Hidden Voices: Generating Counterpoint from Acoustic First Principles

Isha Yerramilli-Rao

Advised by Professor Bernard Chazelle

April 2026

Abstract

Johann Sebastian Bach's six Suites for Unaccompanied Cello (BWV 1007–1012) were written for a monophonic instrument, yet they imply a rich polyphonic texture. While this is a well known fact analyzed by Baroque musicians, the polyphonic melodies have never been formally reconstructed. This paper develops a mathematically grounded framework for generating this implied second voice. We derive a dissonance metric from psychoacoustic first principles using the Plomp-Levelt roughness model, validate it against Bach's four-voice chorale transitions via maximum likelihood estimation under a Boltzmann transition model, and use the resulting cost function to drive a phrase-conditioned Boltzmann sampler over harmonic chord tones. Phrase structure is captured by a CategoricalHMM trained on 433 Bach chorales via Baum-Welch EM, whose hidden states modulate the generation temperature per position based on emission entropy. The approach requires no training data specific to the Cello Suites and generalizes across movements. The algorithm is demonstrated on the Sarabande, Minuet and Prelude from Suite No. 1, with a discussion of successes and limitations of the proposed mathematical framework.

Sheet music outputs and mp3 files for each movement can be found at: https://github.com/ishayrao/hidden-voices.

1. Introduction

Johann Sebastian Bach's six Suites for Unaccompanied Violoncello (BWV 1007–1012), composed circa 1720, are some of Bach's most well-known works. Written for a single-line instrument, they consistently create the impression of multiple simultaneous voices. Pablo Casals, whose 1936 recordings introduced the Suites to modern audiences, remarked that "it is fantastic to think that with one note after the other there can be such richness." Bach was the master of counterpoint, which is the art of combining independent melodic lines, and even in ostensibly monophonic writing, his compositional logic operates polyphonically.

The overlap of music and mathematics is well-explored, and Bach's fugal specialty lends particularly well to explorations in the overlapping domain. Tymoczko [14] formalized the geometric structure of voice leading, showing that smooth motion between chords can be modeled as paths through a high-dimensional pitch space. Davis [3] analyzed implied

polyphony in Bach's unaccompanied string works using auditory stream segregation theory, demonstrating that Bach's melodic construction systematically creates the perception of multiple simultaneous voices through register alternation and pitch proximity. Winold [15] provided detailed harmonic and voice-leading analyses of all six Cello Suites, identifying the implied inner voices measure by measure. However, a generative algorithm that reconstructs these voices from mathematical principles alone has not been suggested.

Data-driven approaches to music generation, including DeepBach [6] and BachBot [8], achieve impressive results on the structure of four-voice chorales as they can be trained on thousands of examples of the desired texture. For the implied voices in the Cello Suites, no such training data exists. The hidden voices are only implied, which requires an algorithm that derives its generative behavior from principles independent of the specific texture being generated.

This paper proposes an experimental model derived from acoustic physics and an analysis of consonance and dissonance throughout the piece. The roughness between two simultaneous tones is a function of their frequencies and can be derived from the psychoacoustic properties of the auditory system [10]. Roughness is a quantitative measure of the interference induced by beating between nearby partials. We use this derivation to construct a cost function that assigns a scalar value to any voice-leading transition, without reference to any corpus of hidden-voice examples. We then validate this cost function against Bach's empirical voice-leading choices in the 433 chorales, using maximum likelihood estimation under a Boltzmann transition model. Finally, we use the validated cost function to drive a generative algorithm for the implied second voice of the Cello Suite movements.

2. Acoustic Dissonance Model

2.1 Psychoacoustic Foundations

The perceptual roughness between two simultaneously sounding pure tones was characterized by Plomp and Levelt [10], who measured it as a smooth function of the frequency difference relative to the critical bandwidth. We use the standard parametrization due to Sethares [13], which fits the Plomp-Levelt empirical data to a difference of exponentials:

$$r(f_1, f_2) = \ell_{12} \left[e^{-b_1 s(f_2 - f_1)} - e^{-b_2 s(f_2 - f_1)} \right], \tag{1}$$

where $b_1 = 3.5$, $b_2 = 5.75$, $s = d^*/(s_1 f_{\min} + s_2)$ with $d^* = 0.24$, $s_1 = 0.0207$, $s_2 = 18.96$, and $\ell_{12} = \min(\ell_1, \ell_2)$ is the loudness of the softer partial. The two exponential form in equation (1) is a phenomenological fit to the empirical data of Plomp and Levelt [10], who measured perceived roughness between pairs of sinusoidal tones across a range of frequency separations. It captures the characteristic shape of the measured response where the second exponential dominates near coincidence ($f_1 \approx f_2$), producing rapid onset of roughness as the partials separate into the beating regime, while the first exponential governs the decay once separation exceeds one critical bandwidth and the partials are resolved by distinct auditory filters. The coefficients b_1, b_2, s_1, s_2 are fit parameters; the underlying psychoacoustic mechanism is the incomplete resolution of nearby partials by the cochlea's bank of overlapping critical-band filters, which produces amplitude modulation perceived as beating or buzzing. Figure 1 shows the resulting curve as a function of frequency difference.

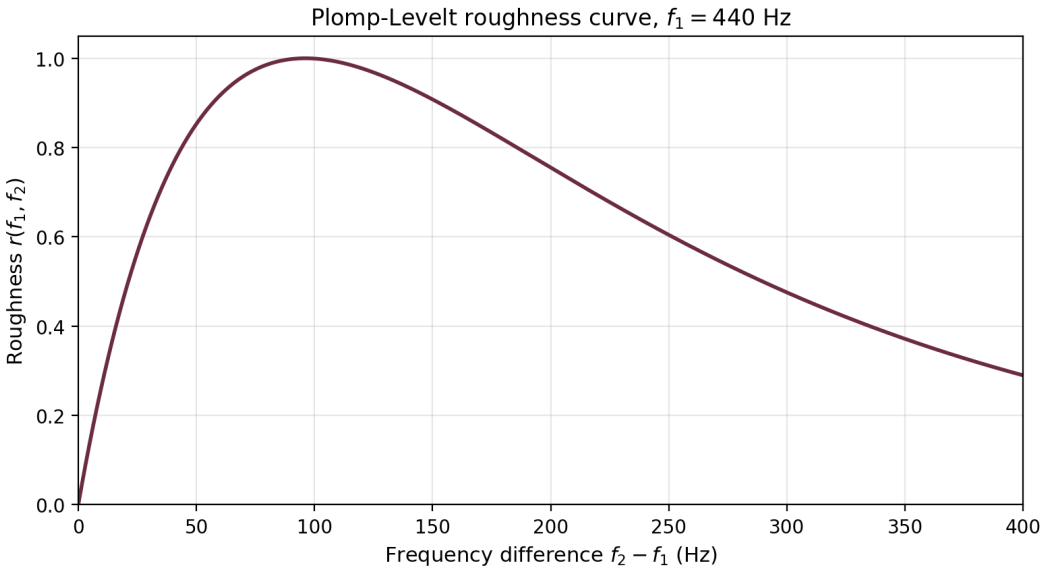

Figure 1: Plomp-Levelt roughness $r(f_1, f_2)$ as a function of frequency difference $\Delta f = f_2 - f_1$, with $f_1 = 440$ Hz fixed. Roughness rises sharply from zero at coincidence, peaks near $\Delta f \approx 96$ Hz (approximately one critical bandwidth at this register), and decays smoothly as the partials resolve into distinct auditory filters.

2.2 Extension to Complex Tones

Musical instruments produce complex tones whose spectra consist of a fundamental frequency and a series of harmonics. For a note with fundamental f, we model the spectrum as the first N harmonics

$$\mathcal{H}(f) \;=\; \{f,\, 2f,\, 3f,\, \ldots,\, Nf\}, \tag{2}$$

For the cello, we use $N = 6$, which captures the dominant partials of the instrument's timbre.

We assign equal amplitude to all N harmonics as a simplification. In general, real instrumental spectra exhibit amplitude rolloff with harmonic number, and partials that are inaudible should contribute less to perceived roughness. However, bowed cello tones in the register used in this work exhibit relatively flat spectra across the first six partials, with typical amplitude variation of order 10 dB under sustained bowing [4], so equal weighting is a reasonable first-order approximation for this instrument. More fundamentally, the absolute scale of $\hat{D}$ does not affect the downstream generative algorithm. Only the relative ordering of candidate notes at each position matters, and the weight λ_2 on the dissonance term is calibrated empirically against Bach's voice-leading choices in Section 4, which absorbs residual error from the equal-amplitude simplification. The limitations of this choice, visible in the upper register of Table 1, are discussed in Section 7.

Given a chord $\mathbf{f} = (f_1, f_2, \ldots, f_M)$ of M simultaneous notes, the total roughness is computed by summing the pairwise roughness over all pairs of harmonics across all pairs of notes:

$$R(\mathbf{f}) \;=\; \sum_{\substack{i<j \\ i,j\in\{1,\ldots,M\}}} \sum_{p=1}^{N} \sum_{q=1}^{N} r(p\, f_i,\, q\, f_j)\,. \tag{3}$$

To facilitate comparison across chords of different sizes, we normalize by the number of note pairs and partial pairs:

$$\hat{D}(\mathbf{f}) \;=\; \frac{R(\mathbf{f})}{\binom{M}{2} \cdot N^2}. \tag{4}$$

This normalization ensures that $\hat{D}$ lies in $[0, 1]$ and is independent of chord size.

2.3 Validation via Recovery of the Consonance Hierarchy

A key test of the dissonance model is whether it recovers the standard ranking of musical intervals without any reference to music theory or training data. For each of the twelve intervals within an octave, we compute $\hat{D}$ between the corresponding pitch pair using equation (4) with $M = 2$.

Table 1: Computed dissonance scores $\hat{D}$ for diatonic intervals, using the Plomp-Levelt roughness model with $N = 6$ equal-amplitude harmonics, root A_4 (440 Hz). Sorted by ascending $\hat{D}$.

Interval	Frequency ratio	$\hat{D}$
Octave	2:1	0.182
Major 7th	15:8	0.256
Unison	1:1	0.276
Minor 7th	16:9	0.292
Major 6th	5:3	0.296
Perfect 5th	3:2	0.302
Minor 6th	8:5	0.315
Tritone	45:32	0.354
Perfect 4th	4:3	0.360
Major 3rd	5:4	0.373
Minor 2nd	16:15	0.383
Minor 3rd	6:5	0.392
Major 2nd	9:8	0.415

The model recovers the broad structure of the consonance hierarchy from acoustic physics alone: narrow intervals (seconds, thirds) cluster at the dissonant end while wide intervals (octave, sevenths, sixths, fifth) cluster at the consonant end, consistent with Plomp-Levelt's prediction that roughness is maximized when partials fall within roughly one critical bandwidth. Two deviations from the textbook ordering are worth noting. The unison computes to a nonzero value because the model sums roughness across all partial pairs between two spectra treated as distinct sources, and the major and minor sevenths rank as less rough than the thirds because the harmonic coincidences that drive the seventh's instability emerge only at $N \geq 15$ [7, 13]. More fundamentally, $\hat{D}$ measures *sensory* roughness, which is distinct from musical dissonance, a contextual phenomenon shaped by tendency-to-resolve and tonal function. $\hat{D}$ therefore serves as a physically grounded but acoustically incomplete proxy that captures the broad hierarchy while missing fine-grained distinctions, motivating the empirical calibration via maximum likelihood estimation in Section 4.

3. Voice-Leading Cost Function

3.1 Two-Voice Transition Cost

Let the anchor voice denote the originally composed cello melody, whose pitch at position t is a_t. The hidden voice occupies pitch n_t at position t. The cost of transitioning the hidden voice from n_{t-1} to a candidate note n at position t is

$$C(n_{t-1} \to n) \;=\; \lambda_1 \, |n_{t-1} - n| + \lambda_2 \, \hat{D}(a_t,\, n) + \lambda_3 \, K(n), \tag{5}$$

where $|\cdot|$ denotes semitone distance, $\hat{D}$ is the normalized dissonance of equation (4) evaluated against the anchor note, and $K(n) \in \{0, 1\}$ is a binary key penalty.

The three terms penalize, respectively:

1. **Large melodic leaps.** The λ_1 term implements the principle of stepwise motion from species counterpoint such that small intervals are preferred.

2. **Dissonance with the anchor.** The λ_2 term penalizes harmonically rough combinations, selecting notes via acoustic model of Section 2.

3. **Out-of-key notes.** The key penalty is

$$K(n) \;=\; \begin{cases} 0 & \text{if } n \bmod 12 \in \mathcal{S}_k, \\ 1 & \text{otherwise,} \end{cases}$$

   where $\mathcal{S}_k$ is the set of seven pitch classes of the active key k. Key changes within a movement (e.g. G major to G minor in the Minuet) are detected automatically and assigned per harmonic position.

3.2 Counterpoint Penalties

In addition to the smooth cost (5), we impose hard penalties of weight $P = 3.0$ on voice-leading violations drawn from species counterpoint rules, outlined by Fux [5]. First, if the interval between anchor and hidden voice is a perfect fifth (7 semitones) or octave (12 semitones) at two consecutive positions, and both voices move in the same direction, a penalty P is added. Second, if the hidden voice rises above the anchor voice, a penalty P is added.

These are implemented as additive terms to C rather than hard constraints, allowing the optimizer to trade them off against other cost components.

3.3 Higher-Order Cost

The first-order cost (5) conditions only on the immediately preceding note. To capture longer-range melodic coherence, we define the order-k cost as

$$C_k(n_{t-k:t-1} \to n) \;=\; C(n_{t-1} \to n) + \gamma \, \mathbf{1}[\mathrm{dir}(n_{t-k:t-1}) \neq \mathrm{dir}(n_{t-1} \to n)], \tag{6}$$

where $\mathrm{dir}(\cdot)$ denotes the sign of the net semitone displacement over the preceding k steps and $\gamma = 0.1$ is a reversal penalty. This implements a weak tendency toward directional consistency, such that a voice moving steadily upward is penalized for an abrupt reversal, while a voice already reversing direction is not additionally penalized. In generation we use $k = 3$.

Formally, for a window of recent hidden-voice notes $n_{t-k:t-1}$ we define

$$\mathrm{dir}(n_{t-k:t-1}) = \mathrm{sign}(n_{t-1} - n_{t-k}) \in \{-1, 0, +1\}, \tag{7}$$

the sign of the net semitone displacement across the window. The direction of a candidate move is $\mathrm{dir}(n_{t-1} \to n) = \mathrm{sign}(n - n_{t-1})$. The reversal penalty in equation (6) is applied when both directions are nonzero and opposite.

4. Markov Chain Validation and Parameter Tuning

4.1 Boltzmann Transition Model

To tune λ_1 and λ_2 empirically, we model Bach's voice-leading choices as samples from a Boltzmann distribution over candidate transitions. Given the current chord state s_i, the probability of moving to state s_j is

$$Q(s_j \mid s_i) \;\propto\; \exp\!\left(-\frac{C(s_i \to s_j)}{\tau}\right), \tag{8}$$

where $\tau > 0$ is a temperature parameter controlling the sharpness of the distribution. Under this model, lower cost transitions are exponentially more probable, which formalizes the intuition that Bach consistently prefers smooth, consonant voice leading.

4.2 Corpus and Transition Extraction

We extract four-voice SATB chord sequences from all 433 Bach chorales available via the `music21` corpus [2], yielding 30,323 chord states across a vocabulary of 647 unique chords. Consecutive pairs (s_i, s_j) constitute the empirical transitions used for estimation, giving 29,895 transitions after filtering repeated chords. The empirical transition matrix P yields a stationary distribution π and entropy rate

$$H \;=\; -\sum_i \pi_i \sum_j P_{ij} \log P_{ij} \;=\; 3.55 \text{ bits/chord}. \tag{9}$$

This quantifies the average uncertainty in Bach's next chord given the current one. A fully deterministic composer would yield $H = 0$, and a composer choosing uniformly among M candidates would yield $H = \log_2 M$. For our vocabulary of $|\mathcal{S}| = 647$ chord states, the maximum entropy would be $\log_2 647 \approx 9.34$ bits. The observed $H = 3.55$ bits corresponds to an effective branching factor of $2^{3.55} \approx 11.7$, meaning that at any given chord Bach's next choice behaves as if drawn uniformly from roughly twelve equally probable options out of 647 possibilities. This is a strong statistical regularity relative to the full vocabulary and is well suited to fitting a parametric model.

4.3 Maximum Likelihood Estimation

The negative log-likelihood of the observed transitions under the Boltzmann model is

$$\mathcal{L}(\lambda_1, \lambda_2) \;=\; -\sum_{(s_i, s_j)} \log Q(s_j \mid s_i;\, \lambda_1, \lambda_2). \tag{10}$$

We minimize Equation (10) using the Nelder-Mead simplex method [9], which is appropriate given the non-smooth cost structure as it is free of derivatives.

The MLE estimates are $\hat{\lambda}_1 = 0.0225$ and $\hat{\lambda}_2 = 0.1855$, with a minimum NLL of 4701.82. The fitted model assigns rank 1 (highest probability) to Bach's actual next chord in 15.1% of transitions and achieves a median rank of 5 out of roughly 50 candidates, as seen in Figure 2.

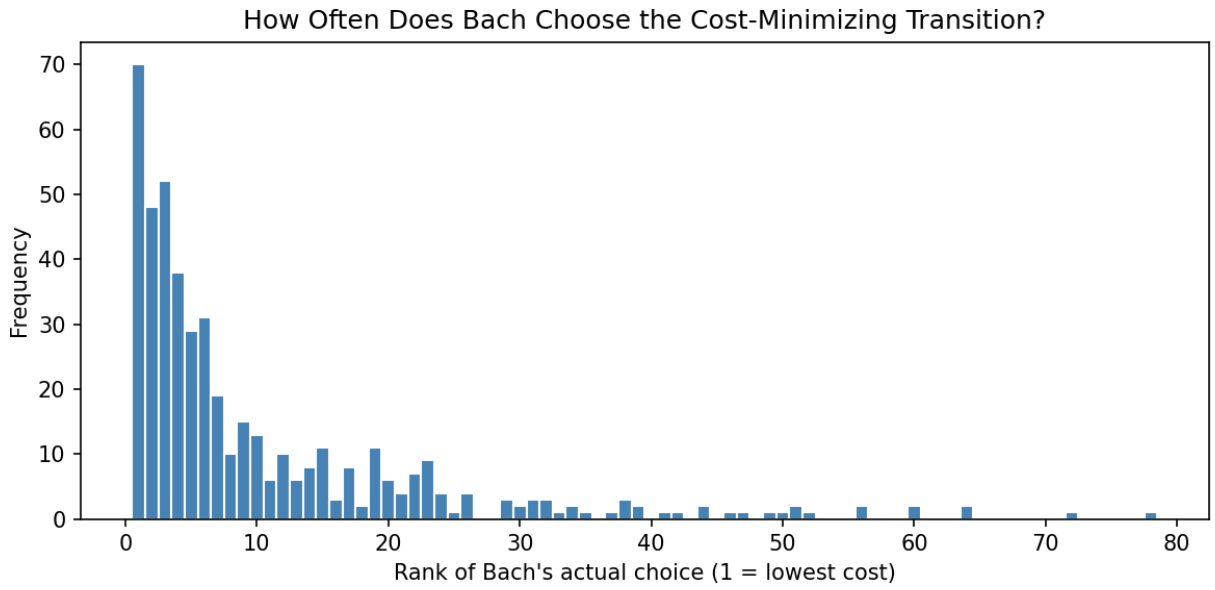

Figure 2: Distribution of Bach's actual next-chord rank under the fitted Boltzmann model across the 29,895 chorale transitions. The residual uncertainty motivates the phrase-level HMM of Section 5.

5. Phrase Level Hidden Markov Model (HMM)

Pure note-by-note optimization under Equation (5) treats every position independently and cannot capture the overarching structure of a movement over extended phrases. We model this variation using a hidden Markov model trained on the chorale corpus.

5.1 Model Definition

The HMM has $K = 8$ hidden states, selected by minimizing the Bayesian information criterion [12] over $K \in \{2, \ldots, 9\}$ (optimal BIC = 163,351.83 at $K = 8$). This balances phrase type resolution against overfitting on the chorale corpus. Each state emits symbols drawn from a vocabulary of 36 elements encoding pitch class and duration category jointly:

$$\text{symbol}(p, d) \;=\; 3p + d_{\text{cat}}, \tag{11}$$

where $p \in \{0, \ldots, 11\}$ is the pitch class and $d_{\text{cat}} \in \{0, 1, 2\}$ encodes duration as short ($\leq$0.5 beats), medium (0.5–1.5 beats), or long ($>$ 1.5 beats). Transition and emission parameters are estimated from the chorales via the Baum-Welch EM algorithm [1, 11].

5.2 Temperature Modulation

The HMM does not produce notes. Instead, it tunes the Boltzmann temperature τ in Equation (8) as a function of the decoded phrase state. For each hidden state i, let $\hat{b}_i(v) = P(o_t = v \mid z_t = i)$ denote the estimated emission probability of symbol v, as learned by Baum-Welch [1]. For each state i, we compute the emission entropy $H_i = -\sum_v \hat{b}_i(v) \log \hat{b}_i(v)$ and the weight on tonic and dominant pitch classes w_i^{td}. High entropy states (exploratory passages) receive a high τ, flattening the Boltzmann distribution and allowing freer voice leading. On the other hand, low entropy states with strong tonic or dominant weighting

(cadential passages) receive a low τ, concentrating probability mass on the most consonant transitions. This gives the generative algorithm phrase-level awareness without requiring any Cello Suite training data.

The state temperature is then defined as

$$\tau_i \;=\; \tau_{\min} + \alpha \cdot \frac{H_i}{\log 36} - \beta \cdot w_i^{\text{td}}, \tag{12}$$

where w_i^{td} is the total emission probability mass on tonic and dominant pitch classes in state i. The constants $\tau_{\min} = 0.15$, $\alpha = 0.5$, and $\beta = 0.2$ are set empirically. $\tau_{\min}$ ensures a minimum level of stochasticity even at cadential positions, α scales the entropy contribution so that τ_i remains in $[0.15, 0.65]$ across all states, consistent with the effective base temperature of $\tau = 0.45$ identified during development. β controls how strongly tonal concentration pulls the temperature downward. The specific magnitudes are design parameters tuned by qualitative evaluation of the generated output.

6. Generative Algorithm

Both the cost function weights (Section 4) and the phrase-state HMM (Section 5) are estimated from the chorale corpus, but they serve distinct roles. The former calibrates the physics-based cost function against Bach's empirical preferences, while the latter captures phrase-level structure to modulate generation temperature. Neither model has access to any Cello Suite data.

6.1 Candidate Generation

At each harmonic position t with anchor pitch a_t, the set of candidate notes for the hidden voice is

$$\mathcal{C}_t \;=\; \big\{ n \in [45, 71] \;\big|\; \hat{D}(a_t, n) \le \hat{D}(a_t, n_{\text{maj6}}) \big\}, \tag{13}$$

where the range $[45, 71]$ (A2 to B4) is the practical register for a second cello voice, and n_{maj6} is the major sixth above a_t. The consonance gate rejects candidates rougher than a major sixth, retaining thirds, fourths, fifths, and octaves as the primary harmonic vocabulary.

6.2 Forward Boltzmann Sampling

Given the candidate sets $\{C_t\}_{t=1}^{T}$, the cost function C_k of equation (6), and the position-specific temperatures τ_t from the HMM phrase state, we generate the hidden voice sequentially. At each position $t \ge 1$, given the rolling history $n_{t-k:t-1}$ of the previously generated notes, we sample

$$n_t \;\sim\; \frac{1}{Z_t} \exp\!\left(-\frac{C_k(n_{t-k:t-1} \to n_t)}{\tau_t} \right), \qquad n_t \in C_t, \tag{14}$$

where $Z_t = \sum_{n \in C_t} \exp(-C_k(n_{t-k:t-1} \to n)/\tau_t)$ normalizes the distribution. The first note n_1 is sampled from the same distribution conditioned on the empty history, with cost reduced to $\hat{D}(a_1, n)$.

This sampling formulation is the natural instantiation of the Boltzmann transition model used to calibrate λ_1 and λ_2 in Section 4: the generative distribution at each position has the same parametric form as the empirical distribution against which the cost function was fit. The phrase state at each position is decoded from the anchor melody via the Viterbi algorithm [11] as described in Section 5, and supplies τ_t. The temperature controls

the sharpness of the distribution such that at low τ_t (cadential phrase states) the sampler concentrates on the most consonant candidates and at high τ_t (exploratory states) it admits a broader range of voice-leading choices.

6.3 Key Change Handling

Multi-section movements such as the Minuet I/II involve modulation between keys (G major and G minor). Key changes are detected automatically by computing the best-fit key signature at each harmonic position using pitch-class profile matching, and the key penalty $K(n)$ in Equation (5) is updated accordingly.

6.4 Output

The hidden voice and anchor melody are exported to MusicXML using `music21`, with the anchor on the treble staff and the hidden voice in bass clef. During this project, I used MuseScore 4[1] to interpret and analyze the outputs of the algorithm.

7. Results and Discussion

7.1 Sarabande (BWV 1007, mvt. 4)

The algorithm performs best on the Sarabande, which has rich harmonic rhythm and clear phrase structure. The generated voice moves primarily in thirds and sixths with the anchor, with contrary motion at phrase boundaries. Cadential positions are correctly identified by the HMM and receive tighter, more consonant voice leading. A representative passage is shown in Figure 3.

7.2 Minuet I/II (BWV 1007, mvt. 5/6)

The Minuet results are comparable in quality. The key change from G major (Minuet I) to G minor (Minuet II) is handled automatically; the key penalty correctly excludes the sharpened sixth and seventh of the major scale after the modulation, and the generated voice tracks the modal shift without manual intervention. A representative passage is shown in Figure 4.

7.3 Prelude (BWV 1007, mvt. 1)

The Prelude is the most challenging movement for the algorithm because its arpeggiated texture concentrates the harmonic information into single triads at each detected position, leaving little for the dissonance term to discriminate over. The generated voice is still musically coherent as the temperature schedule and the rolling history window contribute structure, and the resulting line tracks the harmonic skeleton through the arpeggios with reasonable contour. Cadential moments identified by the HMM produce the expected tightening toward consonant intervals, though the effect is weaker than in the Sarabande or Minuet because the candidate set provides less to choose between.

7.4 Limitations

Two limitations of the framework are worth noting explicitly. First, the forward sampling procedure of equation (14) makes each decision conditioned only on the rolling history window of length k. This produces locally smooth voice leading but cannot enforce global properties such as long-range contour or independence of the hidden voice from the anchor over multi-measure spans. A second voice that is locally consonant at every position may

[1] https://musescore.org

still track the anchor too closely to constitute a fully independent melodic line. Addressing this would require a global objective evaluated over full sequences (for example via simulated annealing or MCMC) or a generative model with longer memory.

Second, the candidate-filtering stage uses only local consonance against the harmonic anchor. When the harmonic vocabulary at a position is itself impoverished (as in the arpeggiated textures of the Prelude) the consonance gate retains nearly every pitch in the register and contributes little to discrimination between candidates. Generation under these conditions falls back almost entirely on the melodic term. This is a limitation of the filtering stage rather than of the cost function or sampler, and it would persist under any local generation procedure. A richer candidate filter that uses structural or contextual information beyond local consonance is an open direction for future work.

8. Conclusion

We have presented a generative model for the implied second voice in Bach's Cello Suites that derives its behavior entirely from acoustic physics and empirical voice-leading statistics, without any training data specific to the target texture. The Plomp-Levelt roughness model provides a physically grounded dissonance measure that recovers the standard consonance hierarchy from first principles. Maximum likelihood estimation on 433 Bach chorales reveals that vertical consonance, not stepwise motion, is the dominant driver of Bach's chord-to-chord transitions in the four-voice chorale texture, exemplified by the fact that the dissonance term contributes roughly 1.8× more to the fitted cost than the voice-motion term at each observed transition. A phrase-level HMM modulates generation temperature to distinguish cadential from exploratory passages, giving the algorithm structural awareness at no additional supervision cost.

The model succeeds on harmonically rich movements and fails predictably on arpeggiated ones, which provides a clean characterization of its scope. The primary open problem is the enforcement of global melodic independence, which local cost minimization cannot achieve.

9. Appendix

Excerpts of music generated through the described algorithm.

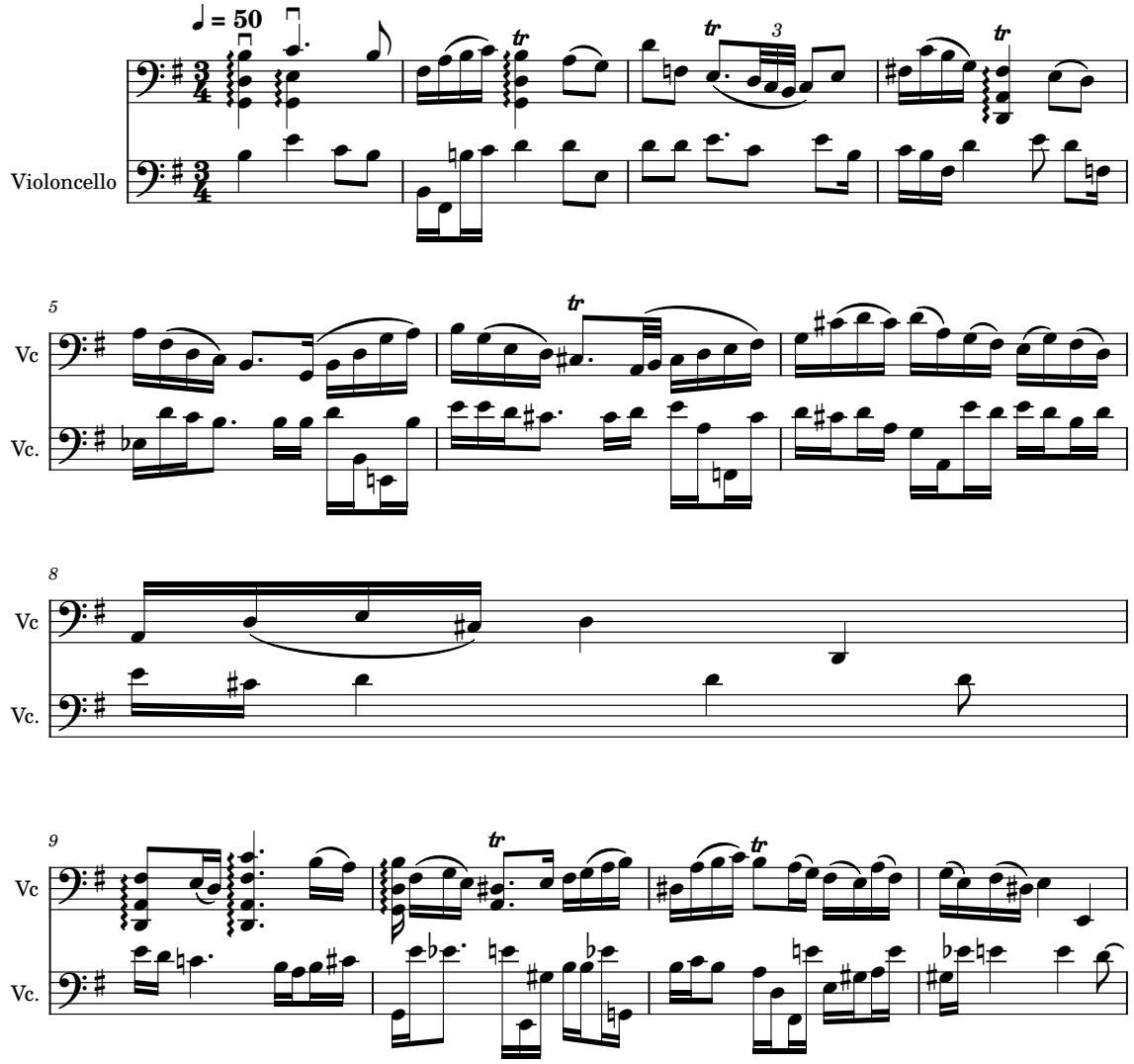

Figure 3: Generated hidden voice (*Vc.* staff, lower) against the original Sarabande melody (*Vc* staff, upper), mm. 1–12, BWV 1007. The hidden voice moves in contrary motion at phrase boundaries and gravitates toward thirds and sixths in the interior of each phrase. The chromatic passing tones in mm. 5–7 reflect the key penalty correctly permitting diatonic inflections in G major.

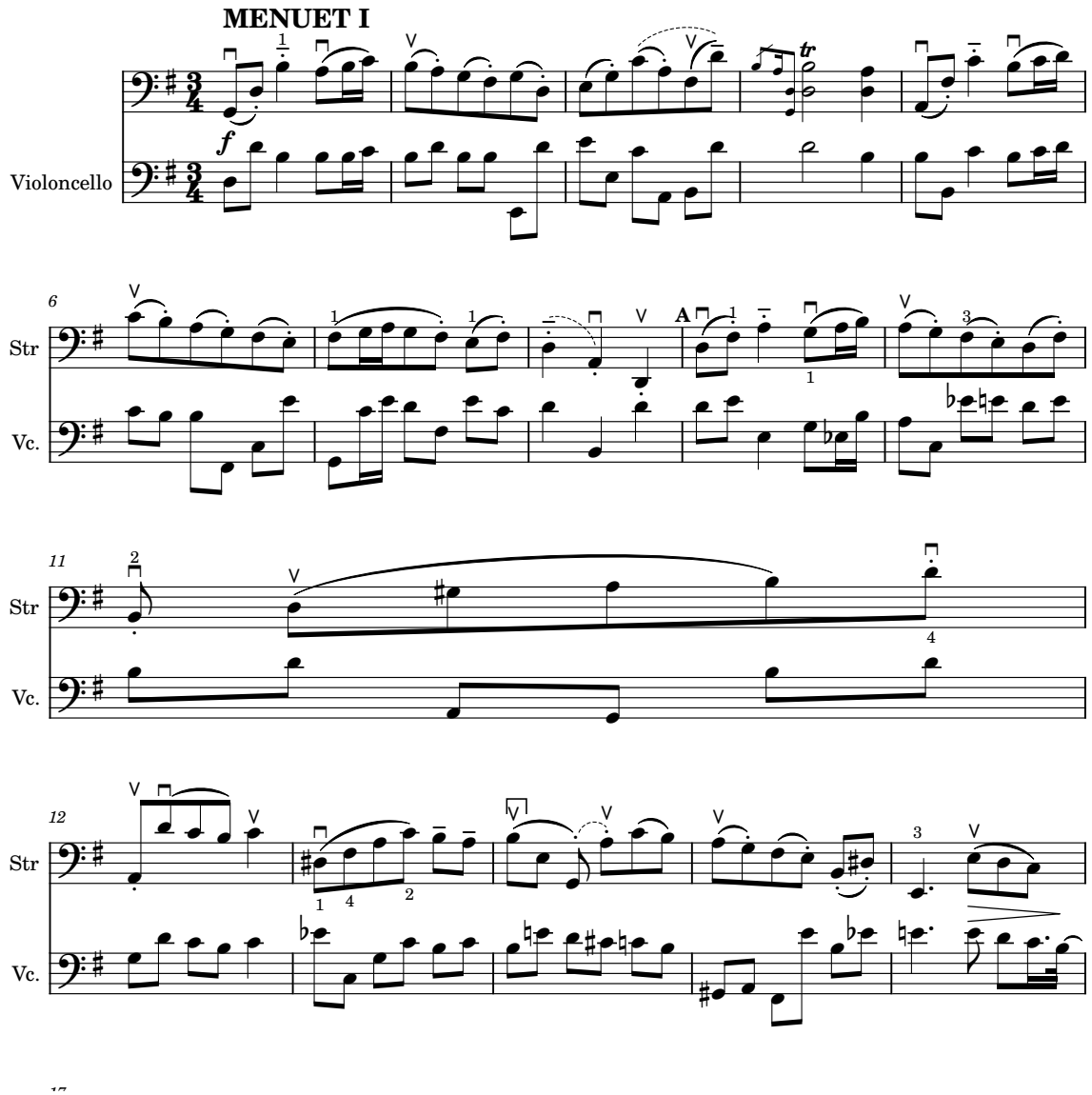

Figure 4: Generated hidden voice (*Vc.* staff, lower) against the original Minuet I melody (*Str.* staff, upper), mm. 1–16, BWV 1007. Both staves are in bass clef. The hidden voice moves primarily in thirds and sixths with the anchor, maintaining registral separation throughout. The chromatic inflections in mm. 9–10 and mm. 12–15 reflect the key penalty correctly admitting accidentals during the modal ambiguity of the B section.

References

[1] L. E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 41(1):164–171, 1970. doi:10.1214/aoms/1177697196.

[2] M. S. Cuthbert and C. Ariza. music21: A toolkit for computer-aided musicology and symbolic music data. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, pages 637–642, Utrecht, Netherlands, 2010.

[3] S. Davis. Stream segregation and perceived syncopation: Analyzing the interplay of implied polyphony and rhythm in bach's unaccompanied string works. *Music Theory Online*, 17(1), 2011. URL: https://mtosmt.org/issues/mto.11.17.1/mto.11.17.1.davis.html.

[4] N. H. Fletcher and T. D. Rossing. *The Physics of Musical Instruments*. Springer, New York, 2nd edition, 1998.

[5] J. J. Fux. *The Study of Counterpoint from Johann Joseph Fux's* Gradus ad Parnassum. W. W. Norton, New York, revised edition, 1965. Translated and edited by Alfred Mann with the collaboration of John Edmunds. Originally published as *Gradus ad Parnassum* (Vienna, 1725).

[6] G. Hadjeres, F. Pachet, and F. Nielsen. DeepBach: A steerable model for Bach chorales generation. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, pages 1362–1371. PMLR, 2017.

[7] W. Hutchinson and L. Knopoff. The acoustic component of Western consonance. *Interface*, 7(1):1–29, 1978. doi:10.1080/09298217808570246.

[8] F. T. Liang, M. Gotham, M. Johnson, and J. Shotton. Automatic stylistic composition of Bach chorales with deep LSTM. In *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR)*, pages 449–456, 2017.

[9] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965. doi:10.1093/comjnl/7.4.308.

[10] R. Plomp and W. J. M. Levelt. Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America*, 38(4):548–560, 1965. doi:10.1121/1.1909741.

[11] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989. doi:10.1109/5.18626.

[12] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978. doi:10.1214/aos/1176344136.

[13] W. A. Sethares. *Tuning, Timbre, Spectrum, Scale*. Springer, London, 2nd edition, 2005.

[14] D. Tymoczko. The geometry of musical chords. *Science*, 313(5783):72–74, 2006. doi:10.1126/science.1126287.

[15] A. Winold. *Bach's Cello Suites: Analyses and Explorations*. Indiana University Press, Bloomington, 2006.