Explaining music with AI: Advancing the scientific understanding of music through computation – Research Article

sempre: Society for Education, Music and Psychology Research | MUSIC & SCIENCE

Music & Science
Volume 7: 1–11

DOI: 10.1177/20592043231225731
journals.sagepub.com/home/mns
Sage

Melodic Differences Between Styles: Modeling Music With Step Inertia

Matt Chiu[1] **and David Temperley**[2]

Abstract

A well-known phenomenon in melodic structure is "step inertia": the tendency for a step to be followed by another step in the same direction. There is strong evidence of step inertia in three corpora of Western common-practice melodies: European folk songs, classical instrumental themes, and English hymn tunes. Surprisingly, modern Western popular music does not reflect step inertia. In Billboard's Hot 100, and *Rolling Stone* magazine's list of "greatest songs," inertial (same-direction) steps are less likely than non-inertial ones. To further explore the role of step inertia in different corpora, we created a generative model that assigns probabilities to melodies, considering just four factors: range, pitch proximity, scale-degree frequency within a key, and step inertia. We optimized the weights of these factors for the Essen Folksong Collection and the Billboard corpus, and compared them with *n*-gram models. The optimal (normalized) weight of the inertia factor is large and positive for the Essen collection (.51) and small for the Billboard corpus (.02). This is further evidence that step inertia plays a much smaller role in popular melodies than common-practice ones, and that non-inertial steps are slightly favored.

Keywords

Artificial intelligence, folk music, inertia, melodic prediction, *n*-gram, popular music, probability, style

Submission date: 2 October 2022; Acceptance date: 23 December 2023

Background

As defined by Steve Larson (2012, p. 22, original emphasis), " 'musical inertia' is the tendency of a pattern of pitches or durations, or both, to continue in the *same* fashion". Musical inertia is, therefore, another way of discussing the phenomenon of continuation (Meyer, 1956; Narmour, 1990). More specifically, "step inertia" is the expectation that a scale step will be followed by another scale step in the same direction (Huron, 2006; von Hippel, 2002).[1,2] However, even though theorists have suggested that step inertia works in both directions, data have suggested that step inertia might only be valid in descending steps (Huron, 2006)—which may be taken to reflect, extending Larson's (2012) terms, a type of pitch-height gravity.

The aim of this work is to research the role of step inertia in melody, focusing especially on differences between styles. Using representative corpora from folk songs, classical themes, hymn tunes, and popular music, we studied how step inertia operates in these styles. We then built a probabilistic model by adding an "inertia factor" to the model first proposed in Temperley (2008). In the discussion section of this article, we hypothesize about other potential influences on inertia in popular music, such as energy, phrase shape, form, and melodic–harmonic divorce.

Corpus Data

Tables 1 and 2 show a preliminary corpus study for the *Rolling Stone* corpus (Temperley & de Clercq, 2013), Billboard corpus (Arthur & Condit-Schultz, 2021;

[1] Baldwin Wallace University, Berea, OH, USA
[2] Eastman School of Music, University of Rochester, Rochester, NY, USA

Corresponding author:
Matt Chiu, Baldwin Wallace University, Berea, OH, USA.
Email: mchiu@bw.edu

Data Availability Statement included at the end of the article

Table 1. Corpus data and probability of step inertia for different corpora, for a step size of 2.

	Rolling Stone	Billboard	Essen	Barlow & Morgenstern	Hymn Tune Index
Pieces or songs	194	214	3,786	9,776	17,683
Total intervals	41,066	68,493	145,381	142,149	901,372
Total ascending	18,638	34,555	66,959	69,481	407,907
Total descending	22,428	33,938	78,422	72,668	493,465
A–A (2)	2,174	2,597	14,805	18,859	153,528
A–D (2)	4,110	4,945	8,876	9,934	123,430
D–A (2)	4,048	4,590	9,057	11,249	117,712
D–D (2)	5,276	4,624	28,068	23,593	263,091
P(step inertia): %	48	43	71	67	63

Table 2. Corpus data and probability of step inertia for different corpora, for a step size of 3.

	Rolling Stone	Billboard	Essen	Barlow & Morgenstern	Hymn Tune Index
A–A (3)	3,950	4,889	17,699	22,076	—
A–D (3)	6,741	8,677	18,591	17,265	—
D–A (3)	6,476	8,129	17,013	16,878	—
D–D (3)	7,932	7,132	35,266	28,083	—
P(step inertia) : %	47	42	60	59	—

Burgoyne et al., 2011), Essen Folksong Collection (Schaffrath, 1995), Barlow and Morgenstern (1948), and Hymn Tune Index (Temperley & Manns, 1983).[3] The *Rolling Stone* and Billboard corpora contain examples of popular music, whereas the Essen Folksong Collection contains folksongs, the Barlow and Morgenstern corpus (encoded by David Huron) contains classical themes, and the Hymn Tune Index contains hymn tunes from 1535 to 1820. Data from these corpora were converted to MIDI encodings—MIDI encodings were used throughout the study. The Hymn Tune Index is the one exception because it is encoded as scale degrees.

This preliminary study compares step inertia with "step reversal"—the probability that a step is followed by a step in the opposite direction. To do this, we remove any pitch repetitions in the data to control for simple pitch prolongations, and isolate instances of steps ("context steps"), where a step is an interval less than or equal to 2 semitones.[4] Using these context steps, we then record the direction of subsequent steps. In Tables 1 and 2, A–A shows the number of ascending steps followed by ascending steps, A–D shows the number of ascending steps followed by descending steps, and so on. The probability of step inertia (in the context of stepwise motion) is calculated as the summed count of A–A with D–D, divided by the summed total of all stepwise motion [*P*(step inertia) in Tables 1 and 2].[5] A probability above 50% suggests that step inertia is more prevalent in that style than step reversal, and a probability below 50% suggests the opposite. As shown, step inertia is more common in Western folk songs, classical themes, and hymns, whereas step reversal is more common is popular music.

Pop music often includes pentatonic melodies, meaning that scale steps are occasionally 3 semitones instead of 2.[6] Table 2 shows the corpus study repeated with the step size set to 3. This decreases the probability of step inertia in folk songs and classical themes but does not notably change the results for pop music.[7]

To show a more detailed perspective, the interval distribution following both ascending and descending steps is shown in Figure 1—histograms are sorted by corpus.[8] The histograms for the Essen and Barlow & Morgenstern corpora and the Hymn Tune Index show clear instances of step inertia: following a descending step (on the left), the most frequent consequent interval is a descending step; following an ascending step (on the right), the most frequent consequent interval is an ascending step. This is, however, not true for the popular music corpora. The *Rolling Stone* corpus exhibits step inertia for descending steps, but the Billboard corpus shows that, following a descending step, descending and ascending steps are equally probable. For ascending steps in both corpora, step reversal is more frequent than step inertia.

The results from these preliminary studies suggest that step inertia is not always a feature within popular music styles—a detail that has been unobserved in the literature on step inertia and popular music. To test the role of inertia in modeling music, we designed a probabilistic model of melody that implements inertia, named KRPI after its contributing factors: key, range, proximity, and inertia. The model was trained and tested on (a subset of) the Essen and Billboard corpora to compare their stylistic differences.

The KRPI Model

To investigate the role of step inertia in style, we built an artificial intelligence model for symbolic music—specifically, a probabilistic gestalt-based model.[9] Unlike a

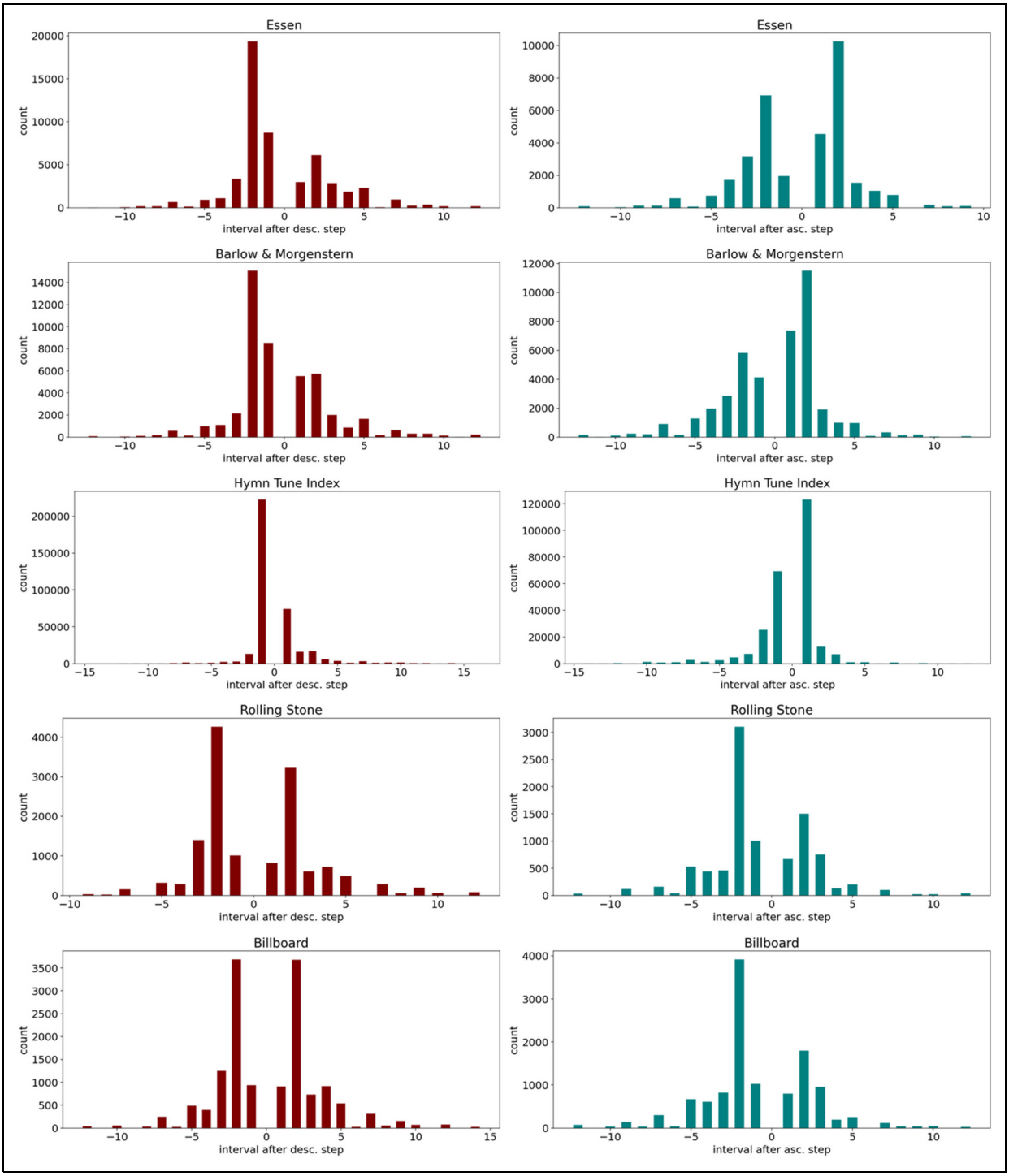

Figure 1. Interval distributions after (left) a descending step and (right) an ascending step. Intervals are measured in semitones.

supervised, machine-learning model, whose behavior emerges from a large number of learned probabilities, our model was designed with a small number of higher-level principles in mind. Because it is inspired by gestalt psychology, the model is also a model of melodic perception.

The model incorporates four factors: range; proximity; key; and inertia. Each factor represents a particular musical characteristic and uses that characteristic for prediction; for each note in an ongoing melody, each factor independently makes predictions about the next pitch. The factors then combine to yield a single overall prediction. The goal of the model is to achieve the highest probability for melodies, given its factors, the logic being that the model that best predicts a melody best describes the musical style that gave rise to it. All predictions are made using MIDI values (between 28 and 85), and

assume discrete pitches; further advances might explore the model's use in the audio domain.

In his original model, Temperley (2008) used range, proximity, and key for melodic prediction. We will first define Temperley's original model and then introduce the additional inertia factor.

Melodies often use a relatively small range and most pitches are drawn to the center of that range. The range factor approximates the probability of pitches around a central pitch by using a normal distribution. A normal distribution requires *mean* and *variance* values to calculate, respectively, *where* the bell curve is centered and *how wide* it flares.[10] The mean value for the normal distribution is the moving average of the previous pitches in the melody; the variance is a parameter that the model optimizes through a training period (see next section).[11]

Most (folk and popular) melodies tend to move by small intervals, an observation supported by corpus data (Huron, 2001, pp. 74–75).[12] The "proximity" factor makes the model prioritize motion by smaller intervals. To account for melodies' use of small intervals, each predicted note is therefore restricted by the previous pitch. This preference for small intervals is approximated using another normal distribution. Whereas the mean of the range factor is determined by a moving average, the mean of the proximity factor's normal distribution is updated using the previous pitch. Just like the range factor, the variance for the proximity factor is optimized through the training period.

The last factor incorporated in Temperley's (2008) model is "key". Folk and popular melodies tend to stay within a key, emphasizing more stable, tonic-triad pitches. Margulis's (2003) stability factor draws from Bharucha and Krumhansl (1983) and Lerdahl (2001), and reflects how "anchored" a pitch class is within a key using hierarchical weighting. Less stable pitches within a key are more likely to proceed to stable pitches; instead, we base the key parameter on pitch-class (PC) distributions from the Essen and Billboard corpora. This corpus approach is more flexible when describing differences between styles.

To derive the PC distribution for the Essen corpus, we first determine each melody's key.[13] For each melody, we apply the default key-finding algorithm from Music21 (based on the Krumhansl–Schmuckler probe-tone algorithm). If a piece receives greater than .8 major key confidence, we transpose that melody to the key of C major and extract its PCs; the PCs for each piece are then totaled into a PC distribution. The PC distribution for the Essen Folksong Collection is given in Table 3.

Melodies from the Billboard corpus have keys encoded at the top of each .krn file (in the CocoPops dataset). Isolating major transcriptions in the Billboard corpus leaves 172 melodies—the PC distribution is given in Table 4. Tables 3 and 4 show that both corpora have similar PC distributions, but the Billboard corpus's distribution is slightly flattened. It also has pronounced scale degrees 6 and flat 7, perhaps a result of pop's frequent use of pentatonicism and certain modal tendencies in pop and rock music.

An innovation of our model over Temperley's (2008) model is the addition of an inertia factor. The inertia factor requires two context pitches before taking effect. If there is a small interval between the two context pitches, inertia predicts a continuation in the same direction. We use the term "small interval" synonymously with scale step. Where w is the inertia factor weight and s is a designated small interval, if the interval between the two context pitches i,j is less than or equal to s (in other terms, $|j - i| \leq s$) and ascending ($j > i$), the inertia weighting function for pitch x after an ascending step is represented as

$$f(x) = \begin{cases} 1 + w & \text{if } x \leq j + s \quad \text{and} \quad x > j \\ 1 - w & \text{if } x \geq j - s \quad \text{and} \quad x < j \\ 1 & \text{if } x > j + s \quad \text{or} \quad x < j - s \end{cases} \tag{1}$$

If the context interval is a descending step, the addition or subtraction signs between j and s are inverted.

The weighting value w alters the impact of the inertia function with respect to other factors, rewarding steps that continue in the same direction ("inertia steps"), and punishing those that change direction ("step reversals"). Values for w range from -1 to 1. The following is a demonstration of how the inertia factor works. All pitches are first assigned "1" as a value. If a step is not present, or the weighting w is set to 0, the distribution is uniformly "1." Assigning the weight 0 is the equivalent of saying that "inertia has no impact." When w is assigned .5, any pitch within a scale step that continues in the same direction is weighted 1.5, reversal steps are weighted .5, and other pitches are still 1. If w is negative, this suggests that step reversals are favored over step inertia. Figure 2 shows the inertia factor (for MIDI pitches 45–75) when given an ascending interval from B3 (59) to C4 (60), the small interval s is set to 2, and

Table 3. Essen Folksong Collection pitch-class distribution for melodies with >.8 key confidence (3,786 melodies).

Pitch class	0	1	2	3	4	5	6	7	8	9	10	11
Probability: %	17	.05	16	.04	20	12	.5	22	.04	7	.1	5

Table 4. Billboard corpus pitch-class distribution (172 melodies).

Pitch class	0	1	2	3	4	5	6	7	8	9	10	11
Probability: %	17	4	11	5	13	8	3	14	3	10	4	8

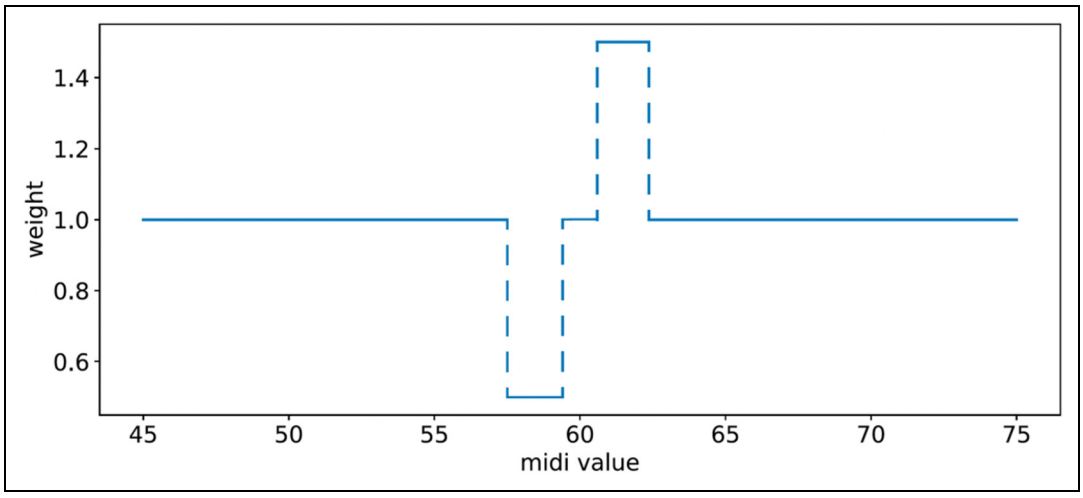

Figure 2. Inertia factor contextualized by previous pitches B3 and C4, $s = 2$, $w = .5$.

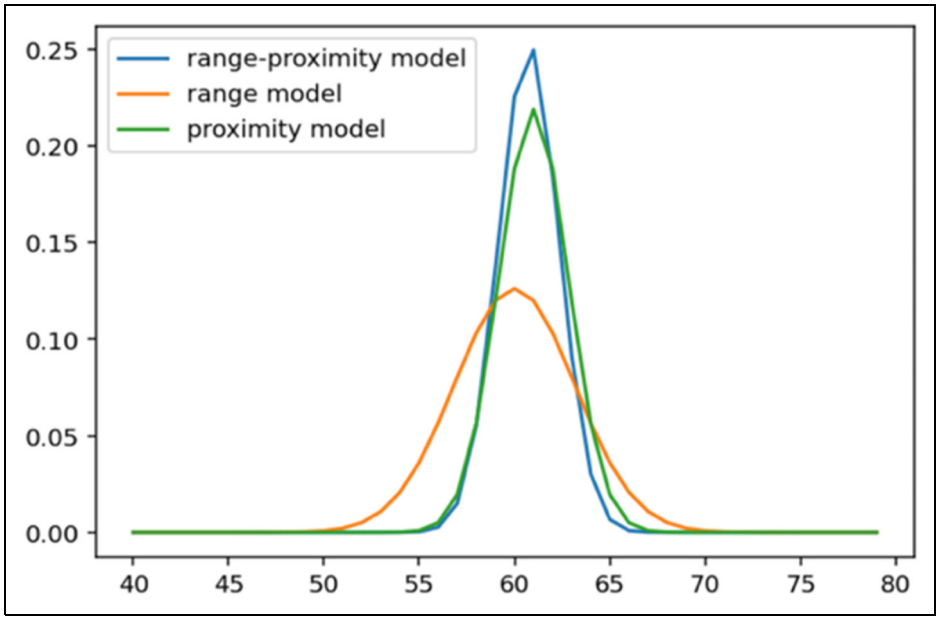

Figure 3. Combining normal distributions for range and proximity models.

w is .5.[14] On its own, the model expects C♯4 and D4 and punishes reversal steps to B♭3 or A3.

Now that we have discussed each factor (KRPI) in detail, let's discuss how they are combined. The point-wise product of the four factors returns a normalized distribution such that all values add to 1. In other words, the model creates a probability distribution, where the highest values are the most expected from the model. As a visual demonstration of the combination of factors, Figure 3 shows a possible combination of the range and proximity factors. The range–proximity model predicts the subsequent pitch of a melody that contains the starting pitches [B3, C#4]. Notice how the range distribution is centered on MIDI 60—the average of B3 (59) and C#4 (61)—and that the proximity distribution is centered on MIDI 61, since the model just encountered C#4. When combined, a new distribution is formed that predicts a melody with a limited range that might move by small intervals. Temperley calls this a range–proximity (RP) profile. Adopting this labeling system, we called the completed gestalt model the KRPI model.

Parameters and Optimization

Each factor uses "parameters" that are adjusted to affect the relative contribution of the factors and the predictions of the overall model. These parameter values are optimized for a given corpus. Because these values are optimized for a particular corpus, by examining the parameters we can interpret which musical features the model finds most useful for prediction within a style: For example, a low value (close to 0) for the inertia's "weight" parameter would

suggest that it has little to no effect for predicting pitches within that style, and, because these parameters are optimized based on data, it is deduced that that step inertia occurs less often in that style. The parameters that were optimized are listed in Table 5, showing a range of initiated values; some parameter values are unchanged by optimization, and some parameters are randomly initiated and optimized.

To adjust parameters, we designed an optimization procedure (Figure 4). The optimization procedure aims to maximize the probability of each melody. The probability of a melody is equivalent to the joint distribution of its individual conditional probabilities (an equivalency known as the *chain rule*):

$$P(e_1, e_2, \ldots, e_n) = P(e_1)P(e_2 \mid e_1)P(e_3 \mid e_1, e_2)\ldots P(e_n \mid e_1, e_2, \ldots, e_{n-1}) \quad (2)$$

The optimization algorithm tries to maximize this value; the higher the joint probability, the better the model performs at predicting pitches overall.

In the optimization procedure, the corpora (Essen Folksong and Billboard) are first split into six parts; 5/6 of the dataset for training and the other 1/6 as the evaluation dataset. The training dataset is used to optimize the model parameters, and the evaluation dataset tests whether the model is generalizable. Separating the corpus into training and evaluation samples avoids overfitting a model to a specific set of melodies rather than learning broader, more applicable parameter values. The algorithm loops through the training dataset. For each melody, a single parameter is randomly moved up or down by a "step."[15] The step size is .01 times the current parameter value for range and proximity variance and the inertia step size is .25 times the current value, meaning that the steps scale with the magnitude of the value. Using the original unchanged parameter value and the newly adjusted parameter value, the models calculate the probability of the melody. If the probability (measured in bits) is higher with the new parameter value, then the adjusted parameter value is kept, replacing the initial value.[16] Otherwise, that parameter reverts to its initial value. Therefore, the model continues to choose parameter values with higher likelihoods. While each of the parameter changes is small at first, the process is repeated for each song in a training set.

Table 5. Parameter values for KRPI model.

	Parameters	
Factor	Constant	Initially randomized within given range
Key	Pitch-class distribution: {corpus}	
Range	Mean: Average of melody	Variance: 1–50
Proximity	Mean: Previous pitch	Variance: 1–50
Inertia	Small interval: 2	Weight: .01–.99

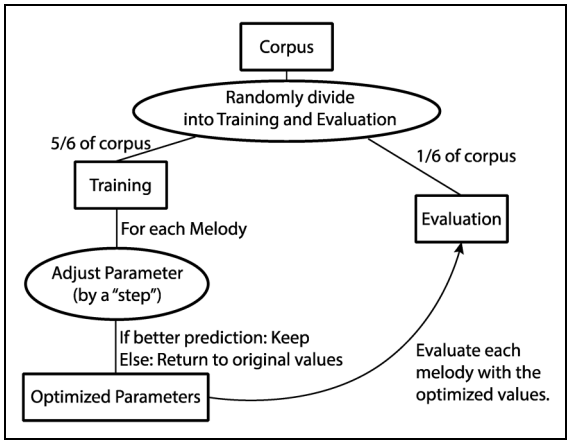

Figure 4. Optimization procedure for KRPI model.

As reported in the section on PC distributions, the dataset consisted of major pieces from the Essen and Billboard corpora. This means that differences in corpus size also affect the number of opportunities for training-set parameter adjustments: Because the Essen corpus contains 3,786 songs, the training set contains 3,155 songs. The Billboard's training dataset contains 144 songs. Therefore, to approximate the same number of optimization steps from the Essen corpus, each training set in the Billboard was duplicated 22 times. For each of the corpora, the optimization process was run three times with different (shuffled) training and evaluation datasets, and with three differently randomized initialized parameters.[17]

As a demonstration, Figure 5 shows how the KRPI model optimized the parameter weights (range variance, proximity variance, and inertia weight) for the Billboard corpus. Figure 5 represents the three repeated optimization procedures on the randomly initiated parameters; each of the parameters started randomly (between 0 and 50 for variance values and between 0 and 1 for inertia weight). In some instances, like Figure 5(b), a parameter starts close to its final value: the inertia parameter starts at .08 and moves to .03. The further away a parameter is from an optimal value, the longer the slope: the proximity parameter in Figure 5(a) starts with a steep decline before leveling out. The same process was implemented for the Essen model's parameters.

Results

Experiment 1: Comparing KRP and KRPI Models

To show the impact of inertia on various styles, we compared the KRPI model with Temperley's original key, range, and proximity (KRP) model. The optimized KRP and KRPI models for the Essen and Billboard corpora are shown in Tables 6 to 8. The first two tables (Tables 6 and 7) show the optimized parameter values for each iteration of the optimization procedure on the Billboard and Essen corpora. The iterations are each ordered top to bottom. Table 8 gives the average optimized values from Tables 6 and 7.[18]

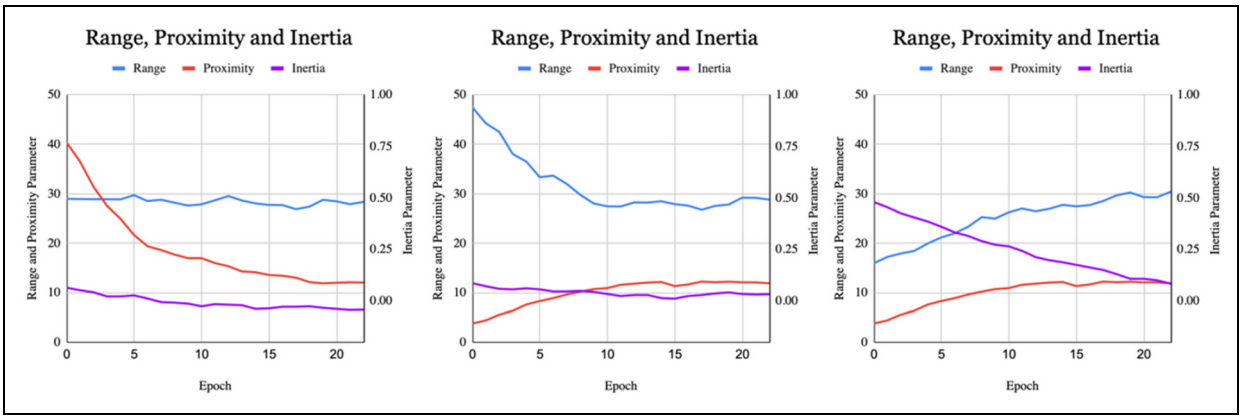

Figure 5. (a–c) Three optimization processes for KRPI model optimized on Billboard corpus.

Table 6. Optimized parameter values for Essen corpus using KRP and KRPI models.

	Range	Proximity	Inertia	Prediction: bits
KRP	{28.03, 31.41, 26.18}	{10.20, 12.21, 10.85}		{2.72, 2.73, 2.70}
KRPI	{25.55, 26.61, 23.68}	{12.39, 10.71, 10.41}	{.58, .42, .52}	{2.67, 2.66, 2.68}

Table 7. Optimized parameter values for Billboard corpus using KRP and KRPI models.

	Range	Proximity	Inertia	Prediction: bits
KRP	{23.97, 28.32 23.28}	{11.59, 11.48, 10.05}		{3.66 3.55, 3.60}
KRPI	{28.34, 28.81, 30.40}	{12.03, 11.91, 11.16}	{−0.04 0.03 0.08}	{3.47, 3.63, 3.65}

Tables 6 and 8 show that, even when the inertia factor is introduced in the KRP Essen model, the proximity's variance remains between 10.2 and 12.39. The variance for the range factor between the KRP and KRPI Essen models decreases slightly on average. The inertia is optimized to .51, meaning that the model is responsive to the observed inertia data discussed previously, further confirming that the Northern European folk song style is reliant on step inertia. Adding inertia to the model also decreases the information per note (improving the prediction) by 1.8%.

Like the Essen model, when inertia is introduced to the KRP Billboard model the variance for both range and proximity change marginally (Table 8). However, unlike the Essen model, the inertia weight optimizes to a negligible number (.02). These results bolster the statistical claim shown earlier that step inertia is not as common in popular music. Even though a step inertia factor improves the average accuracy of the KRP model (by .6%), Table 7 shows that two of the three predictions were worse than some of the predictions made using the KRP model. Step inertia is not a reliable feature for inclusion in a model of popular music.

Experiment 2: N-Gram Model

To give the models more context, we compared the KRP and KRPI models with *n*-gram models. An *n*-gram is a string of characters of length *n* from a given sequence—in our case, a sequence is a string of pitches. Models based on *n*-grams use these previous sequences as "contexts" to make predictions about "target" pitches. One of the most successful and widely implemented *n*-gram models in music is the *Informational Dynamics of Music* (IDyOM) system (Pearce, 2005, 2018; Pearce et al., 2004; Pearce & Wiggins, 2004, 2006).[19] IDyOM takes as its input a sequence of events representing discrete basic musical features, including onset time, duration, or pitch.

For our comparison, we used a simple *n*-gram model trained on pitch data. The *n*-gram model stores contexts of length 1–3 in a "pitch dictionary" and then records the subsequent pitch as a "target." The model compares pitch patterns in the music with stored contexts, and uses a probability distribution of targets to make a prediction. All *n*-gram lengths are combined in a single probability distribution, where each order is weighted equally.[20] Pitch predictions receive an additional smoothing weight of .01 so that the probability of each pitch is never 0. The pitch dictionary is reset for each piece—this is referred to in IDyOM as a short-term memory *n*-gram model.

The *n*-gram model performed worse than the KRP or KRPI models on the Essen folk songs (4.15 bits per note), but it performed better on the Billboard corpus (2.67 bits per note). This suggests an additional stylistic difference between folk and popular music: Folk songs seem

Table 8. Average optimized parameters values and predictions for Essen and Billboard corpora using KRP or KRPI models. Prediction is average information (in bits) per note in evaluation datasets.

	Essen KRP	Essen KRPI	Billboard KRP	Billboard KRPI
Range (variance)	28.54	25.28	25.19	29.18
Proximity (variance)	11.09	11.17	11.04	11.7
Inertia (weight)		.51		.02
Prediction: bits	2.72	2.67	3.60	3.58

to be influenced by more generic musical principles—like those coded into the KRP or KRPI models—whereas popular music melodies might be dictated more by exact repetition of short pitch patterns (captured by the *n*-gram model).

Experiment 3: Melody Length Control

The results from Experiment 2 show that the *n*-gram models outperform the KRPI models in the popular music corpus. However, this is probably due to the notated repeats in the Billboard corpus. On average, songs from the Billboard dataset are ≈404 pitches in length and those from the Essen dataset are ≈51 pitches; the Billboard corpus has its repeated sections, like verses and choruses, written out, whereas in the Essen Folksong Collection, repeats (such as stanzas) are generally omitted. These notated repeats might bias the *n*-gram model, so we reran the Billboard *n*-gram and KRPI models while controlling for the length of the melody.

Using the major melodies in the Billboard corpus, we ran the same *n*-gram model (interpolated with orders 1–3) and KRPI model (with the average optimized parameters) on the first 51 notes of each Billboard melody. The results are given in Table 9. The Billboard *n*-gram model, as expected, performed worse than it did previously because it had fewer notes to build its dictionary. It still performed better than the Essen *n*-gram model (Table 8). This suggests that, regardless of melody length, the popular music corpus contains more pitch repetitions.

Evaluating the Billboard KRPI on the opening 51 notes of each melodies also improves its performance. Not only did it improve, but it outperformed the *n*-gram model. This may be attributed to the standard verse–chorus form of popular music songs—a topic we pick up in the discussion section.

Table 9. Comparison of Billboard KRPI and Billboard *n*-gram model with controlled melody lengths (maximum 51).

	Billboard KRPI	Billboard *n*-gram
Prediction: bits	3.49	3.66

Discussion

Experiment 1 showed a marked difference when using the same model for different musical styles. The KRPI model performed much better for folk songs than it did for popular melodies, and step inertia was not useful for making predictions for the popular music corpus. One explanation for this difference could be related to the energy of different styles: lower-energy styles (on average), such as classical, folk, and hymns, may use more stepwise motion, whereas rock and pop—styles associated with high energy—may use more leaps and less step inertia. This hypothesis seems more plausible when considering the range of each melody in the corpora: the average range for a folk song melody is 13.1 semitones and its average interval size is 2.8 (repetitions removed), whereas the average range for a Billboard melody is 19.4 semitones and its average interval size is 3.8 (minor pieces included and repetitions removed).[21]

Another explanation for the presence of step inertia in some styles and not others attributes step inertia to other musical characteristics, rather than being the generator of pitch behavior; we hypothesize that step inertia and the lack of it in popular music might be the result of a higher-order musical feature that gives rise to inertia. In the transcribed Essen folk songs, many of the melodies follow a common-practice tonal language. Without an accompanying harmony, melodies often walk stepwise and establish tonal, harmonic progressions. In this way, step inertia might be the fallout of trying to convey a particular triad: Moving between triadic pitches fulfills the criteria for step inertia, so it is possible that the high frequency of these patterns results in step inertia. Recently, many popular music scholars have suggested that pop melodies might not be as yoked to their harmonizations (or to convey triads) as a more traditional style (de Clercq, 2019; Nobile, 2015; Temperley, 2007). In popular music, such harmony-reinforcing motion might not be "needed": Because melodies in popular music are "divorced" from the underlying harmony, a melodic reinforcement—that is, step inertia—would be unlikely.

Another potential (and not mutually exclusive) conjecture for step inertia might be phrase shape. Huron (2006) shows that phrases in the Essen Folksong Collection have an arch-like contour. Since melodies tend to move stepwise, motion toward the apex and then down again would fulfill criteria for step inertia. The absence of such phrase shapes in popular music would decrease the amount of step inertia.

We conducted a preliminary study of phrase shapes in popular music to show that such arch shapes are present, but slightly different in popular music. We first extracted "unbroken bounded sequences" (UBSs)—adjacent pitch sequences bounded with rests on both sides—of length 7–9 from the Billboard corpus.[22] Each starting pitch of a UBS was normalized to 0; then, for each position in the UBS, we averaged the interval from the starting note.

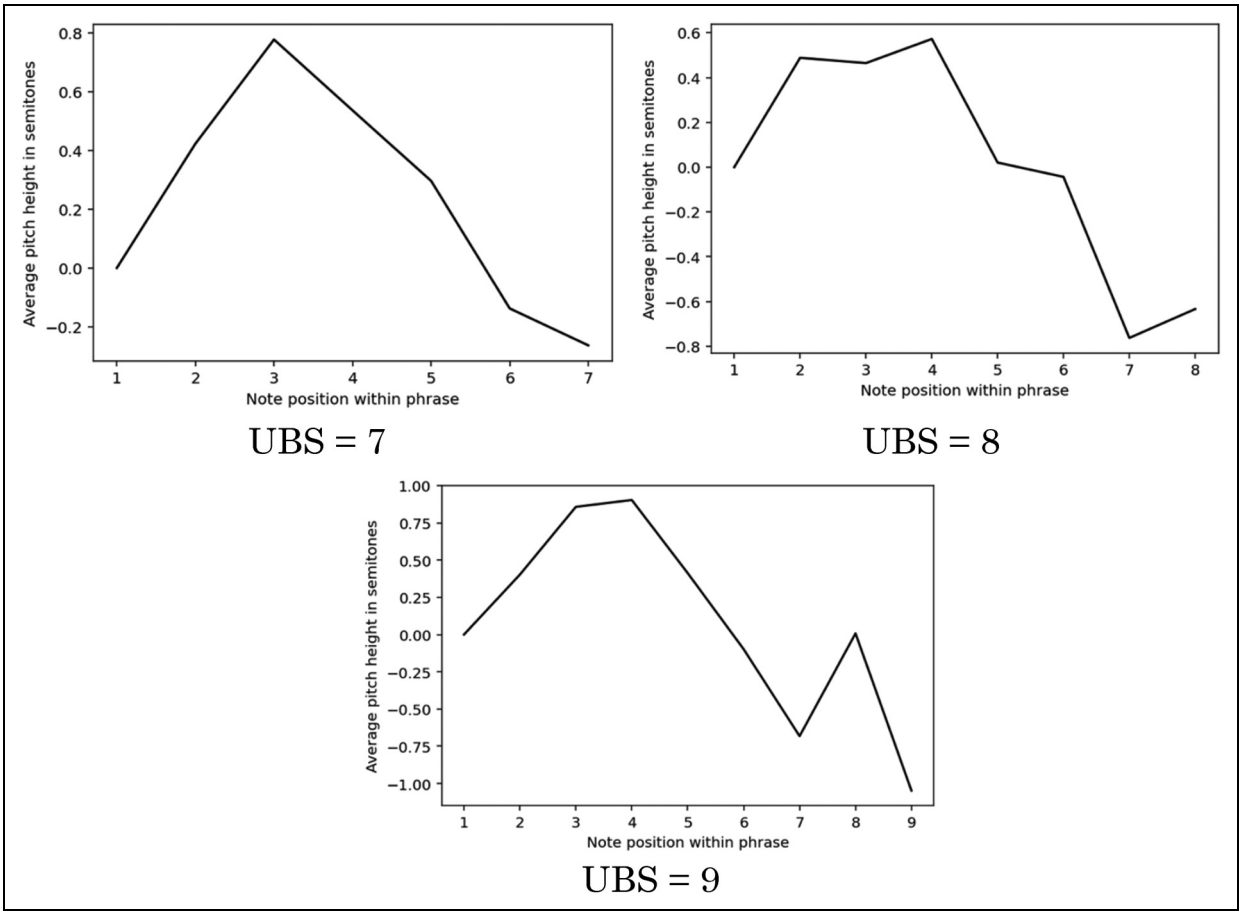

Figure 6. Average pitch height for unbroken bounded sequences in the Billboard corpus.

Though crude, if phrases in popular music establish consistent shapes, this may give approximations of average phrase shapes. Figure 6 shows the results. Interestingly, the UBSs do result in arch-like shapes similar to those shown in Huron (2006), but the apex of the arch slants to the left and the ends of the sequences dip below the initial pitch. This could be the result of the local range: perhaps UBSs in popular music drift first above and then below the starting pitch, elaborating on a pitch in a particular register. These results suggest interesting characteristics about the popular music phrase, but they are not enough to directly connect phrase shape with step inertia.

Lastly, a final discussion on melody length: When melody length is controlled, the KRPI Billboard model outperforms the *n*-gram model. This potentially points to the formal organization of popular songs: The verse and chorus sections in verse–chorus songs need to be differentiated in some way, and one way to do that is with register changes. The KRPI model performs better if a melody changes register less frequently—the combination of range and proximity factors prioritizes small, stepwise motion. It might also be true that verses often start songs, and that verses have a more confined range than choruses. This aligns with our intuition, but it would take additional research into the different properties of formal sections in pop music.

Conclusion

This study reminds us to be cautious with our corpus methodologies: The generalizations we form from data should be carefully contextualized. Though step inertia had been a well-researched and established phenomenon in folk and classical tonality, its effect, as shown by this article, is stylistically constrained. This leads us to question other assumptions we make about music—whether in researching or teaching—and whether such assumptions are truly cross-stylistic; studying these differences in style points to the very characteristics that make styles unique. Future research might then use these characteristics for style or genre classification.

Action Editor

David Meredith, Aalborg University, Department of Architecture, Design and Media Technology.

Peer Review

Two anonymous reviewers.

Author Contributions

MC wrote the code for the model and its optimization and wrote the manuscript. DT developed the original framework for the model, provided guidance for the project, and edited the manuscript.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical Approval

This research did not require ethics committee or institutional review board approval. This research did not involve the use of personal data, fieldwork, or experiments involving human or animal participants, or work with children, vulnerable individuals, or clinical populations.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Matt Chiu https://orcid.org/0000-0002-8186-7241

Data Availability Statement

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Notes

1. For a survey of step inertia, see Huron (2006, pp. 77–80).
2. There have been, however, critiques and skepticism about step inertia. In Eugene Narmour's implication-realization model, he suggests that step inertia is the result of other competing factors (1992). Consider the following: If a melody is in the lowest part of its range and there is a melodic step toward its mean, we might expect the melody to continue towards the middle of its range. Is a subsequent step towards the mean the result of inertia, simply a move to the mean of the melody's range, or somehow both?
3. The Billboard corpus used here is derived from transcriptions in the CocoPops corpus (Arthur & Condit-Schultz, 2021). The CocoPops corpus encodes melodies from the Billboard corpus into .xml and .krn files.
4. Because the Hymn Tune Index is encoded in scale degrees, its scale step is set to 1.
5. Tables 1 and 2 show that, in all corpora, descending intervals occur more than ascending ones; this suggests that D–D overrepresents inertia because of the frequency of descending intervals. However, in the overall calculation of P(step inertia), A–D controls for that imbalance, since it also includes descending intervals, but is not considered step inertia.
6. This pentatonic motion is clearer when comparing the average interval size used in the Essen and Billboard corpora: after repetitions are removed, the average interval between notes in the Essen folksong collection is 2.8, whereas the average interval in the Billboard corpus is 3.8.
7. The Hymn Tune Index is removed for the second study because it is not possible to measure an analogous distance using scale degrees: 1 scale-degree step can be either 1 or 2 semitones (which is why the step distance is initially set to 2), but 2 scale-degree steps can be either 3 or 4 semitones.
8. Repetitions are still removed, step size equals 2, and any intervals with fewer than 20 counts are taken out. Note (once again) that the Hymn Tune Index uses scale degrees, so its x-values represent generic, scale-degree intervals.
9. Temperley's (2008) model is referred to as a gestalt model by Morgan et al. (2019).
10. One may also use a standard deviation value, since variance is the standard deviation squared.
11. In Temperley's original model, the parameter values were based on the Essen corpus prior to evaluation.
12. Such claims are bolstered by perceptual studies in auditory-stream segregation—or the ability to isolate and group sounds based on acoustic and schematic features of a sound pattern (Bregman, 1990; Dowling, 1968). Scale steps allow human beings to group pitches more easily into a melody. However, auditory streams might not necessarily be melodies and melodies are not always single streams. For example, "compound melodies" are considered "melodies," but they consist of several lines.
13. Melodies are taken from the "Europa" folder of the Essen corpus, including a total of 3,786 melodies (after transposition).
14. In this article, the small-interval parameter s accounts for scale steps of 1–2 semitones. Even though a scale step might be 3 semitones in the context of a pentatonic scale, preliminary results suggest that there is no significant difference when using step size = 3.
15. Be aware of the terminological overlap between scale step and an optimization step. The method used here is sometimes known as a "brute force" method.
16. Bits are calculated as $-\log_2(P)$.
17. The three initialized random parameters were used for both the Essen and Billboard models to avoid an initial condition that might bias the optimizations. For example, the first set of randomized parameters (rand1) was used to train the KRP and KRPI models for the first set of shuffled Essen and Billboard datasets, the next set of randomized parameters (rand2) were used for the next set of shuffled datasets, and so on.
18. We used an arithmetic mean because values were not as volatile.
19. IDyOM is a system rather than a single model because it allows users to specify parameters and then construct Markov models.
20. This process is known as interpolation, and future iterations should weight orders differently.
21. We thank Reviewer 2 for suggesting this explanation.
22. This results in 978, 891, and 519 UBSs for lengths 7, 8, and 9, respectively.

References

Arthur, C., & Condit-Schultz, N. (2021). Testing the 'loose-verse, tight-chorus' model: A corpus study of melodic-harmonic divorce [Conference presentation]. Forty-Fourth Annual Meeting of the Society for Music Theory, Virtual.

Barlow, H., & Morgenstern, S. (1948). *A dictionary of musical themes*. Crown Publishers.

Bharucha, J., & Krumhansl, C. L. (1983). The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition*, *13*(1), 63–102. https://doi.org/10.1016/0010-0277(83)90003-3

Bregman, A. (1990). *Auditory scene analysis: The perceptual organization of sound*. MIT Press.

Burgoyne, J. A., Wild, J., & Fujinara, I. (2011). An expert ground truth set for audio chord recognition and music analysis. *Proceedings of the International Society for Music Information Retrieval*, *11*, 633–638.

de Clercq, T. (2019). The harmonic-bass divorce in rock. *Music Theory Spectrum*, *41*(2), 271–284. https://doi.org/10.1093/mts/mtz006

Dowling, W. J. (1968). *Rhythmic fission and perceptual organization* [Unpublished doctoral dissertation]. Harvard University.

Huron, D. (2001). Tone and voice: A derivation of the rules of voice-leading from perceptual principles. *Music Perception*, *19*(1), 1–64. https://doi.org/10.1525/mp.2001.19.1.1

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. MIT Press.

Larson, S. (2012). *Musical forces: Motion, metaphor, and meaning in music*. Indiana University Press.

Lerdahl, F. (2001). *Tonal pitch space*. Oxford University Press.

Margulis, E. (2003). *Melodic expectation: A discussion and model* [PhD dissertation]. Columbia University.

Meyer, L. (1956). *Emotion and meaning in music*. University of Chicago.

Morgan, E., Fogel, A., Nair, A., & Patel, A. D. (2019). Statistical learning and gestalt-like principles predict melodic expectations. *Cognition*, *189*, 23–34. https://doi.org/10.1016/j.cognition.2018.12.015

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realization model*. University of Chicago Press.

Narmour, E. (1992). *The analysis and cognition of melodic complexity*. University of Chicago Press.

Nobile, D. (2015). Counterpoint in rock music: Unpacking the 'melodic-harmonic divorce.'. *Music Theory Spectrum*, *37*(2), 189–203. https://doi.org/10.1093/mts/mtv019

Pearce, M., Conklin, D., & Wiggins, G. (2004). Methods for combining statistical models of music. *Proceedings of International Symposium on Computer Music Modeling and Retrieval*, *2*, 295–312. https://doi.org/10.1007/978-3-540-31807-1_22

Pearce, M., & Wiggins, G. (2004). Improved methods for statistical modelling of monophonic music. *Journal of New Music Research*, *33*(4), 367–385. https://doi.org/10.1080/09298210520000343840

Pearce, M., & Wiggins, G. (2006). Expectation in melody: The influence of context and learning. *Music Perception*, *23*(5), 377–405. https://doi.org/10.1525/mp.2006.23.5.377

Pearce, M. T. (2005). *The construction and evaluation of statistical models of melodic structure in music perception and composition* [PhD dissertation]. City University London.

Pearce, M. T. (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

Schaffrath, H. (1995). *The Essen Folksong collection in Kern format*. Edited by David Huron. CA: Center for Computer Assisted Research in the Humanities.

Temperley, D. (2007). The melodic-harmonic 'divorce' in rock. *Popular Music*, *26*(2), 323–342. https://doi.org/10.1017/S0261143007001249

Temperley, D. (2008). A probabilistic model of melody perception. *Cognitive Science*, *32*(2), 418–444. https://doi.org/10.1080/03640210701864089

Temperley, D., & de Clercq, T. (2013). Statistical analysis of harmony and melody in rock music. *Journal of New Music Research*, *42*(3), 187–204. https://doi.org/10.1080/09298215.2013.788039

Temperley, N., & Manns, C. G. (1983). *Fuging tunes in the eighteenth century*. Information Coordinators.

von Hippel, P. (2002). Melodic-expectation rules as learned heuristics [Conference presentation]. Proceedings of the 7th International Conference on Music Perception and Cognition.