An Intuitive Physics Approach to Modeling Melodic Expectation

Breanna K. Nguyen (breanna.nguyen@yale.edu)
Psychology, Yale University, New Haven, CT, USA

Ilker Yildirim (ilker.yildirim@yale.edu)
Psychology, Yale University, New Haven, CT, USA

Abstract

Humans have an intuitive understanding of music. We can predict the ensuing notes of a melody given the first few notes, but what exactly drives these predictions? Previous research on musical cognition explores probabilistic models of melody perception where a melody's structure can be inferred given its surface. Other research theorizes about "musical forces", forces that are analogous to how we represent the physical world, and which inform the way we form expectations about music. We propose a single model of melodic expectation that combines both ideas using a structured generative model and sequential Monte Carlo inference. The generative model formalizes these musical forces, and combined with inference, enables predicting the last note of a melody given the beginning notes. This model explains human performance in an existing dataset of melodic predictions. The model explains more variance than its ablations, and suggests an "intuitive physics" basis for melodic expectation.

Keywords: melodic expectation; musical cognition; Bayesian inference; machine learning

Introduction

Theoretical Background

Melodic expectation is a fundamental concept of musical cognition, referring to the cognitive processes involved when one is (often spontaneously) predicting the next notes in a melody. Existing research in musical cognition has explored different aspects of melodic expectation including its similarity to linguistic prediction (Fogel, Rosenberg, Lehman, Kuperberg, & Patel, 2015), the cognitive neuroscience of sound expectancy (Koelsch, Gunter, Friederici, & Schröger, 2000), and its structural analysis using mathematical theories (e.g., Gjerdingen & Narmour, 1992).

What computations underlie our perception of melodies and melodic expectation? Existing work considers two possibilities in largely non-overlapping literature: melody perception as a probabilistic process (Temperley, 2008) and the existence of "musical forces" that influence melodic expectation (Larson, 2004). While these two theories may not necessarily coexist currently, their existence presents a promising opportunity to combine both into a single computational model of melodic expectation. We hypothesize that a model that incorporates both probabilistic melody perception and musical forces can explain patterns of human melodic expectation.

Probabilistic approaches have been extensively considered for understanding musical cognition (Temperley, 2007). Most notably, Temperley (2008) proposed a probabilistic model of melody perception where a melody's structure is inferred from its "surface", or the observed notes. Temperley focuses on 'key' as the underlying structure of a melody, but structure can also include meter or "other musical information". The general idea of the model is that listeners of music use "surface processes" or measurements, which include pitch identification, error detection, and expectation, to infer the most probable structure. Through data analysis of a corpus of European folk melodies, Temperley also found that the notes in a melody can be represented as a normal distribution centered around a central pitch, the center of the tonal range. (This observation informs how we build the prior distribution of notes in our probabilistic model.)

The research on musical cognition often observes temporal predictions as a core element, much like the predictive processes implicated in sentence completion (Feld & Fox, 1994). However, existing modeling work in melodic expectation takes primarily a statistical approach, learning conditional distributions of melodic transformations based on existing corpora of melodies. Examples include the Temperley model (Temperley, 2008), the expectation network model (Verosky & Morgan, 2021), and the information dynamics of music model (Pearce, 2018). In other cases, for example, in rhythm perception, the temporal distribution is marginalized over (Fram & Berger, 2023), still focusing on the statistical properties of notes.

A different line of work directly engages to characterize the nature of these temporal predictions, suggesting that melodic expectations may come from "musical forces". Larson (2004) discusses a set of concepts, in analogy to forces between physical objects, as the drivers of our melodic expectations. These include gravity, magnetism, and inertia, which are meant to metaphorically reflect our "experience of physical motions" (Larson, 2004). Listeners (but not necessarily players) of music systematically expect completions of musical phrases where the aforementioned musical forces dictate the "auralized traces", or auditory representation of music notes. In other words, we expect melodies to be completed and we expect musical forces to be at play in these completions. Gravity is the tendency for an unstable note to descend, magnetism is the tendency for an unstable note to move towards the nearest stable pitch, and inertia is the tendency of a melody to continue in the direction it was moving in before the current note. These individual forces come together to create a cumulative force which influences how we perceive note-to-note transitions in a melody. Consider how some emotional songs may feel "heavy" or how some vocal runs start low and gradually climb in pitch. Unlike the probabilistic models, Larson's proposal of musical forces, and similar conceptual frameworks (Margulis, 2005), remain

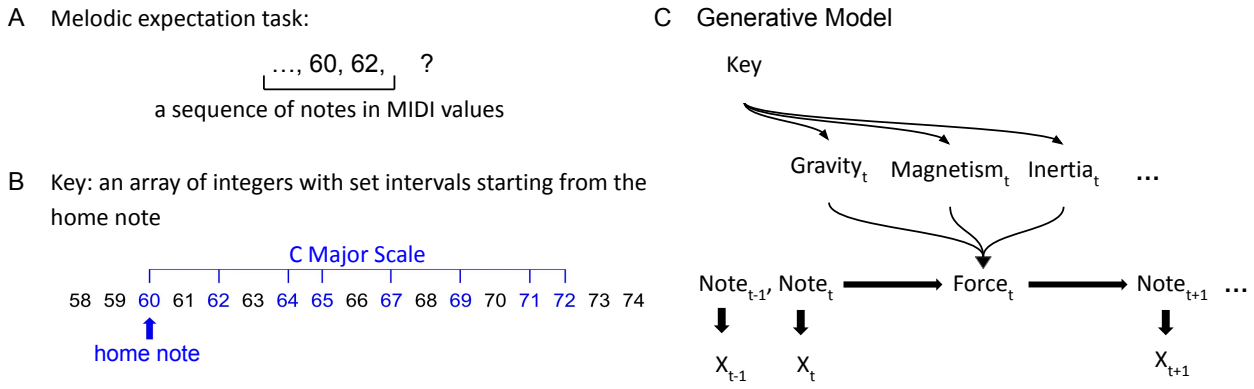

Figure 1: (A) Melodic expectation task involves predicting the final note from a "melodic stem" — a sequence of notes except the final note. We represent notes using the MIDI notation. (B) One part of the structure in the generative model is the musical key. We illustrate key using the example of C major scale. We incorporate a prior over the 15 major keys in the generative model. (C) In addition to the key, the generative model incorporates "musical forces" –the forces due to gravity, magnetism, and inertia– calculated using the previous note (the previous two notes in the case of inertia). These forces are then summed and applied to the previous note to arrive at the next note.

non-computational.

Larson's theory, that we perceive these musical elements in terms of gravity, magnetism, and inertia, has striking parallels to more modern work in cognitive science which proposes that humans have an "intuitive physics engine" (Battaglia, Hamrick, & Tenenbaum, 2013). This is a theory of how we can make fast and often accurate inferences about the physical world, including how objects move and interact. This engine runs on probabilistic simulations which allow us to make these inferences even when all information is not necessarily observed. We wonder if this proposal extends to the musical world and how it can be implemented.

Proposal

We propose a model of melodic expectation that combines both Temperley's work on probabilistic inference of melodic structure and Larson's work on musical forces. This model relies heavily on a melody's key as its underlying structure. It selects each note based on the sequence of notes before it and a calculation of each musical force. Our goal is to algorithmically represent how melodies and melodic expectation are represented in the mind using a probabilistic and physics-based approach.

To solve this inference problem, we first created a generative model that solves a sequence of inference problems where the output of one application is the input of the next. The nature of this inference problem invites the utilization of a sequential Monte Carlo method called particle filtering (Elfring, Torta, & Van De Molengraft, 2021). Once the generative process is established, it is called within a particle filter to generate a series of melodic sequences. At each time step, each sequence is weighed by its likelihood to be correct with respect to an observed note, and resampled according to these weights for the next time step. We evaluate the model by comparing its predictions after the final note (the final note in an observed sequence) to that of human participants in melodic expectation tasks. These behavioral measurements come from an existing study (Morgan, Fogel, Nair, & Patel, 2019). We find that the model's predictions correlates with average human performance. We also analyzed outputs from the full and ablated versions of the model to assess to the extent to which the musical forces we considered in our hypothesis are supported by the data.

Computational Model

Overview of the Generative Model

To understand the representations behind melodic expectation in computational terms, we first establish it as a generative process. Using the two aforementioned theories from the literature, we created a function that generates melodies in a randomly chosen major key with an initial note chosen from a normal distribution centered around the home note of that key. We implemented this model and our inference procedure in Julia using Gen.jl, a state-of-the-art probabilistic programming system (Cusumano-Towner, Saad, Lew, & Mansinghka, 2019).

Before constructing the model, we transformed musical pitches into numerical values using the MIDI note value system where C4 (middle C) is equal to 60 and each semitone up or down from C4 is +1 or -1. Not only does this method allow us to accurately and consistently represent music notes in machine-readable form, we are also easily able to determine the proximity between the model's performance and the human performance.

Following Temperley (Temperley, 2008), we can pose melodic perception as probabilistic inference of musical

structure from surface

$$P(structure|surface) = \frac{P(surface|structure)P(structure)}{P(surface)}$$

where surface is a noisy measurement of a note and the structure includes the key and the actual note played.

Crucially, here, we hypothesize that melodies follow a probabilistic "intuitive physics" like structure (Sanborn, Mansinghka, & Griffiths, 2013; Battaglia et al., 2013):

$$Note_{t+1}|Note_t \sim \mathcal{N}(Note_t + F, \sigma)$$

where F is the cumulative force that moves the previous note $Note_t$ into the current note $Note_{t+1}$. To characterize this force, we adopt Larson's theory (Larson, 2004):

$$F = w_G G + w_M M + w_I I$$

where G, M, and I are values that represent the influence of the individual forces of gravity, magnetism, and inertia, and w_G, w_M, and w_I are weights assigned to each force. $Note_{t+1}$ is calculated by drawing a value from a normal distribution centered around $Note_t + F$ with noise parameter σ. Given an initial note and the key, the generative model aggregates musical force at each time step, and applies it to the current note to arrive at the next note. We illustrate various samples from this generative process in Fig. 2.

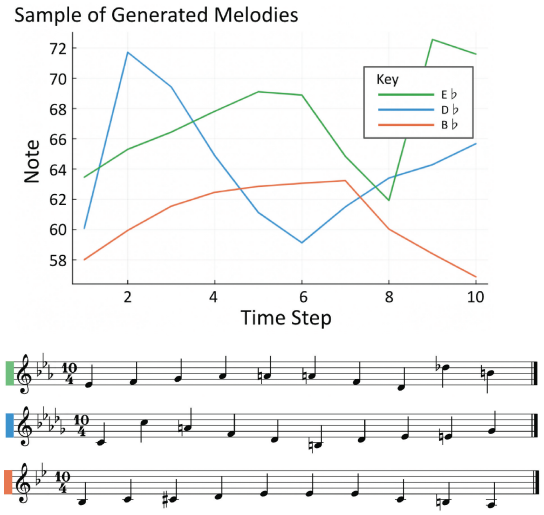

Figure 2: Three melodies sampled from the generative model. We draw a key and initial note at random and then use the temporal kernel of "musical forces" implemented in our generative model to unfold the remaining notes.

Then melodic expectation takes the form of the following posterior under this generative model:

$$P(Key, Note_{1:T}|x_{1:T}) \propto \quad (1)$$
$$P(Key)P(Note_1)\prod P(Note_{t+1}|Note_t)P(x_{t+1}|Note_{t+1})$$

where $P(Key)$ is a uniform prior over notes and $P(Note_1)$ is a normal distribution centered around the corresponding key's home note.

This posterior incorporates the theories of Larson and Temperley by using a selected key and musical forces as the structure of melody and probabilistically relating this to observations (surface). We note that our implementation of either theory is not meant to exactly match what was originally in the literature. For instance, Larson's original theory stated that gravity and inertia should be represented by assigning scores that denote the degree to which a pattern gives in to each force. Temperley's original model also had a key-finding feature whereas, during inference, we assume that the key is given. We seek to preserve the higher-level ideas of each theory but also implement them such that they work together. Therefore, despite these deviations, the generative model preserves the core elements of each theory.

Structure: Key and Musical Forces

We now specify how in the generative model the key and musical forces are implemented.

Key To reflect Temperley's idea that the underlying structure of a melody is its key, the generative model incorporates a single one-octave scale per each of the 15 major keys. Each of these scales starts at the home note (i.e., the "tonic" in music theory, typically the first note on the scale) with the rest of the notes on the scale following the set intervals for major scales. A broad range of evidence suggests that key, including the systematic spacing of notes in a scale, drives our intuitions about correct and incorrect notes (e.g., McDermott & Oxenham, 2008).

The scale establishes which notes are stable and guides the calculation of musical forces as the generative process unfolds. In each of the following musical force functions, the notes of a scale of a given key is the fundamental element. The key, represented as an array of integers, and the current sequence of notes are inputs to each musical force function.

Force due to gravity Gravity, G, is calculated by first finding the difference between the current note and the home note of the scale. The difference is then multiplied by -1 if the current note is higher than the home note in the scale. If the current note is lower than the home note in the scale, then gravity is set to 0. This implementation allows the effect of the gravity to be proportional to the distance between the current note and the start of the scale. In practice, this means that higher notes are more heavily influenced by gravity and notes lower than the start of the scale are not influenced by gravity. (We interpret the -1 as the gravity constant and the distance as the mass.)

Force due to magnetism Magnetism, M, is calculated by taking the inverse square of the distance in semitones to the nearest note in the scale (i.e., stable note), d_{to}, and subtracting the inverse square of the distance in semitones to the closest stable note in the other direction, d_{from}. This formula influ-

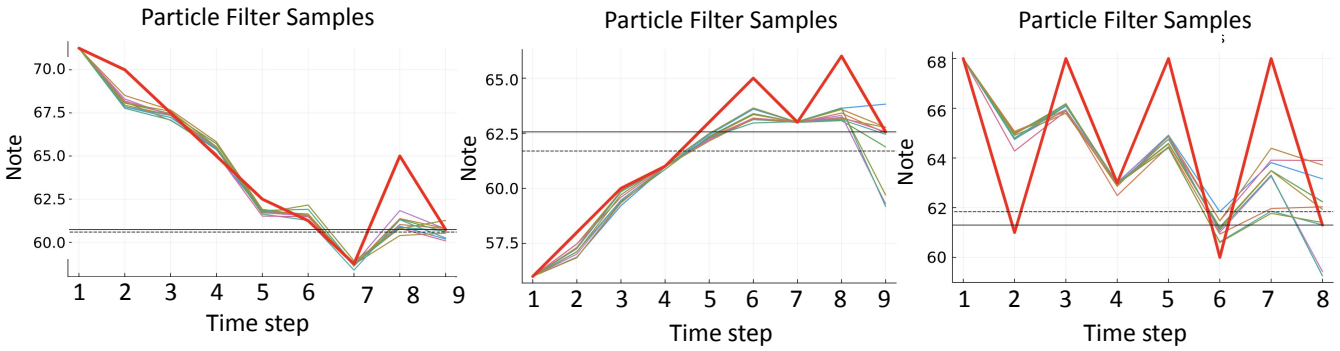

Figure 3: Posterior samples from the particle filter for three typical input melodic stems. The red line represents the actual sequence of observed notes with the mean of participant responses added as the last note. The solid and dashed black lines represent the mean of human predictions and model predictions, respectively, of the final note to complete the melody.

ences notes to move towards stable notes is directly adapted from Larson's existing theory:

$$M = 1/{d_{to}}^2 - 1/{d_{from}}^2$$

Force due to inertia Calculating inertia, I, requires input of the current *and* previous note to determine the direction of the melody. If the current note is the first note in the melody, then inertia is set to 0. Inertia is also set to 0 if the current note is below the start of the scale (i.e., the home note) to guide the melody toward the range of the scale. Otherwise, inertia is set to a positive or negative constant depending on if the current note is above or below the previous note.

Additionally, the generative model allows for a "bounce" to occur with probability 0.1 which negates the inertia value. We implemented this to account for the dominating force of gravity and the periodic rises of pitch in a melody.

In summary, this structure of key and forces form a temporal kernel that calculates each musical force for a given note and the key. Fig. 1 shows a schematic of this generative process, transitioning from previous notes to the next. First, separate functions calculate contributions of gravity, magnetism, and inertia for the current note using the key and current sequence of notes as input (Fig. 1A & 1B). We introduce stochasticity to the outputs of these functions using a normal distribution to perturb these outputs. Respective weights are applied to each force. All forces are summed and applied to the current note. Lastly, the resulting value is used as the center of a normal distribution from which the next note is drawn.

Inference

We implemented a particle filter to approximate the posterior in Equation 1. Particles are collections of samples that represent possible states of an ongoing melody (i.e., the current note) as well as their associated posterior probability scores. At each time step, the model makes predictions by resampling the current collection of particles (according to their posterior scores) and updating the weights based on newly observed data, assigning higher weights to particles that have a higher likelihood of occurring. Ultimately, we use these inferences to draw a last note prediction from the generative model (the posterior predictive distribution), which we compare to human predictions.

Inference takes as input sequentially presented notes that together make up a "melodic stem" (i.e., all notes in a melody except the last note). For each input sequence, we run our particle filter with 100 particles and return 100 unweighted samples from the posterior for further analysis (to compute a prediction over the last note of the stem).

For computational convenience, the particle filter is initialized with the ground truth key of the melodic stem as well as the first note in the melody. (Instead, this initialization can be carried out using a data-driven proposal.) At each following time step, the particle filter state evolves by possibly resampling and updating the weights based on the observed data, which only includes the current note.

Fig. 3 shows several runs of the particle filter on three example input melodic stems ranging in length from 7-8 notes. We simulate the prediction of a final note (the last time step) by simulating the temporal kernel of the generative model on the posterior samples.

Behavioral Task and Model Simulation

An existing experiment from "Statistical learning and Gestalt-like principles predict melodic expectations" (Morgan et al., 2019) parallels the model's task in human participants. We use the behavioral dataset from this to test our model and its ablations. This experiment presented

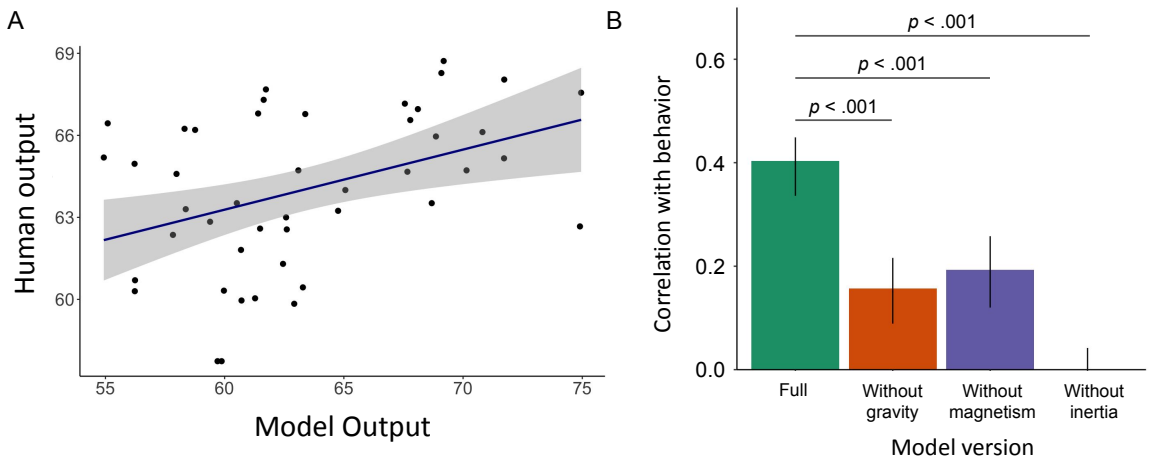

Figure 4: Comparisons between model and behavior. (A) Model vs. human accuracy for full model. (B) Linear correlation between human and model accuracy across full and ablated models. Error bars/regions depict 95% CI.

melodic stems, i.e., the first 6-9 notes of a melody, to 25 participants and asked them to sing the next note. The melodic stems were composed from another study (Fogel et al., 2015). To the best of our knowledge (based on what we can infer reading these papers), we note that the design of the behavioral study and the melody stems used in it were not in any way designed to elicit or be based on Larson's theory of musical forces. Participants were asked to complete 45 distinct melodic stems, all of varying lengths and keys. We coded the melodic stems and the respective mean of participant responses into MIDI note values.

We simulated our particle filter on each of the 45 melodic stems and obtained a posterior predictive distribution on the final note of each. We compare the average predicted note by our model (average of 100 samples) to average human predictions. In our simulations, we set the weights on the forces as $w_G = 0.3, W_I = 0.4$, and $w_M = 12.5$; these values were chosen to ensure that after scaling by these weights, each force contributed in a similar numerical range. Prior to multiplying with these weights, we apply a small Gaussian noise to each force component, with a standard deviation of 1 for gravity, 0.005 for magnetism, and 0.75 for inertia. (Again, these standard deviations were chosen to roughly match the dynamic range of each force component.) Finally, we set the observation noise $\sigma = 0.025$.

Model vs. Behavior Comparisons

We performed a linear regression between the average model predicted notes and human data (Fig. 4A). We found a moderate, positive correlation, $r(43) = 0.40$, $R^2 = 0.162$, $p = .006$, between the model and humans in predicting a final note for each melody. This result suggests that the model's generative structure captures some of the variance in people's melodic expectations.

We also conducted ablation studies to explore the impact of individual musical forces in the generative model. By developing alternative models that incorporate only a subset of two out of the three forces, we could then analyze how the absence of a specific force influenced the model's performance. Fig. 4B shows a visualization of the correlations between the different model versions. Compared to the original model, models that only included magnetism and inertia ("without gravity"; $r(43) = .16, p < .001$), gravity and inertia ("without magnetism"; $r(43) = .19, p < .001$), or gravity and magnetism ("without inertia"; $r(43) = -.02, p < .001$) resulted in a lower correlations to behavior (using direct bootstrap hypothesis testing with bootstrapping participants 1000 times). In other words, removing any one of the musical forces results in a less accurate model of human melodic expectation.

Discussion

We presented a model of human melodic expectations that combines two, previously separate, accounts of this process: Temperley's theory of the probabilistic inference of structure from surface and Larson's theory of musical forces as the drivers of melodic expectations. (We note that Larson in his seminal work has speculated about full implementations of "computer models" of his theory.) Our model appropriately combines these two distinct perspectives. The generative model unfolds melodies using the combined structure of a key and the algorithmic influences of musical forces. Inference under this generative model typically well explains the melodic stems. The predictive distribution under these inferences correlate with average human responses, lending support for our intuitive physics approach to modeling melodic expectation. The results of our ablation studies affirm that all of the musical forces we considered are necessary in modeling melodic expectation in the context of our model.

This model may relate to existing hypotheses about Bayesian surprise and music preference (Itti & Baldi, 2009;

Sarasso et al., 2022; Stupacher, Matthews, Pando-Naude, Foster Vander Elst, & Vuust, 2022). Music has evolved from one-layer (monophonic) Gregorian chants in the 9th century to hundred-layer (polyphonic) electronic music today (Harris, 2024). Perhaps what drives our musical preferences are patterns that *do* deviate from probabilistic or algorithmic norms — perhaps in ways that themselves remain predictable. We consider the "bounce" variable in our generative model (implemented within the inertia force) as one such candidate. In the model, the bounces occur against "invisible" surfaces, affording some degree of algorithmic surprise.

The moderate performance of our model prompts consideration of both its strengths and limitations. In some cases, inferences under the generative model do not accurately recapitulate the observed notes, indicating that the generative model is not sufficiently expressive. The current implementation of the model also does not infer key (and the initial note). A fuller account should address these issues, for example by using data-driven proposals or simply including their inferences in the particle filter.

Another limitation of the model is that musical forces for each note are calculated independently of the longer context of the previous notes; all of the forces consider only the previous step (i.e., Markovian as is typical in physics simulation) with the exception of inertia which considers the previous two steps. But in music (and in physics) there can be temporally elongated, perceptually relevant structures (e.g., a pattern of bouncing of a ball). Future versions of this model should account for the entire melody when calculating a next note. Ideas from other domains of musical cognition (Fram & Berger, 2023) and more generally structured representations in cognitive science should be of interest (Sablé-Meyer, Ellis, Tenenbaum, & Dehaene, 2022).

Lastly, our model represents melodies solely in terms of pitches (MIDI notation is a numerical mapping of pitches) and time steps of equal length. Music is not solely made of observed notes. It also includes rhythm, meter, dynamics, and even qualitative aspects such as emotion and articulation. Extending the model with these more complex aspects of music is of great interest. Similarly, considering different types of keys (i.e., minor or blues) or styles of music (i.e., Eastern or Arabic) could enhance the model's scope of performance and give better insight into the mechanism behind melodic expectation in the mind.

In conclusion, our model introduces a new approach to modeling melodic expectation by combining two distinct theories: probabilistic melody perception and musical forces. The model performed moderately well at predicting human performance in melodic expectation tasks. Where the model does not perform well, there is opportunity to explore different features of music that may contribute to this process of "intuitive physics" or the idea of Bayesian surprise in musical cognition. We hope that future work will build upon our model to further our understanding of musical cognition and melodic expectation.

References

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013, November). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, *110*(45), 18327–18332. doi: 10.1073/pnas.1306572110

Cusumano-Towner, M. F., Saad, F. A., Lew, A. K., & Mansinghka, V. K. (2019). Gen: A general-purpose probabilistic programming system with programmable inference. In *Proceedings of the 40th acm sigplan conference on programming language design and implementation* (pp. 221–236). New York, NY, USA: ACM. doi: 10.1145/3314221.3314642

Elfring, J., Torta, E., & Van De Molengraft, R. (2021, January). Particle Filters: A Hands-On Tutorial. *Sensors*, *21*(2), 438. doi: 10.3390/s21020438

Feld, S., & Fox, A. A. (1994). Music and language. *Annual review of anthropology*, *23*(1), 25–53.

Fogel, A. R., Rosenberg, J. C., Lehman, F. M., Kuperberg, G. R., & Patel, A. D. (2015, November). Studying Musical and Linguistic Prediction in Comparable Ways: The Melodic Cloze Probability Method. *Frontiers in Psychology*, *6*. doi: 10.3389/fpsyg.2015.01718

Fram, N. R., & Berger, J. (2023, December). Syncopation as Probabilistic Expectation: Conceptual, Computational, and Experimental Evidence. *Cognitive Science*, *47*(12), e13390. doi: 10.1111/cogs.13390

Gjerdingen, R. O., & Narmour, E. (1992, December). The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model. *Notes*, *49*(2), 588. doi: 10.2307/897927

Harris, J. (2024, January). The evolution ofpolyphony: From gregorian chants to modern music. *Breve Music Studios*.

Itti, L., & Baldi, P. (2009, June). Bayesian surprise attracts human attention. *Vision Research*, *49*(10), 1295–1306. doi: 10.1016/j.visres.2008.09.007

Koelsch, S., Gunter, T., Friederici, A. D., & Schröger, E. (2000, May). Brain Indices of Music Processing: "Nonmusicians" are Musical. *Journal of Cognitive Neuroscience*, *12*(3), 520–541. doi: 10.1162/089892900562183

Larson, S. (2004, June). Musical Forces and Melodic Expectations: Comparing Computer Models and Experimental Results. *Music Perception*, *21*(4), 457–498. doi: 10.1525/mp.2004.21.4.457

Margulis, E. H. (2005, April). A Model of Melodic Expectation. *Music Perception*, *22*(4), 663–714. doi: 10.1525/mp.2005.22.4.663

McDermott, J. H., & Oxenham, A. J. (2008, August). Music perception, pitch, and the auditory system. *Current Opinion in Neurobiology*, *18*(4), 452–463. doi: 10.1016/j.conb.2008.09.005

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

Pearce, M. T. (2018, July). Statistical learning and probabilistic prediction in music cognition: mechanisms of stylistic enculturation. *Annals of the New York Academy of Sciences*, *1423*(1), 378–395. doi: 10.1111/nyas.13654

Sablé-Meyer, M., Ellis, K., Tenenbaum, J., & Dehaene, S. (2022). A language of thought for the mental representation of geometric shapes. *Cognitive Psychology*, *139*, 101527.

Sanborn, A. N., Mansinghka, V. K., & Griffiths, T. L. (2013). Reconciling intuitive physics and newtonian mechanics for colliding objects. *Psychological review*, *120*(2), 411.

Sarasso, P., Barbieri, P., Del Fante, E., Bechis, L., Neppi-Modona, M., Sacco, K., & Ronga, I. (2022, December). Preferred music listening is associated with perceptual learning enhancement at the expense of self-focused attention. *Psychonomic Bulletin & Review*, *29*(6), 2108–2121. doi: 10.3758/s13423-022-02127-8

Stupacher, J., Matthews, T. E., Pando-Naude, V., Foster Vander Elst, O., & Vuust, P. (2022, August). The sweet spot between predictability and surprise: musical groove in brain, body, and social interactions. *Frontiers in Psychology*, *13*, 906190. doi: 10.3389/fpsyg.2022.906190

Temperley, D. (2007). *Music and probability*. Mit Press.

Temperley, D. (2008, March). A Probabilistic Model of Melody Perception. *Cognitive Science*, *32*(2), 418–444. doi: 10.1080/03640210701864089

Verosky, N. J., & Morgan, E. (2021, October). Pitches that Wire Together Fire Together: Scale Degree Associations Across Time Predict Melodic Expectations. *Cognitive Science*, *45*(10), e13037. doi: 10.1111/cogs.13037