ZGMTH Zeitschrift der Gesellschaft für Musiktheorie

Rohrmeier, Martin (2013): Musical Expectancy. Bridging Music Theory, Cognitive and Computational Approaches. ZGMTH 10/2, 343–371. https://doi.org/10.31751/724

veröffentlicht / first published: 17/10/2014
zuletzt geändert / last updated: 19/02/2016

Musical Expectancy

Bridging Music Theory, Cognitive and Computational Approaches[1]

Martin Rohrmeier

ABSTRACT: This article contributes to an interdisciplinary discussion of ways in which music-theoretical, cognitive, and computational accounts of musical expectancy may be bridged. It introduces some fundamental concepts concerning modeling, computation, representation, and some of their implications for theory building. Taking Markov models as a case in point, this paper illustrates in detail core notions of representation, model structure, parameter estimation, context-dependency, sparsity and overfitting, as well as the distinction between different levels of expectancy (short-term vs. long-term and knowledge-driven vs. data-driven vs. veridical) that interact in the context of musical listening. The final part compares local and hierarchical accounts of music and analyzes phenomena of nested implication-realization patterns, revision, and garden-path effects.

Dieser Artikel leistet einen Beitrag zur interdisziplinären Diskussion darüber, in welcher Form eine Brücke zwischen musiktheoretischen, kognitiven und computationalen Ansätzen zur musikalischen Expektanz geschlagen werden kann. Der Text führt zunächst grundlegende Konzepte der Modellierung, Computation und Repräsentation ein und diskutiert deren Relevanz für musiktheoretische Theoriebildung. Anhand des Beispiels von Markov-Modellen exemplifiziert der Autor wesentliche Aspekte der Repräsentation, Modellstruktur, Parameterschätzung, Kontextabhängigkeit, ›sparsity‹ und ›overfitting‹ sowie die Unterscheidung verschiedener Expektanzebenen (Kurzzeit vs. Langzeit und wissensgesteuert vs. datengesteuert vs. veridisch), die im musikalischen Hören zusammentreffen. Schließlich werden lokale und hierarchische Beschreibungen von Musik verglichen und damit verbundene Phänomene, insbesondere verschachtelte Implikations-Realisations-Patterns, musikalische Revision und Holzwegeffekte, analysiert.

1. Introduction

A mind is fundamentally an anticipator, an expectation-generator.[2]

1 I owe special thanks to Markus Neuwirth for many inspiring discussions about this text and the ongoing exchange about ways of bridging the gap between music theory and music cognition. I am also very grateful to Taiga Abe, Christian Utz, and Jan Philipp Sprick for their numerous suggestions that improved the article to a great deal. Funding for this research has been generously provided by the MIT Department of Linguistics and Philosophy as well as the Zukunftskonzept at TU Dresden funded by the Exzellenzinitiative of the Deutsche Forschungsgemeinschaft.

2 Dennett 1996, 57.

> Once a musical style has become part of the habitual responses of composers, performers, and practiced listeners it may be regarded as a complex system of probabilities […]. Out of such internalized probability systems arise the expectations—the tendencies—upon which musical meaning is built […]. The probability relationships embodied in a particular musical style together with the various modes of mental behavior involved in the perception and understanding of the materials of the style constitute the norms of the style.[3]

Musical expectancy is regarded as one of the most central aspects of music perception[4], and as such has received a great deal of scientific attention. The concepts of expectancy and prediction link static analytical approaches in music theory and analysis with the dynamic temporal aspects of the musical listening experience. The core insight that musical experience is (in part) closely linked with the cognitive processing of patterns of expectancy dates back to Leonard B. Meyer's seminal theory.[5] Prediction and expectancy formation constitute fundamental neurocognitive mechanisms of ongoing, automatic temporal processing of events of all kinds and are coupled with emotional reactions to the forms of expectancy associated with (musical) events.[6] Up to present there is large agreement that a substantial part of musical emotional experience originates in 'side effects' of processing likely and unlikely events, fulfilled and unfulfilled predictions. Given its neurocognitive basis, it is barely surprising that music has evolved to make heavy use of various forms of expectancy and predictive processing to trigger strong emotional effects.[7] Supporting this cognitive account is a large body of interdisciplinary recent research bridging psychology, computational modeling, and the neurosciences.[8] The study of the phenomenon of musical expectancy demonstrates successful ways to bridge theoretical/analytical, psychological, neurocognitive, and computational approaches in order to jointly advance our understanding of the foundations of the dynamic experience of musical listening.

Since the recent growth of literature on musical emotion, tension, and expectancy[9], several extensive reviews of cognitive and computational approaches have been published that explain the psychophysiological, neural, cognitive, and computational underpinnings of musical expectancy.[10] Accordingly, the purpose of this contribution is not to reiterate a review of the cognitive bases of musical expectancy that is up to date with this

3 Meyer 1967, 8f.

4 Cf. Huron 2006.

5 Meyer 1956.

6 Cf. the recent issue of the *International Journal of Psychophysiology*, Todd/Schröger/Winkler 2012.

7 Cf. Huron 2006; Koelsch 2012.

8 E.g., Koelsch 2010, 2012, 2014; Rohrmeier/Koelsch 2012; Farbood 2012; Tillmann 2005, 2012; Pearce/Wiggins 2012; Carrus/Pearce/Bhattacharya 2013; Egermann/Pearce/Wiggins/McAdams. 2013; Hansen/Pearce 2014.

9 For example Juslin/Västfjäll 2008; Juslin/Sloboda 2010; Juslin 2013; Koelsch 2010; Lehne/Rohrmeier/Koelsch 2013.

10 Cf. Rohrmeier/Koelsch 2012; Huron 2006, 2012; Pearce/Rohrmeier 2012; Pearce/Wiggins 2012; Rohrmeier/Rebuschat 2012; Wiggins 2012a, 2012b.

recent series of publications, but to introduce and to relate some of these notions and core underlying ideas to a more general music-theoretical audience in order to illustrate how interdisciplinary interactions between theoretical, computational, and cognitive approaches may be established.

This article is organized as follows: It first introduces core concepts to understand expectancy from a perspective of cognitive model building and discusses theoretic aspects of musical representation and constraints of well-defined models that have important implications for music-theoretical approaches. Departing from this background the article discusses different types of models of expectancy, in particular local Markov models. Using the example of Markov models the article introduces the notions of overfitting and sparsity and illustrates that these notions, that often remain unaddressed in current music-analytical literature, are of fundamental relevance for music-theoretical descriptions in general. The final part of the text analyzes the implications of hierarchical models for the concept of expectancy.

2. Relating Theoretical, Psychological and Computational Perspectives

Fundamentally, expectancy is a core cognitive process (i. e., a partial foundation of musical listening), and not a property of the musical score or of any other representation of musical structure (e. g., MIDI). Although they might be described in terms of music-theoretical constructs, statements about musical expectancy always refer fundamentally to an act of listening associated with underlying neural/mental processes and not to certain structures. Furthermore, discussing expectancy in the context of music analysis implicitly (and inescapably) assumes an underlying cognitive model of listening and expectancy formation that operates in terms of the music-theoretical or music-analytical concepts used. Following on from discussions by Ian Cross, Geraint Wiggins and others, it is crucial to ensure that the foundations of music-theoretical concepts and arguments do not lie in vague forms of *folk-psychology*[11], but are firmly rooted in psychological, neurocognitive, mathematical and computational foundations.

A description that specifies in detail how elements (derived from music-theoretical concepts and described by a *well-defined* language) are organized and combined to predict other elements constitutes a (more or less detailed) formal model, which is in essence analogous to the definition of a *computational* model in the sense of computing well-defined operations on symbolic representations.[12] The notion of formal descriptions that are in essence computational is far older than the arrival of the first electronic computers. Similarly *the intent of a numerous (music-)theoretical descriptions and models is (implicitly) computational*. Music analytical models, "pen-and-paper" models, syntactic models or (psychological) box models may be conceived of as *computational* descriptive or explanatory models that just—and crucially—differ with respect to the level of detail

11 See Cross 1998 and also the discussions in Wiggins 2012a und 2012b.

12 It is important to note that a well-defined, formal computational model is by no means equivalent with a statistical corpus analysis in general.

and consistency in which they are described or specified.[13] The degree of detail in some music-theoretical approaches is indeed very close to computational modeling[14], particularly in cases that parallel some forms of syntactic analysis in linguistics.[15] The potential for the development of a rich understanding of musical expectancy lies in close collaboration between theoretical and computational approaches.[16] The advantage of concise computational models (and their implementations) over mere theoretical accounts is manifold: The details are fleshed out in terms of the precise nature and representation of all elements involved, as well as the precise order and interaction of all processes involved[17]; the process of implementation frequently leads to the identification of notable inconsistencies and redundancies in the theoretical account; the implementation allows one to verify whether the theoretical predictions are indeed predicted; and whether they are fulfilled or not, given evaluation criteria and evaluation with test cases. The following discussion examines some theoretical accounts of musical expectancy from this background.

3. Expectancy and Underlying Models—Basic Concepts

What, When, and Which Forms of Representation?

The following paragraphs introduce and exemplify a number of fundamental issues regarding musical expectancy. Consider a simple first simple example in the domain of harmony (see Fig. 1).

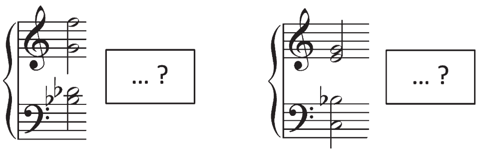


Figure 1. Harmonic implication and prediction illustrated by two predictive contexts.

Figure 1 illustrates two simple cases of harmonic expectancies generated by different chords. One might state that a *sixte ajoutée* chord predicts a subsequent dominant (Fig. 1,

13 Compare Wiggins 2012a.

14 E. g., Caplin 1998; Tymozcko 2011.

15 For one of the rare instances of a music analytic approach in the spirit of linguistic analysis, see Polth 2001, apart from the well-known main traditions following Keiler 1978 and Lerdahl/Jackendoff 1983. For an example of a recent detailed music-theoretical model based on a computational and linguistic framework, see Rohrmeier/Neuwirth in press.

16 For example Rohrmeier/Neuwirth in press; Eerola 2009.

17 Both requirements are not trivial at all. Compare, for instance, the numerous steps required by Fred Lerdahl and Ray Jackendoff's *Generative Theory of Tonal Music* [henceforth GTTM] (Lerdahl/Jackendoff 1983) to instantiate a well-defined theory based on Schenkerian theory (cf. Schenker 1935), and also the remarkable level of complexity encountered by Masatoshi Hamanaka et al. in a project to implement the GTTM (Hamanaka/Hirata/Tojo 2004, 2005, 2006, and 2007); or by Marsden and Smoliar to render a computationally tractable version of Schenkerian theory (Marsden 2010; Smoliar 1980).

left) or, more trivially, that a dominant-seventh chord predicts a major or minor chord the root of which is a fifth below (Fig. 1, right). However, these simple examples involve several implicit assumptions; first of all the 'what' question, with which most analysts are concerned. However, expectancy does also involve the 'when' aspect. Given a *sixte ajoutée* chord, *when* is the dominant expected to enter? At the next bar, the next beat, the (immediate) next event, at an unspecific point in the 'near' future? All of these options suggest different assumptions concerning the underlying model of expectancy: Is the element (the 'what') dependent on metrical structure, metrical hierarchy, rhythm, the mere sequence of events (i. e., independent of meter), or dependent on the following context (does the model allow for other elements to fall in between; and if so, which ones are defined as structurally important or unimportant)? It is important to note that any answer a music theorist may give inescapably involves an underlying theoretical choice concerning such choices of representation even if it may not be made explicit. Every account of expectancy relies on a well-defined model of expectancy. Crucially, if no model is explicitly defined, its parameters are covert in the analytical process and the account is prone to be incomplete, inconsistent or ill-defined. Each of the options listed above (and their combinations) imply different expectancy models with respect to the postulated components and their interaction.

Turning back to the 'what' aspect, does expectancy formation make predictions about the chordal root (C), the chord type (a major chord on C: root and type), the functional category (a dominant), a specific instance of the chord or even a specific voicing? (It is important to keep in mind here that this description of expectancy concerns listening, not the score.) As before, there is no natural, a priori 'right' answer to this, and no 'right' answer that can be discovered through reflection; the answer depends on the underlying model and involves a decision with respect to the level of representation, the fine-grained or rough resolution, and, more specifically, the purpose of the model. Further, are the input and output of the model of the same type? Does a dominant-seventh chord on G imply a continuation with a C root, a generic chord, or a specific C-major chord? At one theoretical level, it may be appropriate to say that *sixte-ajoutée* chords imply dominant-seventh sonorities or a dominant function, while at another level more detail may be required. Purely analytical or theoretical approaches require careful design in order to be globally consistent with respect to the level of representation. These points illustrate the role of the chosen representation, the components of the model, and the interaction between these components.

One frequent counter-argument raised against cognitive or computational models is that they involve accounts of music that are overly naïve or simplistic, lack complexity, or are dismissive of important context or musical (philological or music-theoretical) details. While it is always possible to include more detail or raise the level of abstraction, there is no *a priori* foundation of a 'right' level of analytical detail (or the 'best model') without reference to a specific purpose or evaluation criterion depending on the purpose.

Depending on the evaluation criterion there is also a trade-off between simplicity and accuracy as well as the problem of overfitting (see below). After all, accounts with greater levels of detail are not necessarily better and may gain only little improvement for the price of massive additional complexity, or they may even turn out with worse descriptive/predictive power, while (seemingly) simple models may have strong predictive power. Highly complex models that account for a large number of different factors/components may further be prone to design mistakes and redundancies when specifying their interaction and be difficult to interpret (which interactions did lead to the observed results?).

Like Borges's famous "map" "representing" its territory at a 1:1 scale[18], excessive musical detail may impair the interpretability and the use of a music-theoretical account. In the spirit of Borges's "Del rigor en la ciencia" ("On Exactitude in Science"), imagine the final establishment of an ideal type of a comprehensive 6000-page encyclopedia (any resemblance to real projects or publications purely coincidental) with a finely detailed 'final' description of sonata form classified by an abundantly rich variety of historically, philologically and systematically relevant distinctions and parameters, organized into 35 types and subtypes, differentiated in time by single year, in place by shire, organized by the causal web of mutual influence, amongst many other pieces of information. What would a pattern of interaction between type 3, 8 and 12 in two locations and between 1781, 1785 and 1787 mean—once the effort is comprehensively pursued to excavate such a relation—and what would it entail with respect to an account of expectancy of form? What generalizable insight might be drawn from a description with excessive level of detail? The first step to 'analyze' such a comprehensive account (say, in the case of teaching it to students or to yourself) would be one of simplification, in other words coming up with a map for the map, i.e. a simple *model* to cut a swath through the thicket. One instance of a fruitful outcome of the tension between theoretical complexity and stepwise simplification in consequence of theoretical and empirical explorations is found in the recent history of music cognition in terms of the successive simplification of Eugene Narmour's implication-realization theory that led to fundamental insights into the nature of musical expectancy that are also informative from a music-theoretical perspective.[19]

Incorporating Context-Dependency

Expectancy is further dependent on various kinds of 'context': The *sixte ajoutée* chord or the dominant-seventh chord (Fig. 1) may entail different predictions depending on the underlying tonal or stylistic context. The dominant-seventh chord suggests a much greater variety of possible continuations in works by Schumann or Liszt than in works by Vivaldi, Telemann, or Handel, and may exhibit a much weaker implicative tendency in the context of a Blues scheme. Another common example for style-dependent expectancy is constituted by the added sixth major (or minor) chord, which clearly invokes a subdominant function in the eighteenth century while it may function as a tonic for composers like Claude Debussy or Duke Ellington. This aspect of the overarching (stylistic) context is an implicit or covert assumption in a model of expectancy. How can this context be accounted for in a model?

18 Borges 1996.

19 See Narmour 1990, 1991; Schellenberg 1997; Krumhansl 1995; Eerola/Louhivuori/Lebaka 2009; Eerola 2003; Krumhansl/Louhivuori/Toiviainen/Järvinen/Eerola 1999; Krumhansl/Toivanen/Eerola/Toiviainen/Järvinen/Louhivuori 2000; Hippel/Huron 2000; Huron 2006.

One way would be to posit a different model for each case: one for common-practice music, one for Jazz, one for Rock, one for the Classical Style, one for Schumann, one for Handel, one for early Beethoven, one for late Beethoven, one for Bach's Partitas, one for Beethoven's "Waldstein" sonata, etc. While it may be insightful to study differences or similarities between different stylistic models, this list points at the core distinction between the type of a model and its parameters.

Generally, the definition of a model distinguishes between its *parameters* and the independent structure of the model. The *parameters* represent the information that the model operates on (for instance the information encoded in a Markovian transition matrix, that may represent style-specific knowledge about, e. g., chord transitions). While the parameters may be different for each of the above cases, there may be a single type of underlying model of expectancy that is independent of its parameters (such as a Markov model, e. g., a table of usual root progressions in the sense of Piston, or a tree structure in the sense of hierarchical models). In the case of music-analytical description of features that govern a certain style, it may be beneficial to draw clear distinctions between the model structure and its parameters. Once there is a clear distinction between a model and its parameters, one may examine how the parameters of the model may be inferred from data given. Recent computational models commonly involve methods to infer/learn the parameters from given examples (*training data*) such that the parameters do not need to be specified by hand but may be flexibly adapted to the ecological properties of the data/corpus it operates on.

The clear definition of a model of expectancy, its parameters and the inference process may be closely related to an overarching notion of (an aspect of) musical competence.[20] If one intends the model to represent cognitively relevant representations that govern musical competence and assumes that the corresponding processes are shared across the members of a community, the unity of the model, its parameter space in conjunction with its acquisition process as well as the corresponding stabilized structures of the music of the community characterize the intersubjectively shared medium[21] of musical communication (e. g., Western tonality, Middle Eastern *maqām*, or the North Indian *rāga*).[22] This understanding makes it possible to undermine purely subjectivist or solipsist accounts of aesthetics or musical forms of private languages by a cognitively founded account of intersubjectivity.[23] From this cognitive perspective arguments about subjectivity of musical listening (an imaginary overly post-structuralist critic may insist to *hear* iii to be implied by V rather than I which in turn 'sounds irregular') may point towards questions that are decidable on an empirical basis: Assuming that a Markov model (see below) is a fair model of musical competence and musical expectancy, given the Markov model

20 Cf. Stevens/Byron 2009; Rohrmeier/Rebuschat 2012.

21 See the detailed account by Polth 2001.

22 Communication arises through emergence, autopoietic stabilization and reproduction. Individual competence is a product of interactive social and cognitive adaptation processes. See also related arguments by Luhmann 1992, 2000, or Polth 2001.

23 E. g., Chomsky 1980; Davidson 1989, 2001; DeBellis 2009; Swain 1994; Temperley 1999, 2001, 2009; Wittgenstein 1953.

and its parameters learned from a corpus of common-practice tonal music the context of a V chord predicts a I chord as the most likely continuation.[24]

The Choice of the Type of Model

Following up on the previous account of a model, one core foundation underlying analytical approaches to musical expectancy concerns the structure of the model, i. e. making explicit how expectancy is derived. One simple way is to define a model based on statistical frequency of occurrence (as above): Find instances of 'V–x' in the chosen musical material (the 'corpus'), count their number, count the instances of 'V–I' among them, and divide the latter by the former to arrive at an *estimate* of the predictive probability based on the frequency counts. This is not far from what Walter Piston's early account of the table of common root-progressions (Fig. 2) may express, if it were specified with sufficient level of detail and with explicit accounts of the covert underlying assumptions.

Table of usual root progressions		
I is followed by IV or V,	sometimes by VI,	less often by II or III.
II is followed by V,	sometimes by VI, VI,	less often by I or III.
III is followed by VI,	sometimes by IV,	less often by I, II or V.
IV is followed by V,	sometimes by I or II,	less often by III or VI.
V is followed by I,	sometimes by VI or IV,	less often by III or II.
VI is followed by II, V,	sometimes by III or IV,	less often by I.
VII is followed by III,	sometimes by I.	

Figure 2. Walter Piston's table of usual root progressions (1948).[25] The table constitutes an early instance of an informal Markov model of root progressions based on musical experience and intuition (or 'intuitive statistics').

Note that the notion of such a Markov model rests on the assumption that there is an (accessible) level of representation that allows for comparison and counting (e. g., how to count pitch slides or notes with slight differences in intonation?) and that an *estimate* of the predicted next event can be determined by frequency counts. The account is independent of style and musical representation (i. e., the choice of building block): It is sufficiently general to be applied to melody, harmony, sequences of drum strokes or North Indian *r ga*. Crucially, the above specification of the Markov model contains another underlying assumption: that only the immediate context but no larger context[26] is relevant for expectancy (since prediction is only computed in terms of one predictive

24 Moreover, computational models that infer their parameters from exposure are capable of expressing individual variation.

25 Piston 1948.

26 Note that, here, 'context' refers to the sequence musical events preceding the predicted event.

event, any event preceding this *context* is not taken into account and hence irrelevant to the model; models that incorporate nonlocal dependencies are discussed below). This model is called a Markov model, and the assumption about the relevant context the Markov assumption, given in mathematical terms as:

$$p(c|e_1^i) \approx p(c|e^i)$$

In words: The probability of any event c at the position $i + 1$ given a context e ranging from the beginning of the piece to position i (written as e_1^i) is well approximated by the probability of c, given just the smaller context of the element e^i. Such models, which have been implemented from the 1950s onwards[27], are widespread and still frequently employed today.[28] A recent example of a corpus study of Bach's chorales provides one possible empirical implementation of a heuristic that approximates harmonic counts and provides an empirical *estimate*:[29] V–I is roughly seven times more likely than V–vi. The difference is that Piston's judgments (Fig. 2) about frequency are intuitive[30] ('regular' vs. 'sometimes' without specification of whether 'sometimes' has roughly the *same* meaning across different rows), while the latter are empirical estimates and based on both a corpus and a detailed replicable and comparable process.[31]

Further, well-defined music-theoretical accounts of expectancy such as Markov models lend themselves to empirical testing and evaluation providing further insight into the process of expectancy formation. For instance, the characterization of harmonic expectancy by progression tables has been tested empirically in studies using probe chord and harmonic continuation paradigms, leading to partial confirmation and partial revision of the theoretical predictions made by Piston.[32]

Different levels of musical expectation

So far, a model of expectancy has been characterized as the combination of the model structure and its (potentially learned or acquired) parameters from which predictions are derived. However, in an endeavor to model expectancy in listening (or interactive cognitive tasks such as improvising), different sources of expectancy have to be accounted for: There are differences between expectations based on our general acquired musical competence and expectations based on particular features of the current piece we are

27 See the review by Pearce/Wiggins 2012.

28 Cf. for example Pearce 2005; Conklin 2013; Pearce/Wiggins 2006, 2012; Temperley/de Clercq 2013; Tymoczko 2003.

29 See Rohrmeier/Cross 2008; Rohrmeier 2005; see also Tymoczko 2003 for similar work, and Hedges/Rohrmeier 2011 for an empirical exploration of Rameau's theory of the *basse fondamentale* (Rameau 1722).

30 Neuwirth 2013 refers to these estimations common in the humanities as 'intuitive statistics.'

31 Even though the method of deriving at such estimates may be debatable as are decisions of human analysts and may be revised by improved methods, the computed numbers are internally consistent by being computed using the same algorithm whereas human analyses of such a large corpus may be prone to inconsistencies across different pieces.

32 See Bharucha/Krumhansl 1983; Krumhansl 1990; Schmuckler 1989.

listening to or interacting with. Accordingly, a distinction is required between knowledge driven and online expectancies (or long-term and short-term models according to the terminology by Pearce et al.[33]). These types of expectancy formation may interfere (see below).[34]

Another distinction has to be made regarding levels of expectancy. The cognitive musicology literature commonly assumes three levels of expectancy: data-driven, veridical and schematic/knowledge driven.[35] Data-driven predictions characterize simple musical processes that may not require a foundation in acquisition: an ascending scale, for example, or a simple pattern such as note repetition or alternation.[36] One may further classify accounts of musical processing and prediction based on purely sensory processing as forms of knowledge-free data-driven sources of expectancy.[37] Veridical expectancy refers to cases in which the musical source itself (the piece) is known so that predictions about upcoming events are based on prior knowledge of the (presumed) true source (Fig. 3). Finally, expectancy formation that is neither based on simple patterns nor on prior veridic knowledge of a piece may rely on previously acquired style-specific knowledge or schemata (musical competence), e.g. of harmony or voice leading. The *sixte-ajoutée* chord above (Fig. 1) constitutes an example of knowledge-driven expectancy acquired by previous experience of tonal music. Crucially, knowledge-driven forms of expectancy are bound to an underlying process of expectancy generation (such as a Markov model or a tree-based model) and a complementary process of knowledge inference and acquisition (implicit learning).[38]

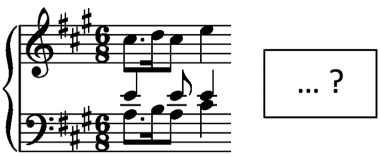


Figure 3. Example of veridical expectancy

The Problems of Overfitting and Sparsity

It is a commonplace in computational modeling that descriptions and models do not necessarily get better by adding more information. One example of a study modeling Jazz harmony shall illustrate this.[39] Using Markovian methodologies similar to the ones introduced above, the study implemented *n-gram* models (amongst others) for the prediction of chord sequences from a large Jazz corpus and compared the performance of

33 Pearce/Wiggins 2012, Conklin/Witten 1995.

34 See also similar distinctions by Huron 2006; Margulis 2007.

35 For this distinction, see Bharucha 1987; Eerola 2003; Huron 2006.

36 Already low-level sensory neural processing is capable of dealing with such regularities without requiring higher-order knowledge-driven processes, see Koelsch 2012.

37 See, e.g., Parncutt 1989; Leman 1997.

38 Rohrmeier/Rebuschat 2012.

39 Rohrmeier/Graepel 2012.

models of different context length: The prediction of the next chord was based on mere single-chord frequencies (*1-gram model*), conditioned on the previous chord (*2-gram model*), the two previous chords (*3-gram model*), or the three or four previous chords (*4-gram* and *5-gram model*). These *n-gram* models were evaluated using a large corpus of harmonic lead-sheet annotations of 1600 Jazz standards.[40] Furthermore, the study compared two forms of evaluation: In the standard case, the corpus of Jazz standards was split into two parts, one of which was used for model training (i. e., for estimating the probability tables used for prediction as outlined above) and one for evaluating each model; in the second *veridical* case, the training set contained the evaluation set.[41] Figure 4 displays the performance of the different *n-gram* models under standard and veridical conditions.

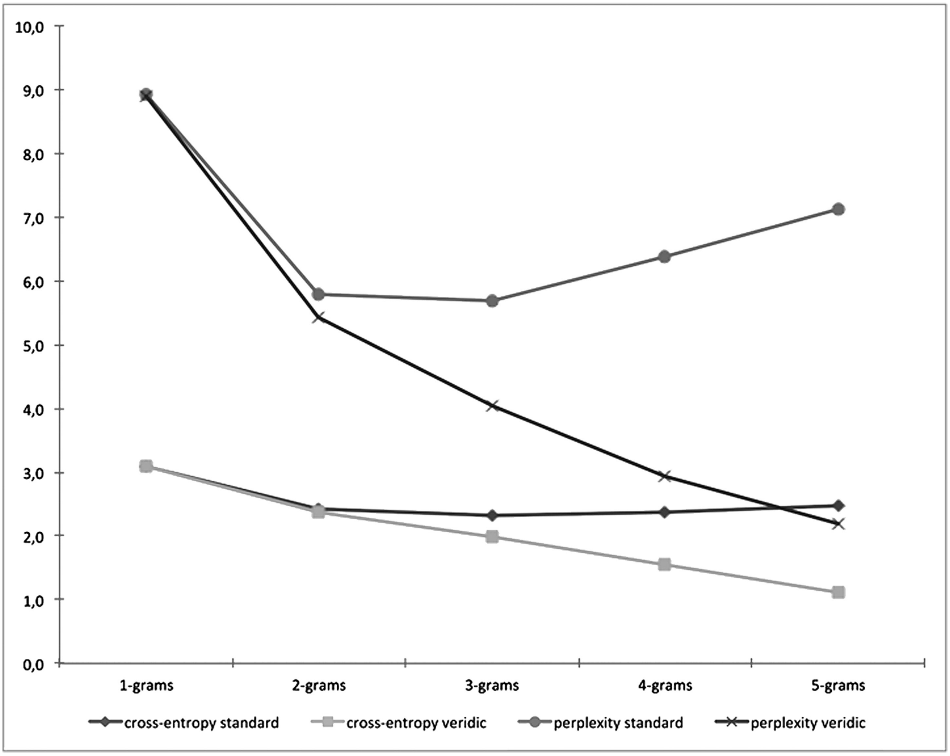


Figure 4. Performance of predictive *n-gram* models of harmony using different context length[42]

40 Haas/Veltkamp/Wiering 2008.

41 Although veridical evaluation is avoided in modern computational modeling because of the problems discussed in the following paragraphs, this case was included to exemplify the effect sizes of overfitting.

42 Reprinted from Rohrmeier/Graepel 2012.

As the figure illustrates, the *n-gram* models of the Jazz corpus reach an optimal performance level for n = 2 or n = 3. This means that a context of one or two previous chords is optimal for predicting the next chord, whereas a model with a larger context possesses lower predictive power. Hence, more information incorporated into the model does not necessarily improve its performance. To the contrary, larger contexts here contain too many chord sequences specific to individual pieces or individual progressions and thus are not generalizable across a larger set of cases.

A second observation can be made comparing the performance for the veridical vs. the standard case. In the veridical case, in which the test set is included in the training set from which the model parameters are estimated, the performance continuously improves with increasing context length. For the cases of *4-* and *5-grams*, predictions for chord contexts of three and four chords are compared. These chord progressions are highly individual; since the models have incorporated all the test cases as well, they are good at predicting individual progressions in the test cases. However, this does not entail that these are 'better' models: They would generalize poorly similar when confronted with novel progressions as the example of the same models in the non-veridical case illustrates. Although they incorporate a large amount of piece-specific knowledge, by no means does this result in neither improved general harmonic knowledge, nor does it improve the model's capacity for predicting harmonic progressions. This case constitutes an example in which more detailed harmonic knowledge is even detrimental for a description of harmony with generalized predictive power.[43]

This result is not limited to harmony and transfers to other musical structures. For instance, Marcus Pearce and Geraint Wiggins report a similar finding for the modeling of melodic prediction with *n-gram* models.[44] This problem is referred to in the computer science literature as the 'overfitting problem.' Some models may be over-trained with too much information that does not improve, but rather impairs the description/the inferred knowledge of the structures.

Another related issue concerns the contrasting common problem of sparsity: Even though a Markov or *n-gram* model may be trained with data from a large number of musical examples, there is still a high likelihood that an application of the model to musical prediction may encounter a context the training materials did not contain and the model has no information about. Given the definition of the Markov model above, it is impossible to derive a prediction for the next event if there is not a single sample case (e.g., how to predict the continuation of [*U V W* …] if there is no instance of [*V W*] in the reference data?). This problem occurs frequently in computational modeling[45] and bears theoretical implications for music theory (see below). Commonly, Markov or n-gram models involve specific techniques of 'zero-escape methods' and 'smoothing' to avoid such cases.[46]

43 Note, however, that there are some models that are less prone to problems overfitting, such as the results of modeling chord progressions with Hidden Markov Models (Rabiner 1989) as reported in the same paper by Rohrmeier/Graepel 2012.

44 Pearce/Wiggins 2004.

45 E.g., Manning/Schütze 1999; Pearce/Wiggins 2004.

The phenomenon of sparsity, related to the 'Zipf distribution,' is a common property of the distribution of events across a large range of natural phenomena, including language and music.[47] Briefly construed, it implies that there is a small number of items that occur very frequently and account for most of the domain while a large number of items occurs highly infrequently[48] (frequency approximates inverse rank). The relevance of this distribution for music has been demonstrated for the cases of pitch and harmony.[49]

Zipf's law relates in two ways to modeling as well as theoretical descriptions: On the one hand, it implies that fair models or descriptions may be achieved employing a small number of rules (capturing the rules governing the most frequent items). On the other hand, however, a complete or comprehensive description requires an exponential number of rules and exponential effort; the accurate completion of this is, without the aid of computational analysis, a virtually intractable task for human analytic scholarship and even with computational methods, a comprehensive description would hold little informational value. Therefore, the problems of overfitting, sparsity and Zipf's law define crucial constraints for descriptions in music theory: Style descriptions (say of Handel's suites, or a structure like sonata form in general) do not benefit from extensive and indefinite addition of detail and description —even if the Sisyphean task of a comprehensive description were tractable and possible for a particular repertoire in decades of analytical scholarship. This problem is further aggravated because the characterization of a musical style frequently deals with corpora that are historical and hence complete. Numerous music-theoretical approaches therefore operate in ways that are analogous to the case of 'veridic' modeling described above. Accordingly, such approaches face the problem of having to draw a careful line between meaningful description and generalization and problematic overfitting by adding excessive details that do not generalize and merely describe random artifacts and coincidences in the corpus—analogous to the case of veridical overfitting described above. The method of withholding a set of data that is not used for theory building and only for theory evaluation constitutes a core and obligatory standard in computational and statistical modeling[50], yet it is still almost entirely ignored in the standards of music-theoretical practice. Reflections on the nature of description, generalizability and model building may suggest the use of such standards of in future music-theoretical endeavors.

To conclude and to tap into the spirit of David Huron's, Michael Cuthbert's research paradigms (amongst others) as well as Ian Quinn's reflection of musicology in the age of Big Data and Digital Humanities, music theoretical scholarship of the 21st century may

46 These techniques attribute a small default value to all possible yet unobserved cases and/or use more frequent and shorter contexts to derive predictions; see Manning/Schütze 1999; Pearce/Wiggins 2004 for a detailed discussion and comparison of such techniques.

47 Zipf 1935, 1949. See, for instance, Piantodosi (in press) for a recent discussion.

48 Frequency is approximately proportional to the inverse rank, i.e. $f(r) \propto \frac{1}{(r+b)^{\boldsymbol{a}}}$ (with constants *a, b; a* > 0) according to the accounts by Zipf 1936 (for b=0) and the refinement by Mandelbrot 1953 (for b > 0).

49 Manaris/Romero/Machado/Krehbiel/Hirzel/Pharr/Davis 2005; Zanette 2006; Rohrmeier 2005; Rohrmeier/Cross 2008; Voss/Clarke 1975, 1978.

50 E.g., Manning/Schütze 1999.

draw great benefit from close interdisciplinary collaboration and from taking on board a number of the issues raised above in theory building. In particular, this concerns aspects such as precise, formal definitions of concepts, operations, methods and notation, making explicit underlying assumptions, grounding theoretical concepts and their operationalization in firm psychological, cognitive, mathematical and computational foundations, defining testable hypotheses and evaluation criteria, and evaluating theoretically derived hypotheses that concern corpora or ways of listening in terms of computational or psychological research.[51]

4. Local and Hierarchical Structure

The discussion above has largely focused on only one specific type of expectancy models, namely local models (such as Markov or *n-gram* models, and to some extent, regular grammars) which share the common assumption that event prediction is only characterized by the local context, consisting of the immediately preceding events. However, a large number of music-theoretical models such as those proposed by Schenker or more the strictly formalized approaches of Alan Keiler, Fred Lerdahl and Ray Jackendoff (GTTM), Eugene Narmour, Mark Steedman, Martin Rohrmeier, and Jonah Katz and David Pesetsky[52], take as their basis hierarchical principles of musical structure, which formally exceed the expressive power of local models and postulate both proximate and distal realizations of implications.[53] With respect to music expectancy, however, nonlocal models come packed with a number of implications that are not self-evident, outlined below.

Hierarchical Structure and Expectancy

One linguistic example will first illustrate the issues involved in hierarchical processing before turning back to music. Speaking about a man, which word or word class is predicted by "the old…"? One might say: "man" or a noun. The continuation, however, is: "the old and…". Was the prediction "man" violated or unfulfilled? What is the new prediction? Is it "humble" (or any adjective) or is it still "man"? In some sense, it is both. Continuing as "the old and humble…", one may now be aware that the prediction may be "man" yet also "but", for instance: "the old and humble, but…" or "the old and humble, but frequently… man" etc. A merely local account (i.e., an account that treats predictions as strictly local) is insufficient for such a case. For instance, arriving at "but frequently" a trigram model would have lost the context of "the…" predicting a noun and may, in contrast, predict the continuation "but frequently *you* [*will*]". A complex expectancy structure like the one illustrated by the present example involves nonlocal

51 Quinn 2014; Huron 1999; Wiggins 2012; Pearce/Rohrmeier 2012; See also, e.g., the Music21 platform by Michael Cuthbert and its endeavor to provide a novel unified platform for computational research in musicology and music theory (Cuthbert/Ariza 2010).

52 Schenker 1935; Keiler 1978; Lerdahl/Jackendoff 1983; Narmour 1992, 1999; Steedman 1984, 1996; Rohrmeier 2007a, 2007b, 2011; Katz/Pesetsky, online draft of 2011.

53 See Temperley 2011 for a contrasting viewpoint regarding the relevance of nonlocal models.

dependencies and predictions that may be interrupted by another structure. This is illustrated by the dependency structure represented in Figure 5. This example illustrates that cases like this require an account that is able to capture such hierarchical, potentially nested dependencies, for instance, by employing representations of a flexible number of simultaneous instances of predictions at different local or nonlocal levels.

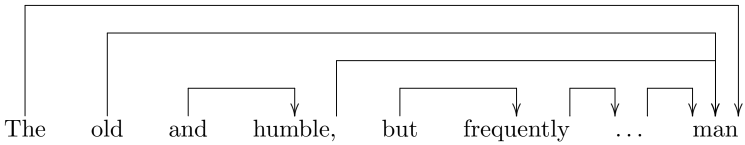


Figure 5. Nested syntactic patterns of expectancy

This case bears a musical analogue. Figure 6 illustrates several ways in which local and nonlocal types of expectancy are linked together. First, there are several local implications: $\mathrm{ii}^{6/5}$ (m. 2) implying V (m. 3), V (mm. 3, 5, and 7) implying I, $\mathrm{VI}^{6/4}$ implying $\mathrm{VI}^{5/3}$ (m. 4), $\mathrm{V}^{6/4}$ implying $\mathrm{V}^{5/3}$ (m. 5), $\mathrm{I}^{7/4/2}$ implying I (m. 8; note that this latter implication is context-dependent; in a neutral, non-cadential context the implication pattern would more likely be $\mathrm{I}^{7/4/2}$–$\mathrm{V}^{6/5}$). All of these local implications are immediately met except for the V–I implication, which is not (immediately!) met all three times. Note further that implication and realization pairs are tightly linked by the fact that realization events and new implicative events are combined: e. g., the V chord in m. 3, which is the expected realization of the $\mathrm{ii}^{6/5}$ in m. 2, in turn sets up a new expectation. The expectation set up by the V chord (m. 3, 5) is violated twice by the same V–$\mathrm{VI}^{6/4}$ deceptive progression (m. 4, 6; note that the second occurrence establishes a stronger implication due to the doubling of the bass note of V in m. 5). In turn the sequence $\mathrm{VI}^{6/4}$–$\mathrm{VI}^{5/3}$–$\mathrm{V}^{6/4}$–$\mathrm{V}^{5/3}$ constitutes a chain of mutual implication-realization patterns (combined with a 6–5–6–5 voice-leading pattern); this chain leads to the reestablishment of the V harmony and raises again the previously unfulfilled expectation of a I sonority, only to interrupt it once again – constituting a "one-more-time pattern."[54] The third time, V proceeds to V^7, demarcating the end of the eight-bar phrase, yet not resolving into I immediately despite the right bass note at m. 8. Accordingly, the final chord is locally implied by the V^7 chord at the end of m. 7 while being interrupted by an even more local implication of the $\mathrm{I}^{7/4/2}$. This constitutes a form of two nested implications. Moreover, the three V chords may be regarded at a higher level as a prolongation of an overarching V function that sets up a strong final V–I implication by virtue of being interrupted and reestablished twice by deceptive progressions, thus reinforcing a chain of implications towards the final V. In this respect the nested implications of the musical example are analogous to the linguistic example above. It is important not to neglect the fact that the first two chords establish the key of F minor almost unambiguously due to their scale membership[55] and thus strengthen the V–I expectation.

54 See Schmalfeldt 1992.

55 Note that already the first F-minor chord is sufficient for almost unambiguously establishing the key of F minor. An *n-gram* model with a padding symbol marking initial silence or a Bayesian model would support this result in straightforward ways based on the distribution of piece beginnings in a corpus.

Furthermore, the I chord may be analyzed as setting up an expectation itself in terms of its implied tonic return at the end of the period. Altogether this example motivates a representation of the different implication-realization patterns or nested expectancies as set up in the example of this sonata movement. Turning back to the linguistic example, the I and V chords set up nonlocal implications that are maintained and interrupted by several other patterns of implication and realization until they are realized.

Figure 6. Wolfgang Amadé Mozart, Piano Sonata F-Major, KV 280, second movement, mm. 1–8

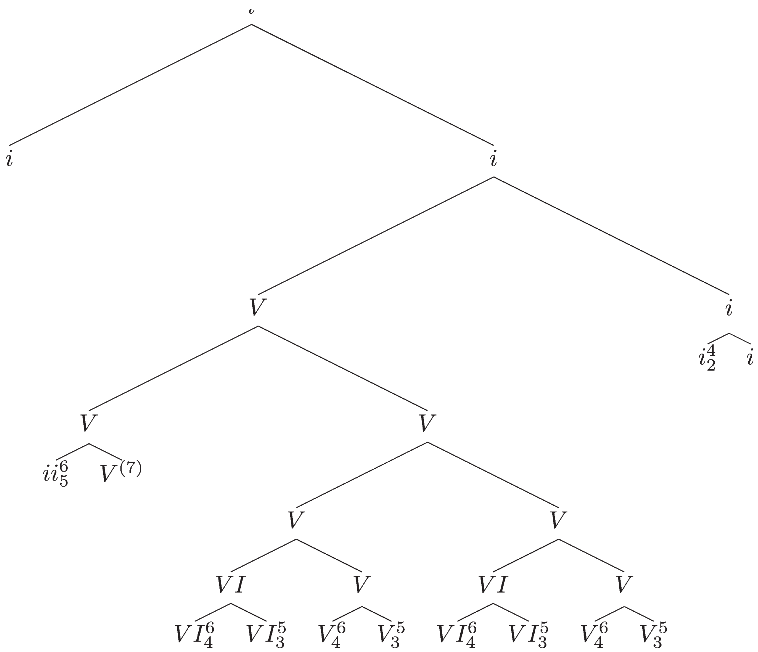
Such hierarchically nested patterns of expectancy involve both local and nonlocal components.[56] Note that, for instance, in Figure 7 the arrows are organized in a way that some may be superordinate to two or more others.[57] Such a hierarchical form of organization entails a tree-based representation. This formalization of the hierarchical understanding of musical dependency and expectancy is useful for casting precise and testable predictions that may not be straightforward under an informal notion of hierarchicality. Specifically, it predicts that crossed patterns of implication-realization events such as those in Figure 8 may not occur.[58] As outlined before, the requirement for processing (and listening to) such a structure is the ability to keep more than one open implication actively in mind, while other intervening events and patterns of implication, realization, and prolongation occur.

Once a hierarchical model of music is involved, the notion of expectancy becomes less straightforward (as outlined in the previous example). One cannot easily predict the next event any more because an interruption of a current implication may occur at a large number of points. Consequently, to say that a I–IV–V progression implies a I now

56 The latter could be understood as a special case of nonlocal dependencies with no intervening material; hence a process that is able to capture nonlocal dependencies will naturally also capture local dependencies.

57 Note that prolongation works in close analogy to coordination in language (the "and" as used above), a claim that has been made already by Mark Steedman (Steedman 1984, 1996).

58 The occurrence of such crossed patterns of dependency would provide evidence for the necessity of an even more complex model of dependency structure and associated forms of expectancy.



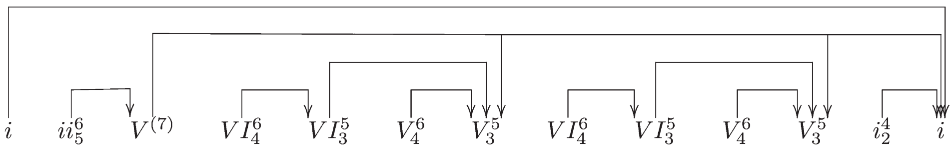


Figure 7. Visualization of the nested patterns of expectancy in the Mozart phrase from Figure 6

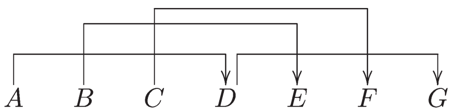


Figure 8. Crossing patterns of expectancy that are predicted to be impossible in musical contexts

entails the awareness that the implied I may occur several measures later after a potentially large series of multiple and recursive interruptions. Examples of such structures may be found in the analysis of deceptive cadences and half cadences amongst other phenomena.[59] In this context, a recent empirical study provides neural evidence that nonmusicians process original and modified versions of two two-part phrases from Bach's chorales differently, depending on whether the second part returned to the initial key

59 See Rohrmeier/Neuwirth in press.

of the entire phrase or not after a (comparably long) intermitting modulation.[60] Such a case of nonlocal prediction provides a prototypical example of predictive processes that cannot parsimoniously be expressed by virtue of plain local, *n-gram* or Markov models and suggests that we possess and employ capacities of nonlocal processing in music, for the least supplementing local processing[61]).

Expectancy Violation and Revision in Hierarchical Models

The example above illustrates that the notions of expectancy fulfillment and expectancy violation need to be reconsidered when taking into account a hierarchical model of structural organization: Returning to the sentence "the old and humble, but frequently … man," the occurrences of "but," "and," "old," or "frequently" might further be regarded as instances of expectancy violation in the context of local predictions. Assuming that the ongoing listening/parsing process maintains an updated version of the best possible analysis[62], the time course of analyzing the sentence involves that, having expected the word "man" to close the noun phrase (NP), the parser is required to adapt the inferred tree model of the NP to accommodate for the newly encountered information. In terms of the predicted tree structure, encountering a less probable, yet grammatically correct option like "but" forces another adaption to the sentence model during online perception. This is an instance of an expectancy violation due to a less probable but grammatical event—a case contrasting the violation through an ungrammatical event such as "the old and but" where there is no structural recovery possible.

In addition, the nested dependency structure requires that all parts are fulfilled: "the old and man" strikes one as ungrammatical even though the nonlocal structural predictive dependency is fulfilled. An analogous musical example of this may be found in the following common-practice harmonic sequences:

– (1*) I V $ii^{6/5}$ I
– (2*) I V $ii^{6/5}$ $V^{6/4}$ I

In both cases, the two overarching implications are fulfilled, but the local contexts involve open implicative dependencies ($ii^{6/5}$ or $V^{6/4}$) that require closure for the sequence to be regular. An analogous example could be constructed for the respective nonlocal dependencies.

– (3*) I […] ii V || vi iii vi V/iii iii
– (4) I […] ii V || vi iii vi V I

60 Koelsch/Rohrmeier/Torrecuso/Jentschke 2013; see also the preliminary results by Woolhouse/Cross/Horton 2006.

61 However, a mechanism that is able to instantiate nested predictive dependencies as the ones described above, is sufficiently powerful to deal with local predictions without requiring a separate 'local processing module.'

62 See Jackendoff 1991 for a discussion of this in the case of musical parsing.

Assuming that these two examples would occur constituting the context of a complete 8-measure period, their difference illustrates the style-specific necessity of the tonic return after a progression to V and particularly after a non-tonic continuation in the second half of the phrase in order to fulfill the nested predictive dependencies—just as illustrated by the empirical study mentioned in the previous paragraph.[63]

An understanding of expectancy in terms of hierarchical structure and multiple (recursively) nested predictive dependency relationships bears further consequences for the concept of expectancy violation: While expectancy and its violation is a mere matter of degrees of continuous probability values for a local model, patterns of expectancy receive a different interpretation in a hierarchical account. Common musical patterns such deceptive cadences, one-more time patterns, interruption, etc., may be accounted for in terms of overarching nonlocal dependencies. Accordingly, what may appear to a local model as a local expectancy violation, may resolve into a regular analysis under the perspective of a hierarchical account. In these terms, an expectancy violation may likely be a local interruption of an established predictive event, which delays or defers the realization of the predicted event and in turn sets up a new nested predictive context (analogous to "but frequently…" in the language example). A very simple example of this is provided by the following harmonic progression:

– (5) I ii V [vi ii V] I

This sequence sets up a V–I implication which is interrupted by a deceptive progression V–vi which in turn initiates two further predictive dependencies to reestablish (and potentially strengthen) the predictive effect of the initial dominant context (as indicated by the brackets). In the tree analysis vi would not be merged with the preceding V, but analyzed as subordinated to ii and hence be merged with ii. This understanding may be useful to recast a variety of deceptive cadences in terms of recursively embedded predictive dependencies rather than understanding them solely in local terms of 'regular' and 'less regular' continuations of a dominant.[64]

Two Case Studies of Expectancy Violation and Revision

The hierarchical understanding of expectancy outlined above further entails a link between expectancy violation and revision. As argued above, the update of the current tree based on the previous context in an instance of Jackendoff's idealized parser may require smaller or larger adaptations of the tree structure based on newly encountered unexpected events. Within the framework of a recursive grammar model such updates of the tree structure may imply further change and revision of previously heard structure.

63 Koelsch/Rohrmeier/Torrecuso/Jentschke 2013.

64 See Rohrmeier/Neuwirth in press for a detailed discussion of this analysis of deceptive and other types of cadences.

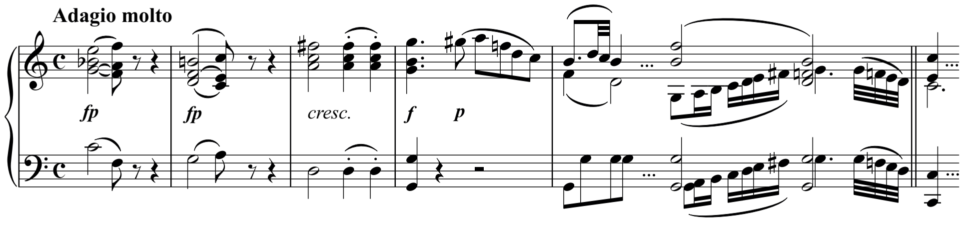


Figure 9. Ludwig van Beethoven, Symphony I, C Major, op. 21, first movement, mm. 1–13

Two well-known examples illustrate this, the beginning of Ludwig van Beethoven's First Symphony (Fig. 9) and the beginning of Robert Schumann's lied "Am leuchtenden Sommermorgen" (Fig. 10).[65] The beginning of Beethoven's symphony initially implies the key of F major by an unusual initial, dynamically reinforced dominant-seventh chord (m. 1). This tonal context is immediately revised to the key of C major by virtue of a deceptive progression (V–vi, m. 2). This weakly established C-major key is in turn revised by the subsequent seventh chord on D setting up a prediction for a (local) tonic G-major chord (m. 3). Finally this G-major context is functionally revised to be the dominant of the C-major key (mm. 4–7), which turns out to be underlying key of the entire passage. This passage can be interpreted in close analogy to linguistic 'garden path' phenomena such as "The horse raced past the barn fell" (which forces a parse of "The horse [(that) raced past the barn] fell" rather than "[The horse raced past the barn] fell", which requires the parser to backtrack and revise the entire constituent structure after encountering the word "fell"). In analogy, the remarkable feature of this segment is that the expectancy violations trigger a reparse and revision of the underlying key and the entire set of assigned scale degrees and tonal functions three times for three different points of tonal reference.

A phenomenon like this suggests that the process of expectancy formation is a by-product of predictive generative parsing and that expectancy violations inform an internal update and revision process of the parser (the likely candidate parse(s) is/are generated on the fly and interactively matched with the incoming stream of events to update the best current candidate parse(s)). In his discussion of this phenomenon, Ray Jackendoff asserts a similar dynamic parsing process for rhythmic or metrical ambiguities.[66] Finally, it is important to note in this context that a mere plain local Markovian account of musical processing (as outlined above) is in principle incapable of capturing such features relating to a parsing process since it does by definition not incorporate a representation of underlying deep structure that may be revised on the fly.

Schumann's lied "Am leuchtenden Sommermorgen" (Fig. 10) illustrates an additional feature of expectancy formation, expectancy violation, and parsing by predictive processing. The opening of the piece creates a strong surprise by continuing what sounds like a dominant-seventh chord with a 6-4 suspension and a semitone descent in the bass. The strong effect is caused by the sudden effort required by the parsing process to revise the

65 Agawu 1994 discusses these examples in the context of musical ambiguity.

66 Jackendoff 1991; see also Temperley 2001.

dominant-seventh tonic expectancy retrospectively towards an unlikely German sixth, a precedential $V^{6/4}$ progression in a different key (B vs. B^b). As in the previous example the parsing process is forced to reinterpret both tonal function and key structure. The formal preconditions of this strong effect are two-fold: Firstly, within the tonal system functionally ambiguous chords are possible and, secondly, the probabilities of the competing interpretations are skewed (i. e., they diverge largely[67]).

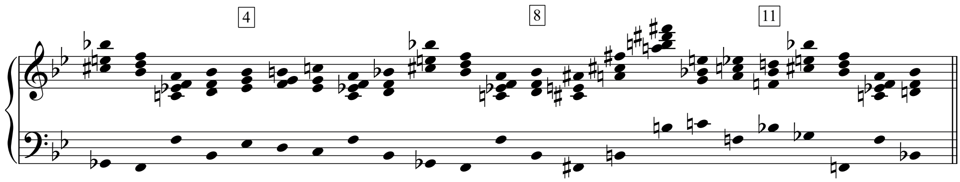


Figure 10. Schumann, "Am leuchtenden Sommermorgen", No. 12 from *Dichterliebe*, mm. 1–13

The example of Schumann's lied provides an additional instance of the interaction between predictive processing, online learning and expectancy violation: When the initial sonority reappears for the third time (m. 8) after a strong stabilization of the Bb-major key, it is continued as a dominant-seventh chord (mm. 8–9) providing the B-major context that was originally expected. However, after having created a strong garden-path effect by subverting the established hearing of the initial dominant seventh chord, Schumann here demonstrates a second comparably strong effect by playing yet another trick on the established expectation: After two occurrences of the German sixth chord reinterpretation, the continuation to B major which was previously the most likely has now turned into an unlikely progression. The basis of this effect is again twofold: Firstly, it takes advantage of the established Bb-major context and the higher likelihood of interpreting the sonority as a German sixth chord in Bb-major rather than expecting a modulation. Secondly, there is an effect of *online learning* of this motivic sequence during the course of the first eight measures.[68] With respect to the first aspect, it is important to note that, without any additional context, the probability of tonal function and key of the pitch class set Gb/F♯–Bb/A♯–Db/C♯–(Fb)/E is highly favoring the interpretation of a dominant-seventh chord, whereas this probability changes once a previous context in the key of Bb is established. Regarding this second aspect, Schumann's piece provides a case of online learning (or what Darrell Conklin as well as Marcus Pearce and Geraint Wiggins refer to as "short-term model,")[69] i. e., learning during the course of a piece. Another example of the strong effects of expectancy violations based on online learning is found in the second movement of Schubert's piano sonata A major D. 959, in which the multiple repetition of the A–G♯ motive in F♯ minor is unexpectedly replaced by the step A–G in D major

67 If the probabilities for the different options of continuation were not skewed but rather similar, we would hear them as two (or more) equally possible or plausible continuations.

68 See Rebuschat/Rohrmeier/Cross 2011 and Rohrmeier/Cross 2014 for an empirical exploration of online learning effects.

69 Conklin 2013; Pearce 2005; Pearce/Wiggins 2006.

(mm. 189–190). While this effect can be simply accounted for in terms of online learning and a short-term model of pitch-class distribution or melodic-harmonic bigram structure, this and the previous example illustrate the strong contribution of online-learning to listening and the interaction of long-term and short-term knowledge to musical experience.

Effects of expectancy, expectancy violation, ambiguity, and revision continue to have an (albeit weaker) effect even over the course of multiple listening. Such effects and their emotional correlates that remain after multiple listening would be difficult to account for considering the ongoing implicit learning and, particularly, the learning of the veridical structure of a piece. One potential explanation that solves this dilemma is proposed by Ray Jackendoff: He suggests that the parser constitutes a separate module that is "'informationally encapsulated' from long-term memory of pieces"[70] and overrides veridical knowledge to some extent by operating independently on the musical input. Accordingly, the same backtracking and revision processes would still operate each time we listen to the opening of Beethoven's first symphony triggering similar emotional effects despite our growing knowledge of the piece.

Altogether, the examples above illustrate how closely musical expectancy is linked to implicit learning and implicit knowledge both in long-term enculturation and short-term musical listening.[71] It is further deeply grounded in the processing of local as well as hierarchical structure, and involves multiple nested dependencies as well as the workings of the recursive parsing mechanism that provides incremental structure building, predictive generation, update and revision processes. Automatic expectancy formation, effects of retardation, anticipation, expectancy violations, deceptive structures, ambiguity, musical garden-path phenomena and revision: Such effects in musical listening and the emotional experience[72] result from the operation of an ongoing parsing mechanism that processes the musical stream, generates likely parses and continuations, and matches them with the ongoing stream of musical events.

5. Conclusion

Generally, perspectives of cognition and modeling may provide a number of contributions to the field of music theory: Apart from demonstrating the necessity of precise specification of covert assumptions and characterizing constraints of theoretical description that arise from problems such as sparsity or overfitting, they illustrate the insight that theoretical models of music and expectancy are intrinsically linked to implicit or explicit underlying formal, computational assumptions.

After all, musical expectancy is intrinsically linked to cognitive accounts of predictive processing. It provides a constructive case for the mutual interaction of music theory and music cognition[73] and illuminates ways in which concepts from music cognition, computational modeling, and (neuro)psychology may help to address music-theoretical

70 Jackendoff 1991, 221; compare also the discussion in Temperley 2001.

71 Rohrmeier/Rebuschat 2012.

72 See, e.g., Meyer 1956; Koelsch 2010, 2012; Rohrmeier/Koelsch 2012.

73 For further discussion, see Pearce/Rohrmeier 2012.

issues from a different perspective. They provide ways to support, adapt, and revise music-theoretical concepts, to clarify theory formation in music analysis and to take into account music-theoretical insights in the formation of cognitive theory.

References

Agawu, Kofi. 1994. "Ambiguity in Tonal Music: A Preliminary Study." In *Theory, Analysis and Meaning in Music*. Edited by Anthony Pople. Cambridge: Cambridge Universtiy Press: 86–107.

Bharucha, Jamshed J. 1987. "Music Cognition and Perceptual Facilitation: A Connectionist Framework." *Music Perception* 5/1: 1–30.

Bharucha, Jamshed J. / Carol L, Krumhansl. 1983. "The Representation of Harmonic Structure in Music: Hierarchies of Stability as a Function of Context." *Cognition* 13: 63–102.

Borges, Jorge L. 1996. "Del rigor de la ciencia." *Arte y Parte* 4: 65.

Caplin, William E. 1998. *Classical Form: A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart, and Beethoven*. Oxford: Oxford University Press.

Carrus, Elisa / Marcus T. Pearce / Joydeep Bhattacharya. 2013. "Melodic Pitch Expectation Interacts with Neural Responses to Syntactic but not Semantic Violations." *Cortex* 49/8: 186–200.

Chomsky, Noam. 1980. *Rules and Representations*, New York: Columbia University Press.

Conklin, Darrell / Ian H. Witten. 1995. "Multiple Viewpoint Systems for Music Prediction." *Journal of New Music Research* 24/1: 51–73.

Conklin, Darrell. 2013. "Multiple Viewpoint Systems for Music Classification." *Journal of New Music Research* 42/1: 19–26.

Cross, Ian. 1998. "Music Analysis and Music Perception." *Music Analysis* 17/1: 3–20.

Cuthbert, Michael S. / Christopher Ariza. 2010. "Music21: A Toolkit for Computer-Aided Musicology and Symbolic Music Data." In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010)*. Edited by J. Stephen Downie and Remco C. Veltkamp. Utrecht: ISMIR: 637–642.

Davidson, Donald. 1989. "The Myth of the Subjective." In *Relativism: Interpretation and Confrontation*. Edited by Michael Krausz. Notre Dame, Ind.: University of Notre Dame Press: 159–172.

——. 2001. *Subjectivity, Intersubjectivity, Objectivity*. Oxford: Oxford University Press.

DeBellis, Mark. 2009. "Perceptualism, Not Introspectionism: The Interpretation of Intuition-based Theories." *Music Perception* 27: 121–130.

Dennett, Daniel C. 1996. *Kinds of Minds: Toward an Understanding of Consciousness*. New York, NY: Harper Collins.

Eerola, Tuomas. 2003. *The Dynamics of Musical Expectancy, Cross-Cultural and Statistical Approaches to Melodic Expectations*. PhD dissertation, University of Jyväskylä.

Eerola, Tuomas / Jukka Louhivuori / Edward Lebaka. 2009. "Expectancy in North Sami Yoiks Revisited: The Role of Data-Driven and Schema-Driven Knowledge in the Formation of Melodic Expectations." *Musicae Scientiae* 13/2: 39–70.

Egermann, Hauke / Marcus T. Pearce / Geraint Wiggins / Stephen McAdams. 2013. "Probabilistic Models of Expectation Violation Predict Psychophysiological Emotional Responses to Live Concert Music." *Cognitive Affective Behavioral Neuroscience* 13/3: 533–553.

Farbood, Morwaread M. 2012. "A Parametric, Temporal Model of Musical Tension." *Music Perception* 29/4: 387–428.

Hamanaka, Masatoshi / Keiji Hirata / Satoshi Tojo. 2004. "Automatic Generation of Grouping Structure Based on the GTTM." In *Proceedings of the International Computer Music Conference*. Ann Arbor, MI: MPublishing, University of Michigan Library: 141–144.

——. 2005. "ATTA: Automatic Time-span Tree Analyzer Based on Extended GTTM." In *Proceedings of the Sixth International Conference on Music Information Retrieval (ISMIR 2005)*. Edited by Reiss, Joshua D. Reiss, and Geraint Wiggins. London: University of London: 358–365.

——. 2006. "Implementing A Generative Theory of Tonal Music." *Journal of New Music Research* 35: 249–277.

——. 2007. "FATTA: Full Automatic Time-Span Tree Analyzer." In *Proceedings of the International Computer Music Conference (ICMC)*. Ann Arbor, MI: MPublishing, University of Michigan Library: 153–156.

Hansen, Niels C. / Marcus T. Pearce. 2014. "Predictive Uncertainty in Auditory Sequence Processing." *Frontiers in Psychology*, 5/1052. doi: 10.3389/fpsyg.2014.01052.

Hedges, Tom / Martin Rohrmeier. 2011. *Exploring Rameau and Beyond: A Corpus Study of Root Progression Theories,* in: *Mathematics and Computation in Music. Third International Conference, MCM 2011, Proceedings*. Edited by Carlos Agon, Moreno Andreatta, Gérard Assayag, Emmanuel Amiot, Jean Bresson, and John Mandereau. Heidelberg: Springer: 334–337.

Hippel, Paul von / David Huron. 2000. "Why do Skips Precede Reversals? The Effect of Tessitura on Melodic Structure." *Music Perception* 18/1: 59–85.

Huron, David. 1999. *Music Research Using Humdrum: A User's Guide*, Stanford, California: Center for Computer Assisted Research in the Humanities.

——. 2006. *Sweet Anticipation: Music and the Psychology of Expectation*, Cambridge, Mass: MIT Press.

——. 2012. "Two Challenges in Cognitive Musicology." *Topics in Cognitive Science* 4/4: 678–684.

Jackendoff, Ray. 1991. "Musical Parsing and Musical Affect." *Music Perception* 9/2: 199–229.

Juslin, Patrik N. / John A. Sloboda. 2010. *Handbook of Music and Emotion: Theory, Research, Applications*. New York: Oxford University Press.

Juslin, Patrik N. / Daniel Västfjäll. 2008. "Emotional Responses to Music: The Need to Consider Underlying Mechanisms." *The Behavioral and Brain Sciences* 31/5: 559–575. Discussion: 575–621.

Juslin, Patrik N. 2013. "From Everyday Emotions to Aesthetic Emotions: Towards a Unified Theory of Musical Emotions." *Physics of Life Reviews* 10/3: 235–266.

Katz, Jonah / David Pesetsky. 2011. "The Identity Thesis for Language and Music", online draft of 2011 at http://ling.auf.net/lingbuzz/000959.

Keiler, Allan. 1978. "Bernstein's 'The Unanswered Question' and the Problem of Musical Competence." *The Musical Quarterly* 64/2: 195–222.

Koelsch, Stefan. 2010. "Towards a Neural Basis of Music-Evoked Emotions." *Trends in Cognitive Sciences* 14/3: 131–137.

——. 2012. *Brain and Music*. Hoboken, NJ: Wiley-Blackwell.

——. 2014. "Brain Correlates of Music-Evoked Emotions." *Nature Reviews Neuroscience* 15/3: 170–180.

Koelsch, Stefan / Martin Rohrmeier / Renzo Torrecuso / Sebastian Jentschke. 2013. "Processing of Hierarchical Syntactic Structure in Music." *Proceedings of the National Academy of Sciences* 110/38: 15443–15448.

Krumhansl, Carol L. 1990. *Cognitive Foundations of Musical Pitch*. New York: Oxford University Press.

——. 1995. "Music Psychology and Music Theory: Problems and Prospects." *Music Theory Spectrum* 17/1: 53–90.

Krumhansl, Carol L. / Frank C. Keil. 1982. "Acquisition of the Hierarchy of Tonal Functions in Music." *Memory and Cognition* 10/3: 243–251.

Krumhansl, Carol L. / Jukka Louhivuori / Petri Toiviainen / Topi Järvinen / Tuomas Eerola. 1999. "Melodic Expectation in Finnish Spiritual Folk Hymns: Convergence of Statistical, Behavioral, and Computational Approaches." *Music Perception* 17/2: 151–195.

Krumhansl, Carol L. / Pekka Toivanen / Tuomas Eerola / Petri Toiviainen / Topi Järvinen / Jukka Louhivuori. 2000. "Cross-Cultural Music Cognition: Cognitive Methodology Applied To North Sami Yoiks." *Cognition* 76/1: 13–58.

Lehne, Moritz / Martin Rohrmeier / Stefan Koelsch. 2013. "Tension-Related Activity in the Orbitofrontal Cortex and Amygdala: An fMRI Study With Music." *Social Cognitive and Affective Neuroscience,* first published online August 22, 2013 doi:10.1093/scan/nst141

Leman, Marc. 1997. Ed. *Music, Gestalt, and Computing: Studies in Cognitive and Systematic Musicology*. Heidelberg: Springer.

Lerdahl, Fred / Ray Jackendoff. 1983. *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.

Luhmann, Niklas. 1992. "What is Communication?" *Communication Theory* 2/3: 251–259.

——. 2000. *Art as a Social System*. Stanford: Stanford University Press.

Manaris, Bill / Juan Romero / Penousal Machado / Dwight Krehbiel / Timothy Hirzel / Walter Pharr / Robert B. Davis. 2005. "Zipf's Law, Music Classification, and Aesthetics." *Computer Music Journal* 29/1: 55–69.

Mandelbrot, Benoit. 1953. "An Informational Theory of the Statistical Structure of Language." *Communication Theory* 84: 486–502.

Manning, Christopher D. / Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

Margulis, Elizabeth H. 2007. "Surprise and Listening Ahead: Analytic Engagements with Musical Tendencies." *Music Theory Spectrum* 29/2: 197–217.

Marsden, Alan. 2010. "Schenkerian Analysis by Computer: A Proof of Concept." *Journal of New Music Research* 39/3: 269–289.

Meyer, Leonard B. 1956. *Emotion and Meaning in Music*. Chicago: University of Chicago Press.

——. 1967. *Music, The Arts, and Ideas*. Chicago: University of Chicago Press.

Narmour, Eugene. 1990. *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*. Chicago: University of Chicago Press.

——. 1991. "The Top-Down and Bottom-Up Systems of Musical Implication: Building on Meyer's Theory of Emotional Syntax." *Music Perception* 9/1: 1–26.

——. 1992. *The Analysis And Cognition Of Melodic Complextiy: The Implication-Realization Model*. Chicago: University of Chicago Press.

——. 1999. *Hierarchical Expectation and Musical Style*, in: *Psychology of Music*. Edited by Diana Deutsch. Second edition. San Diego: Academic Press: 441–472.

Neuwirth, Markus. 2013. *Recomposed Recapitulations in the Sonata-Form Movements of Joseph Haydn and His Contemporaries*. PhD Diss. University of Leuven.

Parncutt, Richard. 1989. *Harmony: A Psychoacoustical Approach*. Heidelberg: Springer.

Pearce, Marcus T. 2005. *The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition*. PhD Diss. Department of Computing, City University, London.

Pearce, Marcus T. / Martin Rohrmeier. 2012. "Music Cognition and the Cognitive Sciences." *Topics in Cognitive Science* 4/4: 468–484.

Pearce, Marcus T. / Geraint A. Wiggins. 2004. "Improved Methods for Statistical Modelling of Monophonic Music." *Journal of New Music Research* 33/4: 367–385.

——. 2006. "Expectation in Melody: The Influence of Context and Learning." *Music Perception* 23/5: 377–405.

——. 2012. "Auditory Expectation: The Information Dynamics of Music Perception and Cognition." *Topics in Cognitive Science* 4/4: 625–652.

Piston, Walter. 1948. *Harmony*. New York: Norton.

Polth, Michael. 2001. "Nicht System – Nicht Resultat: Zur Bestimmung von harmonischer Tonalität." *Musik und Ästhetik* 18: 12–36.

Quinn, Ian. 2014. “Digital and Multimedia Scholarship.” *Journal of the American Musicological Society*, 67/1: 295–307.

Rabiner, Lawrence R.. 1989. “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition.” In *Proceedings of the IEEE, 77/2*. Published by the IEEE (doi: 10.1109/5.18626): 257–286.

Rameau, Jean-Philippe. 1722. *Traité de l’harmonie reduite a ses principes naturels*, Paris.

Rebuschat, Patrick / Martin Rohrmeier / Ian Cross / John Hawkins. 2011. *Language and Music as Cognitive Systems*. Oxford: Oxford University Press.

Rohrmeier, Martin. 2005. *Towards Modelling Movement in Music. Analysing Properties and Dynamic Aspects of Pc Set Sequences in Bach’s Chorales*. Master’s thesis, University of Cambridge.

——. 2007a. “Modelling Dynamics of Key Induction in Harmony Progressions.” In *Proceedings of the 4th Sound and Music Computing Conference*. Edited by Haralambos Spyridis, Anastasia Georgaki, Christina Anagnostopoulou, and Georgios Kouroupetroglou. Athens: National and Kapodistrian University of Athens: 82–89.

——. 2007b. “A Generative Grammar Approach to Diatonic Harmonic Structure.” In *Proceedings of the 4th Sound and Music Computing Conference*. Edited by Haralambos Spyridis, Anastasia Georgaki, Christina Anagnostopoulou, and Georgios Kouroupetroglou. Athens: National and Kapodistrian University of Athens: 97–100.

——. 2011. “Towards a generative syntax of tonal harmony.” *Journal of Mathematics and Music* 5/1: 35–53.

Rohrmeier, Martin / Ian Cross. 2008. “Statistical Properties of Harmony in Bach’s Chorales.” In *Proceedings of the 10th International Conference on Music Perception and Cognition*. Edited by Ken’ichi Miyazaki, Mayumi Adachi, Yuzuru Hiraga, Yoshitaka Nakajima, and Minoru Tsuzaki. Sapporo, Japan: Hokkaido University: 619–627.

——. 2014. “Modelling unsupervised online-learning of artificial grammars: Linking implicit and statistical learning.” *Consciousness and Cognition* 27: 155-167.

Rohrmeier, Martin / Thore Graepel. 2012. “Comparing FeatureBased Models of Harmony.” In *Proceedingsof the 9th International Symposium on Computer Music Modeling and Retrieval (CMMR 2012)*. Edited by Richard Kronland-Martinet, Solvi Ystad, Mitsuko Aramaki, Mathieu Barthet, Simon Dixon, London: Queen Mary University London: 357–370.

Rohrmeier, Martin / Stefan Koelsch. 2012. “Predictive Information Processing in Music Cognition, A Critical Review.” *International Journal of Psychophysiology* 83/2: 164–175.

Rohrmeier, Martin / Markus Neuwirth. In press. „Towards a Syntax of the Classical Cadence.“ In *What is a Cadence? Theoretical and Analytical Perspectives on Cadences in the Classical Repertoire*. Edited by Markus Neuwirth and Pieter Bergé, Leuven: Leuven University Press.

Rohrmeier, Martin / Patrick Rebuschat. 2012. “Implicit Learning and Acquisition of Music.” *Topics in Cognitive Science* 4/4: 525–553.

Rohrmeier, Martin / Patrick Rebuschat / Ian Cross. 2011. "Incidental and Online Learning of Melodic Structure." *Consciousness and Cognition* 20/2: 214–222.

Schellenberg, E. Glenn. 1997. "Simplifying the Implication-Realization-Model of Melodic Expectancy." *Music Perception* 14/3: 295–318.

Schenker, Heinrich. 1935. *Der Freie Satz, Neue musikalische Theorien und Phantasien*. Vienna: Universal Edition.

Schmalfeldt, Janet. 1992. "Cadential Processes: The Evaded Cadence and the 'One More Time' Technique." *Journal of Musicological Research* 12/1–2: 1–52.

Schmuckler, Mark. 1989. "Expectation and Music: Investigation of Melodic and Harmonic Processes." *Music Perception* 7/2: 109–149.

Smoliar, Stephen W. 1980. "A Computer Aid for Schenkerian Analysis." *Computer Music Journal* 4/2: 41–59.

Steedman, Mark J. 1984. "A Generative Grammar for Jazz Chord Sequences." *Music Perception* 2/1: 52–77.

——. 1996. "The Blues and the Abstract Truth: Music and Mental Models." In *Mental Models in Cognitive Science*. Edited by Alan Garnham and Jane Oakhill. Mahwah. NJ: Psychology Press: 305–318.

Stevens, Catherine / Tim Byron. 2009. "Universals in Music Processing." In *Oxford Handbook of Music Psychology*. Edited by Susan Hallam, Ian Cross, and Michael Thaut. Oxford: Oxford University Press: 14–23.

Swain, Joseph P. 1994. "Music Perception and Musical Communities." *Music Perception* 11/3: 307–320.

Temperley, David. 1999. "The Question of Purpose in Music Theory: Description, Suggestion, and Explanation." *Current Musicology* 66: 66–85.

——. 2001. *The Cognition of Basic Musical Structures*: Cambridge, MA: MIT Press.

——. 2009. "In Defense of Introspectionism: A Response to Debellis." *Music Perception* 27/2: 131–138.

——. 2011. "Composition, Perception, and Schenkerian Theory." *Music Theory Spectrum* 33/2: 146–168.

Temperley, David / Trevor de Clercq. 2013. "Statistical Analysis of Harmony and Melody in Rock Music." *Journal of New Music Research* 43/2, 187–204.

Tillmann, Barbara. 2005. "Implicit Investigations of Tonal Knowledge in Nonmusician Listeners." *Annals of the New York Academy of Science* 1060: 100–110.

——. 2012. "Music and Language Perception: Expectations, Structural Integration, and Cognitive Sequencing." *Topics in Cognitive Science* 4/4: 568–584.

Todd, Juanita / Erich Schröger / István Winkler. 2012. Eds. "Special Issue: Predictive Information Processing in the Brain: Principles, Neural Mechanisms and Models." *International Journal of Psychophysiology* 83/2: 199–258.

Tymoczko, Dmitri. 2003. "Function Theories: A Statistical Approach." *Musurgia* 10/3–4: 35–64.

Voss, Richard F. / John Clarke. 1975. "1/f Noise in Music and Speech." *Nature* 258: 317–318.

——. 1978. "'1/f Noise' in Music: Music From 1/f Noise." *Journal of the Acoustical Society of America* 63/1: 258–263.

Wiggins, Geraint. 2012a. "Computer Models of (Music) Cognition." In *Language and Music as Cognitive Systems*. Edited by Patrick Rebuschat, Martin Rohmeier, and John A. Hawkins. Oxford: Oxford University Press: 169–188.

——. 2012b. "Music, Mind and Mathematics: Theory, Reality and Formality." *Journal of Mathematics and Music* 6/2: 111–123.

Wittgenstein, Ludwig (1953), *Philosophical Investigations*, Oxford: Blackwell.

Woolhouse, Matthew / Ian Cross / Timothy Horton. 2006. "The Perception of Non-Adjacent Harmonic Relations." In *Proceedings of the 9th International Conference on Music Perception & Cognition* (ICMPC9). Edited by Mario Baroni, Anna Rita Addessi, Roberto Caterina, and Marco Costa. Bologna: University of Bologna: 1236–1244.

Zanette, Damián H. 2006. "Zipf's Law and the Creation Of Musical Context." *Musicae Scientiae* 10/1: 3–18.

Zipf, George Kingsley. 1935. *The Psycho-Biology of Language*. Boston: Houghton Mifflin.

——. 1949. *Human Behaviour and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley.