereby avoiding treatment of the large number of temporally overlapping endpoints as independent observations. Movement-level correlations were Fisher-z transformed and averaged within each subcorpus. The resulting subcorpus-level $z$estimates were used for inference with two-sided Wilcoxon signed-rank tests against zero; their equal-weight mean across the 26 subcorpora was back-transformed for corpus-level reporting. This analysis evaluates whether D^2 preserves the ordering of posterior changes captured by the KL divergence conventionally used to quantify Bayesian surprise.

3.5 Historical analysis and musical examples

Historical analyses treated the 26 subcorpora as equally weighted units. For each subcorpus, endpoint-level λ values were pooled and summarized by their arithmetic

mean, and the association with subcorpus median composition year was tested by a Spearman correlation. Since the profile family is estimated from common-practice repertoire, and the Bartók and Ravel subcorpora are likely to fall outside typical common-practice norms, the correlation was repeated without those two as a supplementary analysis. Effective evidence strength depends on harmonic density, so the historical trend in subcorpus median h_b was examined separately.

Musical examples were selected to illustrate the tonal-state geometry. The second movement of Mozart's Piano Sonata K. 280 and Wagner's Prelude to *Tristan und Isolde* were chosen to contrast a stable Classical tonal organization with a late-Romantic idiom in which successive posteriors are less concentrated and more displaced.

3.6 Comparison with established tonal descriptors

To determine which aspects of Bayesian tonal-state geometry are represented by established tonal measures, five score-derived descriptors were adapted to the same endpoint resolution as λ, S_Σ, and D^2. All five were defined on the same $2B$-bar endpoint comprising the previous and current B-bar windows, thereby aligning their temporal support with that of the Bayesian tonal-state quantities.

Following Perttu (2007), out-of-key chromatic density was defined using the expert-annotated prevailing local key as a moving reference. Each sounding note was weighted by the portion of its duration falling within the retained complete bars of the endpoint span. The major-key reference comprised the seven diatonic pitch classes; for minor, the natural-minor collection plus the raised seventh was used. Density was defined as out-of-key sounding duration divided by total sounding duration in the endpoint span.

Chordal chromaticity, non-chordal chromaticity, and chordal dissonance were adapted from Guan et al. (2026) and computed at the harmonic-segment level. Rows of the notes table with non-positive duration were excluded before sounding tonal pitch-class (TPC) sets were constructed. TPC represents spelled pitch classes as integer positions along the line of fifths, thereby distinguishing enharmonically equivalent spellings. For each harmonic segment, sounding TPCs identified by the harmony label as chord tones or added tones were classified as chord tones, while all remaining sounding TPCs were classified as non-chord tones. Because note TPCs in the corpus tables are expressed in absolute coordinates, whereas label-derived chord and added tones are specified relative to the local tonic, the latter were converted to absolute TPCs before matching. Annotated added tones were treated as chord tones in the present implementation. Segments with no sounding chord tone or added tone were omitted; segments containing chord tones but no non-chord tones were retained and assigned a non-chordal chromaticity of zero.

Chromaticity was defined in line-of-fifths space as the shortest distance from a sounding TPC to the seven-position diatonic region of the prevailing local key. For major

keys this region spans one fifth below the tonic to five fifths above it; for minor keys it is the region of the relative major, that is, the natural-minor collection (Guan et al. 2026). TPCs inside the region scored zero, and segment-level chordal and non-chordal chromaticity were obtained by summing these distances over the corresponding TPC sets. Chromaticity was thus measured relative to the diatonic region rather than as distance from the tonic, and note duration did not weight individual TPCs within a segment. Chordal dissonance was calculated over all unordered pairs of distinct sounding chord-tone TPCs, with interval classes 1–6 weighted by the coefficients of Guan et al. (2026). The weighted pair count was normalized by the number of sounding chord-tone TPCs rather than by the number of pairs.

For these three measures, segment-level values were extended to each endpoint by overlap-weighted averaging, the weight of a segment being the duration of its intersection with the retained complete bars of the endpoint span. Bars excluded by the completeness rule contributed no weight, so a segment spanning such a bar entered only for the part of it falling inside retained bars. Local mode determined the diatonic reference of each individual segment, and no mode-specific averages were formed. The two minor references differ by source and were left as each source defines them: the three Guan et al. measures use the natural-minor region, whereas out-of-key chromatic density retains Perttu's natural-minor-plus-raised-seventh collection.

Building on the line-of-fifths center-of-gravity representation introduced by Temperley (2000) and used in later corpus studies (Moss et al. 2023; Guan et al. 2026), each sounding TPC within a B-bar window was mapped to its fifth position and weighted by its sounding duration in the retained complete bars. The resulting duration-weighted mean fifth position defined the window's tonal center of gravity. Fifths displacement was then defined as the absolute difference between the centers of gravity of the previous and current windows, providing a non-Bayesian measure of between-window shift along the line of fifths.

An endpoint contributed only when the relevant windows contained sounding duration within retained complete bars and, for the Guan et al. measures, at least one surviving harmonic segment. A movement contributed when at least 20 endpoints had non-missing values for the relevant pair and the descriptor assumed at least three distinct values. Spearman correlations were computed within movement, Fisher-z transformed and averaged within subcorpus, then averaged across the 26 subcorpora with equal weight and back-transformed. Two-sided Wilcoxon signed-rank tests against zero were used for the 26 subcorpus-level estimates. These tests were descriptive and were not subjected to multiple-comparison correction.

3.7 Computational implementation and use of generative AI

Analyses were implemented in Python 3.11 using NumPy 2.4, pandas 3.0, and SciPy 1.17. Score and annotation tables were read from the Distant Listening Corpus v3.2 TSV files. Exact fractional onsets and durations were parsed using Python's fractions module before conversion to floating point.

Generative AI tools—ChatGPT (GPT-5.6 Sol, OpenAI) and Claude (Opus 5 and Fable 5.1, Anthropic)—were used during the development of this study to assist with discussion and refinement of model formulations and analytical strategies; development and debugging of analysis code; and language editing during manuscript preparation. The author takes full responsibility for the accuracy and integrity of the manuscript.

4. Results

4.1 The admissible parameter region and reference

Fig. 1 shows the λ trajectory through each of the two parallel-mode modulations at a series of fixed values of N. In both movements the plateaus rise and the troughs deepen as N increases, so that each requirement of Section 3.3 is satisfied on one side of a critical value and the two together leave an interval. In K. 331/iii, whose harmonic density is 1.0 chords per bar, the trough at m. 27 first lies 0.25 below the surrounding plateau at $N = 4.77$ and first falls to $\lambda = .5$ at $N = 6.59$. In K. 332/ii, at 4.1 chords per bar, the corresponding values for the troughs at mm. 28 and 32 are $N = 6.29$ and $N = 9.23$.

Both bounds are higher for the denser movement, so the two intervals do not coincide and the constants are not separately determined by either movement alone. Mapping the intervals through $N = k\cdot(h_b\cdot B)^{\wedge}\alpha$ gives the two bands of Fig. 2, which differ in slope because the movements differ in density; their intersection is the admissible region. Restricting α to non-negative values, the admissible region extends from $\alpha = 0$ to approximately $\alpha = 0.47$. Projected onto the k axis the region spans roughly 2.50 to 6.59, but α and k are jointly rather than independently constrained, so only a narrow interval of k is admissible at any one α: at $\alpha = 0.30$, for example, k ranges from 3.14 to 3.98.

A reference parameter setting was specified within this region at $\alpha = 0.30$ and $k = 3.5$, with the latter chosen at the centre of the admissible k interval for that α. It applies $N = 5.3$ to K. 331/iii and $N = 8.1$ to K. 332/ii. At these values the three troughs reach $\lambda = .57$, .54 and .53, so each remains between .03 and .07 above the equal-mixture boundary while lying between .29 and .34 below its plateau. The reference parameter setting therefore satisfies both requirements with margin on each side rather than at either limit.

4.2 Validation and illustration of tonal-state geometry

The rank ordering of successive tonal-state changes was robust to the choice of displacement measure. The Fisher-z-aggregated correlation between D^2 and $D_{KL}(q\|p)$ was .911 (Wilcoxon $p = 3.0 \times 10^{-8}$), and lay between .90 and .92 at every setting of the

sensitivity analysis. D^2 thus preserves nearly the same ordering of posterior changes as the KL divergence conventionally used to quantify Bayesian surprise, while retaining the bounded quadratic form required by the concentration–displacement geometry. The agreement is close but cannot be exact: D_{KL} is unbounded, whereas D^2 is capped by the simplex.

Fig. 3 illustrates the geometry. In Mozart's Piano Sonata K. 280, ii, mm. 1–8, both windows commit to F minor and the pair sits high in the plane with almost no displacement. In the Prelude to *Tristan und Isolde*, mm. 3–10, neither window commits to a single key — the first divides its weight chiefly between A major and F♯ minor, the second between C minor, A minor and C major — and the two are spread over different regions of key space, so the pair is at once less concentrated and more displaced. The two examples therefore lie on opposite sides of the equal-mixture boundary $S_\Sigma = D^2 + 1/12$, which 8.46% of all endpoints fall below.

4.3 Historical change in leakage thresholds

Mean λ declined with subcorpus median composition year (Spearman $\rho = -.817$, $p = 3.6 \times 10^{-7}$; Fig. 4). The decline is large relative to the range of the measure: subcorpus means run from .83 in J. C. Bach's sonatas to .58 in Bartók's Bagatelles. The change is one of spread as well as of level, since subcorpora with median years before 1850 occupy a narrow band between .72 and .83, whereas later ones range from .58 to .76. Excluding Bartók and Ravel, whose harmonic language is likely to fall outside typical common-practice norms, the association was weaker but still clearly present ($\rho = -.778$, $p = 7.5 \times 10^{-6}$); the lowest remaining subcorpus mean is .65, in Liszt's *Années de pèlerinage*.

Across the settings of the sensitivity analysis the correlation remained between −.81 and −.83, with p below 6.2×10^{-7} throughout (Supplementary Table S2). Subcorpus median h_b shows no detectable trend over the period (Spearman $\rho = -.218$, uncorrected $p = .28$).

4.4 Established tonal descriptors and Bayesian-state geometry

The five descriptors were asymmetrically associated with the two components of Bayesian tonal-state geometry (Table 1). Out-of-key chromatic density, chordal chromaticity and chordal dissonance were all more strongly associated with S_Σ than with D^2, by factors of roughly two to five; for chordal dissonance the association with D^2 was effectively absent ($\rho = +.04$). Non-chordal chromaticity showed the same asymmetry with weaker associations throughout, indicating that posterior concentration tracks chromaticism in chord tones more closely than chromaticism carried by non-chord tones. Fifths displacement showed the converse, correlating with D^2 at +.54 and with S_Σ at only −.17. Both orderings hold at every setting of the sensitivity analysis (Supplementary Table S3).

For three of the five descriptors the magnitude of the association with λ fell between those with S_Σ and with D^2; for the other two it exceeded the larger of the two by less than .01,

and no association with λ reached .55. Together with the component-wise asymmetry of the individual descriptors, this indicates that no single descriptor captures the joint concentration–displacement geometry represented by λ.

5. Discussion and conclusion

Although research on pitch-related musical expectancy has typically operationalized uncertainty and surprise at the level of upcoming events (Pearce and Wiggins 2012; Cheung et al. 2019; Gold et al. 2019; Pearce 2018), recent work has renewed emphasis on tonality as a fundamental representational system in music cognition (Mehr 2025; Vuust et al. 2022). The present study builds on this perspective by representing tonality as an evolving probabilistic belief over candidate key states, allowing tonal trajectories to be described in terms of Bayesian inference over time. The between-context displacement measure D^2 closely tracked KL-based posterior change, supporting its use as a Bayesian-surprise-like measure of tonal displacement. As formalized in equation (2), D^2 and S_Σ (the concentration counterpart of within-context uncertainty) jointly determine the leakage threshold λ. Across the 26 subcorpora, mean λ declined with composition year. By relocating uncertainty and surprise from individual pitch events to probabilistic tonal beliefs, the framework redefines these quantities at the level of tonality and introduces λ as a quantitative index of tonal continuity across successive contexts.

Endpoints below the $\lambda = .5$ boundary were uncommon, but their geometry is theoretically revealing. When $\lambda < .5$, the current tonal evidence favors beginning afresh from a key-neutral prior over carrying forward the inherited tonal context. Endpoints below this boundary satisfy $D^2 > S_\Sigma - 1/12$ and therefore reflect a trade-off between posterior concentration and between-context displacement: relatively concentrated posteriors require large displacement to cross the boundary, whereas lower concentration requires less displacement to do so. The concentration–displacement trade-off is structurally reminiscent of the interaction between uncertainty and surprise reported for chord expectancy by Cheung et al. (2019), although the representational levels differ: their measures concern individual chord predictions, whereas the present quantities concern latent tonal-state beliefs. This parallel raises the possibility that event-level expectancy and tonal-state inference may exhibit related uncertainty–change geometries without implying a shared mechanism.

Because λ is a stimulus-defined leakage threshold, its psychological relevance lies in how listener-side attenuation of prior tonal context relates to that threshold. Musically, an abrupt modulation to a remote key provides a plausible situation in which reliance on the preceding tonal context may need to be reduced. In a recent fMRI study of the diminished-seventh chord, the chord was introduced after three bars of stable tonal context and served as a pivot into a remotely related key. This tonal modulation elicited an early suppression of bilateral intraparietal sulcus (IPS) activity, followed by a later rebound (Tsai et al. 2026).

Under the present model, such a transition is expected to produce a pronounced λ trough, making this biphasic response a possible neural analogue of reduced reliance on the preceding tonal context followed by renewed evidence accumulation for the emerging key. Related evidence comes from an earlier fMRI study by the same group, in which chromatic harmony elicited stronger activity than diatonic harmony in a frontoparietal network including the IPS, precuneus, and prefrontal regions (Li et al. 2021). These findings motivate the question of which component of tonal-state dynamics—displacement, uncertainty, or attenuation of prior context—is most closely related to executive-control recruitment.

An important distinction in addressing this question concerns transition direction. Because λ is a monotonic function of posterior overlap, it is symmetric in the pair of tonal-state posteriors and is therefore not a directional transition measure. Likewise, S_Σ records their combined concentration without indicating whether the tonal-state distribution becomes more or less concentrated over time. Such directional change could instead be characterized by the signed difference $S(q) - S(p)$. Listener-side updating, by contrast, may be strongly directional and time dependent. A parallel-mode switch from major to minor, for example, could transiently broaden the inferred tonal-state distribution and so reduce posterior concentration, increasing uncertainty and the demands placed on cognitive control; the reverse switch need not follow the same trajectory. Consistent with this possibility, parallel-mode switches from major to minor have elicited increased activity in frontoparietal regions implicated in conflict monitoring and working memory, whereas parallel-mode switches from minor to major have not (Tsai et al. 2023). Such asymmetry may therefore reflect differences in how tonal-state beliefs evolve over time rather than differences in λ.

While λ provides a stimulus-side estimate of tonal continuity, event and formal structure may influence whether and how strongly prior carry-over is actually attenuated by listeners. Event Segmentation Theory proposes that substantial contextual change and transient increases in prediction error promote updating of the current event model (Zacks et al. 2007), while experimental work shows that information preceding an event boundary becomes less accessible to subsequent processing (Swallow et al. 2009). In music, boundaries may be signaled by pauses and changes in harmony, register, texture, dynamics, motivic material, and sectional structure (Clarke and Krumhansl 1990; Bruderer et al. 2009). The same value of λ therefore need not correspond to the same degree of listener-side attenuation. Where low tonal continuity coincides with cues marking an event or formal boundary—a configuration that may be termed *boundary-assisted leakage*—listener-side attenuation of the preceding tonal context may be stronger than in passages with comparable λ but weaker boundary cues.

Low tonal continuity, however, need not imply perceptual discontinuity. Classical sonata-form development provides an important precedent: although development sections are marked by tonal instability, they can preserve formal coherence through sequence, motivic recurrence, and continued transformation of previously established thematic material (Caplin 1998). Tsai (2024) similarly argued that the sequential modulations characteristic of development sections require repeated updating of the tonal model, while motivic and phrase repetition maintain melodic predictability. Tonal instability can therefore coexist with continuity supported by other musical dimensions. Wagner extended this developmental logic into a more continuous nineteenth-century musical process, generalizing procedures associated with development beyond a bounded formal section and drawing on the developmental discourse of Beethovenian symphonic music (Berger 2016; Bribitzer-Stull 2015). In the *Tristan und Isolde* Prelude, λ falls below .5 at m. 10, yet motivic recurrence and sequential motion provide countervailing cues that bind successive tonal contexts into a single unfolding process. This configuration exemplifies *continuity-supported persistence*: despite low λ, continuity in other musical dimensions can sustain integration when tonal continuity becomes fragile.

A methodological contribution of this study is the bracketing of the two constants that set effective evidence strength, together with an explicit assessment of whether the principal results hold across the region the bracketing leaves open. Because those constants govern posterior sharpness, every comparison the model supports inherits whatever values are chosen. More generally, probabilistic models that convert losses or likelihoods into posterior support require a choice of evidence scale or learning rate (Bissiri et al. 2016; Holmes and Walker 2017), with related scaling problems arising in perceptual categorization and Bayesian cue integration (Nosofsky 1986; Ernst and Banks 2002). In tonal modeling, Gonzalez-Espinoza and Plotkin (2026) confronted a closely related distance-to-probability scaling problem and fixed the corresponding scale from tonal-space geometry. The present study instead constrains the scale using an explicit musical-plausibility criterion: a parallel-mode modulation should register as a distinct event without registering as a break in tonal continuity, and two such passages of differing harmonic density suffice to confine both constants to a bounded region. The formulation does not claim a uniquely correct value; it makes the scaling of musical evidence explicit, grounds it in a stated perceptual requirement, and leaves its analytical consequences auditable.

The present findings should be considered in light of several limitations. First, all repertoire was evaluated with a single major/minor profile family. Holding the observer constant gives a common representational scale across the corpus, but whether the same historical trend would emerge under a profile family other than TKP remains untested here. Second, subcorpus median year covaries with composer, genre, instrumentation, and repertoire composition. Expanding the analysis to a larger and more diverse set of subcorpora would provide greater leverage for disentangling temporal change from these

correlated repertoire factors, and would permit more reliable multivariable or hierarchical modeling of their respective contributions to the observed historical trends.

Beyond expanding the empirical scope of the analysis, a promising theoretical direction is to situate the leakage threshold within broader accounts of adaptive inference that ask when an established representation should be maintained, updated, or replaced. Change-point and hierarchical Bayesian models adjust the relative influence of prior and new information according to environmental change, uncertainty, and volatility (Nassar et al. 2010; Mathys et al. 2014). Complementary accounts ask whether discrepant information should update an existing representation or establish a new one, as in memory-updating and latent-cause models (Sevenster et al. 2014; Gershman et al. 2014, 2017). Recent neuroimaging evidence further suggests that prediction precision may influence whether unexpected information updates an existing memory representation or is encoded separately (Boeltzig et al. 2026).

Applied to tonal inference, a λ trough marks a moment at which listeners may attenuate the preceding tonal context or, in more extreme cases, reset the current tonal representation. A useful computational analogy is the hierarchical multiscale long short-term memory network (HM-LSTM; Chung et al. 2016), in which learned boundary detectors regulate whether representations are maintained, updated, or reset across multiple temporal scales. The same architecture has been used to conceptualize the integration of nested musical events across levels of song structure (Tsai 2026). Although the present single-span analysis does not yet test this multiscale perspective, this analogy suggests how λ might operate across scales. At a one-bar span, low λ can mark local tonal reorientation within a phrase, as when an applied dominant prepares a tonicized half cadence. At an eight-bar span, low λ would instead signal fragile continuity between more sustained tonal contexts, as in the transition to a closely related secondary key area in a sonata exposition. This multiscale hypothesis warrants direct investigation in future work.

This account also suggests a concrete experimental strategy. Future studies could construct stimuli that vary systematically in λ while independently manipulating boundary cues or motivic-textural continuity. At comparable levels of λ, stronger boundary cues should promote attenuation of the preceding tonal context, whereas stronger motivic or textural continuity may favor persistence.

If individual listeners differ in effective evidence strength, then the listener-specific tonal posteriors—and hence the corresponding λ values—would also differ, even for the same musical passage. A precedent comes from change-point learning: Nassar et al. (2010) found substantial individual differences in belief updating, with some participants weighting new observations more heavily and others their prior beliefs, part of this variability being attributable to subjective expectations about the rate of environmental change. Related work in perceptual decision-making shows stable individual differences

in Bayesian evidence weighting and in evidence-accumulation efficiency (Goodwin et al. 2024; Nunez et al. 2015), and trial-to-trial fluctuations in attention can alter the accumulation rate even when the physical stimulus is unchanged (Kelly and O'Connell 2013). Listener differences in evidence weighting may correspond, at the model level, to different effective evidence-strength settings.

The harmonic density h_b also warrants further consideration because it should not be interpreted as a direct perceptual quantity. It measures the density of annotated harmonic events, whereas the effective contribution of those events to tonal inference may depend on the listener and the listening strategy. A listener with less implicit knowledge of common-practice harmonic syntax, for example, may make less effective use of increases in harmonic-event density, which would be represented in the present parameterization by a smaller α; a listener who draws less from harmonic evidence in general, at any density, would instead be represented by a smaller k. Such individual differences in the effective weighting of harmonic evidence warrant direct investigation. A separate issue is that harmonic density also depends on the granularity of harmonic annotation, so the same passage may yield different values under different analytical representations. A granularity that rescales every movement's h_b by the same factor is absorbed into k. Only granularity that varies systematically across repertoire would distort those comparisons. The extent to which alternative annotation schemes produce such non-uniform changes in harmonic density is therefore another question for future study.

In conclusion, the present framework defines a tonal-state geometry in which posterior concentration and between-context displacement jointly determine λ, an index of tonal continuity. Across the corpus, mean λ was lower in later repertoire. At the level of individual transitions, however, a low λ need not imply a single perceptual outcome: continuity may be sustained by other musical cues, or attenuation of the preceding tonal context may be reinforced by boundary cues. By framing tonal continuity as an issue of sequential inference, the framework links harmonic organization, formal structure, and music cognition within a common explanatory scheme.

References

Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS (2007) Learning the value of information in an uncertain world. Nature Neuroscience 10(9):1214–1221. https://doi.org/10.1038/nn1954

Berger K (2016) The secret of music-dramatic form: music drama as opera. In: Beyond Reason. https://doi.org/10.1525/california/9780520292758.003.0002

Bissiri PG, Holmes CC, Walker SG (2016) A general framework for updating belief distributions. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 78(5):1103–1130. https://doi.org/10.1111/rssb.12158

Boeltzig M, Siestrup S, Bramão I, Schubotz RI (2026) To update or to separate: Neural signatures and consequences of latent cause inference in episodic memory. NeuroImage 332:121898. https://doi.org/10.1016/j.neuroimage.2026.121898

Bribitzer-Stull M (2015) Understanding the leitmotif: From Wagner to Hollywood film music. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781316161678

Bruderer MJ, McKinney MF, Kohlrausch A (2009) The perception of structural boundaries in melody lines of Western popular music. Musicae Scientiae 13(2):273–313. https://doi.org/10.1177/102986490901300204

Caplin WE (1998) Classical form: A theory of formal functions for the instrumental music of Haydn, Mozart, and Beethoven. Oxford University Press, New York. https://doi.org/10.1093/oso/9780195104806.001.0001

Cheung VKM, Harrison PMC, Meyer L, Pearce MT, Haynes J-D, Koelsch S (2019) Uncertainty and surprise jointly predict musical pleasure and amygdala, hippocampus, and auditory cortex activity. Current Biology 29(23):4084–4092.e4084. https://doi.org/10.1016/j.cub.2019.09.067

Chung J, Ahn S, Bengio Y (2016) Hierarchical multiscale recurrent neural networks. https://doi.org/10.48550/arXiv.1609.01704

Clarke EF, Krumhansl CL (1990) Perceiving musical time. Music Perception 7(3):213–251. https://doi.org/10.2307/40285462

Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415(6870):429–433. https://doi.org/10.1038/415429a

Farbood M M (2016) Memory of a tonal center after modulation. Music Perception 34(1):71–93. https://doi.org/10.1525/mp.2016.34.1.71

Gershman SJ, Monfils M-H, Norman KA, Niv Y (2017) The computational nature of memory modification. eLife 6:e23763. https://doi.org/10.7554/eLife.23763

Gershman SJ, Radulescu A, Norman KA, Niv Y (2014) Statistical computations underlying the dynamics of memory updating. PLOS Computational Biology 10(11):e1003939. https://doi.org/10.1371/journal.pcbi.1003939

Glaze CM, Kable JW, Gold JI (2015) Normative evidence accumulation in unpredictable environments. eLife 4:e08825. https://doi.org/10.7554/eLife.08825

Gold BP, Pearce MT, Mas-Herrero E, Dagher A, Zatorre RJ (2019) Predictability and uncertainty in the pleasure of music: A reward for learning? Journal of

Neuroscience 39(47):9397–9409. https://doi.org/10.1523/JNEUROSCI.0428-19.2019
Gonzalez-Espinoza A, Plotkin JB (2026) Quantifying the evolution of harmony and novelty in western classical music. PLoS One 21(8):e0356548. https://doi.org/10.1371/journal.pone.0356548
Goodwin I, Hester R, Garrido MI (2024) Temporal stability of Bayesian belief updating in perceptual decision-making. Behavior Research Methods 56(6):6349–6362. https://doi.org/10.3758/s13428-023-02306-y
Guan X, Hentschel J, Moss FC, Rohrmeier M (2026) Studying the diachronic development of chromaticism and dissonance in Western music over three centuries. Humanities and Social Sciences Communications. https://doi.org/10.1057/s41599-026-08811-7
Hentschel J, Rammos Y, Neuwirth M, Rohrmeier M (2025) A corpus and a modular infrastructure for the empirical study of (an)notated music. Scientific Data 12(1):685. https://doi.org/10.1038/s41597-025-04976-z
Holmes CC, Walker SG (2017) Assigning a value to a power likelihood in a general Bayesian model. Biometrika 104(2):497–503. https://doi.org/10.1093/biomet/asx010
Huron D, Parncutt R (1993) An improved model of tonality perception incorporating pitch salience and echoic memory. Psychomusicology: Music, Mind, and Brain 12(2):154–171. https://doi.org/10.1037/h0094110
Itti L, Baldi P (2009) Bayesian surprise attracts human attention. Vision Research 49(10):1295–1306. https://doi.org/10.1016/j.visres.2008.09.007
Kelly SP, O'Connell RG (2013) Internal and external influences on the rate of sensory evidence accumulation in the human brain. Journal of Neuroscience 33(50):19434–19441. https://doi.org/10.1523/JNEUROSCI.3355-13.2013
Kopp D (2002). Chromatic transformations in nineteenth-century music. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511481932
Krumhansl C, Kessler E (1982) Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. Psychological Review 89:334–368. https://doi.org/10.1037/0033-295X.89.4.334
Krumhansl C (1990) Cognitive foundations of musical pitch. Oxford University Press, Oxford.
Li C-W, Guo F-Y, Tsai C-G (2021) Predictive processing, cognitive control, and tonality stability of music: An fMRI study of chromatic harmony. Brain and Cognition 151:105751. https://doi.org/10.1016/j.bandc.2021.105751
López NN, Arthur C, Fujinaga I (2019) Key-finding based on a hidden Markov model and key profiles. In DLfM '19: Proceedings of the 6th International Conference on Digital Libraries for Musicology (pp. 33–37). Association for Computing Machinery. https://doi.org/10.1145/3358664.3358675
Mathys CD, Lomakina EI, Daunizeau J, Iglesias S, Brodersen KH, Friston KJ, Stephan KE (2014) Uncertainty in perception and the Hierarchical Gaussian Filter. Front Hum Neurosci 8:825. https://doi.org/10.3389/fnhum.2014.00825
Mehr SA (2025) Core systems of music perception. Trends in Cognitive Sciences 29(8):763–777. https://doi.org/10.1016/j.tics.2025.05.013

Miller JW, Dunson DB (2019) Robust Bayesian inference via coarsening. Journal of the American Statistical Association 114(527):1113–1125. https://doi.org/10.1080/01621459.2018.1469995
Mizener B, Dowling WJ (2022) Real-time modulation perception in Western classical music. Music Perception 39(5):484–502. https://doi.org/10.1525/mp.2022.39.5.484
Moss FC, Lieck R, Rohrmeier M (2024) Computational modeling of interval distributions in tonal space reveals paradigmatic stylistic changes in Western music history. Humanities and Social Sciences Communications 11(1):684. https://doi.org/10.1057/s41599-024-03168-1
Moss FC, Neuwirth M, Rohrmeier M (2023) The line of fifths and the co-evolution of tonal pitch-classes. Journal of Mathematics and Music 17(2):173–197. https://doi.org/10.1080/17459737.2022.2044927
Murphy KP (2022) Probabilistic machine learning: An introduction. MIT Press, Cambridge, MA.
Nassar MR, Wilson RC, Heasly B, Gold JI (2010) An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. Journal of Neuroscience 30(37):12366–12378. https://doi.org/10.1523/JNEUROSCI.0822-10.2010
Nosofsky RM (1986) Attention, similarity, and the identification-categorization relationship. Journal of Experimental Psychology: General 115(1):39–57. https://doi.org/10.1037/0096-3445.115.1.39
Nunez MD, Srinivasan R, Vandekerckhove J (2015) Individual differences in attention influence perceptual decision making. Frontiers in Psychology 6:18. https://doi.org/10.3389/fpsyg.2015.00018
Pearce MT (2018) Statistical learning and probabilistic prediction in music cognition: Mechanisms of stylistic enculturation. Annals of the New York Academy of Sciences 1423(1):378–395. https://doi.org/10.1111/nyas.13654
Pearce MT, Wiggins GA (2012) Auditory expectation: The information dynamics of music perception and cognition. Topics in Cognitive Science 4(4):625–652. https://doi.org/10.1111/j.1756-8765.2012.01214.x
Perttu D (2007) A quantitative study of chromaticism: Changes observed in historical eras and individual composers. Empirical Musicology Review 2:47–54. https://doi.org/10.18061/1811/24822
Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77(2):257–286. https://doi.org/10.1109/5.18626
Rényi A (1961) On measures of entropy and information. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol 1. University of California Press, Berkeley, pp 547–561.
Sevenster D, Beckers T, Kindt M (2014) Prediction error demarcates the transition from retrieval, to reconsolidation, to new learning. Learning & Memory 21:580–584. https://doi.org/10.1101/lm.035493.114
Swallow K, Zacks J, Abrams R (2009) Event boundaries in perception affect memory encoding and updating. Journal of Experimental Psychology: General 138:236–257. https://doi.org/10.1037/a0015631

Temperley D (1999) What's key for key? The Krumhansl-Schmuckler key-finding algorithm reconsidered. Music Perception 17(1):65–100. https://doi.org/10.2307/40285812
Temperley D (2000) The line of fifths. Music Analysis 19(3):289–319. https://doi.org/10.1111/1468-2249.00122
Temperley D (2004) Bayesian models of musical structure and cognition. Musicae Scientiae 8(2):175–205. https://doi.org/10.1177/102986490400800204
Temperley D (2007) Music and probability. MIT Press, Cambridge, MA.
Temperley D (2008) A probabilistic model of melody perception. Cognitive Science 32(2):418–444. https://doi.org/10.1080/03640210701864089
Toiviainen P, Krumhansl CL (2003) Measuring and modeling real-time responses to music: The dynamics of tonality induction. Perception 32(6):741–766. https://doi.org/10.1068/p3312
Tsai C-G (2024) Predictive processing within music form: Analysis of uncertainty and surprise in different sections of sonata form. Music & Science 7:20592043241267076. https://doi.org/10.1177/20592043241267076
Tsai C-G (2026) The musical-poetic mind: A journey through the cognitive science of popular songs (1st ed.). Jenny Stanford Publishing. https://doi.org/10.1201/9781042002351
Tsai C-G, Chien L-Y, Goh JOS (2026) Temporal dynamics of intraparietal sulcus activation in response to tonal uncertainty: An fMRI study of the diminished seventh chord. Psychology of Music https://doi.org/10.1177/03057356261442576
Tsai, C-G, Fu Y-F, & Li C-W (2023) Prediction errors arising from switches between major and minor modes in music: An fMRI study. Brain and Cognition 169:105987. https://doi.org/10.1016/j.bandc.2023.105987
Vuust P, Heggli OA, Friston KJ, Kringelbach ML (2022) Music in the brain. Nature Reviews Neuroscience 23(5):287–305. https://doi.org/10.1038/s41583-022-00578-5
Weiß C, Schreiber H, Müller M (2020) Local key estimation in music recordings: A case study across songs, versions, and annotators. IEEE/ACM Transactions on Audio, Speech, and Language Processing 28:2919–2932. https://doi.org/10.1109/TASLP.2020.3030485
Woolhouse M, Cross I, & Horton T (2016) Perception of nonadjacent tonic-key relationships. Psychology of Music 44(4):802–815. https://doi.org/10.1177/0305735615593409
Zacks J, Speer N, Swallow K, Braver T, Reynolds J (2007) Event perception: A mind–brain perspective. Psychological Bulletin 133:273–293. https://doi.org/10.1037/0033-2909.133.2.273

Figures

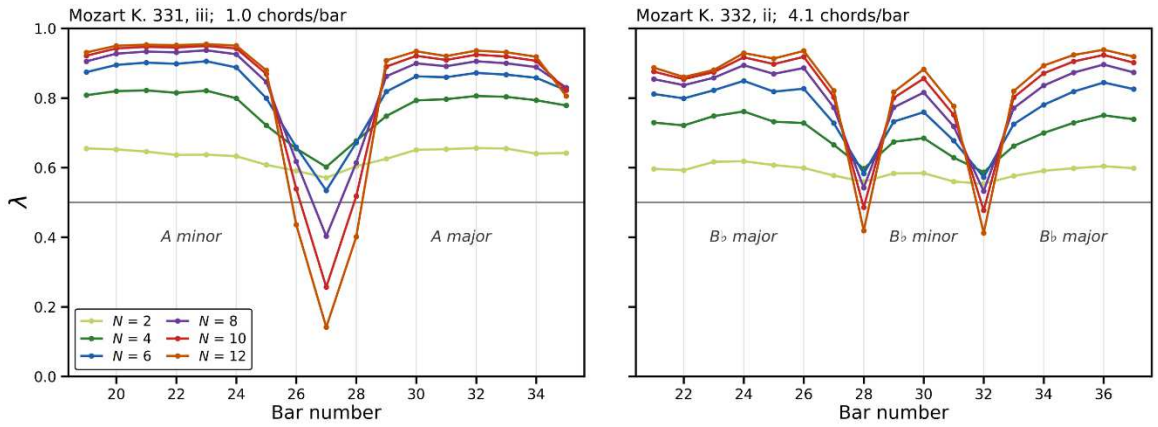


Fig. 1. Bracketing the evidence-strength constants. Leakage threshold λ at successive 4-bar endpoints through a parallel-mode modulation in each of two Mozart movements, at fixed values of *N*. Left, K. 331/iii, A minor to A major; right, K. 332/ii, B♭ major to B♭ minor and back; panel titles give each movement's harmonic density $h_{\flat}$. Each point is plotted at the last bar of the current window, the point at bar 27 thus comparing mm. 20–23 with mm. 24–27. The grey line marks the equal-mixture boundary λ = .5.

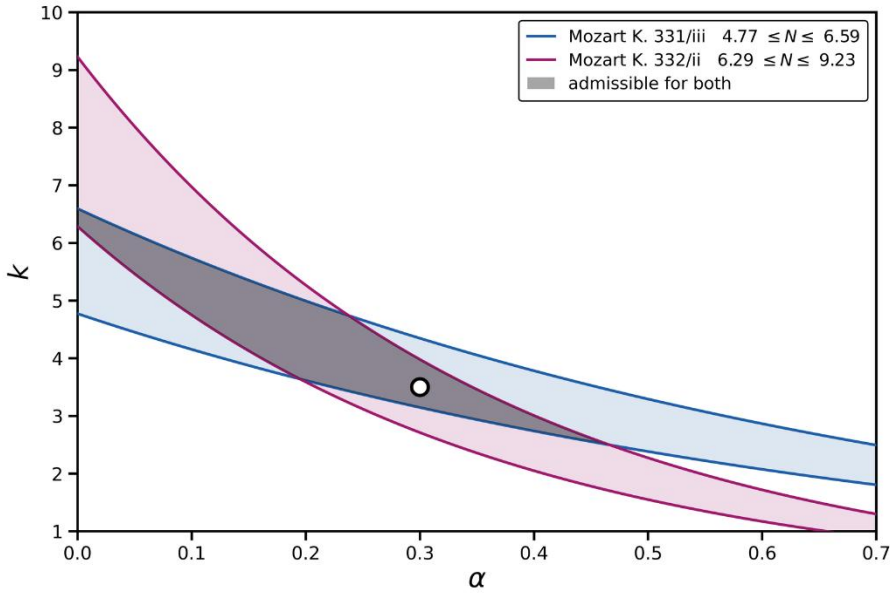


Fig. 2. The admissible region for the evidence-strength constants. Each band shows the values of (α, *k*) that satisfy both requirements of Section 3.3 for one movement. The bands differ in slope because the two movements differ in harmonic density. Their intersection, shaded, is the admissible region for the pair. The open circle marks the reference parameter setting, α = 0.30 and *k* = 3.5.

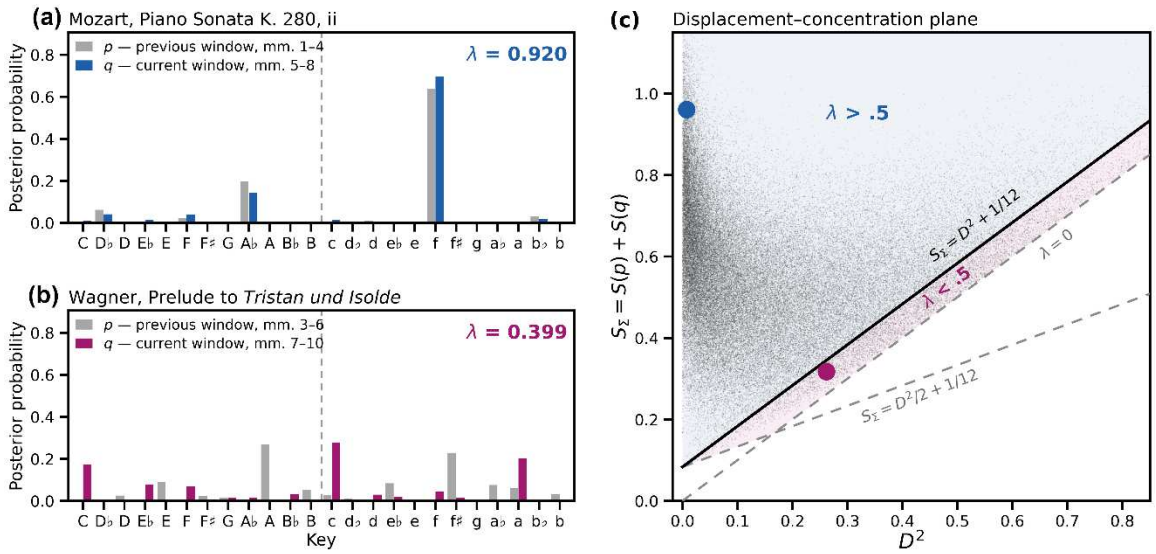


Fig. 3. Posterior concentration and displacement jointly determine the leakage threshold. (a) Mozart, Piano Sonata K. 280, ii: tonal-state posteriors for the previous window (mm. 1–4) and the current window (mm. 5–8), both committing to F minor. (b) Wagner, Prelude to *Tristan und Isolde*, mm. 3–6 and mm. 7–10: neither window commits to a single key, and the two occupy different regions of key space. (c) The two endpoints, marked by filled circles, in the displacement–concentration plane, against the pooled distribution of all endpoints at the 4-bar span. The solid line marks $\lambda = .5$. The remaining theoretical boundaries delimit zero overlap and the probability-simplex lower bound.

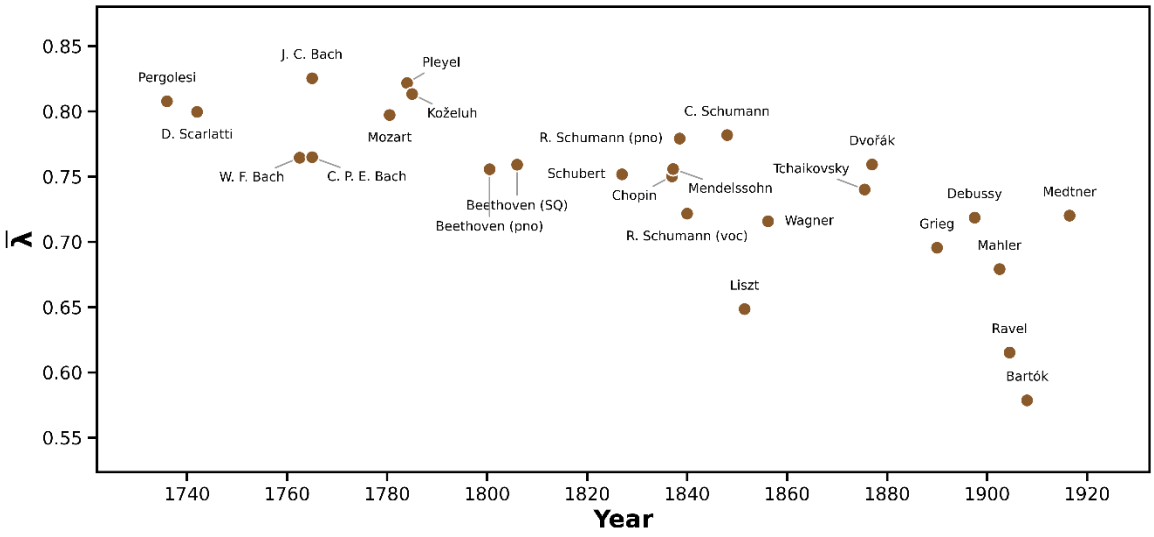


Fig. 4. Association between 4-bar mean λ and subcorpus median composition year. Each point represents one of the 26 subcorpora. Abbreviations: SQ, string quartet; pno, piano; voc, vocal. Subcorpus composition is reported in Supplementary Table S1.

Table

Table 1. Endpoint-level within-movement Spearman associations between five tonal descriptors and Bayesian tonal-state quantities.

Descriptor	λ	S_Σ	D^2
Out-of-key chromatic density	−.501 (< .001)	−.553 (< .001)	+.232 (< .001)
Chordal chromaticity	−.392 (< .001)	−.383 (< .001)	+.197 (< .001)
Non-chordal chromaticity	−.171 (< .001)	−.212 (< .001)	+.058 (.006)
Chordal dissonance	−.160 (< .001)	−.222 (< .001)	+.043 (.071)
Fifths displacement	−.547 (< .001)	−.167 (< .001)	+.542 (< .001)

Note. All quantities are evaluated at an integration span of 4 bars, with the observer at the reference values $\alpha = 0.30$ and $k = 3.5$. Spearman correlations were computed across endpoints within each movement, Fisher-z transformed, averaged within movement and then within subcorpus, averaged across the 26 subcorpora with equal weight, and back-transformed. Movements are those contributing at least 20 endpoints with non-constant values of the descriptor. Values in parentheses are two-sided Wilcoxon signed-rank p values testing the 26 subcorpus means against zero.

Data availability

The musical scores and expert harmonic annotations analysed in this study are openly available from the Distant Listening Corpus v3.2 (**https://doi.org/10.5281/zenodo.13844105**). The datasets derived in this study are deposited at Zenodo (**https://doi.org/10.5281/zenodo.23083815**), accessible to editors and reviewers during peer review and openly available from the date of publication.

Code availability

The analysis code required to reproduce all reported results and figures is deposited in the same Zenodo record (**https://doi.org/10.5281/zenodo.23083815**), under the same access terms.

Funding

The author declares that no funds, grants, or other support were received during the preparation of this manuscript.

Competing interests

The author declares no competing interests.

Ethics statements

Ethical approval was not required because this study analyzed existing musical-score and expert-annotation data and did not involve human participants, personal data, or animal subjects.

Author contributions

C.-G. Tsai conceived the study, developed the theoretical and computational framework, conducted the data processing and formal analyses, interpreted the results, prepared the figures and tables, and wrote and revised the manuscript.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SItonalcontinuity.docx