Tonalness: A Signal-Domain Transform for Virtual Pitch, with Implications for Musical Harmony

Gregory J. Strohman

ORCID: 0009-0008-8077-3842

Abstract

This paper defines a transform that inputs a sound pressure signal and returns a single time-varying function of frequency, the tonalness function. It folds critical band filtering, frequency response, and subharmonic projection of virtual pitch into the Fourier transform such that every quantity in the framework is read from that one function. The framework includes a function that maps a continuous landscape of candidate fundamentals generalizing the chord root, a companion landscape of fluctuation, a dissonance measure, and a function measuring the migration of acoustic mass between successive sonorities. Two novel mechanisms are introduced. The first is a coincidence metric that is applied in two directions and conserves salience between them, making a virtual pitch as strong as the harmonics corroborating it while withdrawing from each spectral pitch exactly the salience it supplies to the fundamentals beneath it. The second is that the summation of virtual pitch contributions includes phase. This models virtual pitches that beat at a rate set by the alignment of the spectral pitches that contribute to them. Roughness and harmonicity are read from one function at two timescales rather than modeled separately and combined by fitted weights. Beating, timbre-dependent dissonance, the register dependence of harmonic dissonance, and the differing fusion of sonorities with identical intervallic content arise as natural consequences of the framework. Five predictions are stated, several distinguishing the model from existing models.

Keywords: virtual pitch, tonalness, consonance, roughness, tonal fusion, harmonic distance, psychoacoustics

---

1. Introduction

A theory of harmony that begins from Western notation makes several unstated assumptions: that a harmony is a set of discrete pitches drawn from a fixed scale, tuned in a particular way, played on an instrument with a harmonic or near-harmonic timbre. While such a theory was practical for the repertoire addressed by Rameau (1971 [1722]) in the 18th century, it has been under strain ever since. A generalizable contemporary theory of harmony should be able to be applied directly to any acoustic signal.

Fourier analysis is indispensable for describing the frequency content of an acoustic signal, but the limits of pure spectral models of human hearing are well established in empirical studies. Listeners discriminate frequency ratios in ways spectral models do not entirely predict (Schellenberg and Trehub 1994), and dissonance perception can be lost if damage occurs to the auditory cortex (Peretz et al. 2001).

This paper defines a tonalness function $\tau(f, t)$ of frequency and time whose input is an arbitrary sound pressure signal $P(t)$. This feature enables τ to analyze sound pressure signals with discrete or continuous spectra, consisting of harmonic or inharmonic timbres, under both static and dynamic conditions. τ is the framework's central quantity and the sole object from which its results are read: a steady-state landscape $\psi_1(f, t)$, a fluctuation landscape $\psi_2(f, t)$, a dissonance metric $\eta(t)$, and a progression function $\Omega(t_1, t_2)$.

The aim of this paper is a small set of assumptions from which many psychoacoustic phenomena relevant to musical analysis follow, rather than a model in which each phenomenon is separately stipulated. This model should be judged not by the number of effects it addresses but by the ratio

of those effects to the number of assumptions required to reach them. This paper is organized as follows: Section 2 states what the model inherits and from whom, Section 3 defines the tonalness function $\tau(f, t)$, Section 4 states what is novel in this framework, Sections 5 and 6 give formal properties and the closed forms that make it computable, Section 7 shows the consequences of the model, and Section 8 states what could falsify it.

This paper offers no quantitative comparison against competing models, as such a comparison could only be properly addressed in a separate study. It also does not address the synoptic comprehension of musical works. The tonalness framework concerns the immediate apprehension of a sonority—what the auditory system generates before memory, expectation, and learned syntax can be processed. It is a model of one layer, and synoptic theories of music require several.

2. What the Model Inherits

The tonalness framework consists in part of the following six components, each taken from prior research with little to no modification. This list separates borrowed from novel concepts and locates the model within a body of literature it does not attempt to replace.

Fourier analysis. The transform of Equation 1 is a point of departure for the tonalness model, not a contribution.

Critical bandwidth. The width parameter of Equation 4 approximates the equivalent rectangular bandwidth models of Glasberg and Moore (1990; Moore 1998), scaled by $\kappa = 0.25$ to approximate the quarter critical band threshold of Plomp and Levelt (1965).

Frequency response. Equation 5 is the standard A-weighting curve, which approximates the inverse of the 40-phon equal-loudness contour descending from Fletcher and Munson (1933). As

a single fixed curve, it is independent of level. This is a simplification, since equal-loudness contours flatten at higher levels. It is adopted here for algebraic convenience, and the choice is modular, since R enters only through pointwise computation. Level-dependent models such as ISO 226:2023 can be substituted without altering any closed form.

Subharmonic projection of virtual pitch. The claim that the human auditory system infers virtual fundamentals beneath spectral pitches was made by Terhardt (1979), further modeled by Terhardt, Stoll, and Seewann (1982), and adapted to musical harmony by Parncutt (1988, 1989), whose proposal that a chord project a profile of candidate roots rather than a single root is the most comparable to the tonalness framework described in this paper. Leman (2000) pursued a related account through temporal theory and periodicity rather than place theory and harmonic templates.

Roughness as interference between partials. Helmholtz (1954 [1877]) first described dissonance as the sensory roughness produced when partials of two tones lie close enough in frequency to beat rapidly. Plomp and Levelt (1965) verified the account experimentally and tied it to the critical band, locating maximal roughness near a quarter of a critical bandwidth. Hutchinson and Knopoff (1978) created a computable model. Sethares (1993, 2005) established that roughness depends on the spectrum of the tones sounding within an interval, that consonance depends on tuning and timbre rather than interval class, and that the dissonance minima for inharmonic timbres can migrate away from simple ratios entirely. Vassilakis (2001) refined the amplitude dependence of roughness. The tonalness framework can reproduce the same findings as Sethares for inharmonic timbres. This is not a new finding but rather a check on the tonalness model. Harrison and Pearce (2020) evaluated sixteen consonance models against four behavioral datasets and three corpora and concluded that consonance arises from interference, harmonicity, and cultural familiarity and is not reducible to any one metric.

Peak rather than mean salience. Equations 9 and 10 take maxima over a temporal window. This follows evidence that auditory perception tracks peak rather than average activity (Penagos, Melcher, and Oxenham 2004; Engell et al. 2016; Song et al. 2025).

Three further bodies of work influenced the progression function described in Section 3.4 without being incorporated directly into it. Lerdahl (2001) assigned distances within a cognitive-structural framework. Tymoczko (2011) developed a geometry in which voice-leading distance is a metric on a chord space. Milne et al. (2011) modeled the similarity of pitch collections through spectral expectation fitted to behavioral data.

3. The Tonalness Transform

Frequencies are expressed in hertz (Hz) and times in seconds (s) throughout.

3.1 The psychoacoustic Fourier function

Equation 1 is the ordinary Fourier transform.

$$X(f) = 2\int_{-\infty}^{\infty} e^{-2\pi i t f} P(t)\, dt \tag{1}$$

Equation 2 defines a family of transforms indexed by a harmonic number h. Each member describes how much evidence there is at harmonic h of f for a candidate fundamental at f.

$$Y_h(f,t) = 2R(hf)\int_{-\infty}^{\infty} \frac{e^{-\frac{(u-t)^2}{2\hat{\sigma}_h(f)^2}} e^{-2\pi i (u-t) h f} P(u)}{\sqrt{2\pi}\hat{\sigma}_h(f)}\, du \tag{2}$$

$$\hat{\sigma}_h(f) = \frac{1}{2\pi\sigma_h(f)} \tag{3}$$

$$\sigma_h(f) = \kappa\left(\frac{24.7}{h} + 0.107939 f\right) \tag{4}$$

$$R(f) = \frac{12194^2 f^4}{\left(f^2 + 20.6^2\right)\left(f^2 + 12194^2\right)\sqrt{\left(f^2 + 107.7^2\right)\left(f^2 + 737.9^2\right)}} \tag{5}$$

The Gaussian window makes the transform local in time, so Y_h and everything derived from it varies as the signal varies. Its width $\hat{\sigma}_h$ is fixed by the reciprocal relation of Equation 3 to the critical bandwidth, so frequency uncertainty and temporal resolution are not independently tunable. The model commits to a single scale parameter κ and takes the tradeoff as given. The demodulation term removes the carrier introduced by evaluating at hf rather than f. Equation 2 carries no weighting by harmonic index. The amplitude prior of A2 enters through Equation 7 instead, for reasons stated in Section 3.5.

3.2 The coincidence gate and tonalness function

$$\tau(f,t) = \left| \sum_{h=1}^{\infty} W_h(f,t) \left| Y_h(f,t) \right| \exp\left(\frac{i\Phi_h(f,t)}{h} \right) \right| \tag{6}$$

$$W_h(f,t) = \begin{cases} \sqrt{1 - \sum_{k=2}^{\infty} W_k(f/k,t)^2 \frac{|Y_1(f/k,t)|}{|Y_1(f/k,t)| + \zeta}} & \text{if } h = 1 \\ \frac{1}{h}\left(1 - \frac{|Y_h(f,t)|}{\sum_{g=1}^{\infty} |Y_g(f,t)| + \delta}\right) & \text{otherwise} \end{cases} \tag{7}$$

$$\Phi_h(f,t) = \arg\left(Y_h\left(f,t_0\right)\right) + \int_{t_0}^{t} \Im\left(\frac{\frac{\partial Y_h(f,u)}{\partial u}}{Y_h(f,u)} \right) du \tag{8}$$

Equation 7 is the heart of the framework. It is a single measure of coincidence applied in two directions, building virtual pitch out of spectral pitch at $h > 1$ and withdrawing from each spectral pitch at $h = 1$ exactly what it has supplied elsewhere. Everything the model says about fusion, and about why a harmonic timbre is heard as one pitch rather than many, follows from those two readings of one quantity.

Virtual pitch is read upwards ($h > 1$) by coincidence. The bracketed factor returns one minus the share of the total harmonic support at f contributed by the harmonic index h alone. This

measures how many values of h contribute to the candidate fundamental at f. When many harmonics converge there, each contribution is a small fraction of the whole and the factor approaches 1 for all of them. When only one pitch supports the candidate, the factor closes to zero. A virtual pitch is thus governed by the breadth of its corroboration rather than by the magnitude of any one partial. As a result, the deliberate over-generation of candidate fundamentals in Equation 2 does not produce a forest of spurious virtual pitches. The prefactor $1/h$ is the amplitude prior of A2, applied here rather than inside Equation 2 for a reason the next paragraphs make clear.

Spectral pitch is read downward ($h = 1$) by conservation. A spectral pitch at f is the k-th harmonic of every candidate fundamental at f/k. The $h = 1$ branch sums the squared gate values of those fundamentals and subtracts the total from 1. Squared salience is conserved across the relation. Whatever a partial gives to the fundamentals beneath it, it loses from its own standing as an independent pitch. A partial doing no work below itself keeps its full salience.

3.3 Steady state tonalness and tonalness variance

The tonalness function τ is the central quantity of the framework. It represents the point at which every psychoacoustic component has been combined: critical band, frequency response, subharmonic projection, coincidence, and phase all enter before Equation 6 and none after it. Everything the framework reports is contained within τ. The central analytic tools of the framework, each read from τ, are defined by Equations 9–13. This includes the steady state tonalness (SST) function ψ_1, the tonalness variance (TV) function ψ_2, the dissonance function η, and the tonalness progression function (TPF) $\boldsymbol{\Omega}$.

$$\psi_1(f,t) = \max_{|u-t|<\frac{T_c}{2}} \tau(f,u) \tag{9}$$

$$\psi_2(f,t) = \sqrt{\frac{1}{f} \max_{|u-t|<\frac{T_c}{2}} \left| \frac{\partial}{\partial u} \tau(f,u)^2 \right|} \tag{10}$$

The window $T_c \approx 0.05$ s follows from taking 20 Hz as the approximate lower limit of pitch perception, and it partitions the temporal behavior of τ into two regions. Fluctuations slower than $1/T_c$ are heard as rhythm and survive into ψ_1. This includes beating, tremolo, articulation, vibrato, etc. Faster fluctuations register instead as unsteadiness of the sonority itself and are captured by ψ_2.

This is how the framework distinguishes between beating and roughness. They are modeled as a continuum of a single phenomenon: variation in τ. A slightly mistuned unison beats slowly and audibly as a rhythm. The same mistuning widened to a semitone varies too quickly to follow and is heard instead as roughness. Nothing changes except the rate. Section 7.3 shows this differentiation doing consequential work in lower registers where a semitone is narrow enough in hertz that it beats rather than roughens. Both ψ_1 and ψ_2 carry the same units as τ. The square root in Equation 10 is what makes the landscapes commensurable and the ratio of Equation 11 dimensionally coherent.

3.4 Dissonance and tonalness progression

$$\eta(t) = \frac{\int_0^\infty \frac{\psi_2(f,t)^2}{\sigma_1(f)}\, df}{\int_0^\infty \frac{\psi_1(f,t)^2}{\sigma_1(f)}\, df + \delta} \tag{11}$$

$$\mu(t) = \sqrt{\int_0^\infty \frac{\psi_1(f,t)^2}{\sigma_1(f)}\, df} + \delta \tag{12}$$

$$\Omega\left(t_1, t_2\right) = \frac{1}{2}\int_0^\infty \frac{1}{\sigma_1(f)}\left(\frac{\psi_1\left(f, t_1\right)}{\mu\left(t_1\right)} - \frac{\psi_1\left(f, t_2\right)}{\mu\left(t_2\right)}\right)^2 df \tag{13}$$

Equation 11 defines dissonance as total fluctuation energy normalized by total steady-state energy. It represents the proportion of a sonority's tonalness that fails to hold still. Since both

integrands derive from the same function, the harmonicity and interference accounts of dissonance are not two models to be weighted together but two readings of one landscape.

Equations 12 and 13 collectively define the tonalness progression function (TPF) Ω. Equation 12 gives the magnitude of a steady-state landscape in the critical-band-weighted norm. Equation 13 is half the squared distance between two landscapes once each has been divided by its own magnitude. Normalizing first is what makes Ω a comparison of shape rather than of size. Two sonorities whose landscapes differ only by a constant factor return $\Omega = 0$, no matter how far apart their absolute tonalness. Expanding the square shows Ω to be one minus the normalized overlap of the two landscapes. This is equivalent to a cosine distance, reaching 1 only when they share no harmonicity anywhere in the spectrum.

All three integrals are weighted by $1/\sigma_1(f)$ and taken over positive frequencies. The weighting matters. Since $\sigma_1(f)$ is the critical bandwidth at f, the quantity $df/\sigma_1(f)$ is the differential of the auditory frequency scale, so the integrals accumulate energy per critical band rather than per hertz. A measure integrated on the physical scale would count an octave in the bass as a negligible sliver of the spectrum. The weighted form treats equal auditory distances as equal, which is what a perceptual measure requires.

3.5 Assumptions

A1. Virtual pitch is generated at every subharmonic of every spectral pitch. Equation 2 admits a candidate at $f = b/h$ for each spectral pitch b and index h. The over-generation is deliberate and corrected by Equation 7.

A2. Virtual pitch salience scales as $1/h$, and the architecture constrains how. The weighting enters through Equation 7 rather than Equation 2, and its form is bounded by the conservation law rather than merely assumed. Because the $h = 1$ branch subtracts a sum of squared gate values, what must converge is $\sum w_h^2$ and not $\sum w_h$. The harmonic series diverges while its square does not, and that is what makes a $1/h$ prior admissible at all. The constraint restricts

shape as well as steepness. Since $\sum w_h^2$ is minimized, at fixed total weight, when the weights are equal, a prior concentrating weight on a few indices spends the budget faster than one spreading it across many. A peaked prior is therefore admissible only when scaled down, and scaling is not free, because $h = 1$ carries no such factor and the balance between virtual and spectral pitch shifts with it. $1/h$ is the simplest decreasing prior the architecture allows, and a decreasing prior is what one expects of a system developed among predominantly harmonic sources. It is not a model of the dominance region, and does not need to be. Because W_h carries $1/h$ while Y_h carries $R(hf)$, the two compose into a function of the partial's absolute frequency alone. The harmonic index that dominates therefore shifts with the fundamental while the absolute frequency that dominates does not. This is the form Dai (2000) reports, against the fixed-harmonic-number reading of Ritsma (1967) and Plomp (1967). The dominance region is a consequence of this model rather than an input to it, and the mistuning experiments that measure it (Moore, Glasberg, and Peters 1985) measure a sensitivity of τ that the model computes directly. Whether some other admissible weighting would improve matters is left open.

A3. A virtual pitch is as strong as the harmonics corroborating it. A spectral pitch is weakened by exactly the corroboration it supplies. This is expressed in Equation 7, read in both directions. No independent mechanism governs fusion.

A4–A7. Inherited parameters. Critical bandwidth follows Glasberg and Moore scaled by $\kappa = 0.25$. Frequency response follows the A-weighting curve, with the modularity noted in Section 2. The window $T_c \approx 0.05$ s separates rhythm from roughness. Salience over that T_c window is the maximum rather than the mean.

A8. Sums over h and k are truncated at 10. This is a numerical approximation, not a modeling choice. Bounds and predictions are stated for the untruncated sums.

A9. Spectral masking is not a separate stage. The gate performs the function masking would otherwise serve, suppressing weakly supported candidates in dense spectra and, at $h = 1$, withdrawing partials subsumed by a sounding series.

A10. The model describes what the auditory system generates, not what a listener attends to. τ reports many simultaneous virtual pitches, of which a listener notices few. It should be read as the field from which attention selects.

A11. Nonlinear distortion products are not modeled. The model generates no combination tones, but those of musical significance coincide with virtual pitches it predicts on independent grounds. For a just interval $p : q$ the quadratic difference tone falls at $(q - p)f_0$, equal to the projected fundamental exactly when $q - p = 1$ — the 2:3 fifth, 3:4 fourth, 4:5 major third, 5:6 minor third. Hence the resultant bass falls where the model already places a virtual pitch.

Scope. Outside the model as formulated: cultural familiarity and learned expectation, which Harrison and Pearce (2020) identify as a substantial component of consonance judgment; auditory streaming; adaptation and time-course effects beyond T_c; and binaural and spatial effects.

4. What Is New

Six properties distinguish this framework. The first four concern the model's structure, the last two its mechanisms.

N1. The input is a signal, not a list of pitches. Equations 2–8 are defined for any $P(t)$ whatsoever, where the models of Section 2 require that a sonority first be idealized into components — a step the analyst performs before the model runs, using judgment the model

does not contain. The tonalness framework applies to a recording, and to signals with no notated equivalent.

N2. The output is a landscape, not a ranked list. Terhardt's algorithm and Parncutt's adaptation identify a small set of pitch candidates with salience values. $\tau(f, t)$ is a continuous distribution over frequency. A landscape supports questions a list cannot, such as whether a chord's harmonicity is concentrated or dispersed, where two sonorities agree, how a fundamental's support migrates. This means any direct comparison with the template models must be made with care, since the objects returned are not of the same kind.

N3. Time is intrinsic. The models of Section 2 return a value for a sonority considered as a fixed object. Here τ varies in t by construction, which permits Equations 9 and 10 and with them a treatment of beating, attack, decay, and vibrato as consequences rather than as effects outside the model.

N4. Roughness and harmonicity share a mechanism. These have generally been modeled separately and combined, if at all, by weights fitted rather than derived. Here harmonicity is read from ψ_1 and roughness from ψ_2, both functions of τ, and Equation 7 is the single mechanism behind both. Harrison and Pearce's finding that consonance is composite is accommodated structurally rather than by adding a term.

N5. The gate conserves salience, and tonal fusion is the consequence. In the template models spectral pitches are admitted unconditionally, and whatever fusion occurs is either stipulated or assigned to a separate mechanism. The $h = 1$ branch of Equation 7 replaces that exemption with a conservation law. What a partial loses as an independent pitch is exactly what it has contributed to the fundamentals beneath it. A complex tone is heard as one pitch rather than n because the same measure that builds a fundamental out of many harmonics subtracts from each harmonic that built it. Tonal fusion, masking within a harmonic series, and the perceptual unity of timbre become three descriptions of one gate read in two directions.

N6. Virtual pitch contributions are summed with phase. Terhardt's and Parncutt's algorithms sum saliences as non-negative scalars, so contributions can only accumulate. Equation 6 sums complex quantities whose angular positions come from Equation 8, so projections converging on one candidate fundamental interfere.

To these may be added Ω, which addresses the question Lerdahl, Tymoczko, and Milne et al. address but measures it on the tonalness landscape rather than on chord space, pitch-class sets, or fitted expectation tensors. It is therefore defined between sonorities of different cardinality, in different tunings, and with no notated equivalent at all. These differences describe what the model is built to do, not how well it does it.

5. Formal Properties

Boundedness of the gate. For $h > 1$ each $|Y_g|$ is non-negative and the denominator includes $|Y_h|$, so the bracketed ratio lies in [0,1] and $W_h \in [0,1/h]$. For $h = 1$ the presence check lies in [0,1], so the subtracted sum is bounded by $\pi^2/6 - 1 \approx 0.6449$. The radicand is bounded below by 0.3551, the radical is real for every signal, and $W_1 \in [0.596,1]$. The gate can neither amplify a term nor reverse its sign. With $\delta > 0$ the ratios are defined even where P vanishes almost everywhere and $\tau \to 0$.

Range. Ω is bounded in [0,1]. Because τ is a modulus (Equation 6), ψ_1 is non-negative everywhere (Equation 9). Let $a = \psi_1(f, t_1)/\mu(t_1)$ and $b = \psi_1(f, t_2)/\mu(t_2)$, with inner product $\langle a, b\rangle = \int_0^\infty ab/\sigma_1(f)\, df$. Equation 12 gives $\|a\|, \|b\| \leq 1$, with equality as $\delta \to 0$. Expanding the square in Equation 13 gives $\Omega = \frac{1}{2}(\|a\|^2 + \|b\|^2) - \langle a, b\rangle$. By Cauchy–Schwarz this is at least $\frac{1}{2}(\|a\| - \|b\|)^2 \geq 0$. Because a and b are non-negative, $\langle a, b\rangle \geq 0$, so $\Omega \leq 1$. The endpoints have direct readings: $\Omega = 0$ when two landscapes are proportional, and $\Omega = 1$ when they share no harmonicity anywhere. Proportional is weaker than identical, since Ω compares the

distribution of harmonicity, not its total. η is non-negative, since both integrands in Equation 11 are squares, but it has no upper bound by construction. Equation 10 measures how fast τ changes relative to f, so η grows with the rate of fluctuation compared with the register in which it occurs. It stays well below 1 for every sonority in Section 7, whose largest value is about 0.57. It approaches or exceeds 1 only for dense clusters far below the range of pitch perception.

Level. The $h > 1$ branch is a ratio of like quantities, unaffected by uniform gain and exactly scale-free as $\delta \to 0$. The $h = 1$ branch is not, because ζ carries the units of $|Y_1|$. The model is therefore level-dependent, though only near the floor: where partials sit well above ζ the presence check saturates and gain simply rescales τ. So η and Ω are strictly comparable only across sonorities at a common level. All demonstrations in Section 7 satisfy this, and the model makes a claim no formulation admitting spectral pitches unconditionally can make, that there is a level below which a partial is not absorbed into a sounding series, locating the transition at ζ. P3 describes how to establish that value empirically.

Transposition. η is deliberately not invariant under transposition, since transposing alters σ_h through Equation 4 and the component loudnesses through Equation 18. The register dependence of Section 7.3 is a substantive prediction, and a transposition-invariant measure could not make it.

Cardinality. No term in Equations 11–13 depends on n, so dyads, trichords, and dense sonorities are evaluated by the same expression and are directly comparable.

Ω is symmetric, and $\sqrt{2\Omega}$ is a metric. Symmetry is immediate from Equation 13. Ω itself is half a squared distance and so does not satisfy the triangle inequality, but the quantity beneath it does. $\sqrt{2\Omega(t_1, t_2)}$ is the distance between two unit-magnitude landscapes in the critical-band-weighted norm, and distance in a normed space is a metric. Progression values may therefore be accumulated along a path, provided the accumulation is performed on $\sqrt{2\Omega}$ and not on Ω.

6. The Sinusoidal Specialization

For an additive sinusoidal signal the integrals close in elementary form, which is what makes the model computable and comparable to existing consonance models on their own inputs. Equation 14 converts MIDI pitch to frequency, with $s_0 = 69$ and $b_0 = 440$ Hz. Sonorities are specified in MIDI throughout, and MIDI supplies the axis wherever musical interval structure is at issue, being logarithmic in frequency. Equations 15–19 specialize the definitions of Section 3 to that case. Equations 20–24 give the resulting closed forms.

$$b(s) = b_0 2^{\frac{1}{12}(s - s_0)} \tag{14}$$

$$P(t) = \sum_{i=1}^{n} a_i \cos\left(2\pi b_i t + c_i\right) \tag{15}$$

$$\tau(f,t) = \left| \sum_{h=1}^{\infty} \sum_{i=1}^{n} W_h(f,t) U_{h,i}(f) e^{\frac{i\left(2\pi t b_i + c_i\right)}{h}} \right| \tag{16}$$

$$U_{h,i}(f) = l_i e^{-\frac{(hf - b_i)^2}{2\sigma_h(f)^2}} \tag{17}$$

$$l_i = a_i R\left(b_i\right) \tag{18}$$

$$W_h(f,t) = \begin{cases} \sqrt{1 - \sum_{k=2}^{\infty} W_k(f/k,t)^2 \frac{\sum_{i=1}^{n} U_{1,i}(f/k)}{\sum_{i=1}^{n} U_{1,i}(f/k) + \zeta}} & \text{if } h = 1 \\ \frac{1}{h}\left(1 - \frac{\sum_{i=1}^{n} U_{h,i}(f)}{\sum_{i=1}^{n} \sum_{g=1}^{\infty} U_{g,i}(f) + \delta}\right) & \text{otherwise} \end{cases} \tag{19}$$

Each $U_{h,i}$ is a Gaussian in f centered at b_i/h with width σ_h/h. The SST and TV landscapes follow:

$$\psi_1(f,t) = \sum_{h=1}^{\infty} \sum_{i=1}^{n} W_h(f,t) U_{h,i}(f) \tag{20}$$

$$\psi_2(f,t) = \sqrt{\frac{2\pi}{f} \sum_{h_1=1}^{\infty} \sum_{h_2=1}^{\infty} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} W_{h_1}(f,t) W_{h_2}(f,t) U_{h_1,i_1}(f) U_{h_2,i_2}(f) \left| \frac{b_{i_2}}{h_2} - \frac{b_{i_1}}{h_1} \right|} \tag{21}$$

Equation 21 states the model's account of roughness in its most transparent form. Every ordered pair of h and i combinations contributes a term proportional to the product of their contributions to the candidate fundamental at f, weighted by $|b_{i_2}/h_2 - b_{i_1}/h_1|$ — the rate at which the two inferred fundamentals beat. If two fundamentals are projected to exactly the same place, nothing is contributed. However, if two fundamentals are projected to slightly different locations, they contribute in proportion to their mistuning and to how strongly both are gated. Roughness is thus not a separate calculation over pairs of partials but the interference of virtual pitches with each other and with spectral pitches.

Integrating Equations 20 and 21 against the critical-band density yields a closed form for dissonance:

$$\eta(t) = \frac{2\pi \sum_{h_1=1}^{\infty} \sum_{h_2=1}^{\infty} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \frac{W_{h_1}\left(\frac{b_{i_1}}{h_1}, t\right) W_{h_2}\left(\frac{b_{i_2}}{h_2}, t\right) \gamma_{h_1,h_2,i_1,i_2} \left| \frac{b_{i_2}}{h_2} - \frac{b_{i_1}}{h_1} \right|}{\phi_{h_1,h_2,i_1,i_2}}}{\sum_{h_1=1}^{\infty} \sum_{h_2=1}^{\infty} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} W_{h_1}\left(\frac{b_{i_1}}{h_1}, t\right) W_{h_2}\left(\frac{b_{i_2}}{h_2}, t\right) \gamma_{h_1,h_2,i_1,i_2}} \tag{22}$$

$$\gamma_{h_1,h_2,i_1,i_2} = \frac{l_{i_1} l_{i_2} \sigma_{h_1}\left(\frac{b_{i_1}}{h_1}\right) \sigma_{h_2}\left(\frac{b_{i_2}}{h_2}\right) \exp\left(- \frac{\left(b_{i_2} h_1 - b_{i_1} h_2\right)^2}{2\left(h_2^2 \sigma_{h_1}\left(\frac{b_{i_1}}{h_1}\right)^2 + h_1^2 \sigma_{h_2}\left(\frac{b_{i_2}}{h_2}\right)^2\right)} \right)}{\sigma_1\left(\phi_{h_1,h_2,i_1,i_2}\right) \sqrt{h_2^2 \sigma_{h_1}\left(\frac{b_{i_1}}{h_1}\right)^2 + h_1^2 \sigma_{h_2}\left(\frac{b_{i_2}}{h_2}\right)^2}} \tag{23}$$

$$\phi_{h_1,h_2,i_1,i_2} = \frac{b_{i_2} h_2 + b_{i_1} h_1}{h_2^2 + h_1^2} \tag{24}$$

The roughness factor γ is the overlap integral of two Gaussian contributions divided by the critical bandwidth at the pair's centre frequency ϕ. γ is large when two projections land in the same place and falls off as their separation increases. The factor $|b_{i_2}/h_2 - b_{i_1}/h_1|/\phi$ in Equation 22 normalizes the beat rate by the register in which the beating occurs, producing the model's register dependence. The closed form for Ω is given in the Appendix, Equations 25–31.

7. Consequences

Two reference timbres recur below: a **harmonic timbre** with partials at ratios 1, 2, 3, 4, 5 and relative amplitudes 1, 1/2, 1/3, 1/4, 1/5; and an **inharmonic timbre** with partials derived from $n^{\log_2 2.1}$ for n = 1–5 (1, 2.1, 3.2412, 4.41, 5.59977) at the same amplitudes. Pure tones are used where the point concerns the model rather than the timbre. We will assume the following parameter values through these examples: $\kappa = 0.25$, $T_c = 0.05$ s, $\delta = 10^{-7}$, $\zeta = 10^{-3}$, sums truncated at 10, and all sonorities are at a common level with unit component amplitudes unless stated otherwise.

7.1 What a tonalness landscape contains

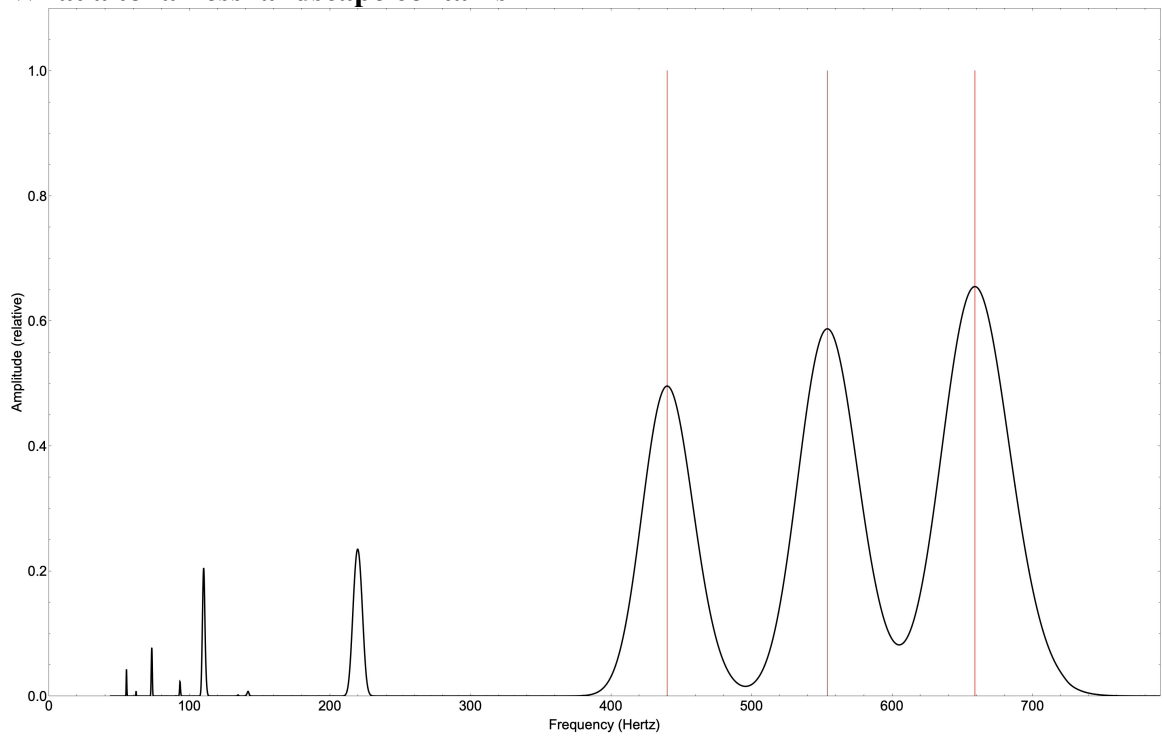


FIGURE 1. *Spectral intensity* $|X(f)|$ *(red) against* $\tau(f,0)$ *(black) for three equal-amplitude pure tones at 440, 554, and 659 Hz — an equal-tempered triad.*

Figure 1 contrasts what the ear receives with what the model says the auditory system makes of it. The Fourier spectrum has three lines of equal height. The tonalness function has those three at unequal heights — 0.496, 0.588, 0.655 — plus bell curves at the subharmonics of each. The spectral peaks are exactly $a_i R(b_i)$, the A-weighted amplitudes. The $h = 1$ gate returns unity because nothing sounds at 220, 277, or 330 Hz, so these partials surrender nothing.

Two subharmonic peaks dominate, at 220 Hz (0.235) and 110 Hz (0.205). Read against 110 Hz all three tones participate, at harmonics 4, 5, and 6 — an approximate 4:5 major third stacked with a 5:6 minor third. Read against 220 Hz only two participate, at harmonics 2 and 3. Breadth of corroboration favors the first reading and the $1/h$ prior the second. The two nearly cancel, and the fifth-based reading emerges marginally stronger. This is a chord root computed rather than assigned, and a distribution rather than a choice.

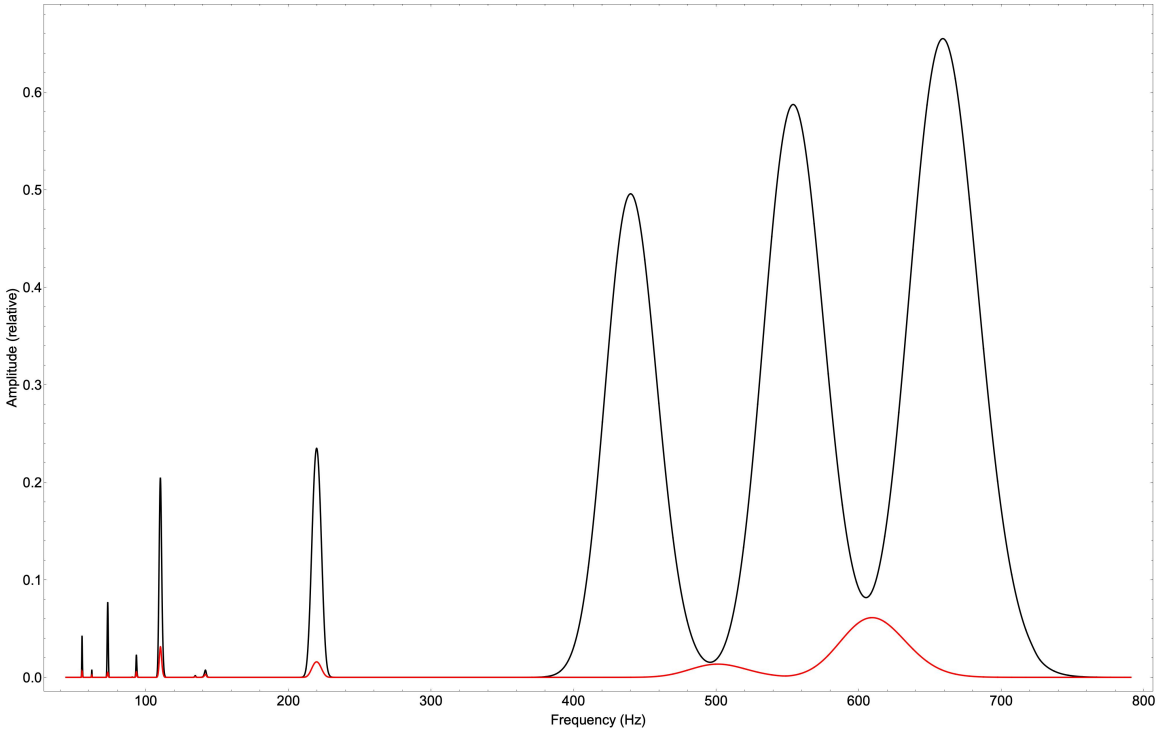


FIGURE 2. *ψ_1 (black) and ψ_2 (red) for the same triad.*

The steady-state curve tracks τ closely, because most of the oscillation in τ is too fast to survive the maximum of Equation 9. The fluctuation curve is small everywhere and structured where it is not. Its largest excursion, about 0.060, falls near 610 Hz — between the upper two partials, which lie roughly one critical band apart at that register and beat against one another. A secondary excursion near 500 Hz does the same for the lower pair. Two further peaks, at 110 Hz (0.032) and 220 Hz (0.018), have a different origin. The tempering of the triad's thirds leaves the projections converging on those fundamentals slightly out of agreement, so the virtual pitches beat. Both are the same equation, only the rate differs.

Note that ψ_2 vanishes nowhere, even where the intervals are wide. Any interval can be heard as a mistuning of some other nearby one, and the residue registers. This is the model's clearest departure from the standard roughness account, and Section 8 states it as a prediction. Section 7.4 shows the converse limiting case, where exact harmonic alignment drives ψ_2 to zero.

7.2 Timbre determines where the consonances are

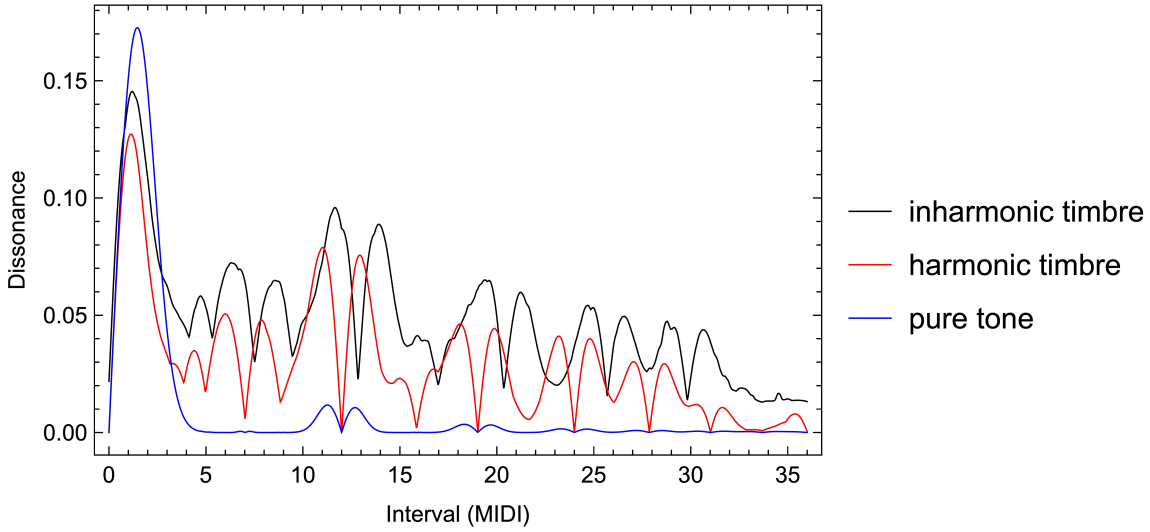


FIGURE 3. *η for dyads spanning MIDI intervals 0–36 above MIDI 60, for pure tones (blue), the harmonic timbre (red), and the inharmonic timbre (black).*

The three curves do not differ by a scale factor, and the way they differ is the point. The pure-tone curve has a single dominant maximum of 0.173 near one and a half semitones when the two tones fall inside a critical band, and falls to 0.005 by four semitones, retaining thereafter only residual excursions never exceeding 0.012. The harmonic-timbre curve reaches a lower maximum, 0.128, but sustains structure across the entire three octaves, with minima that deepen as the interval widens. We see η values of 0.010 at the fifth, 0.006 at the octave and twelfth, 0.004 at the double octave. We also observe secondary maxima of 0.079 and 0.076 near eleven and thirteen semitones, where partials of the two tones collide. The inharmonic curve peaks at 0.146 and never descends below 0.014 with its minima conspicuously shallower than those of the harmonic timbre. At the octave it reaches 0.068 against the harmonic timbre's 0.006, an order of magnitude higher. At the stretched 1:2.1 octave near 12.8447 semitones, the inharmonic curve reaches a minimum of 0.022 — about a third of its value at the octave.

Adding partials therefore does not simply amplify the pure-tone curve. It lowers the critical-band maximum, because the sonority's total tonalness, the denominator of Equation 11, grows faster than its fluctuation, and it creates structure across the whole interval range where pure tones offer

almost none. Consonance becomes a property with musical extension only once the tones have partials.

This is Sethares's result, recovered from a different starting point. It should serve as a check on the model rather than a contribution. It also explains the corollary, consistent with McDermott, Lehr, and Oxenham (2010), that an inharmonic spectrum cannot reach the minima a harmonic one reaches. The mechanism differs from Sethares's. In his account the minima move because the roughness of the interval moves with the spectrum. Here they move for that reason and also because the candidate fundamentals move. An inharmonic spectrum projects subharmonics that fail to coincide, so the gate never opens fully and both ψ_1 and ψ_2 change. Consonance and fusion degrade together, from one cause.

7.3 Dissonance depends on register

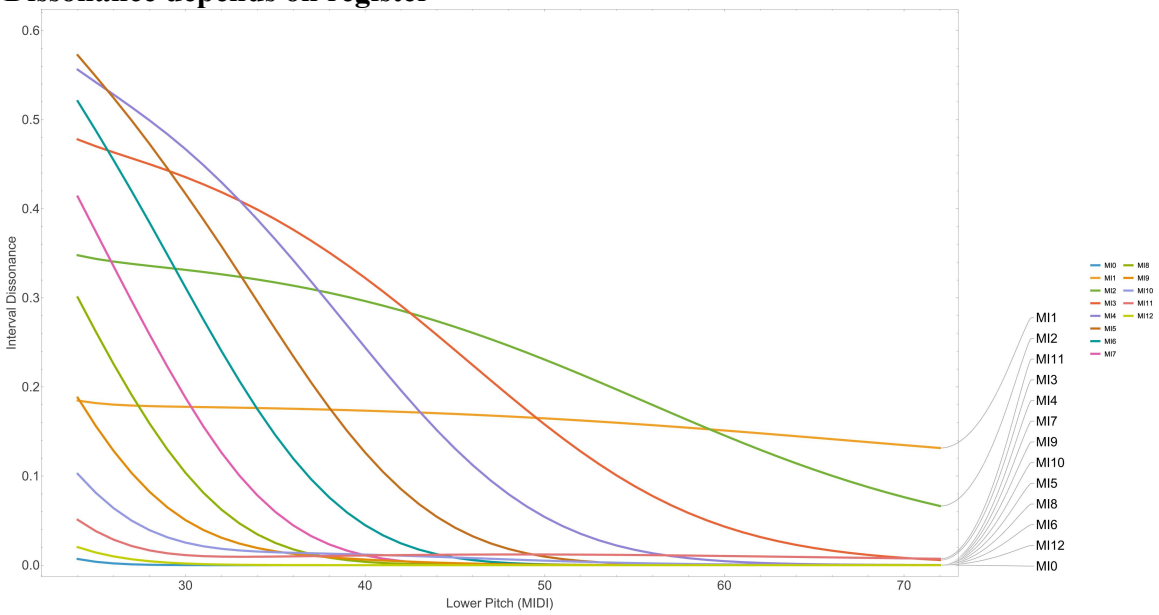
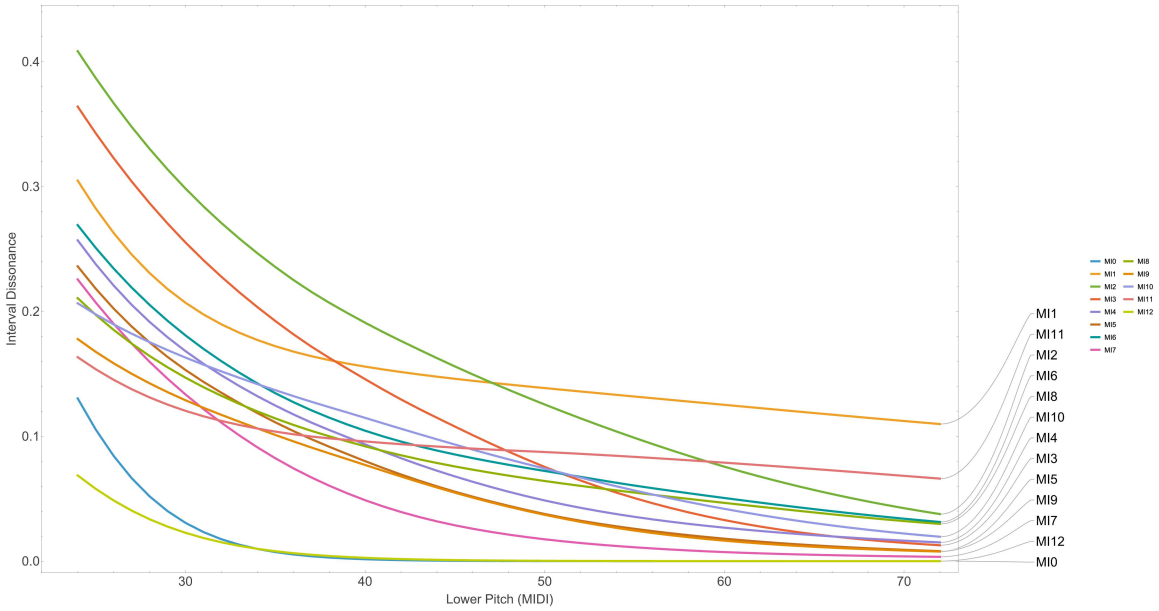


FIGURE 4. *η for dyads of MIDI interval 0–12 with lower pitch spanning MIDI 24–72, one curve per interval. Panel (a), pure tones.*



Panel (b), the harmonic timbre.

Dissonance rises steeply as an interval descends, but the more consequential result is that the ordering of intervals is not fixed. For pure tones at MIDI 24 the most dissonant interval in the set is the perfect fourth, at about 0.57, followed by the major third and the tritone. The semitone ranks eighth, near 0.19. By MIDI 72 the ordering has inverted. The semitone is the most dissonant and the fourth has fallen to ninth. No transposition-invariant measure can produce this.

The inversion follows from Equation 22, which weights each pair's contribution by its beat rate relative to register, $|b_{i_2}/h_2 - b_{i_1}/h_1|/\phi$. At MIDI 24 a semitone spans only 1.9 Hz. Its two tones overlap almost completely but beat about twice a second, so they contribute little fluctuation. A fourth in the same register spans 11 Hz. It is still well within a critical band but beats nearly six times faster. Higher up, the semitone's beat rate grows in proportion to frequency, while the wider intervals separate beyond the region where their tones overlap, and the ordering inverts. The model thus predicts what a listener can verify at a piano: a semitone in the lowest octave wavers rather than roughens. The received advice to space bass voices widely is recovered, and with it the less obvious point that the narrowest bass intervals escape roughness by beating too slowly to produce it.

Panel (b) shows the harmonic timbre compressing this behavior. The extremes are less severe. The maximum falls from about 0.57 to about 0.41, but the curves converge much less at the top of the range, since a complex tone carries partials that interfere regardless of register. The unison is instructive. It is essentially zero everywhere for pure tones, but it begins near 0.13 in the bass for the harmonic timbre. This occurs because a single complex tone's own partials interfere when the fundamental is low enough.

7.4 Fusion is not interval content

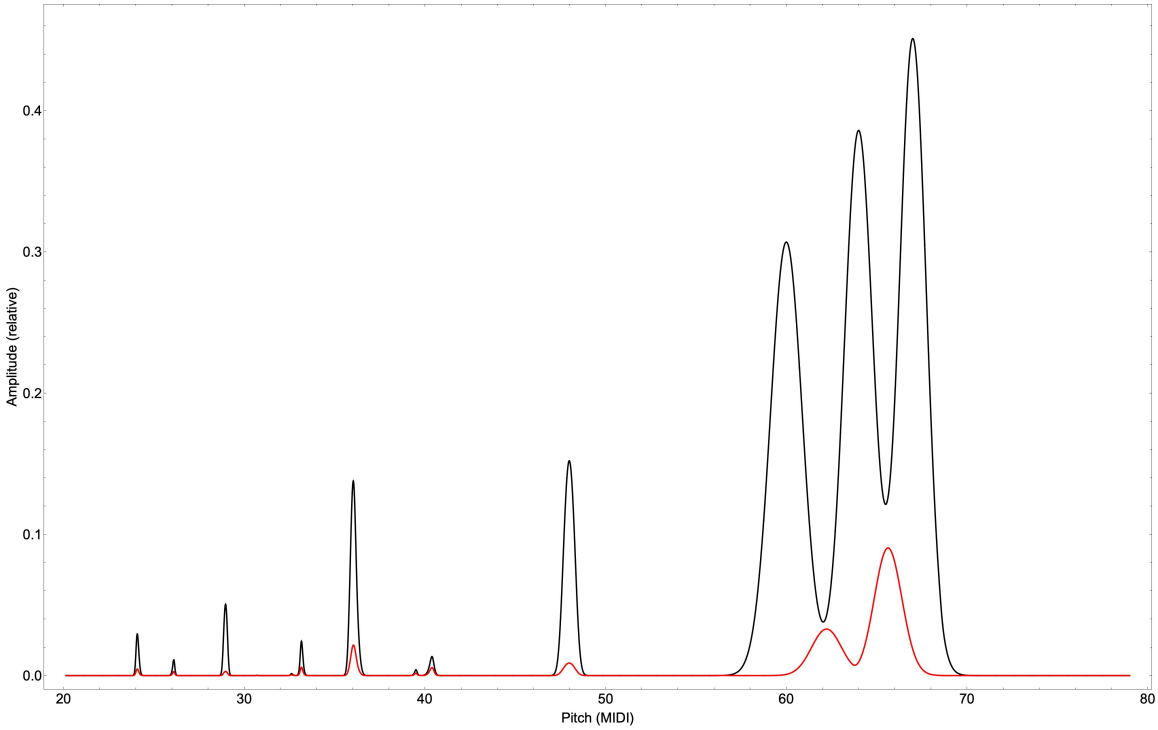
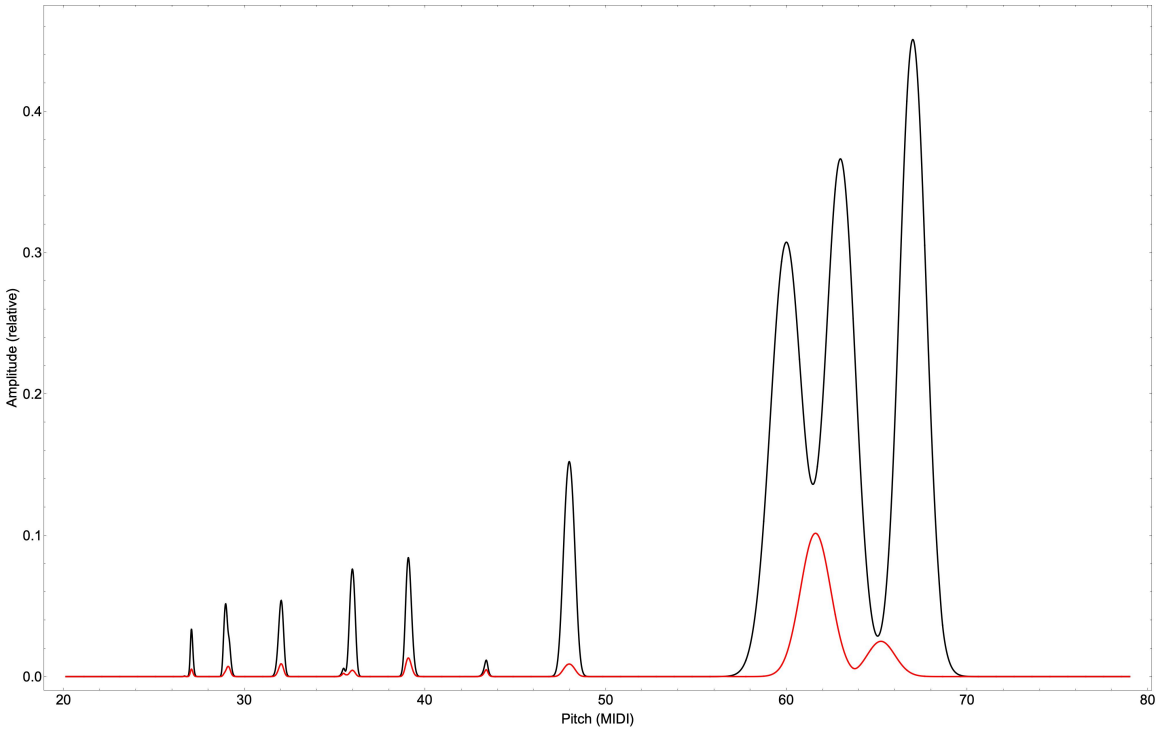
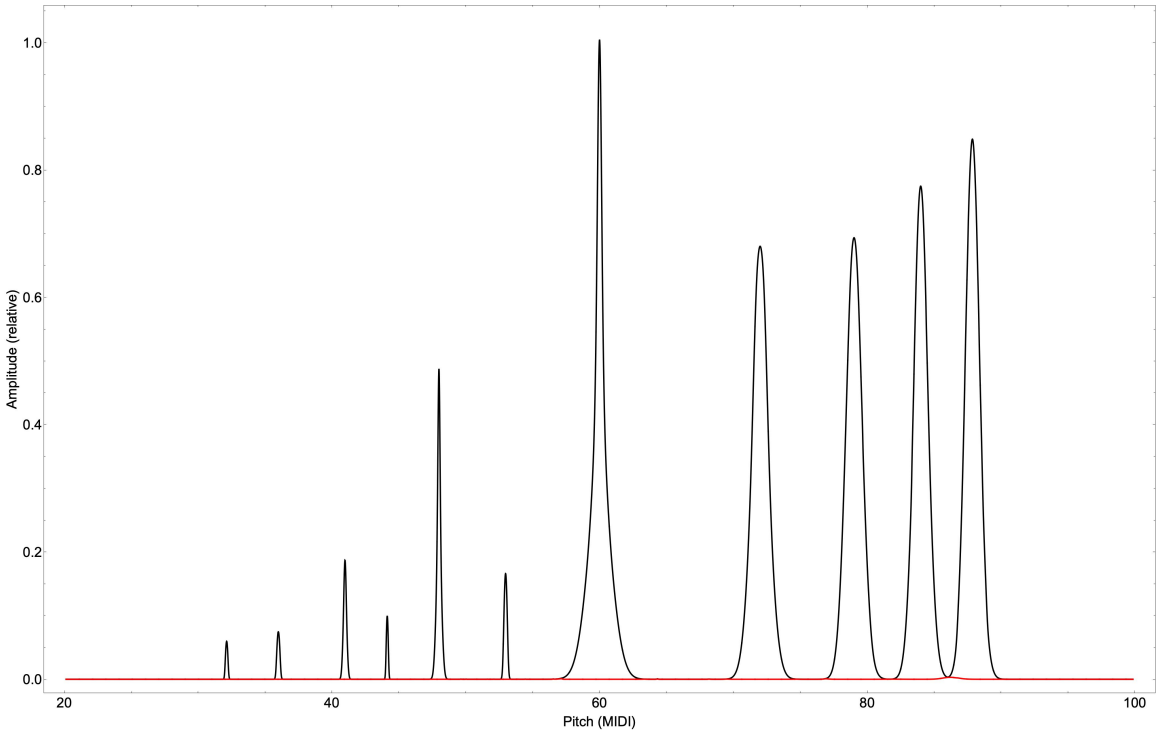


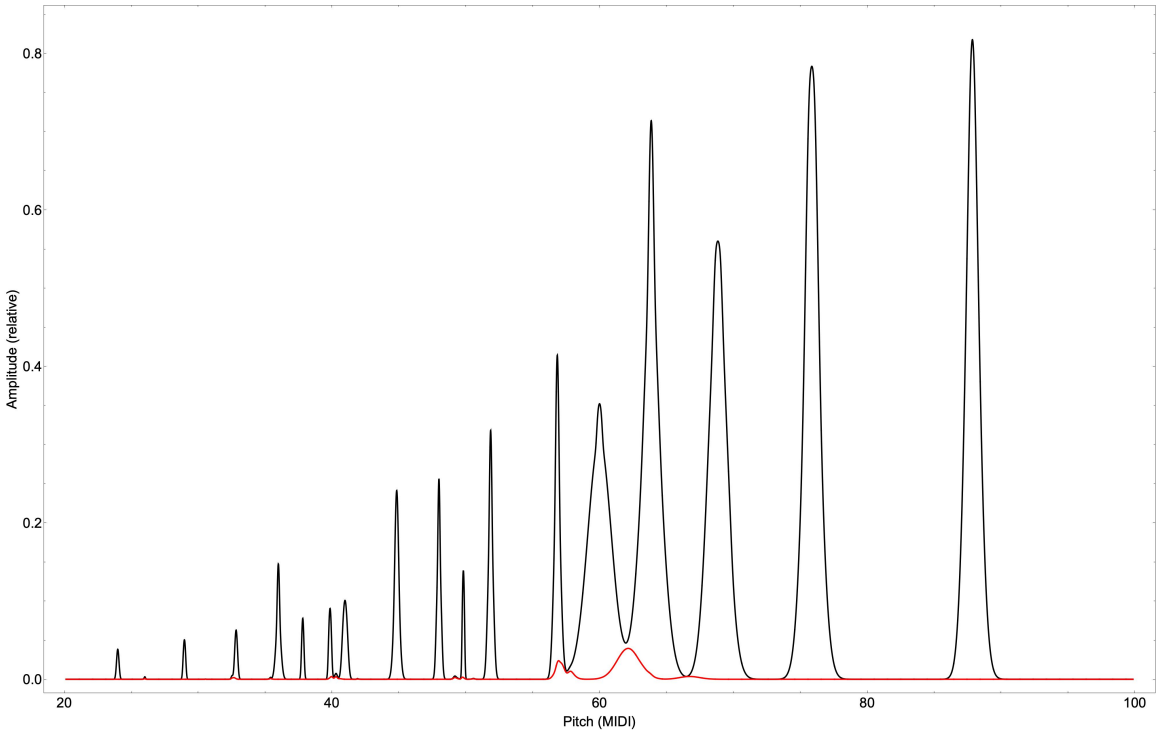
FIGURE 5. *ψ_1 (black) and ψ_2 (red) for four sonorities of equal-amplitude pure tones. Panel (a), major triad, MIDI {60, 64, 67}.*



Panel (b), minor triad, MIDI {60, 63, 67}.



Panel (c), overtone chord, MIDI {60, 72, 79.0196, 84, 87.8631}.



Panel (d), undertone chord, MIDI {60, 63.8631, 68.8435, 75.8631, 87.8631}.

Each pair shares its interval content exactly, and the members are distinguished by nothing a set-theoretic or interval-vector description would register. Their landscapes are not alike.

Both triads project an identical peak of 0.151 at MIDI 48. Both contain C4 and G4, whose 2:3 relation projects there and which the third does not disturb. The pair separates an octave lower. At MIDI 36 the major triad projects 0.138, its three tones falling at harmonics 4, 5, and 6 of that fundamental. The minor manages only a weaker peak of 0.077, since only two of its tones participate, and disperses the remainder across other weak peaks at MIDI 27, 29, 32, and 39, none exceeding 0.083.

The overtone and undertone chords make the point in the extreme. The overtone chord's tallest peak by a clear margin falls at MIDI 60. MIDI 60 is simultaneously its lowest tone and the fundamental of the series to which all five belong. Every other feature stands below it. The undertone chord, identical intervals inverted, has no such peak. Its three tallest features are

competing spectral pitches of comparable height near MIDI 64, 76, and 88, and its subharmonic energy scatters across a dozen moderate peaks.

The measured dissonances are:

Sonority	η
Major triad	0.0201
Minor triad	0.0244
Overtone chord	5.45×10^{-6}
Undertone chord	0.00123

The minor triad measures about 21% more dissonance than the major while differing in their principal fundamental by nearly a factor of two. Fusion and dissonance are therefore separable. A model returning a single scalar could not distinguish these chords as a listener does, and the major triad should be heard as more unified and more clearly rooted on grounds largely independent of its being slightly the smoother.

Separable is not independent, as the second pair shows. η for the overtone chord is 5.45×10^{-6}, some two hundred and twenty-five times below the undertone chord's, and its ψ_2 curve is indistinguishable from zero across the entire range. This is the beating principle of Section 3.3 at its limiting case. The five tones are exact harmonics of one fundamental, every subharmonic projection agrees, the terms of Equation 6 hold fixed relative phase, and nothing fluctuates at all. Perfect fusion and near-perfect smoothness are one fact about one landscape, reported by two readings of it.

7.5 Progression

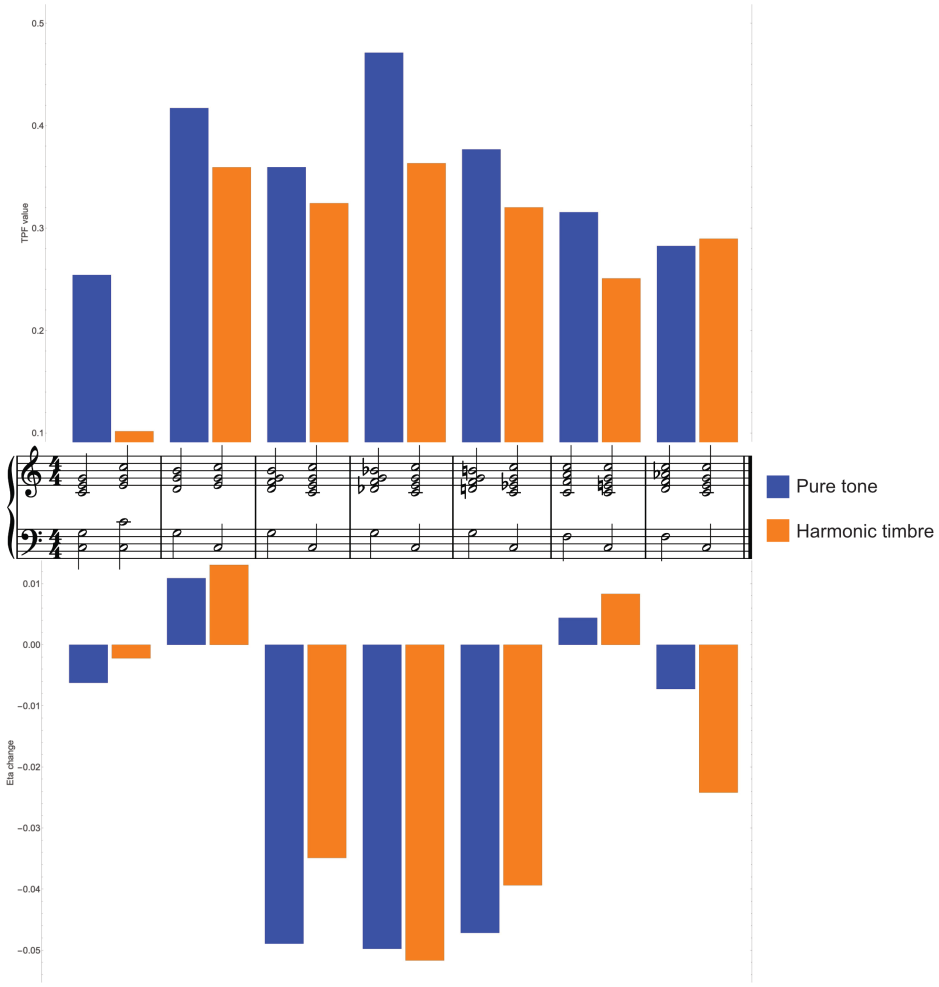


FIGURE 6. *Ω, the tonalness progression function (upper), and change in η (lower) for the progressions of Table 1, with pure tones (blue) and the harmonic timbre (orange). The realizations are notated between the two panels.*

Because the tonalness progression function operates on landscapes rather than notated pitches, it counts agreement among virtual pitches the two chords do not share as notes. Voicing is accordingly not incidental, and the progressions are realized in full rather than as parallel chord shapes. The set of Table 1 is chosen to put Ω and $\Delta\eta$ to work on two questions that functional theory answers by convention rather than by sound: why sevenths are added to dominants, and why chromatic harmonizations substitute so freely for diatonic ones.

Progression	First sonority	Second sonority	Ω	Δη
I–I′	48, 55, 60, 64, 67	48, 60, 64, 67, 72	0.254/0.102	−0.006/−0.002
V–I	55, 62, 67, 71	48, 64, 67, 72	0.417/0.359	+0.011/+0.013
V^7–I	55, 62, 65, 67, 71	48, 60, 64, 67, 72	0.359/0.324	−0.049/−0.035
V^{7}alt–I	55, 61, 65, 67, 70	48, 60, 64, 67, 72	0.471/0.363	−0.050/−0.052
V^7–i	55, 62, 65, 67, 71	48, 60, 63, 67, 72	0.377/0.320	−0.047/−0.039
IV–I	53, 60, 65, 69, 72	48, 60, 64, 67, 72	0.316/0.251	+0.004/+0.008
ii^{ø7}–I	53, 62, 65, 68, 72	48, 60, 64, 67, 72	0.283/0.290	−0.007/−0.024

Table 1. *Realizations in MIDI, low voice to high.* Ω *and* Δη *are given as pure tone/harmonic timbre.*

Why dominants take sevenths. Adding the seventh to V moves both measures at once, and in the same direction. Ω falls, from 0.417 to 0.359 under pure tones and from 0.359 to 0.324 under the harmonic timbre. Δη reverses sign, from +0.011 to −0.049 and from +0.013 to −0.035. The plain triadic cadence slightly raises dissonance, and the seventh chord's resolution sharply lowers it. The seventh is doing two things simultaneously. It raises the dominant's own fluctuation, so the resolution has further to fall, and it draws the dominant's harmonicity landscape nearer the tonic's. On this account the seventh is not a dissonance the cadence must overcome but the thing that gives the cadence its work to do.

Why chromatic substitutes pass for diatonic ones. The altered dominant of the fourth progression — G, B-flat, D-flat, F against the diatonic G, B, D, F — returns Δη of −0.050 and −0.052 against the diatonic chord's −0.049 and −0.035. The dissonance drop is the same or larger. What the alteration costs is a wider Ω under pure tones, 0.471 against 0.359, and that gap closes to 0.363 against 0.324 once the tones carry partials. The same pattern holds at the predominant. IV–I returns 0.316 and 0.251 with Δη positive, while the borrowed ii^{ø7}–I returns 0.283 and 0.290 with Δη negative. The chromatic predecessor is the nearer of the two under pure tones and the smoother under both. Substitutions that functional theory licenses by root relation turn out to be licensed psychoacoustically as well.

Two further readings are worth recording. V^7–i returns 0.377 and 0.320 with Δη of −0.047 and −0.039, very similar to V^7–I throughout. And I–I′ — the same C major triad revoiced — is the

smallest motion in the set at 0.254 and 0.102, which is the right answer, but it is not zero, where every pitch-class-based measure says it must be. Under pure tones a revoicing is about four fifths the size of a plagal cadence.

What none of this supplies is the asymmetry of tonal syntax — the sense in which V–I closes and I–V does not. That is substantially learned expectation, excluded here by construction. Ω and $\Delta\eta$ describe the substrate on which such syntax operates, not a substitute for it.

8. Testable Predictions

A model of this kind earns its formalism by making claims that could turn out to be false.

P1. Roughness persists for pure-tone intervals outside the critical band, and compounds with the number of tones. Plomp and Levelt's account predicts that roughness vanishes once two pure tones are separated by more than a critical band. This model predicts small but nonzero fluctuation at all separations. The pure-tone curve of Figure 3 retains excursions of order 0.01 out to three octaves — because any interval can be heard as a mistuning of a nearby simpler one — and predicts that these residues accumulate as tones are added. A discrimination experiment on widely spaced pure-tone sonorities of increasing cardinality would separate the accounts.

P2. Fusion differs between sonorities with identical interval content, in the direction of virtual-pitch concentration. The pairs of Figure 5 differ in no property available to interval-content description. The model predicts the major triad over the minor, and the overtone chord over the undertone, on ratings of fusion, unity, or clarity of root. For the triads, the model predicts that the fusion difference will exceed the dissonance difference, the two η values differing by only about 20%.

P3. A partial is absorbed into a harmonic series only when the fundamental beneath it is audibly present. The conservation law makes this quantitative and bounded. Holding an upper partial fixed and varying the level of the fundamental, the model predicts the partial is separately audible when the fundamental is quiet, becomes progressively less available as it rises, with the transition at ζ, and can never lose more than about forty percent of its salience however loud the fundamental grows. A model admitting spectral pitches unconditionally predicts no such dependence.

P4. Interval dissonance reverses rank with register, and the narrowest bass intervals escape roughness rather than maximizing it. Figure 4 shows the fourth as the most dissonant interval of its set at MIDI 24 and the semitone the most dissonant at MIDI 72. The second half is the sharper test. A semitone in the lowest octave should be judged less rough — though not necessarily less unpleasant — than a fourth in the same octave, its beat rate too slow to produce roughness.

P5. Adding a seventh to a dominant reduces its distance from the tonic and reverses the direction of the dissonance change. Figure 6 returns $\Omega = 0.417$ with $\Delta\eta = +0.011$ for V–I under pure tones, against 0.359 and −0.049 for V^7–I, with the same pattern under the harmonic timbre. Listeners should therefore rate V^7–I the smoother and more conclusive resolution. The sharper half of the claim is that the plain triadic cadence should be heard as very slightly roughening where the seventh chord's resolution is heard as smoothing. A model treating the seventh purely as an added dissonance to be discharged predicts no such reversal.

9. Limitations

The exclusions of Section 3.5 are real limitations, not deferrals of detail. Cultural familiarity is a substantial component of consonance judgment and this model contains none of it. Its predictions

should be read as claims about a substrate that learning then shapes, and behavioral tests designed accordingly. The noise floor ζ is chosen for numerical robustness rather than fitted. The demonstrations here are insensitive to its value, but any application in which partials approach it would require calibrating it against data. And the model has not been evaluated against behavioral data. The comparison with the sixteen models surveyed by Harrison and Pearce (2020) is the necessary next step.

10. Conclusion

Folding critical band, frequency response, and subharmonic projection into the Fourier integral produces a single function from which a surprising amount of harmonic experience can be read. Dissonance emerges as the fraction of a sonority's tonalness that will not hold still. Beating and roughness emerge as one variation observed on either side of the threshold of temporal resolution. The distinction matters musically, since the narrowest bass intervals beat too slowly to roughen. Tonal fusion emerges from a gate that withdraws from a partial exactly the salience it supplies to the fundamentals beneath it, conserving the total. The same gate, run in reverse, builds those fundamentals. Timbre becomes a determinant of where the consonances are, and the chord root becomes a landscape rather than a label.

What rigorous modeling of harmonic perception offers is not a theory of music, but consistency in the application of a stated set of assumptions. With it comes a principled basis for distinguishing what in harmonic experience belongs to musical culture from what belongs to the shared human auditory endowment. That line is where the interesting questions are, and it cannot be drawn without a framework such as this one.

References

Dai, Huanping. 2000. “On the Relative Influence of Individual Harmonics on Pitch Judgment.” *Journal of the Acoustical Society of America* 107 (2): 953–59.

Engell, Alva, Markus Junghöfer, Alwina Stein, Pia Lau, Robert Wunderlich, Andreas Wollbrink, and Christo Pantev. 2016. “Modulatory Effects of Attention on Lateral Inhibition in the Human Auditory Cortex.” *PLOS ONE* 11 (1): 1–15.

Fletcher, Harvey, and W. A. Munson. 1933. “Loudness, Its Definition, Measurement and Calculation.” *Journal of the Acoustical Society of America* 5: 82–108.

Glasberg, Brian R., and Brian C. J. Moore. 1990. “Derivation of Auditory Filter Shapes from Notched-Noise Data.” *Hearing Research* 47 (1–2): 103–38.

Harrison, Peter M. C., and Marcus T. Pearce. 2020. “Simultaneous Consonance in Music Perception and Composition.” *Psychological Review* 127 (2): 216–44.

Helmholtz, Hermann von. 1954 [1877]. *On the Sensations of Tone as a Physiological Basis for the Theory of Music*. Translated by Alexander J. Ellis. New York: Dover.

Hutchinson, William, and Leon Knopoff. 1978. “The Acoustic Component of Western Consonance.” *Interface* 7 (1): 1–29.

International Organization for Standardization. 2023. ISO 226:2023 *Acoustics — Normal Equal-Loudness-Level Contours*. Geneva: ISO.

Leman, Marc. 2000. “An Auditory Model of the Role of Short-Term Memory in Probe-Tone Ratings.” *Music Perception* 17 (4): 481–509.

Lerdahl, Fred. 2001. *Tonal Pitch Space*. New York: Oxford University Press.

McDermott, Josh H., Andriana J. Lehr, and Andrew J. Oxenham. 2010. “Individual Differences Reveal the Basis of Consonance.” *Current Biology* 20 (11): 1035–41.

Milne, Andrew J., William A. Sethares, Robin Laney, and David B. Sharp. 2011. “Modelling the Similarity of Pitch Collections with Expectation Tensors.” *Journal of Mathematics and Music* 5 (1): 1–20.

Moore, Brian C. J. 1998. *Cochlear Hearing Loss*. London: Whurr Publishers.

Moore, Brian C. J., Brian R. Glasberg, and Robert W. Peters. 1985. “Relative Dominance of Individual Partials in Determining the Pitch of Complex Tones.” *Journal of the Acoustical Society of America* 77 (5): 1853–60.

Parncutt, Richard. 1988. “Revision of Terhardt’s Psychoacoustical Model of the Root(s) of a Musical Chord.” *Music Perception* 6 (1): 65–93.

———. 1989. *Harmony: A Psychoacoustical Approach*. Berlin: Springer-Verlag.

Penagos, Hector, Jennifer R. Melcher, and Andrew J. Oxenham. 2004. “A Neural Representation of Pitch Salience in Nonprimary Human Auditory Cortex Revealed with Functional Magnetic Resonance Imaging.” *Journal of Neuroscience* 24 (30): 6810–15.

Peretz, Isabelle, Anne J. Blood, Virginia Penhune, and Robert Zatorre. 2001. “Cortical Deafness to Dissonance.” *Brain* 124 (5): 928–40.

Plomp, Reinier. 1967. “Pitch of Complex Tones.” *Journal of the Acoustical Society of America* 41 (6): 1526–33.

Plomp, Reinier, and Willem J. M. Levelt. 1965. “Tonal Consonance and Critical Bandwidth.” *Journal of the Acoustical Society of America* 38 (4): 548–60.

Rameau, Jean-Philippe. 1971 [1722]. *Treatise on Harmony Reduced to Its Natural Principles*. Translated by Philip Gossett. New York: Dover.

Ritsma, Roelof J. 1967. “Frequencies Dominant in the Perception of the Pitch of Complex Sounds.” *Journal of the Acoustical Society of America* 42 (1): 191–98.

Schellenberg, E. Glenn, and Sandra E. Trehub. 1994. “Frequency Ratios and the Discrimination of Pure Tone Sequences.” *Perception & Psychophysics* 56 (4): 472–78.

Sethares, William A. 1993. “Local Consonance and the Relationship between Timbre and Scale.” *Journal of the Acoustical Society of America* 94 (3): 1218–28.

———. 2005. *Tuning, Timbre, Spectrum, Scale*. 2nd ed. London: Springer.

Song, Peirun, Haoxuan Xu, Hangting Ye, et al. 2025. “Temporal Merging into Pitch with Click Train in the Macaque Auditory Cortex.” *National Science Review* 12 (5): 1–13.

Terhardt, Ernst. 1979. “Calculating Virtual Pitch.” *Hearing Research* 1 (2): 155–82.

Terhardt, Ernst, Gerhard Stoll, and Manfred Seewann. 1982. “Algorithm for Extraction of Pitch and Pitch Salience from Complex Tonal Signals.” *Journal of the Acoustical Society of America* 71 (3): 679–88.

Tymoczko, Dmitri. 2011. *A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice*. New York: Oxford University Press.

Vassilakis, Pantelis N. 2001. “Perceptual and Physical Properties of Amplitude Fluctuation and Their Musical Significance.” PhD diss., University of California, Los Angeles.

Appendix. Closed Form for the Progression Function

For two sonorities occurring at times t_1 and t_2, with n_1 and n_2 components respectively, where t_x represents either of them and (x, y) any pair of them:

$$\psi_1\left(f, t_x\right) = \sum_{h=1}^{\infty} \sum_{i=1}^{n_x} W_h\left(f, t_x\right) U_{h,i,x}(f) \tag{25}$$

$$W_h(f, t_x) = \begin{cases} \sqrt{1 - \sum_{k=2}^{\infty} W_k(f/k, t_x)^2 \frac{\sum_{i=1}^{n_x} U_{1,i,x}(f/k)}{\sum_{i=1}^{n_x} U_{1,i,x}(f/k) + \zeta}} & \text{if } h = 1 \\ \frac{1}{h}\left(1 - \frac{\sum_{i=1}^{n_x} U_{h,i,x}(f)}{\sum_{i=1}^{n_x} \sum_{g=1}^{\infty} U_{g,i,x}(f) + \delta}\right) & \text{otherwise} \end{cases} \tag{26}$$

$$U_{h,i,x}(f) = l_{i,x} e^{-\frac{\left(hf - b_{i,x}\right)^2}{2\sigma_h(f)^2}} \tag{27}$$

$$\mu\left(t_x\right) = \sqrt{\sum_{h_1=1}^{\infty} \sum_{h_2=1}^{\infty} \sum_{i_1=1}^{n_x} \sum_{i_2=1}^{n_x} W_{h_1}\left(\frac{b_{i_1,x}}{h_1}, t_x\right) W_{h_2}\left(\frac{b_{i_2,x}}{h_2}, t_x\right) \gamma_{h_1,h_2,i_1,i_2}(x, x)} \tag{28}$$

$$\Omega\left(t_1, t_2\right) = 1 - \frac{1}{\mu\left(t_1\right)\mu\left(t_2\right)} \sum_{h_1=1}^{\infty} \sum_{h_2=1}^{\infty} \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} W_{h_1}\left(\frac{b_{i_1,1}}{h_1}, t_1\right) W_{h_2}\left(\frac{b_{i_2,2}}{h_2}, t_2\right) \gamma_{h_1,h_2,i_1,i_2}(1,2) \tag{29}$$

$$\gamma_{h_1,h_2,i_1,i_2}(x, y) = \frac{l_{i_1,x} l_{i_2,y} \sigma_{h_1}\left(\frac{b_{i_1,x}}{h_1}\right) \sigma_{h_2}\left(\frac{b_{i_2,y}}{h_2}\right) \exp\left(-\frac{\left(h_2 b_{i_1,x} - h_1 b_{i_2,y}\right)^2}{2\left(h_2^2 \sigma_{h_1}\left(\frac{b_{i_1,x}}{h_1}\right)^2 + h_1^2 \sigma_{h_2}\left(\frac{b_{i_2,y}}{h_2}\right)^2\right)}\right)}{\sigma_1\left(\phi_{h_1,h_2,i_1,i_2}(x, y)\right) \sqrt{h_2^2 \sigma_{h_1}\left(\frac{b_{i_1,x}}{h_1}\right)^2 + h_1^2 \sigma_{h_2}\left(\frac{b_{i_2,y}}{h_2}\right)^2}} \tag{30}$$

$$\phi_{h_1,h_2,i_1,i_2}(x,y) = \frac{h_2 b_{i_2,y} + h_1 b_{i_1,x}}{h_2^2 + h_1^2} \tag{31}$$

Equations 23 and 30 both derive from the same Gaussian overlap integral, with its constant factor $\sqrt{2\pi}$ omitted to simplify the expressions. Each critical bandwidth is frozen at the centre of the contribution whose width it sets. The omitted factor cancels in both η (Equation 22) and Ω (Equation 29), so μ in Equation 28 equals that of Equation 12 only up to the factor $(2\pi)^{1/4}$.