



Surprisal, Liking, and Musical Affect

Noah R. Fram^(✉)

CCRMA, 660 Lomita Drive, Stanford, CA 94305, USA
nfram@ccrma.stanford.edu
<http://ccrma.stanford.edu>

Abstract. Formulation and processing of expectation has long been viewed as an essential component of the emotional, psychological, and neurological response to musical events. There are multiple theories of musical expectation, ranging from a broad association between expectation violation and musical affect to precise descriptions of neurocognitive networks that contribute to the perception of surprising stimuli. In this paper, we propose a probabilistic model of musical expectation that relies on the recursive updating of listeners' conditional predictions of future events in the musical stream. This model is defined in terms of cross-entropy, or information content given a prior model. A probabilistic program implementing some aspects of this model with melodies from Bach chorales is shown to support the hypothesized connection between the evolution of surprisal through a piece and affective arousal, indexed by the spread of possible deviations from the expected play count.

Keywords: Affect · Entropy · Music · Perception · Probabilistic programming · Surprisal

1 Introduction

Meyer's seminal work, *Emotion and Meaning in Music* (1956) formed the basis of a widely held conceptualized of musical expectation as a cognitive process in a broadly Bayesian mold. Meyer described the crucial role that expectation violations play in constructing musical affect, thus implying that expectation mediates between musical perception and emotional response. Other scholars soon incorporated this claim into their own approaches to a cognitive analysis of musical expectation and surprisal. Narmour's implication-realization model, which was crafted as an alternative to the Schenkerian orthodoxy in music analysis, represents the first such theory to explicitly draw on Gestalt principles of perception (Narmour 1989). It would later be amended to include distinctions between “top-down” (or semantic) and “bottom-up” (or phonetic) expectations (Narmour 1991).

Since then, theorists have also begun representing musical expectation using analogies to physical processes or established measures from information theory. Margulis's delineation of three distinct kinds of musical tension (Margulis 2005)

and Larson's representation of melodies as obeying the musical equivalent of gravity, inertia, and magnetism (Larson 2002; 2004) are indicative of this trend, and offer ample room for computational exploration. Other researchers have taken more directly computational approaches. For instance, Temperley applied Bayesian probability theory directly to describe the cognitive process by which humans predict underlying structural features from musical surfaces (Temperley 2004) and devised a musical version of Shannon entropy (Temperley 2007), and Margulis and Beatty investigated its usefulness as an analytic tool (Margulis and Beatty 2008).

Recently these two general approaches have converged. Agres, Abdallah, and Pearce demonstrated that information-theoretic measures such as those developed by Margulis and Temperley are related to musical memory (Agres et al. 2018). Bayesian modelling has been applied, with varying amounts of success, to neurological correlates of surprisal, notably the mismatch negativity (MMN) (Lieder et al. 2013) and early right anterior negativity (ERAN) (Broderick et al. 2018). In addition, there is a theorized inverse-U relationship between musical complexity and reported preference (Güçlütürk et al. 2016; McMullen and Arnold 1976), and recent research has led to descriptive models of this correlation (Agres et al. 2017). However, a generative model capable of directly predicting behavioral correlates of musical liking or preference based on measures of musical expectation and surprisal is lacking.

This study aims to fill precisely this gap. Its specific aims are twofold: to devise a Bayesian model on a cognitive level of analysis that will capture both the underlying construction of musical expectations and how those expectations relate to behavior surrounding liking; and to determine what measures of liking or preference are feasible and useful for training and assessing such a two-level model.

2 Model Foundations

Extant theories suggest four primary characteristics of musical expectation: it is recursive, in that the match or mismatch between an expectation and an observed event is used to construct future expectations; it is dynamic, in that the mental model used to make predictions is not fixed; it is based on musical tendencies, as opposed to physical or cultural pressures; and it is related to information content. Furthermore, as indicated both by neurophysiological research and theories of musical semiotics, music is perceived as a sequence of events, rather than a continuous stream of sound. Given the symbolic nature of musical perception, two approaches to modelling music expectation – Huron's ITPRA model from his book *Sweet Anticipation* (2006), and Agres, Abdallah, and Pearce's two-level model of informational expectation (2018) – offer advances in this area of scholarship that are particularly relevant to the present project.

2.1 The ITPRA Model

Huron (2006) theorizes that musical expectations are constructed in five overlapping stages that are organized into two epochs: the pre-event epoch, consisting of the imagination and tension stages; and the post-event epoch, consisting of the prediction, reaction, and appraisal portions. This structure is shown in Fig. 1. First, listeners imagine possible outcomes of an event. The approach of the imagined event leads to an increase in tension, as the listener waits to discover whether or not their prediction is valid. Once the event occurs, there is an immediate assessment of the accuracy of the prediction and an immediate reactive assessment, followed by a more thorough appraisal and adjustment of the predictive parameters.

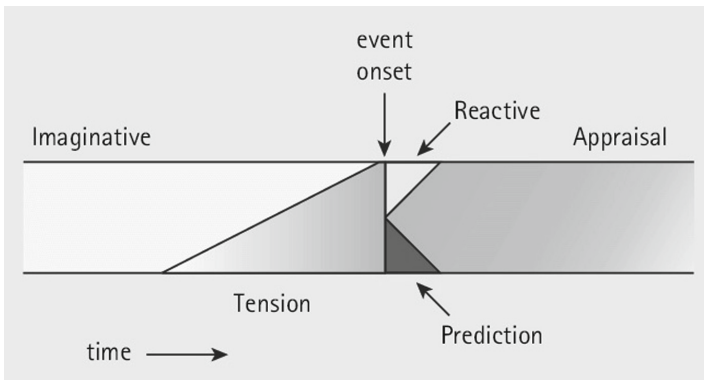


Fig. 1. The ITPRA model of musical expectation

This approach highlights both the affective nature of musical expectation proposed by Meyer (1956) and the plasticity of prediction. To Huron, expectation is based more on the listener's beliefs about musical structure than the music's actual structure; to borrow from Lerdahl (1988), expectation has more to do with the listener's grammar than with the composer's. Therefore, Huron's construction of musical expectation is fundamentally Bayesian, as it entails a listener making non-deterministic inferences about a piece of music and, by extension, about its creator.

2.2 Probabilistic Expectation

As described by Margulis and Beatty (2008), entropy and other information-theoretic measures have become increasingly popular in music analysis. These measures were shown by Agres et al. (2018) to be related to musical expectation. Specifically, they showed that a probabilistic model relying on measures of information content, coding gain, and predictive information could accurately simulate memory for musical sequences. In addition, they demonstrated the benefits

of training both a long-term (top-down) and a short-term (bottom-up) predictive model simultaneously, and using both to generate an expectation.

This approach is not necessarily resource-rational, and does not reflect abstract perception of musical types. But it does represent an important advance in modelling musical expectation. Most notably, it recasts the problem of building expectation, which had previously been a question of pure prediction, as an optimization problem on information-theoretic measures of musical structure. This approach is potentially much more efficient, and lends itself very well to expansions into more ecologically-valid models of musical expectation.

3 Surprisal Model of Musical Liking

Huron's (2006) model is explicitly broken into pre-event and post-event epochs, but the post-event epoch entails two distinct processes: the immediate affective response, represented by the Prediction and Reaction segments; and the reassessment of the listener's mental representation of the piece of music, represented by the Appraisal portion. Although the ITPRA construction entails five separate cognitive processes, this implies a functional division into an expectation formulation portion, an affective response portion, and a prediction assessment portion. If true, then a functional model of the ITPRA framework should explicitly encode that division in such a way that the output of the affective response mirrors empirically assessed correlates of musical affect.

To this end, a three-stage model of musical expectation was constructed. This surprisal model of musical liking relies on Temperley's construction of cross-entropy to generate both a short-term, extrinsic (from the perspective of the listener) model of musical expectation and a long-term, intrinsic collection of models for expectations on inferred musical types.

3.1 Mathematical Formulation

Extrinsic Model. For a naïve listener, the initial condition is assumed to be generated from a flat Dirichlet distribution:

$$\mathbf{p} = (p_0, \dots, p_{k-1})$$

$$\sum \mathbf{p} = 1 \tag{1}$$

Here, \mathbf{p} is any element of the open standard $k - 1$ -simplex.

$$P_0(x|\mathbf{p}) = \prod_{i=0}^{k-1} p_i^{[x=i]} \tag{2}$$

This initial distribution is used in the pre-event epoch to generate an intrinsic musical expectation. The initial (empty) sequence of musical events is denoted

D_0 , and the sequence of the first n events is D_n . At each musical event, the listener makes an observation and assesses its surprisal given this expected intrinsic distribution:

$$H_{n-1}(D_n, P_{n-1}) = -\frac{1}{n} \log P_{n-1}(D_n | D_{n-1}), \quad n = 1, 2, \dots \quad (3)$$

The surprisal vector $\mathbf{H} = (H_0, \dots, H_{k-1})$ will be used to generate a liking rating at the end of the process. During the post-event epoch, the probability distribution is updated within the bounds of a malleability tolerance r , to prevent over-fitting early in the process:

$$P_n = \arg \min_{|P - P_{n-1}| < r} H_{n-1}(D_n, P) \quad (4)$$

Here, $|P - P_{n-1}|$ is determined by a distance measure on distributions such as the Kullback-Liebler divergence, and r governs how much the intrinsic distribution can change in any time-step.

Intrinsic Model. This intrinsic process runs in parallel to an extrinsic, long-term process, which operates according to similar principles. The primary mathematical difference between the extrinsic and intrinsic models is that while the intrinsic model constructs an expectation based on observed data from *within* a piece of music, the extrinsic model predicts which type or category of music a piece belongs to and constructs an expectation based on an archetypal distribution for that type. Since pieces are sorted into categories, if the observed dataset consists of only one piece, there is by definition only one available type. As a result, the distribution over the number of categories is initialized after encountering a second piece. For a naïve listener, this distribution is sampled from a Dirichlet distribution where, for piece k with the distribution over n categories at time step $k - 1$ denoted $\mathbf{q}_{k-1} = (q_0, \dots, q_{n-1})$, the following holds:

$$\begin{aligned} \alpha &= (\alpha_0, \dots, \alpha_n) \\ \alpha_i^{[x < n]} &= \frac{1}{(n-1)q_i}, \quad \alpha_n = 1 \\ \mathbf{q}_k &= \text{sample}(\text{Dir}(\alpha)) \end{aligned} \quad (5)$$

This construction maintains a bias toward the same number of categories as in the previous step, with a neutral possibility of adding another category. The distribution over the number of categories is optimized after each piece of music, so the expected number of categories $E(\mathbf{q}_k)$ remains constant while a piece of music is playing.

Each category has an associated probability distribution, denoted Q_i . At each time step, the listener selects the most likely category C_i given the musical context by minimizing cross-entropy:

$$i = \arg \min_{i \in 0, \dots, E(\mathbf{q}_k)} H_{n-1}(D_n, Q_i) \quad (6)$$

After a musical piece concludes, the number of categories is inferred. If a category is added, the new piece is assumed to be the seed. If there are fewer categories than before, the most similar pair (by Kullback-Leibler divergence) are combined:

$$\mathbf{P} = (p_0, \dots, p_j), \mathbf{Q} = (q_0, \dots, q_j)$$

$$\mathbf{R} = \left\{ \left(\frac{p_i + q_i}{2} \right) \right\}_{i \in 0, \dots, j} \quad (7)$$

After the categories are set, the last categorization choice C_i is preserved and the corresponding distribution is updated using the last short-term distribution P to minimize distribution cross-entropy:

$$Q_i^* = \arg \min_{|Q - Q_i| < R} \left| \sum_{x \in X} Q_i(X = x) \log Q(X = x) \right| \quad (8)$$

This intrinsic model reaches a series of locally stable sets of categories, which are only perturbed by unusual or uncategorizable pieces of music. However, the terms of genres drift over time, which reflects the fundamentally dynamic nature of genre proposed by contemporary theorists such as Brackett (2016), Sturm (2014), Frow (2005), and Bhatia (2004; 2016).

Musical Resolution. Since D_n is a sequence of events, where “events” are arbitrarily defined, this structure allows for variation in perceptual resolution depending on what kind of events are analyzed. Some *a priori* possibilities include onsets (isochronous or otherwise), beats, harmonic shifts, phrases, or melodic cycles. In human cognition, these “events” are likely even more abstract, as they are learned subdivisions of continuous auditory signals.

Liking. The relationship between musical affect and liking is complicated by the strong connection between preference for certain categories or genres of music and exemplars of those categories or genres (Rentfrow and Gosling 2007). However, the inverse-U relationship discussed earlier indicates a relationship between surprisal and preference may emerge when controlling for other extramusical associations or characteristics of music. One possible explanation, which will be assessed in the experimental portion of this paper, is that features of the surprisal vector are related to affective arousal, but not valence. This would imply that information content and the accuracy of expectations are connected to the strength of the affective response, but not necessarily its direction.

3.2 Bayesian Formulation

This procedure is somewhat simpler to depict using a Bayesian network, such as the one in Fig. 2. In this depiction, the precise probabilistic descriptions of the inference steps are hidden in favor of showing the connections among the various

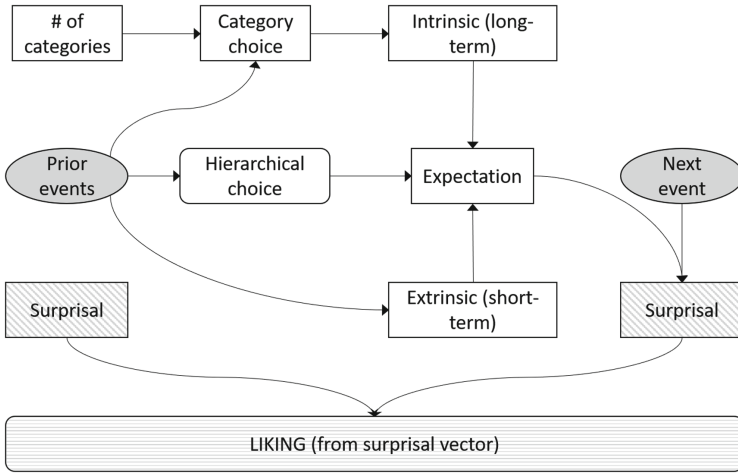


Fig. 2. Schematic of the Bayesian model connecting liking with musical surprisal.

models. Two notable features of this approach stand out: first, although the extrinsic and intrinsic models are probabilistically dependent, there is no causal dependence between the two; second, the only decision made by the listener is the hierarchical choice between the extrinsic and intrinsic predictions. Since music unfolds in time, the network in Fig. 2 represents the activity in one time-step, and does not include the process of adjusting the extrinsic model after a piece has concluded.

Since there is limited data being passed, and the resource-heavy portion occurs after the piece has concluded, this framework is likely resource-rational, especially since the only preserved information is a finite set of distributions.

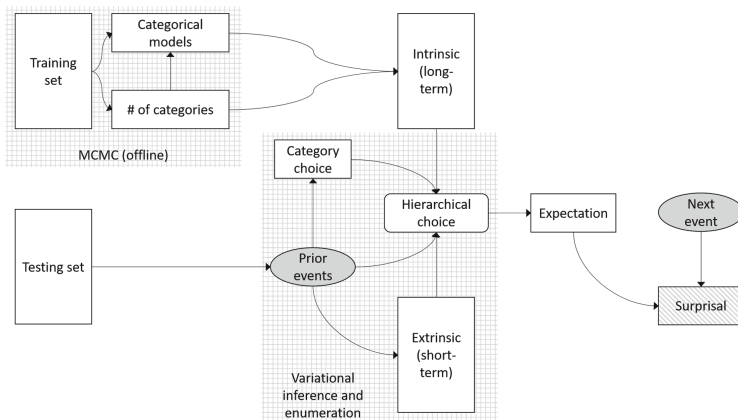


Fig. 3. Visual schematic of implementation in Python and WebPPL.

4 Assessment Methods

The explicit implementation of the surprisal model in this study, shown in Fig. 3, required several restrictions. First and foremost, rather than learn the extrinsic and intrinsic portions in parallel, the intrinsic model was implemented offline and then applied as a static set of distributions. This was to avoid determining an order in which the melodies in question were encountered. Second, the domain of “musical events” was restricted to melodic pitch onsets, rather than a more general description of musical possibilities. Third, the number of possible categories in the extrinsic model was assumed to be between 1 and 10, and the length of a musical “word” was held to two consecutive onsets when constructing the melodic language of thought. These restrictions were to conserve computational power, and do not reflect assumptions about the genuine cognitive process that this model approximates. Lastly, the intrinsic model was trained using a Markov chain Monte Carlo (MCMC) approach, although the theorized methodology is closer to variational inference. This last alteration was due to interference between the inferences on the number of categories and the distributions that characterize those categories.

The adjusted model was implemented in WebPPL using a data management and cross-validation procedure written in Python. Once finished, this process generates a series of surprisal vectors, with a new cross-entropy value generated for each melodic onset.

The probabilistic program described above was applied to a set of 371 melodies from Bach chorales, extracted from the KernScores database hosted by the Center for Computer-Assisted Research in the Humanities (CCARH) at the Stanford University Music Department (Sapp 2005). Post-hoc statistical methods, most notably hierarchical cluster analysis, were used to analyze the resulting data. These methods are not thought to reflect actual cognitive behavior, but are rather an attempt to uncover structures that are embedded in the surprisal vectors.

5 Expected Outcomes

If the ITPRA model does produce something similar to the inverse-U relationship between complexity and liking, the surprisal vectors should be connected to a measure of musical liking. Since the dataset consists entirely of Bach chorales, this should effectively control for valence effects of musical genre, and imply that the inferred categorization relates more to elements such as key or mode, meter, or intended performance venue.¹ Without making assumptions about which features contribute to higher liking values, clustering surprisal vectors using a time-warping or trajectory analysis algorithm should produce groups

¹ There are notable differences between pieces written for casual or amateur musicians and those written for professional musicians, and the rise in virtuosic compositional practice was contemporaneous with Bach’s career (Baron 1998; Lott 2015; Radice 2012).

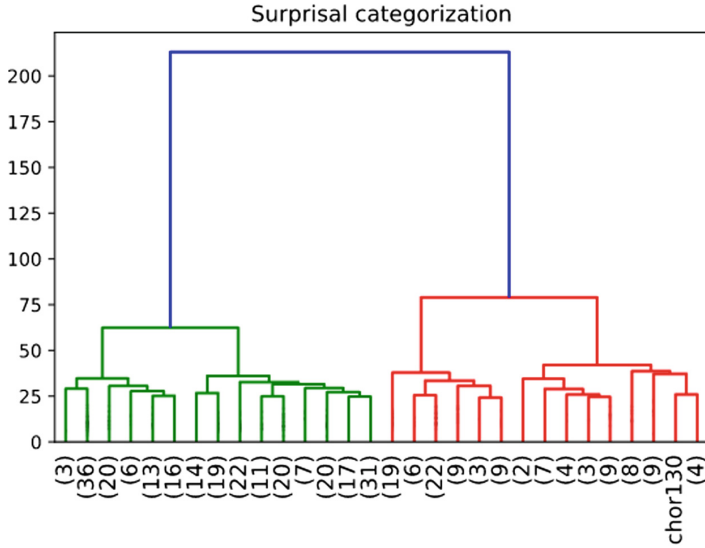


Fig. 4. Dendrogram showing clusters in the surprisal vectors for Bach chorale melodies.

of pieces with different preference ratings. These preference ratings should then be reflected in behavioral measures such as consumption or recommendation frequency.

6 Results

As the melodies are all of different lengths, the surprisal vectors were resampled to include the same number of data points. Distance between surprisal vectors was then calculated using the Keogh lower bound on the dynamic time-warping distance to preserve contour features of their development in time, and those distances were used to cluster the melodies with Ward's algorithm (Fig. 4).

This clustering approach has been used in automatic genre recognition (AGR) research, and has been shown to have a favorable accuracy over other methods when dealing with pieces of significantly different tempi (Holzapfel and Stylianou 2008). In addition, time-warping methods have been shown to ameliorate the effects of noise when searching for periodicity in time series signals (Elfeky et al. 2005), which provides further support for its validity as an analytical method here. Play counts for a subset of these chorales were extracted from Spotify. Since the play counts decay logarithmically as the album progresses, deviation from the expected play count was computed and plotted against category in the surprisal analysis (Fig. 5).

This plot indicates that categories 1 and 3 were much more concentrated around no deviation, while categories 2 and 4 were more likely to overshoot the expectation. However, an ANOVA indicated no significant shift in the mean deviation from expected play time by category.

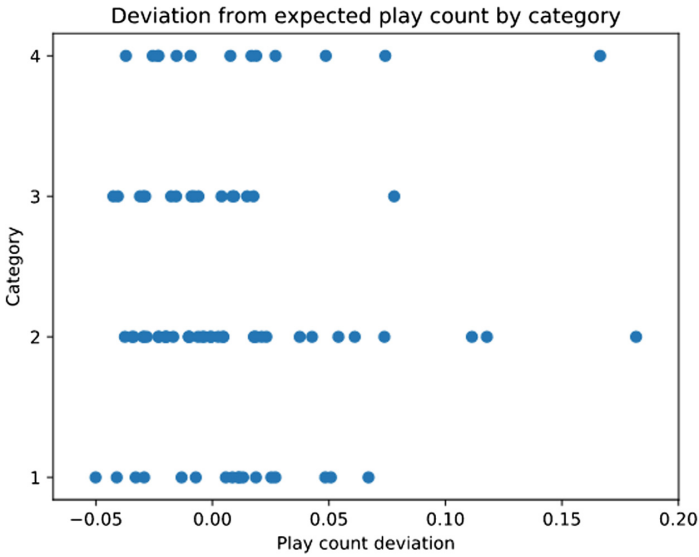


Fig. 5. Plot of $\log(\text{playCount}) - E(\log(\text{playCount}))$ against surprisal vector category.

7 Discussion

Although these data are imprecise, and the results are not statistically clear, there are a few potentially interesting conclusions that may be drawn. Most notably, these data suggest a relationship between features of the surprisal vector and the variation in liking, especially upwards variability. This specific connection is derived from the distance metric used; dynamic time-warping distance preserves similarities in trajectories that are not preserved by more standard metrics such as Euclidean distance. Such expanded variability suggests a possible link between features of surprisal’s movements in time and affective arousal, although further study is necessary for confirmation.

Most promising is that these trends follow the theoretical prediction that expectation violation is directly related to musical affect, and that this effect is visible even with the severe limitations of this particular implementation. The dataset in this case, consisting solely of soprano melodies from Bach chorales, is limited in size, does not have much internal variation, and is underconsumed in the current musical market. Similarly, using deviation from expected play counts as a proxy for liking suffers from multiple assumptions, most notably the assumptions that more-played tracks are better-liked and that deviation from a logarithmic trendline is a robust measure of deviation in play counts.

Further research requires a more complete dataset, including a larger range of musical idioms and a more thorough language of musical symbols, and a more accurate measure of liking. One possibility is to use MIDI realizations of pop songs, build symbols consisting of note simultaneities (a generalization of chords and harmonies), dynamic levels or motion, and basic instrumentation, and corre-

late the resulting surprisal categories with measures of chart performance. This would allow the model to train sequentially rather than building the intrinsic model offline, much like a real listener would.

Acknowledgments. My thanks to Noah Goodman, Ben Peloquin, Robert Hawkins, Malcolm Slaney, and Jonathan Berger for their assistance in the development and implementation of this research.

References

- Agres, K., Abdallah, S., Pearce, M.: Information-theoretic properties of auditory sequences dynamically influence expectation and memory. *Cogn. Sci.* **42**(1), 43–76 (2018)
- Agres, K., Herremans, D., Bigo, L., Conklin, D.: Harmonic structure predicts the enjoyment of uplifting trance music. *Front. Psychol.* **7**, 1999 (2017)
- Baron, J.H.: *Intimate Music: A History of the Idea of Chamber Music*. Pendragon Press, Stuyvesant (1998)
- Bhatia, V.K.: *Worlds of Written Discourse: A Genre-Based View*. Continuum International, London (2004)
- Bhatia, V.K.: Genre as interdiscursive performance in public space. In: Reiff, M.J., Bawarshi, A. (eds.) *Genre and the Performance of Publics*, pp. 25–42. University Press of Colorado, Boulder (2016)
- Brackett, D.: *Categorizing Sound*. University of California Press, Oakland (2016)
- Broderick, M.P., Anderson, A.J., Di Liberto, G.M., Cross, M.J., Lalor, E.C.: Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. *Curr. Biol.* **28**(5), 803–809 (2018)
- Elfeky, M.G., Aref, W.G., Elmagarmid, A.K.: WARP: time warping for periodicity detection. In: *Fifth IEEE International Conference on Data Mining (ICDM 2005)*, p. 8, November 2005. <https://doi.org/10.1109/ICDM.2005.152>
- Frow, J.: *Genre: The New Critical Idiom*. Routledge, New York (2005)
- Güçlütürk, Y., Jacobs, R.H.A.H., van Lier, R.: Liking versus complexity: decomposing the inverted U-curve. *Front. Hum. Neurosci.* **10**, 112 (2016)
- Holzappel, A., Stylianou, Y.: Rhythmic similarity of music based on dynamic periodicity warping. In: *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2217–2220, March 2008. <https://doi.org/10.1109/ICASSP.2008.4518085>
- Huron, D.: *Sweet Anticipation: Music and the Psychology of Expectation*. The MIT Press, Cambridge (2006)
- Larson, S.: Musical forces, melodic expectation, and jazz melody. *Music Percept.* **19**(3), 351–385 (2002)
- Larson, S.: Musical forces and melodic expectations: comparing computer models and experimental results. *Music Percept.* **21**(4), 457–498 (2004)
- Lerdahl, F.: Cognitive constraints on compositional systems. *Contemp. Music Rev.* **6**(2), 97–121 (1988)
- Lieder, F., Klaas, S.E., Daunizeau, J., Garrida, M.I., Friston, K.J.: A neurocomputational model of the mismatch negativity. *PLoS Comput. Biol.* **9**(11), e1003288 (2013)
- Lott, M.S.: *The Social Worlds of Nineteenth-Century Chamber Music: Composers, Consumers, Communities*. University of Illinois Press, Urbana (2015)

- Margulis, E.H.: A model of melodic expectation. *Music Percept.* **22**(4), 663–714 (2005)
- Margulis, E.H., Beatty, A.P.: Musical style, psychoaesthetics, and prospects for entropy as an analytic tool. *Comput. Music J.* **32**(4), 64–78 (2008)
- McMullen, P.T., Arnold, M.J.: Preference and interest as functions of distributional redundancy in rhythmic sequences. *J. Res. Music Educ.* **24**(1), 22–31 (1976)
- Meyer, L.B.: *Emotion and Meaning in Music*. University of Chicago Press, Chicago (1956)
- Narmour, E.: The “genetic code” of melody: cognitive structures generated by the implication-realization model. *Contemp. Music Rev.* **4**(1), 45–63 (1989)
- Narmour, E.: The top-down and bottom-up systems of musical implication: building on Meyer’s theory of emotional syntax. *Music Percept.* **9**(1), 1–26 (1991)
- Radice, M.A.: *Chamber Music: An Essential History*. The University of Michigan Press, Ann Arbor (2012)
- Rentfrow, P.J., Gosling, S.D.: The content and validity of music-genre stereotypes among college students. *Psychol. Music* **35**(2), 306–326 (2007). <https://doi.org/10.1177/0305735607070382>
- Sapp, C.S.: Online database of scores in the Humdrum file format. In: *ISMIR*, pp. 664–665 (2005)
- Sturm, B.L.: The state of the art ten years after a state of the art: future research in music information retrieval. *J. New Music Res.* **43**(2), 147–172 (2014)
- Temperley, D.: Bayesian models of musical structure and cognition. *Musicae Scientiae* **8**(2), 175–205 (2004)
- Temperley, D.: *Music and Probability*. The MIT Press, Cambridge (2007)